SDPD Crime Laboratory – Forensic Biology Unit

Validation of the STRmix^{TM} Software

MCMC
Markov Chain Monte Carlo

Introduction
The goal of DNA mixture interpretation should be to identify the genotypes of the contributors that comprise the mixture. DNA mixture results can often be explained by multiple possible genotype combinations. Given how many loci there are in the GlobalFiler amplification kit, the number of possible genotype combinations is prohibitively large, and deduction of the component genotypes that comprise the mixture (called a deconvolution) becomes a very complex problem. The Markov Chain Monte Carlo (MCMC) describes a standard statistical methodology that dominates modern analysis of statistical problems across disciplines. STRmix uses MCMC to approach the complex problem of DNA mixture interpretation. Below is an overview of MCMC.

The MCMC process involves thousands to millions of iterations, and three main steps in each iteration (Figure 1). *First*, based on the number of contributors input by the analyst, various genotype combinations that could possibly describe the mixture are determined. This is the prior distribution of genotype sets that could describe the data and all are all sets at each markers are systematically, randomly, and independently sampled to ensure that all combinations are considered. The set of variables describing the amount of DNA in the profile are collectively known as the mass variables, which are: DNA template amount, degradation, and amplification efficiency of each locus. DNA template amount and degradation are variables assigned to each contributor in the mixture, whereas locus specific amplification efficiency is applied to each contributor. An expected DNA profile is built using the possible genotypes combinations and mass variables. Allele-specific stutter is then applied to adjust peak heights. *Second*, a probability of the expected DNA peaks given the selected mass parameters is calculated by comparing them to the observed peaks in the data. In addition to the mass parameters and genotype sets, STRmix will also select variance values from the distributions determined from ModelMaker. This comparison of expected to observed takes into account the selected allele and stutter variance values. *Third*, the proposed set of variables are either accepted or rejected depending on whether they are a good description or a poor description of the observed DNA profile data as compared to another expected profile given a different set of mass variables and a genotype set that is different at a single locus.

The first 100,000 iterations (termed burn-in) of the MCMC are dedicated to reaching an equilibrium state where a smaller distribution of values for mass parameters and a more limited number of genotype sets are being regularly chosen in accordance with how well they describe the observed data. This prevents putting too much weight on the more random guess that occur

in the beginning. During the post-burn-in, deconvolution is the process of creating the list of genotype sets and assigning weights to each set that reflect how well they 'fit' the evidence profile. If the proposed set of genotypes from each contributor is less likely to lead to the observed evidence profile then that set will be given a low weighting (close to zero), and if the proposed genotype set is more likely to lead to the observed evidence profile then that set will be given a high weighting (close to 1).

Within STRmix, the typical use of MCMC is to ultimately provide weights for genotype sets that might explain some evidence, given the biological model used to describe DNA profile behavior. This process describes a fully continuous probabilistic genotyping approach to DNA profile interpretation.
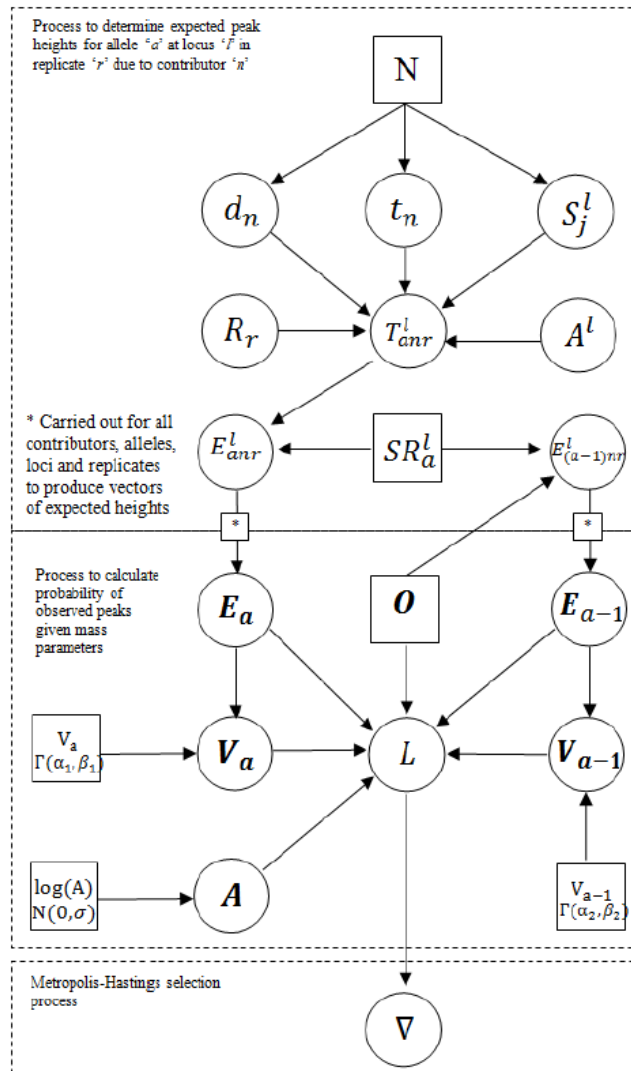

Figure 1 – A description of the MCMC process used by STRmix. Figure taken from the User's manual (1). The circles are modeled by STRmix, and the squares are input parameters by the user (See ModelMaker Validation Study for details about settings input by the user).

Starting with the process to build a profile with expected peak heights, the steps are as follows: the input file provides STRmix with detected peaks and peak heights. The first thing it does is to determine the total possible genotype combos (determined by the formula $(n+1)^{2N}$, where n = the number of alleles detected, and N = the number of contributors, input by user). It then removes duplicates, and goes through and assesses overall peak heights to determine what must be allelic and what could be stutter (based on the global stutter cutoff -a setting in STRmix - 30%), or possible drop-in. STRmix then generates a list of genotype combinations that are possible, given the input file, stutter and drop-in considerations, and the number of contributors input by the analyst. This list is considerably smaller than what it started with, given these constraints.

A set of genotypes from the list of possible genotypes is randomly assigned. Test values for the amount of DNA (t_n), degradation (d_n), and locus specific amplification efficiencies (A^l) are applied. These are known as the mass parameters. Replicates would also be taken into account at this step, but they are not being assessed as part of the internal validation in this lab. STRmix determines the total allelic product (TAP) by applying the test values for these mass parameters.

The mass parameters
The total allelic product (TAP) at a locus is equal to the locus specific amplification efficiency multiplied by the template amount multiplied by the allele count multiplied by the degradation equation. This is done for every contributor, for every allele detected, at every locus in the sample. The following is the equation describing the TAP.

$$T^l_{an} = A^l_r t_n X^l_{an} \times e^{d_n \times \left(mwt^l_a - offset \right)}$$

A^l_r = Locus offset, or locus specific amplification efficiencies (LSAE); r = replicate factor (=1 when a single PCR reaction).

t_n = template for DNA contributor n

X^l_{an} = count of allele 'a' at locus 'l' in contributor 'n' (2 for homozygotes and 1 for heterozygotes)

$e^{d_n \times \left(mwt^l_a - offset \right)}$ is the exponential formula that incorporates the base size of the allele.
- d_n = degradation slope for contributor n
- mwt = molecular weight (nucleotide length)
- $offset$ = smallest size of a detected peak in the electropherogram

Degradation is dependent on fragment size, so as size increases the amount of degradation increases. The degradation reported in the output file is a linear approximation of the exponential curve that is the true degradation factor.

Each of these mass parameters are selected from sliding windows of possible values, and they are varied at each iteration.

Stutter

The total allelic product is now calculated. Stutter is then taken into account. Some of the total allelic product expected from DNA amount and degradation will become stutter. STRmix does this by apportioning total allelic product into allelic and stutter height, using the following equations:

$$\text{Allele} \qquad\qquad \text{Stutter}$$

$$E_{an}^{\ell} = \frac{T_{an}^{\ell}}{1 + SR_a^{\ell}} \qquad\qquad E_{(a-1)n}^{\ell} = \frac{SR_a^{\ell}\left(T_{an}^{\ell}\right)}{1 + SR_a^{\ell}}$$

To get the allele and locus specific stutter, STRmix first looks to the Stutter Exceptions file. If STRmix sees a non-zero value for stutter ratio for any allele at a locus it will use that as the stutter ratio expectation. If, instead, a zero is encountered, it will look to the stutter file and use the stutter ratio value obtained from the regression line for that locus and the observed allele (see STRmix ModelMaker study for more details about the composition of these two files).

STRmix then calculates the probability of obtaining the expected profile, given the genotype set and mass parameters if the values above were true.

Metropolis Hastings - probability of an expected profile

If the randomly chosen mass parameters are correct, then the any differences between the calculated expected (based on the proposed mass parameters) and our observed (the data from the 3500) are only based on random PCR/injection variation (essentially stochastic or sampling variation). We have estimates of how much variation we can expect from ModelMaker: allele and stutter variance parameters (c^2 and k^2), and LSAE variance. STRmix now compares this expected profile to the observed profile and gives it a probability. Through modelling, it is known that the log[observed /expected] has a normal distribution with a mean of 0 and a variance that is inversely proportional to the expected peak height. There are three assumptions in this calculation: 1) an approximate normal distribution with a mean of zero, 2) a variance of c^2/E^l_{an} for the allele model, 3) a variance of k^2/E^l_{an} for the stutter model:

$$\log\left(\frac{O_{(a-1)}}{E_{(a-1)n}^l}\right) \sim N\left(0, \frac{k^2}{E_{an}^l}\right) \text{ for stutter}$$

$$\log\left(\frac{O_a}{E_{an}^l}\right) \sim N\left(0, \frac{c^2}{E_{an}^l}\right) \text{ for alleles}$$

What these equations allow is to determine the likelihood for the observed locus given the parameters chosen by the MCMC.

$$\Pr(O_a^l | E_a^l) \sim N\left(\log\left(\frac{O_a^l}{E_a^l}\right), 0, \frac{c^2}{E_a^l}\right)$$ (or k^2 instead of c^2)

It is important to include E_a in the variance, because we know from the GlobalFiler validation that the variance is dependent on peak height. For example, the peak height balance is more variable for lower peak heights than for taller ones. The same is true with stutter. Dividing by E_a means that we expect less spread (variation) for high RFU alleles (i.e., more template) and more spread for lower RFU alleles. The probability of the observed peak given the expected peaks (mass variables) is the "probability" from the normal distribution, centred on zero. Really these probabilities are determined from the probability density function (curve). This can be done in excel with the NORM.DIST(test value, average value, 0, standard deviation) function.

These probabilities are calculated for each allele. For the locus 'l' we can calculate the likelihood by multiplying each individual allele likelihoods of the 'a' alleles. For the entire profile we can calculate the likelihood by multiplying the likelihood of each of the 'l' loci:

$$\Pr(profile \mid M) = \Pr(locus1 \mid M) \times \Pr(locus2 \mid M)$$

$$= \prod_l \prod_a N\left(\log\left(\frac{O_a^l}{E_a^l}\right); 0, \frac{k}{E_a^l}\right)$$

(in the equation above, k is replaced with c^2 or k^2)

Ultimately, STRmix uses the log of the probability for each allele. This can be calculated in two ways: take the log(probability) for each peak and then sum them, or multiply the probabilities across loci and take a log of the product.

Before these probabilities are used for determining whether to accept or reject the current guess, there are some penalties that are applied. The MCMC will not allow the parameters of allele, stutter, and LSAE variance stray too far off from reasonable values. Penalties are built in when random variables are unlikely. A drop-in penalty also occurs if one of the peaks in the profile is considered to be drop-in in that iteration. Drop-in is considered a rare event and any combinations that require drop-in for the combinations to occur will be penalized.

LSAE penalties: during the MCMC iterations, LSAE is selected from a distribution. This normal distribution is centered on 0 with a variance based on the ModelMaker value of LSAE variance. An LSAE penalty is added to the $\sim N(\log(E), 0, c^2)$ for departures that are far off from 1. The $\Pr(LSAE) = P_{LSAE} = N(LSAE, 0, \sqrt{LSAE\ Variance})$; $\log(P_{LSAE})$ is added to the $\log(P_{peaks})$ for the alleles and the stutter.

Allele and stutter variance penalties: The (stutter and allele) variance is randomly selected from the gamma distribution during each iteration. The gamma distribution is determined by α and β and determined during Model Maker. Any extreme values selected for the variance are

(deviations away from the mode) are penalized. Since variance is modeled by a gamma distribution, deviations that are too far off from the mode (maximum point) are penalized. Penalty = ~Γ(iteration variance,α,β).

Drop-in penalties: The drop-in penalty is based on the probability density function of a particular peak being drop-in.

Metropolis Hastings – accept or reject

Metropolis–Hastings algorithm is a Markov chain Monte Carlo method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. The key is that the accept/reject criterion gives a sample from the desired probability in the long run as long as the accept/reject is proportional to the true probabilities. STRmix is comparing the modelled (expected) profile to the observed profile, and determining if the model built in the current iteration is better than the model it built in the previous iteration. The goodness of fit is the heart of the acceptance/rejection step. This is something that can be assigned a numerical value and can be calculated. STRmix treats each observed and expected peak height comparison independently of the others. It combines Prob (Observed | Expected) for each peak for all loci with a few penalties along the way.

If the proposed model is better than the current model then STRmix accepts it, and if the proposed model is worse than the current model then STRmix accepts it only a certain percentage of the time. This percentage is determined by the probability of the proposed divided by the probability of the current guess. In other words: it is a comparison of where the model has been compared to the new model.

> M-H*new* / M-H*old* $\geq$ 1 = take the step
> M-H*new* / M-H*old* < 1 = take the step in proportion to the ratio
> Any proportion >0.5 means that it is more likely going to take that next step.

MCMC iterations

For each step of the MCMC chain, the mass parameters and a genotype set that differs at one locus are independently chosen (component-wise MCMC). The MCMC is set of algorithms that act like a calculator for solving very complex equations (those that would take too long to solve using standard methods). Eventually the MCMC will reach equilibrium where: 1) DNA amount, degradation, and locus specific amplification efficiency are stable; and 2) Limited number genotypes are chosen in proportion to their probability. In STRmix the MCMC is 'solving' the equation for genotype weights.

MCMC weightings

There are hundreds of thousands to billions of iterations before reaching the required number of MCMC accepts (500,000 total accepts; 400,000 post burn-in). During that time STRmix may spend multiple iterations on the same guess before moving to a better guess. The amount of iterations STRmix spends on one guess will be proportional to how good a guess it is. STRmix

turns this proportion into the weight of that guess. There is some variability associated with the MCMC process, and this can be assessed. Each time a sample is run, STRmix gives a different weighting. Run over and over, these different answers all cluster around each other and the amount that they would vary is small in relation to the magnitude of the answer.

Burn-in

The MCMC starting point is random and therefore the MCMC chains will (likely) start in a very bad sample space (bad guess). It takes some time for them to reach good sample space. Tallying how much time the chain has spent on genotypes isn't done until 100,000 accepts have been reached. This "burn-in" period allows the chain to reach better samples space and keeps the bad genotype combinations from being overrepresented.

Dropout

Dropout is defined as the absence of the observation of a peak above an analytical threshold where one is expected. Dropout can be considered an extreme form of imbalance. Dropout is a possibility if one or more of the contributors are providing low levels of DNA to the amplification reaction. Dropout is designated as a Q allele within STRmix. There are few important things to note about Q alleles as they are encountered in the MCMC process. Q alleles within and between donors never sum their RFUs. For the purpose of profile modeling, Q alleles are always treated as a new allele unlike all others, even though the size will overlap. Also, Prob(O|E) calculated separately for every Q (drop-out) allele.

Diagnostic Tools

In the summary output of STRmix, there are numerous diagnostics that *may* indicate that a deconvolution has not converged on the best sample space. Any of these, on their own do not indicate a problem with the deconvolution, but can be helpful in identifying aspect of the sample to go back and double check.

Gelman-Rubin convergence diagnostic: informs the user whether the MCMC analysis has likely converged. STRmix uses multiple 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 diagnostic (GR), is a ratio of the stationary distribution and within-chain variances. For a converged analysis the GR will be 1. It has been recommended that if the GR is above 1.2 then there exists the possibility that the analysis hasn't converged. We would suggest that if the GR value is above 1.2 the results of the analysis be closely scrutinized. Running the analysis for a larger number of iterations will likely reduce the GR in these instances to below 1.2.

Effective sample size (ESS): the number of independent samples the MCMC has taken from the posterior distribution of all parameters. 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. An ESS of NaN indicates that there might be a problem with the input data. For example, a stutter peak it expects to see is not present in the input data).

Average log(likelihood): this value shows the average log10(likelihood) for the entire post burn-in MCMC. 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.

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 the variance constant has increased markedly from the mode of the prior distribution, then this may indicate that the DNA profile is sub-optimal or that the number of contributors is incorrect. Used in conjunction with the average log10(likelihood), a large variance or stutter variance constant can indicate poor PCR.

In summary, STRmix creates a list of genotype sets and assigns weights to each set that reflect how well they 'fit' the evidence profile. If the proposed set of single sourced genotypes is unlikely to lead to the observed evidence profile then that set will be given a low weighting (close to zero), and if the proposed genotype set is likely to lead to the observed evidence profile then that set will be given a high weighting (close to 1). STRmix uses information provided by the user combined with optimized values for properties of the DNA profile being analyzed to deconvolute a profile and calculate weights.

Purpose
Knowing that STRmix is a fully continuous probabilistic genotyping approach that incorporates the biological model, the purpose of this study was to assess mixture deconvolution by the MCMC process. Two different approaches were taken to assess the MCMC.

The first approach utilized one single source sample to examine the results in detail to the level that they can be reproduced. The extended output provides the iteration-by-iteration detail. In this extended output each of these results were examined: the genotypes and LSAE at each locus, template and degradation for each contributor, locus amp probability, allele variance, allele variance penalty, stutter variance, and stutter variance penalty.

The second approach utilized samples with DNA from more than one person. For mixtures, the most straightforward way to do this was to use mixtures designed and created in the lab ("ground

truth" mixtures), and compare the STRmix results to known genotype sets of the ground truth mixtures.

The developmental validation studies for this software included (but not limited to) extensive evaluation of:
- Expected allele and stutter heights given mass parameters
- Expected peak heights of drop-out alleles given mass parameters
- Probabilities given expected and observed peak heights and varying analytical thresholds
- Locus specific amplification efficiency calculations
- Summation of probabilities for each allele in a locus and across a profile
- Summation of probabilities across multiple replicate profiles
- Informed priors on mixture proportion
- LR values where there are no assumed contributors
- LR values with varying theta values
- LR values for propositions with assumed contributors
- LR HPD interval values
- Sampling from the Beta distributions for theta
- Gaussian walk
- Gelman-Rubin statistic, ESS, weight resampling
- Drop-in function
- Model maker

These are described in the manual and in multiple peer reviewed publications. The assessment of the MCMC in *this* study is aimed at validating STRmix (v2.3.06) for mixture interpretation at the SDPD. For the purposes of our laboratory, the MCMC process was assessed by evaluating the genotype weights determined by STRmix deconvolution. This study included a very wide range of mixture combinations and template amount so as to assess the MCMC in a variety of contexts (i.e. in the presence of dropout, in balanced mixtures, and in both high and low template samples).

Materials and Methods
Two, three, four, and five-person mixtures (a total of 186 mixtures) were created as part of the GlobalFiler Mixture Study. These were mixtures designed for STRmix that had a range of contributor compositions – from balanced mixtures to mixtures where there are one or two contributors that are the source of most of the DNA in the mixture. There are also mixtures in every set that have at least one contributor dropping out. These mixtures were created and amplified with the original GlobalFiler master mix formulation. After these results were assessed in STRmix, a reformulation of GlobalFiler was released. A subset of the mixtures were selected for re-amplification with the new GlobalFiler product. See the ModelMaker Study for more details about the new Allele, Stutter and LSAE variance parameters that were collected from the

reformulated GlobalFiler master mix. After determining that new STRmix models were necessary post-reformulation, a larger subset of the mixtures below were re-deconvoluted with the new STRmix settings in-order to validate both the new amplification kit, and the new settings for STRmix. See Table 1 for a list of all mixtures.

Table 1 – A list of mixtures broken down but input level and contributor ratio

Two Person Mixtures			
Ratio	Sample ID Target High level (3K-10K RFU)	Sample ID Target Mid level (1K-3K RFU)	Sample ID Target Low level (>1.5K RFU)
8:1	2-1	2-15	2-29
5:1	2-2	2-16	2-30
2:1	2-3	2-17	2-31
1:1	2-4	2-18	2-32
1:2	2-5	2-19	2-33
1:5	2-6	2-20	2-34
1:8	2-7	2-21	2-35
8:1	2-8	2-22	2-36
5:1	2-9	2-23	2-37
2:1	2-10	2-24	2-38
1:1	2-11	2-25	2-39
1:2	2-12	2-26	2-40
1:5	2-13	2-27	2-41
1:8	2-14	2-28	2-42

Three Person Mixtures			
Ratio (% contribution)	Sample ID Target High level (3K-10K RFU)	Sample ID Target Mid level (1K-3K RFU)	Sample ID Target Low level (>1.5K RFU)
33.3:33.3:33.3	3-1	3-23	3-45
	3-2	3-24	3-46
70:20:10	3-3	3-25	3-47
	3-4	3-26	3-48
60:30:10	3-5	3-27	3-49
	3-6	3-28	3-50
50:40:10	3-7	3-29	3-51
	3-8	3-30	3-52
50:30:20	3-9	3-31	3-53
	3-10	3-32	3-54
45:45:10	3-11	3-33	3-55
	3-12	3-34	3-56
40:40:20	3-13	3-35	3-57
	3-14	3-36	3-58
35:35:30	3-15	3-37	3-59
	3-16	3-38	3-60
60:20:20	3-17	3-39	3-61
	3-18	3-40	3-62
50:25:25	3-19	3-41	3-63
	3-20	3-42	3-64
40:30:30	3-21	3-43	3-65
	3-22	3-44	3-66

Four Person Mixtures			
Ratio (% contribution)	Sample ID Target High level (3K-10K RFU)	Sample ID Target Mid level (1K-3K RFU)	Sample ID Target Low level (>1.5K RFU)
25:25:25:25	4-1	4-23	4-45
	4-2	4-24	4-46
60:20:10:10	4-3	4-25	4-47
	4-4	4-26	4-48
50:20:20:10	4-5	4-27	4-49
	4-6	4-28	4-50
70:10:10:10	4-7	4-29	4-51
	4-8	4-30	4-52
40:20:20:20	4-9	4-31	4-53
	4-10	4-32	4-54
40:40:15:5	4-11	4-33	4-55
	4-12	4-34	4-56
35:35:20:10	4-13	4-35	4-57
	4-14	4-36	4-58
40:40:10:10	4-15	4-37	4-59
	4-16	4-38	4-60
35:35:25:5	4-17	4-39	4-61
	4-18	4-40	4-62
30:30:20:20	4-19	4-41	4-63
	4-20	4-42	4-64
30:30:30:10	4-21	4-43	4-65
	4-22	4-44	4-66

Five Person Mixtures			
Ratio (% contribution)	Sample ID Target High level (3K-10K RFU)	Sample ID Target Mid level (1K-3K RFU)	Sample ID Target Low level (>1.5K RFU)
20:20:20:20:20	5-1	5-5	5-9
	5-2	5-6	5-10
60:10:10:10:10	5-3	5-7	5-11
	5-4	5-8	5-12

All of these mixtures were evaluated assuming the number of contributors the mixture was designed to have. All of the two and three person mixtures were evaluated extensively. A subset of the 4 person mixtures were chosen for evaluation of deconvolution but these were studied more extensively in the STRmix Comparison to Known Contributors study. The five person mixtures haven't been run due to a limit in java, unless conditioned on one of the balanced contributors. The mixtures were assessed for the percent contribution of each contributor, whether the correct genotypes included in the genotype probability distribution, whether correct combination was in the top 99%, and whether the STRmix genotype possibilities were intuitive.

Target percentage of contribution for each mixture, and manual calculation of contributor percentage (see GlobalFiler mixture study for more details) was compared against STRmix calculations for percent contributor.

Results and Discussion

Single Source MCMC reproducibility

The *locus amp probability* was reproduced by calculating the normal distribution of the log of the LSAE for each locus in that iteration (centered around zero with a standard deviation of the square root of the LSAE variance determined by ModelMaker; equation = NORMDIST(LOG(LSAE for locus X),0,SQRT(0.01556),0)), and then taking the log of the product of those normal distributions for every locus. In the single source sample, the locus amplification probability for three different iterations was recorded from the STRmix extended output. For iteration 0, it was 10.598; for iteration 28,716 it was 8.351; for iteration 78,292, it was 8.756. Each of these values was reproduced in excel to at least the 9th decimal place using the above formula.

The *allele variance penalty* was reproduced by taking the log of the gamma distribution of the allele variance for that iteration using the α and β parameters (determined by ModelMaker; 6.6346 and 1.6553). Equation =LOG(GAMMADIST(allele variance, 6.6346, 1.6553, 0)). In the single source sample, the allele variance penalty for three different iterations was recorded from the STRmix extended output. For iteration 0, it was -0.9998; for iteration 28,716 it was -0.9998; for iteration 78,292, it was -1.0369. Each of these values was reproduced in excel to at least the 9th decimal place using the above formula.

The *stutter variance penalty* was reproduced by taking the log of the gamma distribution of the stutter variance for that iteration using the α and β parameters (determined by ModelMaker; 7.09 and 2.4927). Equation =LOG(GAMMADIST(stutter variance, 7.09, 2.4927, 0)). In the single source sample, the stutter variance penalty for three different iterations was recorded from the STRmix extended output. For iteration 0, it was -1.194; for iteration 28,716 it was -1.5835; for iteration 78,292, it was -2.5416. Each of these values was reproduced in excel to at least the 9th decimal place using the above formula.

MCMC performance on mixtures

Results for every two and three person mixture were carefully scrutinized. A subset of the 4 person mixtures were run, and results from these mixtures are described more fully in the STRmix Comparison to Known Contributors Study.

For this study, 42 two-person mixtures were deconvoluted with the number of contributors set at 2. Originally, these mixtures were run with the Allele, Stutter, and LSAE variance parameters determined from the original GlobalFiler master mix. Subsequently, a GlobalFiler reformulation

was released and a small subset of these were re-amplified. There was essentially no change in the way that mixtures were amplified with the reformulated GlobalFiler master mix. However, new Allele, Stutter, and LSAE variance parameters were determined after the master mix reformulation. A larger subset of these mixtures were re-deconvoluted with the new STRmix model settings. The results in Table 2 summarize the most updated results of the deconvolution.

2 person mixtures:
Of the 42 two-person mixtures, 8 of them had alleles from at least one contributor dropping out. Only one of these mixtures (2-29) had a diagnostic value that warranted a closer look. The Gelman-Rubin Convergence number was 1.31. This mixture had three instances of allelic dropout, but full assessment of this sample did not indicate any other problem. Each known contributor's genotype fell into the top 99% of weights in the Component Interpretation section, and all genotypes and weights were intuitive for both contributors.

Only one of the mixtures (2-39) in which one of the contributors genotypes at one locus was not in the top 99%. This is a low-level balanced mixture where each contributor was contributing ~50%. Upon further inspection of the genotypes, one of the contributors in this mixture had types that completely dropped out at D5S818, and full dropout of that genotype was considered, but only with a weighting of 0.40%, which did not make the top 99% cutoff that was investigated in this study.

All other mixtures, even low level, balanced and imbalanced mixtures were deconvoluted by STRmix in a way that was intuitive and genotypes from the known contributors fell in the top 99% of weights.

Table 2 – Two person mixture deconvolution results – high level target input amount

People	Mixture #	Gelman-Rubin Convergence	Effective Sample Size	Amount of Dropout	Reference IDs	Target % contribution	Manual % contribution	STRmix % contribution	Correct combination in 99%?	STRmix Genotype possibilities intuitive?
2	1	1.02	10914		150-TD	89	86	89	Yes	Yes
					156-RK	11	14	11	Yes	Yes
2	2	1.04	9672		150-TD	83	81	84	Yes	Yes
					156-RK	17	19	16	Yes	Yes
2	3	1.01	7353		150-TD	67	68	68	Yes	Yes
					156-RK	33	32	32	Yes	Yes
2	4	1.05	24250		150-TD	50	51	50	Yes	Yes
					156-RK	50	49	50	Yes	Yes
2	5	1.01	9071		150-TD	33	31	30	Yes	Yes
					156-RK	67	69	70	Yes	Yes
2	6	1.01	11052		150-TD	17	19	18	Yes	Yes
					156-RK	83	81	82	Yes	Yes
2	7	1.05	12928		150-TD	17	12	11	Yes	Yes
					156-RK	83	88	89	Yes	Yes
2	8	1.02	23326		62-BS	89	85	84	Yes	Yes
					31-BB	11	15	16	Yes	Yes
2	9	1.01	1445		62-BS	83	82	79	Yes	Yes
					31-BB	17	18	21	Yes	Yes
2	10	1.03	4866		62-BS	67	63	57	Yes	Yes
					31-BB	33	36	43	Yes	Yes
2	11	1.03	5867		62-BS	50	46	36	Yes	Yes
					31-BB	50	54	64	Yes	Yes
2	12	1.03	3830		62-BS	33	30	23	Yes	Yes
					31-BB	67	70	77	Yes	Yes
2	13	1.01	8052		62-BS	17	16	10	Yes	Yes
					31-BB	83	84	90	Yes	Yes
2	14	1.09	12488		62-BS	11	11	8	Yes	Yes
					31-BB	89	89	92	Yes	Yes

Table 2 – Two person mixture deconvolution results, continued – mid level target input amount

People	Mixture #	Gelman–Rubin Convergence	Effective Sample Size	Amount of Dropout	Reference IDs	Target % contribution	Manual % contribution	STRmix % contribution	Correct combination in 99%?	STRmix Genotype possibilities intuitive?
2	15	1.01	9522		150-TD	89	88	90	Yes	Yes
					156-RK	11	12	10	Yes	Yes
2	16	1.03	7261		150-TD	83	84	86	Yes	Yes
					156-RK	17	16	14	Yes	Yes
2	17	1.01	6982		150-TD	67	75	77	Yes	Yes
					156-RK	33	25	23	Yes	Yes
2	18	1.03	19222		150-TD	50	51	50	Yes	Yes
					156-RK	50	49	50	Yes	Yes
2	19	1.01	9812		150-TD	33	27	23	Yes	Yes
					156-RK	67	73	77	Yes	Yes
2	20	1.01	13582		150-TD	17	16	13	Yes	Yes
					156-RK	83	84	87	Yes	Yes
2	21	1.04	8261		150-TD	11	12	11	Yes	Yes
					156-RK	89	88	89	Yes	Yes
2	22	1.05	18054	1	62-BS	89	86	83	Yes	Yes
					31-BB	11	14	17	Yes	Yes
2	23	1.02	10016		62-BS	83	80	77	Yes	Yes
					31-BB	17	20	23	Yes	Yes
2	24	1.01	11771		62-BS	67	71	67	Yes	Yes
					31-BB	33	29	33	Yes	Yes
2	25	1.01	15689		62-BS	50	45	38	Yes	Yes
					31-BB	50	55	62	Yes	Yes
2	26	1.02	9433		62-BS	33	23	16	Yes	Yes
					31-BB	67	77	84	Yes	Yes
2	27	1.06	10079		62-BS	17	14	10	Yes	Yes
					31-BB	83	86	90	Yes	Yes
2	28	1.09	21662	1	62-BS	11	11	6	Yes	Yes
					31-BB	89	89	94	Yes	Yes

Table 2 – Two person mixture deconvolution results, continued – low level target input amount

People	Mixture #	Gelman–Rubin Convergence	Effective Sample Size	Amount of Dropout	Reference IDs	Target % contribution	Manual % contribution	STRmix % contribution	Correct combination in 99%?	STRmix Genotype possibilities intuitive?
2	29	1.31	4823	3	150-TD	89	85	89	Yes	Yes
					156-RK	11	15	11	Yes	Yes
2	30	1.01	9447		150-TD	83	80	84	Yes	Yes
					156-RK	17	20	16	Yes	Yes
2	31	1.03	9794		150-TD	67	67	67	Yes	Yes
					156-RK	33	33	33	Yes	Yes
2	32	1.08	8543	1	150-TD	50	51	51	Yes	Yes
					156-RK	50	49	49	Yes	Yes
2	33	1.12	1835		150-TD	33	36	46	Yes	Yes
					156-RK	67	64	54	Yes	Yes
2	34	1.01	10190		150-TD	17	18	17	Yes	Yes
					156-RK	83	82	83	Yes	Yes
2	35	1.02	2381		150-TD	17	13	10	Yes	Yes
					156-RK	83	87	90	Yes	Yes
2	36	1.06	24371		62-BS	89	85	87	Yes	Yes
					31-BB	11	15	13	Yes	Yes
2	37	1.02	8052	1	62-BS	83	80	78	Yes	Yes
					31-BB	17	20	22	Yes	Yes
2	38	1.04	17822		62-BS	67	64	56	Yes	Yes
					31-BB	33	36	44	Yes	Yes
2	39	1.01	2954	3	62-BS	50	47	51	No	Yes
					31-BB	50	53	49	Yes	Yes
2	40	1.05	5974		62-BS	33	35	28	Yes	Yes
					31-BB	67	66	72	Yes	Yes
2	41	1.08	7284	1	62-BS	17	16	14	Yes	Yes
					31-BB	83	84	86	Yes	Yes
2	42	1	4376	3	62-BS	11	10	8	Yes	Yes
					31-BB	89	90	92	Yes	Yes

3 person mixtures:
Of the 67 three-person mixtures, 39 of them had alleles from at least one contributor dropping out. All of the low level mixtures had alleles dropping out, and in one mixture, all of the alleles from one of the contributors dropped out. This mixture set allowed us to test a wide range of scenarios (Table 3).

The mixture ratios provided by STRmix were a lot more accurate than the manual calculation, and reflected the target values in most mixtures (see Table 3). This is likely because STRmix is able to use all the loci for this estimate, while for the manual calculation, we were limited only to

a few loci where the genotypes had sufficient separation, and stutter was not taken into account for the manual estimate of percent contribution from one contributor.

Seventeen of the 67 mixtures had a diagnostic value that warranted a closer look. Of these, all had Gelman-Rubin convergence numbers greater than 1.2. In looking closer at these 17 mixtures, 11 of them had no problem with the deconvolution into the known contributors. Many of these mixtures had two or more contributors that were very balanced, and that ambiguity can cause an increase in the GR number. The other 6 with a GR number higher than 1.2 did have one or more contributors whose known genotype at one locus did not make the list of the top 99%. Four of these were 14% contributors (or less) with some dropout associated with them. There simply wasn't enough data to deconvolute that minor contributor. The other two contributors in each of these mixtures were deconvoluted with no problems. The 5th of these had two balanced 10% contributors where dropout was not adequately accounted for at one locus, and the final of these 6 mixtures with a GR number above 1.2 was a 3 person, low level, balanced mixture with dropout where none of the contributors were accurately deconvoluted at one locus (due to dropout not being accounted for highly enough.

Fourteen of the 67 mixtures had only one of three contributors with a known genotype weight of less than 99% (from the Component Interpretation section of the STRmix results). Ten of these 14 contributors were contributing only 15% or less to the mixture, and there was dropout in all but one of these. The other 4 were balanced contributors in mixtures with dropout (in one of these 4, all of that contributors genotypes were dropped from the low level mixture, so it was not surprising that their genotype was not deconvoluted with sufficient weight).

Three of the 67 mixtures had two of the three contributors that had a genotype at only one locus not falling into the top 99% of weights. In two mixtures, both contributors were estimated to provide only 10% & 11%, and 8% and 8%, respectively. The third was a very low level balanced three person mixture with dropout.

Two of the 67 mixtures had problems with the genotypes of all three contributors. One of these (3-48) had a deconvolution at TH01 that was not intuitive for the major contributor (70%), and this resulted in the two known minor genotypes not falling in the top 99%. The other mixture (3-64) was a low level, balanced mixture in which dropout was not being sufficiently accounted for at only one locus.

The results presented above are consistent with the idea that the less someone's DNA is contributing to a mixture, the more unreliable the genotype weights are. That said, there were 51 contributors that were contributing 20% or less from a total of 201 different contributors making up the three person mixtures. Only 16 of these 51 ≤ 20% contributors had genotypes that did not fall into the top 99%. Likewise, 22 of these 201 contributors only contributed ≤ 10% to the

mixture and only 8 of them had a problem with a genotypes not falling into the top 99%. So, just because someone is only contributing a small fraction of total DNA to the mixture doesn't make their deconvolution results unusable, they should just be interpreted with more caution than more robust contributors.

Table 3 – Three person mixture deconvolution results – high level target input amount

People	Mixture #	Gelman-Rubin Convergence	Effective Sample Size	Amount of Dropout	Reference IDs	Target % contribution	Manual % contribution	STRmix % contribution	Correct combination in 99%?	STRmix Genotype possibilities intuitive?
3	1	1.14	6480		131-SB	33	34	36	Yes	Yes
					142-VM	33	30	33	Yes	Yes
					40-DB	33	35	32	Yes	Yes
3	2	1.04	10089		79-PV	33	31	30	Yes	Yes
					71-LW	33	27	29	Yes	Yes
					127-DF	33	42	41	Yes	Yes
3	3	1.47	11593		125-CN	70	74	80	Yes	Yes
					106-BL	20	11	10	Yes	Yes
					155-JK	10	11	10	Yes	Yes
3	4	1.49	5399	3	34-WD	70	71	68	Yes	Yes
					28-GL	20	20	17	Yes	Yes
					35-JD	10	10	15	No	Yes
3	5	1.03	16187	8	39-DC	30	22	23	Yes	Yes
					151-TR	60	76	73	Yes	Yes
					81-AR	10	3	3	Yes	Yes
3	6	1.37	10378	4	41-AD	60	67	73	Yes	Yes
					49-JB	30	24	20	Yes	Yes
					47-ER	10	8	6	Yes	Yes
3	7	1.07	23374		71-LW	50	40	43	Yes	Yes
					93-RC	40	52	44	Yes	Yes
					81-AR	10	10	12	Yes	Yes
3	8	1.04	12996		56-TB	50	44	45	Yes	Yes
					39-DC	40	44	42	Yes	Yes
					81-AR	10	13	13	Yes	Yes
3	9	1.1	2691		23-JS	50	48	47	Yes	Yes
					39-DC	30	35	38	Yes	Yes
					79-PV	20	14	15	Yes	Yes
3	10	1.02	30188		30-SS	50	55	51	Yes	Yes
					52-KK	30	24	24	Yes	Yes
					23-JS	20	24	25	Yes	Yes
3	11	1.6	10755	1	30-SS	45	53	60	Yes	Yes
					161-GZ	45	32	30	Yes	Yes
					88-KB	10	13	10	Yes	Yes
3	12	1.04	18313	1	150-TD	45	52	46	Yes	Yes
					127-DF	45	39	46	Yes	Yes
					155-JK	10	8	8	Yes	Yes
3	13	1.03	15959		61-KM	40	54	41	Yes	Yes
					102-EB	40	16	15	Yes	Yes
					156-RK	20	28	44	Yes	Yes
3	14	1.05	12793		51-HH	40	38	37	Yes	Yes
					62-BS	40	44	37	Yes	Yes
					94-RL	20	16	26	Yes	Yes
3	15	1.07	18439		79-PV	35	29	30	Yes	Yes
					159-MS	35	30	30	Yes	Yes
					156-RK	30	40	40	Yes	Yes
3	16	1.07	5727		39-DC	35	32	32	Yes	Yes
					93-RC	35	42	42	Yes	Yes
					142-VM	30	27	26	Yes	Yes
3	17	1.04	10139	4	127-DF	60	76	81	Yes	Yes
					125-CN	20	16	13	Yes	Yes
					103-KD	20	7	6	No	Yes
3	18	1.08	54318	6	131-SB	60	80	81	Yes	Yes
					106-BL	20	6	6	Yes	Yes
					152-LM	20	16	12	No	Yes
3	19	1.01	13385		125-CN	50	72	77	Yes	Yes
					115-RM	25	11	11	Yes	Yes
					79-PV	25	17	11	Yes	Yes
3	20	1.12	20513		61-KM	50	46	40	Yes	Yes
					30-SS	25	17	19	Yes	Yes
					69-CL	25	36	40	Yes	Yes
3	21	1.02	14090		71-LW	40	32	42	Yes	Yes
					115-RM	30	27	29	Yes	Yes
					39-DC	30	36	29	Yes	Yes
3	22	1.06	8040		150-TD	40	51	52	Yes	Yes
					74-LM	30	29	27	Yes	Yes
					132-JF	30	17	20	Yes	Yes

Table 3 – Three person mixture deconvolution results, continued– mid level target input amount

People	Mixture #	Gelman-Rubin Convergence	Effective Sample Size	Amount of Dropout	Reference IDs	Target % contribution	Manual % contribution	STRmix % contribution	Correct combination in 99%?	STRmix Genotype possibilities intuitive?
3	23	2.14	12750		131-SB	33	36	40	Yes	Yes
					142-VM	33	27	30	Yes	Yes
					40-DB	33	34	30	Yes	Yes
3	24	1.18	11706		79-PV	33	25	34	Yes	Yes
					71-LW	33	33	33	Yes	Yes
					127-DF	33	41	33	Yes	Yes
3	25	1.01	3813	6	125-CN	70	75	77	Yes	Yes
					106-BL	20	15	11	No	Yes
					155-JK	10	12	12	Yes	Yes
3	26	1.19	10744	5	34-WD	70	70	67	Yes	Yes
					28-GL	20	22	18	Yes	Yes
					35-JD	10	11	15	Yes	Yes
3	27	1.07	15108	4	39-DC	60	20	21	Yes	Yes
					151-TR	30	76	74	Yes	Yes
					81-AR	10	4	5	No	Yes
3	28	1.08	5388	3	41-AD	60	68	70	Yes	Yes
					49-JB	30	26	21	Yes	Yes
					47-ER	10	10	9	No	Yes
3	29	1.03	4926		71-LW	50	44	45	Yes	Yes
					93-RC	40	47	45	Yes	Yes
					81-AR	10	11	10	Yes	Yes
3	30	1.03	22711	3	56-TB	50	49	43	Yes	Yes
					39-DC	40	37	42	Yes	Yes
					81-AR	10	14	15	No	Yes
3	31	1.09	3949	1	23-JS	50	47	43	Yes	Yes
					39-DC	30	37	35	No	Yes
					79-PV	20	16	22	Yes	Yes
3	32	1.02	8018		30-SS	50	46	47	Yes	Yes
					52-KK	30	29	27	Yes	Yes
					23-JS	20	26	26	Yes	Yes
3	33	1.03	8525	4	30-SS	45	41	44	Yes	Yes
					161-GZ	45	46	47	Yes	Yes
					88-KB	10	13	9	Yes	Yes
3	34	1.08	7131		150-TD	45	52	48	Yes	Yes
					127-DF	45	36	44	Yes	Yes
					155-JK	10	9	8	Yes	Yes
3	35	1.4	10889		61-KM	40	48	44	Yes	Yes
					102-EB	40	13	14	No	Yes
					156-RK	20	41	42	Yes	Yes
3	36	1.78	23236		51-HH	40	44	38	Yes	Yes
					62-BS	40	38	32	Yes	Yes
					94-RL	20	17	30	Yes	Yes
3	37	1.02	24905		79-PV	35	33	33	Yes	Yes
					159-MS	35	37	34	Yes	Yes
					156-RK	30	34	33	Yes	Yes
3	38	0	5915	3	39-DC	35	37	46	Yes	Yes
					93-RC	35	50	47	Yes	Yes
					142-VM	30	10	7	Yes	Yes
3	39	1.39	31232		127-DF	60	66	69	Yes	Yes
					125-CN	20	24	23	Yes	Yes
					103-KD	20	9	8	Yes	Yes
3	40	2.09	5121	3	131-SB	60	69	79	Yes	Yes
					106-BL	20	9	10	No	Yes
					152-LM	20	22	11	No	Yes
3	41	1.02	22650	1	125-CN	50	51	49	Yes	Yes
					115-RM	25	16	25	Yes	Yes
					79-PV	25	28	26	Yes	Yes
3	42	1.08	15005		61-KM	50	49	44	Yes	Yes
					30-SS	25	15	15	Yes	Yes
					69-CL	25	36	42	Yes	Yes
3	43	1.05	6743		71-LW	40	41	44	Yes	Yes
					115-RM	30	26	27	Yes	Yes
					39-DC	30	32	28	Yes	Yes
3	44	1.01	9075		150-TD	40	59	61	Yes	Yes
					74-LM	30	22	19	Yes	Yes
					132-JF	30	16	19	Yes	Yes

Table 3 – Three person mixture deconvolution results, continued– low level target input amount

People	Mixture #	Gelman-Rubin Convergence	Effective Sample Size	Amount of Dropout	Reference IDs	Target % contribution	Manual % contribution	STRmix % contribution	Correct combination in 99%?	STRmix Genotype possibilities intuitive?
3	45	1.09	3524	16	131-SB	33	n/a	39	Yes	Yes
					142-VM	33	n/a	30	Yes	Yes
					40-DB	33	21	30	Yes	Yes
3	46	1.22	2413	18	79-PV	33	39	37	Yes	Yes
					71-LW	33	31	30	Yes	Yes
					127-DF	33	30	33	Yes	Yes
3	47	1.06	34222	7	125-CN	70	80	84	Yes	Yes
					106-BL	20	11	8	No	Yes
					155-JK	10	17	8	No	Yes
3	48	1.03	4687	5	34-WD	70	70	70	No	No
					28-GL	20	19	15	No	Yes
					35-JD	10	17	15	No	Yes
3	49	1.02	18647	5	39-DC	60	15	20	Yes	Yes
					151-TR	30	83	76	Yes	Yes
					81-AR	10	3	4	Yes	Yes
3	50	1.02	19454	6	41-AD	60	60	68	yes	yes
					49-JB	30	30	24	Yes	Yes
					47-ER	10	10	8	No	yes
3	51	1.33	5404	2	71-LW	50	46	44	Yes	Yes
					93-RC	40	39	42	Yes	Yes
					81-AR	10	14	14	Yes	Yes
3	52	1.61	15737	6	56-TB	50	46	52	Yes	Yes
					39-DC	40	45	25	Yes	Yes
					81-AR	10	14	23	No	Yes
3	53	1.09	17252	9	23-JS	50	49	35	Yes	Yes
					39-DC	30	24	32	Yes	Yes
					79-PV	20	25	33	Yes	Yes
3	54	1.79	5606	7	30-SS	50	27	32	Yes	Yes
					52-KK	30	40	37	Yes	Yes
					23-JS	20	29	32	Yes	Yes
3	55	1.52	18569	8	30-SS	45	32	43	Yes	Yes
					161-GZ	45	50	48	Yes	Yes
					88-KB	10	8	9	No	Yes
3	56	1.15	33017	5	150-TD	45	53	46	Yes	Yes
					127-DF	45	35	45	Yes	Yes
					155-JK	10	11	9	Yes	Yes
3	57	1.02	25388	14	61-KM	40	48	34	Yes	Yes
					102-EB	40	n/a	33	No	Yes
					156-RK	20	52	33	Yes	Yes
3	58	1.06	1232	3	51-HH	40	25	33	No	Yes
					62-BS	40	45	35	Yes	Yes
					94-RL	20	25	32	No	Yes
3	59	n/a	7764	8	79-PV	35	23			
					159-MS	35	24	n/a	n/a	n/a
					156-RK	30	58			
3	60	n/a	6277	11+ all of contributor 3	39-DC	35	35	33	Yes	Yes
					93-RC	35	50	36	Yes	Yes
					142-VM	30	27	31	No	Yes
3	61	1.05	7308	8	127-DF	60	63	71	Yes	Yes
					125-CN	20	19	14	Yes	Yes
					103-KD	20	12	14	Yes	Yes
3	62	1.03	15562	17	131-SB	60	57	40	Yes	Yes
					106-BL	20	21	25	Yes	Yes
					152-LM	20	38	35	Yes	Yes
3	63	0	29244	9	125-CN	50	54	37	Yes	Yes
					115-RM	25	22	31	Yes	Yes
					79-PV	25	26	32	Yes	Yes
3	64	2.04	2681	3	61-KM	50	43	34	No	Yes
					30-SS	25	22	33	No	Yes
					69-CL	25	37	33	No	Yes
3	65	1.7	18189	13	71-LW	40	n/a	33	Yes	Yes
					115-RM	30	39	33	Yes	Yes
					39-DC	30	37	33	Yes	Yes
3	66	1.18	6843	16	150-TD	40	n/a	35	Yes	Yes
					74-LM	30	n/a	33	Yes	Yes
					132-JF	30	20	33	Yes	Yes

Conclusion

The MCMC and Metropolis-Hastings are central processes to STRmix. This study was designed to test STRmix in deconvolution of ground truth mixtures. The internal validation of this software package was done by providing a wide range of mixture samples designed, amplified, and electrophoresed in the SDPD crime lab following ModelMaker. Each of these mixtures was examined in detail to record the known genotype weight of every contributor. Of the 296 contributors making up these 2 and 3 person mixtures, all but one of them had results in which the known genotype was intuitive when the electropherogram was examined closely for peak height balance, mixture ratio, and locus specific amplification efficiency. There were several instances where the GlobalFiler amplification resulted in peak heights at one locus (often SE33) that were inconsistent from other loci, but given the electropherogram, STRmix was able to provide reliable, consistent, and robust results for the contributors to those mixtures. The less a person contributes to the mixture (20% or less), the lower the genotype weight *can* be associated with the mixture. Also, when two or more contributors in a mixture are balanced (contributing equal amounts), there is more ambiguity in their possible genotype combinations. Finally, the more dropout there is associated with a mixture, the more ambiguous the results can be. These three principles are to be expected, and are things that have had to be accounted for in the past. This study was effective for determining some of the interpretation limits within STRmix, which are important to keep in mind as STRmix results are interpreted. Even with these limits in mind, STRmix allows interpretation of many more contributors in many more mixtures than was previously possible in the lab. The level of consistency that STRmix provides is very high, and is one of the largest benefits in moving to probabilistic genotyping for interpretation of mixed DNA results. Another benefit in using this software is that different weights are associated with each genotype choice, and it is dependent on the observed electropherogram generated in the lab. Having a number associated with a particular genotype allows a very precise calculation for a likelihood ratio. This provides reliable results as well as clarity for evidence items examined in the SDPD crime lab.

References

1. STRmix v2.3 Users Manual. Issued by Institute of Environmental Science and Research Limited; Date of Issue: 20 January 2015