SDPD Forensic Science Section – Forensic Biology Unit **July 14, 2017**

STRmix™ modification and performance check: v2.4.06

Introduction

STRmix is an expert system that applies a fully continuous approach to DNA profile interpretation. STRmix v2.3.07 is currently being used in casework. New versions of the software have been released by NichVision since implementing v2.3.07 for casework. STRmix versions 2.4.02, 2.4.05, and subsequently 2.4.06 have been tested internally at SDPD in 2016 and 2017 as each of these versions was released. The goal of this study is to evaluate results from this testing and to implement v2.4.06 for use in casework.

The biggest change from v2.3 to v2.4 is the ability to model forward stutter (N+1 repeat). Other changes were also made regarding likelihood ratio (LR) calculations and user interface. The highlights from the summary of changes (all of which can be found in the 2.4. Operation Manual) are:

- Changes to biological model:
  - Incorporates forward stutter modeling
- Changes to the graphical user's interface:
  - New ability to add multiple files at one time (relevant for replicant amplifications or multiple references)
  - "add to batch" is now disabled once a batch starts running
  - When there is a calculation failure, the program returns to the main screen instead of exiting.
  - Added progress label for indicating the number of loci that have been processed
- Changes to outputs and reports:
  - On the Advanced Report for LR comparisons, best contributor order is shown on the first page (instead of having to go to results file)
  - Calculation time is now reported
  - Can turn off Advanced Report for specific analyses in a batch
- Changes to coding:
  - Corrects an error in how expected peak heights for Q alleles are calculated where multiple contributors have putative dropped alleles at a locus.
  - Java configuration change to use all of the computer's available memory
  - Enhancements to memory use at the start of MCMC
  - New ability to run samples using Low Memory Mode.

The testing done here includes extended output calculations on samples that have drop-out or drop-in, replicate analyses on samples with forward stutter (incorporated at three different stutter percentage levels), LRs, and database searching.

Purpose

The initial validation of the STRmix software at SDPD was done using v2.3.06. A performance check of v2.3.07 was done before implementation of that version. Utilizing the newer v2.4.06 for casework requires not only a performance check for the minor software changes that were made, but also a modification of the forward stutter modeling. The purpose of this study is to test the functionality of this software upgrade at SDPD and ensure that the upgrade to STRmix™ v2.4.06 does not negatively impact interpretation and incorporates the new elements of modeling correctly.

Materials and Methods

STRmix v2.4.02, then v2.4.05, and then v2.4.06 were initially installed on a computer with 32 GB of RAM and Intel Core i5-4570, 3.2 GHz, running Java Version 7 Update 79. The same input files were used to test this software update as were previously used in the internal validation of v2.3.06 and v2.3.07, for performance check purposes. The contributors in these samples had varying template levels. Thirteen samples were chosen for comparison: five single source (SS) (one incorporating dropout), four 2 person mixtures (2M), two 3 person mixtures (3M), one 4 person mixture (4M), and 1 five person mixture. Additonal five person mixtures were attempted on this computer without success. STRmix v2.4.06 was then installed on a computer with 128 GB of RAM and Intel Xeon CPU E56-2640 v3, 2.6 GHz and 2 processors running Java Version 8 Update 72. Twelve different 5 person mixtures were attempted on this computer (each run at least twice).

Additional sample files were also incorporated to go beyond the standard software performance check, and ensure that the modeling of forward stutter was validated. This was accomplished with two new groups of samples: the first group of samples were those used for the previous performance check, and were modified to include realistic N+1 peaks. In order to be as consistent as possible, the forward stutter peaks were added *in silico* to the same bank of evidence .txt files instead of amplifying or analyzing completely different samples. Three sets of sample input files were created from the original panel of modification study samples to test the limits of forward stutter modeling (one single source, two 2 person mixtures, and one 3 person mixtures):

- The first set of samples had N+1 stutter peaks added in at three loci where N+1 was more frequently observed. The heights of these *in silico* peaks were designed based on the average stutter observed, and were added to the STRmix sample input file (similar to how an unresolved peak is added to the evidence .txt file.
- The second set created from the original sample files had the same N+1 stutter peaks, but one of those three was elevated such that it was about double the average, but still below the 15% forward stutter cap (see below for more information about the 15% stutter cap).
- The third set created from the original sample files had the same N+1 stutter peaks, but one of those three was elevated so much that it exceeded the 15% forward stutter cap.

The second group of samples were completely new samples. This set of samples included proficiency test samples, one three person mixture in which one of the minor contributors alleles had a height and position that caused it to be filtered as N+1 stutter by our current GMID-X analysis settings, and one four person mixture that had a problem calculating the probability of multiple instances of dropout at one locus.

The STRmix kit settings required a change to assess the forward stutter capabilities of the software. In order to incorporate the modeling of forward stutter, two changes were made: 1) adding a forward stutter file, and 2) specifying the forward stutter cap. Information about forward stutter had already been compiled from the initial GlobalFiler modification, and updated with the 2016 addendum. Table 1 shows the forward stutter values compiled during the 2016 stutter study. The forward stutter cap was determined using the maximum observed values. There was one locus (D2S1338) with the maximum observed stutter of 13.2%, so the STRmix forward stutter cap was set to be 0.15 (15%), which is above the maximum observed stutter.

Locus	N+1 Stutter (%)			
	Maximum	Average	SD	N
D3S1358	2.77	0.96	0.61	45
vWA	6.29	1.57	0.67	20
D16S539	3.01	1.29	0.79	20
CSF1PO	3.35	1.60	0.93	17
TPOX	0	0	0	0
D8S1179	5.83	1.43	1.19	36
D21S11	9.39	1.53	0.86	78
D18S51	8.97	1.67	0.95	27
DYS391	0	0	0	0
D2S441	2.21	0.95	0.39	35
D19S433	2.63	2.10	0.76	2
TH01	1.70	0	0	1
FGA	4.07	1.42	0.72	37
D22S1045	7.52	4.06	0.93	629
D5S818	9.06	1.13	0.62	90
D13S317	3.34	0.94	0.68	71
D7S820	2.33	1.05	0.60	9
SE33	9.08	1.70	0.92	78
D10S1248	2.81	1.28	0.54	32
D1S1656	4.46	1.26	0.79	76
D12S391	2.09	1.45	0.51	7
D2S1338	13.20	2.06	1.19	20

Table 1 – Forward stutter values collected from single source samples amplified with Globalfiler.

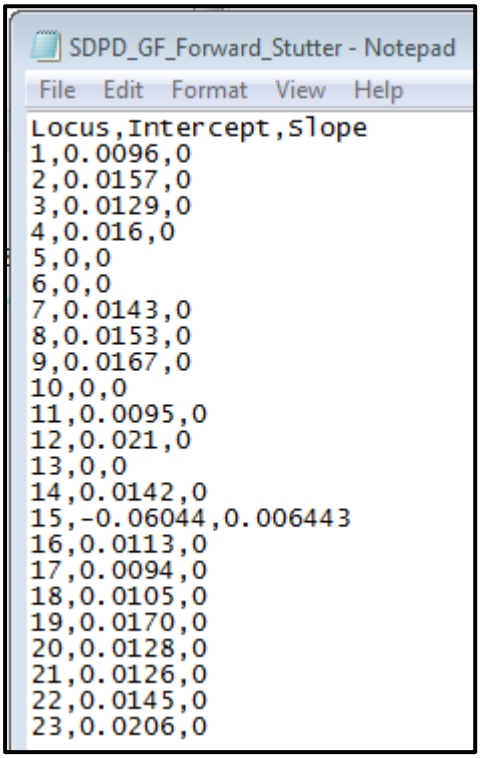

Figure 1 – Forward stutter file for STRmix v2.4.06

The forward stutter file specifies slope and y-intercept for each autosomal locus. Because most of the forward stutter we observe does not appear to vary with allele size, the slope was zero, and

the y-intercept was the average[1] of all observations for that locus (see Figure 1). The only exeption to the slope = 0 was trinucleotide repeat locus D22. It is the only locus with a non-zero slope, and the slope and y-intercept are based on the linear regression of observed stutter (see sutter graphs). Initially in the modification, a stutter file in which the y-intercept was the average plus 2 standard deviations was also tested.

In addition to changing the stutter settings in STRmix, another setting change was made. In moving to v2.4, STRmix technical release notes stated that the new default value for the Var > mode parameter is 0.5. This setting specifies the value that STRmix will not allow the variance to drop below during the MCMC. The actual minimum value is calculated by the Var > mode parameter multiplied by the mode of the kit's variance distributions (see release notes for an example). Increasing this number from the previous value of 0.1 prevents STRmix from underoptimizing the variance values for both allele variance and stutter variance. During our modification, this change was made at the same time as the incorporation of forward stutter files and settings. Changing the Var > mode parameter to 0.5 was done in all three SDPD specific kits. This change required samples to be analyzed with SDPD GlobalFiler and SDPD Minifiler kits using v2.4.06. The same samples described above were used for GlobalFiler. For Minifiler: five single source samples (three of those with dropout), four 2 person mixtures, four 3 person mixtures, and two 4 person mixtures were used to assess Minifiler deconvolution.

Several outputs were examined for each sample run. A database search against modification samples including known contributors and known non-contributors was run after every deconvolution to get an LR for each of the contributors. The same allele frequencies (Caucasian frequencies reported in this study), theta values and database file were used across the entire study. The genotype weights are expected to be the same for complete single source samples between software versions, and, thus, these samples should result in identical LRs. Mixed DNA profiles should result in different, but similar LRs due to the expected variability within the MCMC, depending on which software versions are being compared. Contributor proportions, LRs of all contributors to the mixtures between versions, and LRs of all the contributors to the mixtures when forward stutter is incorporated vs. when it is not incorporated were assessed and recorded.

Results

Software performance check

The first goal was to verify that the new versions of the software were obtaining consistent results with v2.3.07 (currently being used in casework), independent of forward stutter modeling. Table 2 lists results of 12 different samples. None of these samples have forward stutter peaks

[1] **It should be noted that in order to best fit the majority of the data, the average was calculated after excluding data points statistically determined to be outliers at vWA, D21, D18, D5, SE33, and D2S1338.*

included in the input .txt file. The results in the table reflect the MCMC process (contributor percentages in parenthesis), as well as the LRs from the database search.

Column 2 (blue) in Table 2 shows results from STRmix v2.3.07. This version does not have the option to model forward stutter. In 2016, v2.4.05 was assessed before entering any forward stutter modeling settings, and those results are shown in Column 3 (1st yellow). The single source samples all have identical LR values except sample SS5, which has three alleles dropping out and much lower peak heights than the other single source samples. This sample was also run with extended output turned on for more information about the MCMC process with this sample.

Sample ID		v2.3.07	v2.4.05 no N+1 stutter file	v2.4.05 N+1 stutter file	v2.4.06 N+1 stutter file
SS1		$1.2x10^{28}$	$1.2x10^{28}$	$1.2x10^{28}$	$1.2x10^{28}$
SS2		$1.61x10^{34}$	$1.61x10^{34}$	$1.61x10^{34}$	$1.61x10^{34}$
SS3		$5.63x10^{31}$	$5.63x10^{31}$	$5.63x10^{31}$	$5.63x10^{31}$
SS4		$4.31x10^{32}$	$4.31x10^{32}$	$4.31x10^{32}$	$4.31x10^{32}$
SS5 (dropout)		$1.2x10^{28}$	$7.63x10^{26}$	$6.4x10^{26}$	$4.03x10^{26}$
2M1	C1	(81%) $1.2x10^{28}$	(82%) $1.2x10^{28}$	(81%) $1.2x10^{28}$	(81%) $1.2x10^{28}$
	C2	(19%) $7.4x10^{29}$	(18%) $6.87x10^{29}$	(19%) $6.74x10^{27}$	(19%) $6.74x10^{27}$
2M2	C1	(53%) $1.43x10^{16}$	(53%) $1.44x10^{16}$	(52%) $7.19x10^{15}$	(52%) $7.19x10^{15}$
	C2	(47%) $9.38x10^{17}$	(47%) $9.44x10^{17}$	(48%) $4.7x10^{17}$	(48%) $4.7x10^{17}$
2M3	C1	(55%) $3.51x10^{26}$	(55%) $1.35x10^{25}$	(55%) $3.55x10^{24}$	(56%) $3.06x10^{25}$
	C2	(45%) $4.28x10^{26}$	(45%) $1.59x10^{25}$	(45%) $3.58x10^{24}$	(44%) $3.23x10^{25}$
2M4	C1	(92%) $1.99x10^{33}$	(92%) $1.99x10^{33}$	(93%) $1.99x10^{33}$	(92%) $1.99x10^{33}$
	C2	(8%) $7.84x10^{24}$	(8%) $1.14x10^{25}$	(7%) $7.4x10^{23}$	(8%) $2.29x10^{23}$
3M1	C1	(45%) $1.43x10^{22}$	(43%) $2.81x10^{22}$	(43%) $3.72x10^{21}$	(43%) $3.72x10^{21}$
	C2	(35%) $8.72x10^{16}$	(37%) $5.78x10^{17}$	(36%) $4.02x10^{16}$	(36%) $4.02x10^{16}$
	C3	(21%) $2.42x10^{26}$	(20%) $3.65x10^{27}$	(22%) $1.21x10^{26}$	(22%) $1.21x10^{26}$
3M2	C1	(49%) $6.75x10^{19}$	(51%) $6.01x10^{19}$	(52%) $3.32x10^{19}$	(52%) $2.66x10^{19}$
	C2	(37%) $1.2x10^{13}$	(34%) $4.22x10^{12}$	(34%) $7.56x10^{12}$	(34%) $7.05x10^{12}$
	C3	(14%) $1.24x10^{13}$	(15%) $6.84x10^{12}$	(14%) $2.85x10^{12}$	(14%) $4.68x10^{12}$
4M1	C1	(51%) $4.91x10^{12}$	(50%) $4.34x10^{12}$	(49%) $2.04x10^{12}$	(49%) $2.21x10^{12}$
	C2	(22%) $3.64x10^{23}$	(21%) $4.02x10^{23}$	(23%) $1.79x10^{23}$	(23%) $1.75x10^{23}$
	C3	(16%) $9.82x10^{13}$	(17%) $1.03x10^{14}$	(16%) $1.96x10^{13}$	(16%) $1.73x10^{13}$
	C4	(12%) $1.74x10^{9}$	(12%) $1.53x10^{9}$	(11%) $4.05x10^{8}$	(12%) $3.46x10^{7}$

Table 2 – Database search LRs and contributor percentages from mixtures analyzed in different STRmix software versions with and without the ability to model forward stutter. No forward stutter peaks were observed in the sample files for this part of the study.

With regard to the mixtures, every single one was deconvoluted similarly between STRmix versions. The contributor proportions from one version to the next are within 3% of each other, and the genotype weights for each contributor are consistent from run to run. Because genotype weights are consistent, every contributor is included with a likelihood of very similar magnitude. The similar results in v2.4.05 with no forward stutter settings establish the acceptable performance of the minor software changes from v2.3.07 to v2.4.05 affecting both the MCMC and LR functionality.

STRmix v2.4 also has the option of modeling forward stutter. To do this, two STRmix kit settings were changed, as mentioned above: the forward stutter cap was set to 0.15, and a forward stutter file was added. The exact same input files were analyzed (i.e., no forward stutter peaks were added to samples yet), and those results are shown in Column 4 (2nd yellow column). Similarly, the contributor percentages, genotypes weights and LRs were very similar from run to run. The biggest difference here was seen with C2 in mixture 2M1 (LR went from ~10^{29} to ~10^{27}) because genotype weights change when some minor contributor peaks are modeled as possible forward stutter. Results from this set of samples establish that just modifying the settings to model forward stutter (entering a forward stutter cap and loading a forward stutter file containing locus specific average observed stutter) does not compromise the ability to deconvolute samples when no forward stutter peaks are observed/detected.

Column 5 (green) demonstrates that the new software version 2.4.06 can deconvolute samples consistently with previous versions and have consistent inclusions of contributors. Even fewer differences were observed in the results after running the same samples in STRmix v2.4.06 with the forward stutter file settings loaded. Contributor percentages, genotype weights and LRs were reproducible to a very high degree of similarity. Known non-contributors were never included in any of the database searches for any of the runs. Results from this set of samples establish that the minor changes made between v2.4.05 and v2.4.06 do not affect the ability to deconvolute samples in which no forward stutter peaks were observed.

<u>Modeling forward stutter</u>

Once the forward stutter settings were added to the SDPD GlobalFiler STRmix kit, the goal was to test the limits of its ability to interpret samples that had forward stutter peaks in the input file. As described in the Methods section, stutter peaks were added in silico (three different peaks per sample) to four different samples. These mixtures were analyzed three times each with the same seed, each time with increasing peak height for one of the three added N+1 peaks. This was done three times: once using v2.4.05, once using v2.4.06 with a forward stutter file having average+2SD for the y-intercept, and once using v2.4.06 with a forward stutter file having the average stutter as the y-intercept. Results shown in Table 3 are only from v2.4.06 using the average stutter as the y-intercept in the forward stutter file, because they were all very similar to one another (within 2% for contributor percentage and LRs no more than ~1 order of magnitude

different for each contributor), and the same results were achieved when the forward stutter was over the 15% cap, as far as exclusions of known contributors. Since the y-intercept in the forward stutter file did not negatively affect results for samples that had elevated stutter, this served to validate both versions of the stutter file. However, only the forward stutter file with the average stutter ratio (with the exception of D22) will be used in casework.

The single source sample (SS3) had the exact same LR, even when the N+1 peak was elevated to twice the average percentage at that locus (D8). When the N+1 stutter peak was added such that it was over 15% of the homozygous parent peak, it became an obligate allele through the deconvolution (because it was also well over the drop-in cap). This caused an exclusion at that locus because the genotypes did not match. This verifies that the deconvolution is taking the stutter cap into account. Although the two peaks at D8 in this sample are very imbalanced, it still accepts them at 100% genotype weight. The allele variance was higher (33% vs 25%) in the sample with N+1 stutter over the 15% cap, but the other diagnostics did not indicate a problem with the deconvolution.

Sample ID		v2.4.06 N+1 stutter in sample	v2.4.06 N+1 elevated stutter in sample	v2.4.06 N+1 stutter over cap in sample
SS3		$5.63x10^{31}$	$5.63x10^{31}$	0
2M1	C1	(81%) $1.2x10^{28}$	(81%) $1.2x10^{28}$	(82%) $1.2x10^{28}$
	C2	(19%) $2.22x10^{28}$	(19%) $1.75x10^{29}$	0
2M2	C1	(53%) $1.03x10^{16}$	(52%) $8.04x10^{15}$	(55%) $1.81x10^{15}$
	C2	(47%) $6.74x10^{17}$	(48%) $5.26x10^{17}$	(45%) $3.32x10^{17}$
3M1	C1	(43%) $4.57x10^{22}$	(44%) $1.41x10^{23}$	wouldn't run as a 3 person mixture
	C2	(35%) $1.28x10^{18}$	(36%) $3.77x10^{18}$	
	C3	(22%) $7.93x10^{27}$	(20%) $3.45x10^{27}$	

Table 3 – Databse search LRs and contributor percentages from a subset of samples had three forward stutter peaks added to each input file.

The same thing happened in sample 2M1. Both contributors were included in the samples that had N+1 stutter peaks added in, even at an elevated level. But, when that N+1 peak was elevated to be higher than 15% of the parent peak at D5, one of the known contributors was excluded for the same reason – the elevated stutter peak from C1 became an obligate allele for C2 (the ~18% contributor), whose types now did not match the reference genotype at that locus. As the stutter was elevated, the allele and stutter variance diagnostics increased, and the effective sample size and average (log) likelihood decreased, but none of these values fell into a range that would indicate that there was likely an issue, just values high enough that warrant a second look into the input file.

In sample 2M2, the N+1 stutter peak was elevated at D22. Because it was a balanced mixture, and the combination of genotypes at this locus (C1 = 16,18, C2 = 15,16 and N+1 was in the 19 bin), both contributors were still included. When N+1 stutter is below the cap (even when it is elevated), the N+1 stutter peak is not considered in the genotype weights. When it is over the

cap, it is included as a possible genotype for both contributors, and both have an inclusive LR in a database search. However, when both of the known contributors are compared at the same time using the LR from previous analysis function, the LR is 0 because the N+1 stutter peak is too high to be considered drop-in, so it has to be attributed to one of the contributors. As the stutter was elevated, the allele and stutter variance diagnostics increased, and the effective sample size and average (log) likelihood decreased. Unlike mixture 2M1, the stutter variance increased considerably (to 46.7) when N+1 stutter was over 15% at one locus.

In sample 3M1, the analysis could not even run with N+1 > 15%, because the N+1 stutter peak was added at SE33 where all of the contributors were heterozygous. Attempting to run this as a three person mixture resulted in an error message, which occurred immediately after starting the analysis, giving an indication that there were too many alleles for the number of contributors specified. In reality, if this scenario were observed, it would likely not be interpreted as a three person mixture due to the height of the "N+1 peak" at SE33, but would be interpreted as a 4 person mixture. This part of the study demonstrated the functionality of the forward stutter max and STRmix's ability to model forward stutter as a possibility when it is below 15% of the parent peak. As the stutter was elevated, the allele and stutter variance diagnostics increased, and the effective sample size and average (log) likelihood decreased, but none of these values fell into a range that would indicate that there was likely an issue, just values high enough that warrant a second look into the input file.

The abilitiy to assume a contributor was also tested with three N+1 peaks added in (at a level that was average for those loci). Mixtures 2M1, 2M2 and 3M1 were interpreted with one reference file specified as a contributor in both Hp and Hd. The mixture was deconvoluted as expected, with LR values increased accordingly, especially in the more balanced mixtures (2M2 and 3M1).

<u>Modeling forward stutter in evidence-like samples</u>

To further test the capabilities of modeling N+1 stutter, additional samples were assessed. Several proficiency test samples and an evidence mixture in which a pitfall was encountered with the current analysis methods/v2.3.07; specifically, the evidence mixture had a contributor allele filtered in GMID-X as forward stutter.

Performance check samples: Three single source samples from proficiency tests in which analysts indicated the presence of elevated N+1 stutter (based on the Evidence_2016 GMID-X filters) were examined in GMID-X for the presence of *additional* peaks filtered as N+1 stutter. All N+1 peaks above 100 RFU were added back into the STRmix input file. All N+1 peaks were modeled as such in the single source samples. Genotype weights were 100%, and did not include the N+1 peak as an allele.

Four 2-person mixtures from proficiency test samples were analyzed, and the results for each are described below:

- Test sample 2.1 had a contributor ratio of 96:4. At D22, one peak was originally filtered by GMID-X as N+1 stutter. When it was added back in, it was considered as an allele, but not obligated to be one (as evidenced by the genotype weights). At SE33, a peak was left in the mixture because it was elevated slightly above the N+1 filter, and there were other indications of another contributor with peaks about the same height. STRmix also considered this peak as an allele, but now other options (N+1/drop-in) were also possible in the list of genotype weights that include that peak.
- Test sample 2.2 (Ratio 97:3) only had one peak (at SE33) that was possible N+1 stutter. It was about double the average N+1 stutter at that locus. STRmix preferred this peak as an allele as indicated by the genotype weight, but it was not an obligate allele, so it would not result in an excludion if it truly was N+1 stutter.
- Test sample 2.3 (Ratio 58:42) had one peak initially filtered by GMID-X as N+1 stutter at D12. When this peak was added to the STRmix input file, it was not considered as an allele. This result was as expected because of the number and heights of the other peaks detected at this locus.
- Test sample 2.4 (Ratio 96:4) had two peaks originally filtered as N+1 stutter by GMID-X that were considered as alleles, but not obligate alleles. There were also two peaks not filtered or edited out that were also considered, but, again, not obligate alleles.

This result highlights the benefit of not having to filter out N+1 stutter, and having STRmix v2.4.06 model those types of peaks. In all cases the modelting was robust and appropriate for each of the different evidence sample types.

Three person mixture: This sample was originally assessed as an evidence sample with v2.3.07. The problem that occurred was that a minor contributor allele was filtered as N+1 stutter. A mixture of three people (one of which can be assumed based on visual comparison of types and it being an intimate sample) has the following detected peaks at D8: 11 (124 RFU), 12 (1615 RFU), 13 (19816 RFU), 14 (509 RFU), and 15 (645 RFU). Filtering N+1 stutter and not modeling it in STRmix would cause the 14 to be filtered out as N+1 stutter. The assumed contributor has a 13, 13 genotype. The 12 peak falls within stutter expectations for the 13 allele. It is possible that the 11 is N-2 stutter, but other peaks with similar height exist at other loci, so it is reasonable to leave this in for STRmix analysis.

When the 14 peak is filtered as N+1 stutter and the mixture conditioned on a 13,13 genotype at that locus, comparison to a POI with a 14,15 genotype results in an exclusionary LR at that locus. The contributor ratio is 90:10:1 (C1 is the 90% contributor) and the overall LR for this POI falls in the inconclusive range (LR Total = 0.095). Additional loci also support exclusion, although most support inclusion.

When forward stutter peaks are not filtered by GMID-X, the 14 peak at D8 (and another peak at D19) are included in the STRmix evidence input file. Analysis with v2.4.06 gives a different result for this sample. Still assuming the same contributor and comparing to POI the mixture ratio is 95:4:<1, with C1 being the 95% contributor and POI now included with strong support as the 4% contributor. Interestingly, the LR at D8 still supported exclusion (based on the genotype weight), but it is the only one, so the overall LR is overwhelmingly inclusionary (99% HPD = $1.42x10^{11}$)

Thus, modeling forward stutter allows more genotype possibilities to be considered allowing a more robust interpretation in some cases. This comes at a cost of having to consider peaks as N+1 when determining most likely number of contributors.

Expected heights for dropout alleles

Four person mixture: One observation made in STRmix versions prior to 2.4.05 was that when multiple contributors had putative dropped alleles at a locus, the expected peak heights for those Q alleles were summed during the portion of the MCMC when the probability of the expected profiles were calculated. This caused the height of the Q allele to be higher than it should have been, resulting in the lower probabilities for profiles where the genotype sets contained multiple contributors with Q alleles. As such, when LRs are performed it results in each contributor being included individually, but not together. A change was made in going from STRmix v2.4.04 to STRmix v2.4.05 that considers the heights of Q alleles separately, and the four person mixture described here exhibited this issue when run with v2.3.07. This mixture of 4 people originally had a ratio of 68:15:10:6 and included POI A alone as the 10% contributor (moderate support for inclusion) and included POI B alone as the 15% contributor (limited support for inclusion). The calculation that considered both POIs together was exclusionary (LR=0). Upon further inspection, SE33 was the only locus with an LR=0. Both POI A and POI B are heterozygote at that locus, but in the mixture only one of each of their alleles is detected (one allele would be dropping out for each of them at this locus in order for them to both be included). It is important to note that this did not alter the assumption about the number of contributors. STRmix v2.3.07 did not allow for simultaneous dropout of each of these two contributors. To investigate further, two more analyses were performed: 1 – assume POI A and compare to POI B; 2 – assume POI B and compare to POI A. Both of these comparisons resulted in strong support for inclusion for both POI A and POI B because conditioning the profile assumed the presence of the dropped allele of the assumed POI, thus allowing the other POI to have a Q allele. Additionally, a comparison of POI A and POI B together was run excluding SE33, and the LR indicated that both were included with strong support. It is possible that a replicate MCMC could have resulted in accepted genotype weights that would have allowed inclusion of both contributors, but running a replicate is not our standard practice for addressing this type of issue, and given the way Q alleles were being summed in the calculatation.

This mixture was then analyzed with STRmix v2.4.06 (using the same seed). The results indicated a contributor ratio of 69:15:10:7. Similarly, POI A alone is included as the 10% contributor (moderate support for inclusion), and POI B alone is included as the 15% contributor. In the previous verion, this contributor was included with limited support, and with STRmix v2.4.06, POI B is included with strong support. When these two contributors were compared together, they are both included together with strong support, as STRmix considered the simultaneous dropout of an allele from both contributors at SE33.

Thus, the change in v2.4 to how expected peak heights for Q alleles are calculated where multiple contributors have putative dropped alleles at one locus prevents multiple troubleshooting steps to arrive at the conclusion that best reflects the DNA result.

Likelihood ratio calculations

One of the changes made in v2.4.06 was a code change that affects the database search LR and HPD calculations in samples with drop-in and/or forward stutter modeled. The Release and Testing Report indicated that this was a very small change with only very small differences only after the first significant figure. Nevertheless, we tested both both the database search LR function (many times comparing one version to another) on the samples with forward stutter, and performed LR from previous analysis calculations in the samples with forward stutter. Both the LR Total and HPD were calculated using v2.4.06 regardless of which version the mixture was originally analyzed with. Results are shown in Table 4.

Sample ID	v2.4.05 DB LR	v2.4.06 DB LR
SS3	$5.63x10^{31}$	$5.63x10^{31}$
SS3 w/ elevated (+) stutter	$5.63x10^{31}$	$5.63x10^{31}$
2M1	$1.2x10^{28}$	$1.2x10^{28}$
2M1 w/ elevated (+) stutter	$1.2x10^{28}$	$1.2x10^{28}$
2M2	$4.7x10^{17}$	$4.7x10^{17}$
2M2 w/ elevated (+) stutter	$4.56x10^{17}$	$4.56x10^{17}$
3M1	$3.72x10^{21}$	$3.72x10^{21}$
3M1 w/ elevated (+) stutter	$1.56x10^{21}$	$1.56x10^{21}$

Table 4 – Database search LR values for samples with either average N+1 stutter peak height, or elevated N+1 stutter peak height.

Database search LRs are identical between software versions using samples for which forward stutter was modeled. No drop-in peaks were included in these samples, but the height of the N+1 peaks were below the drop-in cap, so could presumably be modeled as drop-in. A single source sample with drop-in was also run (extended output turned on). As expected, the LR Total, and 99% lower HPD LR values were identical when the same seed was set for those samples.

Analysis time

Run time was also recorded for these samples. A new file is created in the results folder called "Progress Dialog output.txt" that includes all the green text that shows up while a sample is being analyzed and a summary at the bottom of the time for each step of the process. Analysis time can also be found in the results file. Analysis time is sample dependent, as can be seen in

Sample ID	v2.4.05 no N+1 stutter file	v2.4.05 N+1 stutter file	v2.4.05 N+1 stutter in sample	v2.4.06 N+1 stutter file	LR 2.4.06 N+1 stutter in sample
SS1	15 sec	18 sec		18 sec	
SS2	15 sec	16 sec		16 sec	
SS3	15 sec	15 sec	18 sec	15 sec	16 sec
SS4	11 sec	11 sec		11 Sec	
2M1	5 min	3 min	3 min	3 min	3 min
2M2	59 sec	45 sec	43 sec	44 sec	43 sec
2M3	56 sec	53 sec		53 sec	
2M4	4 min	4 min		4 min	
3M1	25 min	14 min	17 min	14 min	18 min
3M2	4 min	4 min		5 min	
4M1	30 min	46 min		31 min	

Table 5 – Total analysis time in samples with/without forward stutter settings and with/without forward stutter peaks in the evidence input file.

Table 5. Single source samples are analyzed very quickly, and analysis time increases as the number of contributors increases.

Analysis time can also vary greatly with the same order mixture depending on which alleles are present, and the number and height of the peaks detected. This can be seen clearly in the three person mixtures; 3M1 took 4-5 times longer than 3M2. The run time between software versions had a much lower variability than sample to sample variability. In some cases, run time decreased (see mixture 3M1).

Improvements to extended output mode were also made. Extended output files were created for two single source samples in v2.4.05 and in v2.4.06, and both took less than 5 minutes to complete.

Low Memory Mode and 5 person mixtures

The Low Memory Mode option was added to conserve memory. During the post burn-in process, larger order mixtures can run out of memory and the analysis of the mixture in STRmix may not complete. With the type of computers we currently have in the lab and using v2.3.07, this occurs frequently during analysis of 5 person mixtures. The computer may run out of memory because of the large amount of data (like GR values, log(likelihoods), allele and variance constants) being written to a file during the post-burn-in. In Low Memory Mode, STRmix writes less to the computer memory, but it has to recreate that information later when it needs to used it again. In this way, STRmix can conserve memory, but at the cost of extra execution time.

Sample ID	v2.4.06	v2.4.06 Low Memory Mode
3M1	14 min	35 min
3M2	5 min	13 min
4M1	31 min	1.25 hrs
5-1	3 hrs	4.25 hrs
5-2		
5-3	7.5 hrs	13 hrs
5-4	21 hrs	6.5 hrs
5-5	2 hrs	3.5 hrs
5-6	2.5 hrs	3.5 hrs
5-7	14.5 hrs	22 hrs
5-8	2.5 hrs	3.75 hrs
5-9	1 min	2.5 min
5-10	1 hr	1.25 hrs
5-11		4 hrs
5-12	10.5 hrs	2.75 hrs

Table 6 – Analysis time increases when Low Memory Mode is selected.

Low Memory Mode was tested on several higher order samples during the modification study (using only v2.4.06). For example, on a computer with 32 GB of RAM, Mixture 3M1 took 14 minutes to deconvolute. When this same sample was run again (same seed) with Low Memory Mode selected, it took 35 minutes. Mixtures 3M2 and 4M1 did not have any forward stutter, and the analysis time increased in Low Memory Mode. Even with Low Memory Mode on this computer, only one 5 person mixture was analyzed, and it was done surprisingly quickly, after only ~7 hours (results not shown here). None of the other 5 person mixtures attempted were completed on the 32 GB computers.

The five person mixtures were then tested on the computer with 128 GB of RAM (same seed). There was a lot of variability in run time between the mixtures, but there was a relatively consistent increase in the analysis time going from "regular mode" to Low Memory Mode (see Table 6), with two exceptions. The mixtures colored in green are the robust 5 person mixtures, the yellow mixtures had a lower total DNA input, and lower peak heights. The mixtures in red were the low DNA input, and not all of these presented as 5 person mixtures in the electropherogram. For example, mixture 5-9 was a low level, partial mixture that had a

Sample ID	Contributor Ratio	DB LR using LMM	Sample ID	Contributor Ratio	DB LR using LMM	Sample ID	Contributor Ratio	DB LR using LMM
5-1	27% 22% 20% 17% 14%	6.48×10^{12}	5-5	53% 18% 12% 10% 7%	2.02×10^{25}	5-9	34% 23% 22% 11% 11%	2.72×10^{5}
		7.15×10^{9}			8.29×10^{9}			119.70
		7.45×10^{7}			5.21×10^{6}			44.19
		1.50×10^{6}			4.94×10^{7}			1.11
		1.28×10^{5}			8.59×10^{7}			18.81
5-2			5-6	26% 22% 20% 17% 15%	1.71×10^{12}	5-10	30% 24% 20% 16% 10%	8.39×10^{7}
					2.52×10^{12}			2.50×10^{15}
					9.93×10^{7}			280.08
					2.69×10^{6}			1.44×10^{7}
					4.66×10^{7}			265.85
5-3	64% 12% 10% 8% 6%	3.28×10^{27}	5-7	61% 14% 11% 9% 6%	1.09×10^{27}	5-11	70% 12% 8% 7% 3%	3.56×10^{27}
		2.29×10^{11}			3.79×10^{8}			7.66×10^{7}
		2.40×10^{10}			5.13×10^{7}			4067.97
		6.54×10^{7}			1.52×10^{12}			2.01×10^{6}
		3.99×10^{6}			5.17×10^{7}			1.42×10^{6}
5-4	73% 11% 9% 6% 0%	1.06×10^{28}	5-8	64% 16% 12% 6% 2%	1.05×10^{28}	5-12	63% 17% 12% 9% 0%	7.16×10^{27}
		2.45×10^{12}			1.06×10^{16}			3.07×10^{12}
		1.23×10^{11}			3.09×10^{7}			1.03×10^{4}
		1.91×10^{9}			1.10×10^{8}			1.26×10^{10}
		4.22			2.21			2.23

Table 7 – Database search LRs from samples analyzed with Low Memory Mode selected. When Low Memory Mode was not selected, LRs were identical.

maximum of 4 detected alleles at any locus, which explains the deconvolution time of 1 *minute*. Interestingly, one of the mixtures could only be interpreted if Low Memory Mode was selected.

Low Memory Mode did not affect the quality of the interpretation. Because the same version of the software was used on the same computer, the exact same MCMC results could be replicated by setting the same seed value. Genotype weights are identical, and so the LRs for these mixtures are identical when deconvoluted in Low Memory Mode (only one set of values for the 5 person mixtures are shown in Table 7).

Because this is the first time 5 person mixtures have been consistently successful, these LR results give insight into the quality of the deconvolution of these mixtures by STRmix. Samples in the top half of Table 7 were designed to be balanced mixtures. Samples in the bottom half of Table 7 were designed to have one "major" contributor, and the same color scheme as Table 6 is being employed in that the total target amount of DNA in the green mixtures is high, and low in the red mixtures. Not all known contributors are included. This is due to dropout in the mixture, and, in the case of mixture 5-9, extreme dropout, because only one contributor would be included with our current verbal scale. Because LR from Previous Analysis was not done for each of these contributors at this time, the database search LR is not paired up with an individual contributor, but the order of the LRs for the contributors is consistent from the high to low input mixtures. This table serves to give a summary of the results from the 5 person mixture deconvolutions.

Sample ID	Ratio v2.3.06	DB LR v2.3.06	Ratio v2.4.06	DB LR v2.4.06
Mix 2-18	(50:50)	2.97E+16	(52:48)	5.67E+17
		7.31E+14		8.65E+15
Mix 2-34	(82:18)	3.08E+24	(83:17)	3.20E+23
		6.93E+26		1.16E+28
Mix 2-36	(88:12)	3.03E+30	(87:13)	5.90E+32
		4.72E+28		1.09E+27
Mix 2-41	(87:13)	1.31E+28	(87:13)	1.58E+29
		7.15E+31		1.98E+33
Mix 3-31	(44:42:13)	1.10E+19	(48:38:13)	2.24E+20
		2.65E+13		5.54E+14
		2.31E+18		2.77E+19
Mix 3-43	(38:34:29)	2.06E+19	(39:34:27)	6.34E+13
		7.86E+14		8.08E+13
		2.19E+16		2.77E+16
Mix 3-44	(65:21:17)	3.57E+12	(65:20:15)	2.01E+16
		1.18E+25		6.92E+28
		1.74E+20		8.07E+22
Mix 3-50	(72:21:7)	4.37E+32	(71:22:7)	2.09E+33
		1.25E+18		2.29E+16
		1.37E+26		8.35E+26
Mix 4-39	(38:36:22:5)	2.40E+21	(42:35:20:2)	6.51E+19
		1.73E+19		8.91E+14
		3.39E+23		1.71E+23
		5.85E+17		8.32E+16
Mix 4-47	(43:26:17:14)	3.57E+08	(41:27:19:13)	1.14E+09
		8.45E+01		1.82E+16
		3.34E+08		7.65E+09
		7.06E+10		1.45E+11

Table 8 – Mixture results from samples injected for 24 seconds, and analyzed with the SDPD 24s GlobalFiler kit collected during the original validation compared to results from v2.4.06.

<u>24 second injection Performance Check Results</u>

Because the Var>Mode default (and recommended value) was changed in the other two kits, deconvolutions also needed to be done with the 24s GlobalFiler kit and Minifiler kit. This setting prevents STRmix from underoptimizing the variance values by increasing the minimum variance within each MCMC iteration from the previous setting. In order to check the Var>Mode setting change and ensure forward stutter modeling doesn't negatively affect results, mixtures injected for 24 seconds were analyzed with the SDPD GlobalFiler 24 second STRmix kit. All samples

injected for 24 seconds were originally analyzed with STRmix v2.3.06. This was the version originally used for SDPD casework. Only a subset were run for the v2.3.07 performance check, so results from v2.3.06 are shown in Table 8.

Four 2 person mixtures, four 3 person mixtures, and two 4 person mixtures were analyzed again with v2.4.06. Database search LRs for each known contributor are shown in the table. There were many changes between these two software versions, but the contributor ratios are very similar, even in the 3 and 4 person mixtures. The genotype weights are consistent between versions. The LR values are slightly higher in version 2.4.06, and that is largely due to the last locus (D2S1338) not being analyzed in v2.3.06. The kit setting changes (particularly the Var > mode parameter) and LR calculation changes do affect the results slightly. For the most part, the conclusions about each contributor are not affected. The one exception to this is one of the contributors in Mixture 4-47. The second listed known contributor has in inconclusive LR in the original v2.3.06 as several loci indicated exclusion, and fit best with the genotypes associated with the 17% contributor. When this mixture was re-analzed with the newest v2.4.06, the contributor was strongly included, and associated with the 41% contributor, which better reflects the designed mixture.

Sample ID	Ratio v2.3.06	DB LR v2.3.06	Ratio v2.4.06	DB LR v2.4.06
Mix 2-22	(86:14)	1.09E+12	(86:14)	1.09E+12
		4.73E+10		4.69E+10
Mix 2-30	(84:16)	8.98E+11	(84:16)	8.98E+11
		2.14E+10		2.08E+10
Mix 2-37	(79:21)	1.09E+12	(79:21)	1.09E+12
		2.92E+10		3.24E+10
Mix 2-41	(90:10)	4.50E+10	(89:11)	4.39E+10
		1.20E+11		1.20E+11
Mix 3-33	(48:42:10)	1.54E+07	(48:40:12)	4.68E+07
		4.27E+05		8.48E+05
		4.95E+10		4.36E+10
Mix 3-38	(47:47:6)	1.33E+09	(51:43:6)	7.37E+08
		9.96E+04		6.83E+04
		8.45E+06		5.81E+06
Mix 3-63	(42:29:28)	1.52E+04	(53:28:19)	9.03E+03
		5.96E+05		3.50E+07
		5.08E+05		1.61E+06
Mix 3-64	(38:31:30)	1.76E+04	(49:29:22)	3.11E+04
		5.57E+07		5.55E+08
		8.77E+03		1.01E+04
Mix 4-48	(63:24:7:6)	2.47E+04	(64:22:10:4)	8.22E+03
		7.73E+02		1.51E+02
		8.11E+07		7.98E+07
		1.02E+11		1.02E+11
Mix 4-47	(43:30:16:11)	2.18E+06	(45:27:20:8)	1.48E+05
		5.99E+04		3.85E+04
		8.58E+04		5.47E+04
		3.99E+03		2.41E+03

Minifiler Performance Check Results

The same number of mixtures were run with the SDPD Minifiler kit. Similarly mixture results obtained during the original validation with v2.3.06 were compared to results from the newest version 2.4.06. Results are shown in Table 9.

Results are highly reproducible betwee the two versions with regard to contributor percentage as well as LR.

Table 9a (below) and 9b (right) – Mixture results from Minifiler samples collected during the original validation compared to results from v2.4.06. Results from single source samples shown on the right.

Sample ID	LR v2.3.07	LR v2.4.06	
SS1	5.99E+12	5.99E+12	(all weights 100%)
SS2	1.26E+13	1.26E+13	(all weights 100%)
SS3 (dropout)	1.74E+04	1.68E+04	
SS4 (dropout)	9.69E+09	9.62E+09	
SS5 (dropout)	2.67	2.68	only 1 called peak

Conclusions

Results obtained across different versions of STRmix were comparable. STRmix™ v2.4.06 deconvoluted mixtures as expected with and without the incorporation of forward stutter. Going to this method frees us from having to filter N+1 in GMID-X, so each peak in an N+1 position can be considered as a possible N+1 stutter peak during the MCMC. STRmix v2.4.06 has been tested using many samples analyzed with all three SDPD specific STRmix kits. This newest verion is deemed suitable for casework. Several things will change upon implementation, and several things will remain the same. They are outlined below:

GlobalFiler:

- Evidence samples (regardless of injection time) analyzed in GMID-X will no longer have a forward stutter filtered, which means these stutter peaks will be labeled with an allele number like N-1 stutter peaks are.
- No change to how Evidence and Reference files are exported from GMID-X.
- Minimal changes to the STRmix User interface.
- STRmix will now use 8 chains and will incorporate forward stutter modeling.
- No change to LR from Previous analysis.

Minifiler:

- GMID-X analysis does not change; forward stutter will still be edited out.
- STRmix will now use 8 chains.
- No change to LR from Previous analysis.

References

1. STRmix v2.4 Users Manual; December 1, 2015
2. STRmix v2.4.06 Release and testing report; February 2, 2017
3. STRmix v2.4 Operation Manual; February 17, 2017
4. STRmix Technical and Scientific Support notes on http://support.strmix.com

No new electropherograms were generated for this STRmix software modification/performance check. The two new evidence samples and seven proficiency samples were incorporated at the step of the STRmix input file, so all data including STRmix input files, output files and Advanced Report PDFs, and summary of results can be found on the SDPD Forensic Biology CODIS network
H:\QA-QC files\Validations\3500 and GlobalFiler and STRmix\STRmix

Melissa Strong
Criminalist

Shawn Montpetit
DNA Technical Manager