SDPD Forensic Science Section – Forensic Biology Unit **September 30, 2015**

STRmix™ Validation Summary

Introduction

STRmix™ uses possible genotype combinations and a biological model to create expected profiles that are compared to the quantitative data from the electropherogram (for example peak heights). STRmix™ then calculates the probability of the peak heights given the selected mass parameters values. Using Markov chain Monte Carlo (MCMC) methods with a Metropolis-Hastings algorithm, STRmix solves for genotype combinations that explain the observed data set, generating a list of genotype sets and assigning weights to each set that reflect how well they 'fit' the observed evidence data. If the proposed combination of 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 performs the deconvolution of for mixtures containing any number of contributors without reference to any known contributors or persons of interest (POI), unless specified by the defense hypothesis.

STRmix requires an initial process to determine laboratory-specific parameters that will be used to perform the deconvolution of samples. This process is used to inform the biological model used within STRmix. The parameters that STRmix requires a set of laboratory specific variables that need to be optimized. These include stutter ratios, analytical threshold (or limit of detection), capillary electrophoresis (CE) instrument's saturation limit, drop-in parameters, variance constants for stutter and allele, locus specific amplification efficiency parameter, and population settings including allele frequencies and theta values. Stutter ratios, analytical threshold (or limit of detection), capillary electrophoresis (CE) instrument's saturation limit, drop-in parameters were all determined through the validation of the specific kits on the 3500.

Variance constants for stutter and allele, locus specific amplification efficiency are parameters that define the variation to be expected within the laboratory processes from sampling for PCR to electrokinetic injection of samples onto the CE instrument. The variance parameters are used by STRmix when determining goodness of fit of the expected profile to the observed profile and are determined by Model Maker.

Model Maker

The GlobalFiler Stutter Study can be referenced for details about the compilation of this data, and for graphs of allele-specific stutter at every locus. The *stutter file* was created from the n-1 repeat data plotted in these graphs in the GlobalFiler study (presented as ratio instead of percentage). Specifically, it was created by taking a linear regression of all n-1 repeat stutter.

The *stutter exceptions file* provides a look-up table for the stutter ratio based on the longest uninterrupted sequence (LUS) or the lengths of a multiple core repeats within an allele (multi-

sequence model). These loci were given stutter exceptions for the SDPD stutter exceptions file: (**LUS**) - vWA, D21, TH01, FGA, (**MSM**) - D3, D8, D2S441, D19, D22, SE33, D1, D12, D2S1338.

The analytical threshold was determined in the GlobalFiler baseline study to be 100 RFU for all channels. A saturation limit of the 3500 was observed to be ~32,000 RFU. The drop-in cap was scaled up for data collected on the 3500 instruments to be 390 RFU.

ModelMaker is used to determine numbers that describe allelic variance (c^2), which is modeled using a $\Gamma(\alpha_1,\beta_1)$ prior, and stutter variance (k^2), which is modeled using a $\Gamma(\alpha_1,\beta_1)$ prior. This variance can be expressed numerically so that these variance parameters can be input as settings into STRmix. These values are obtained from a gamma distribution (Figure 1).

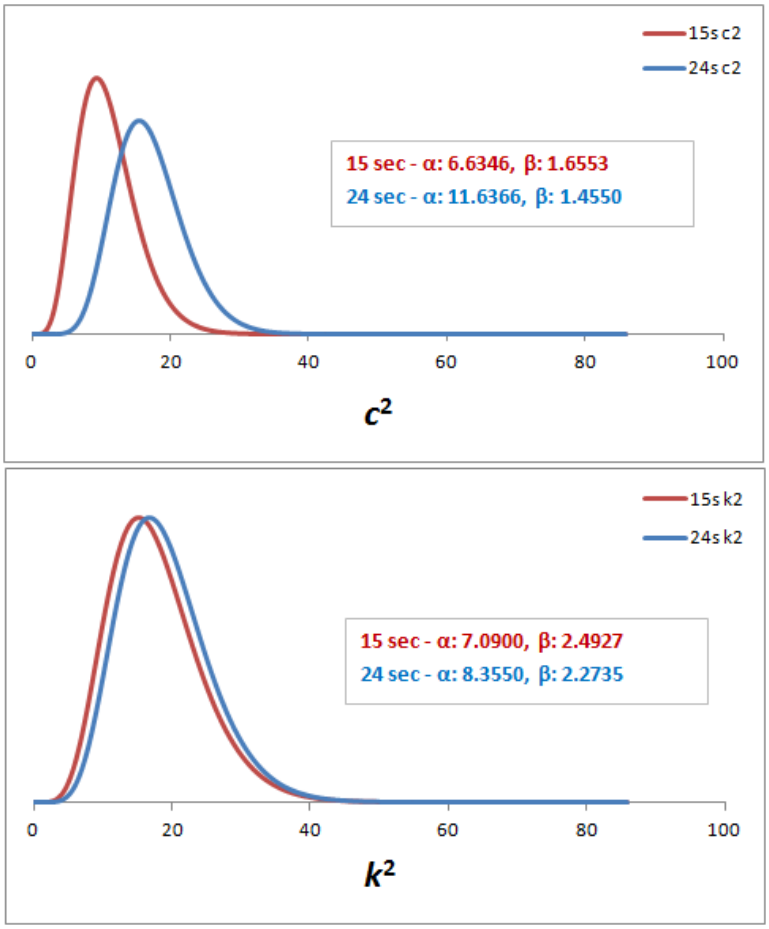

Figure 1 – The allele and stutter variance parameters.

Locus specific amplification efficiency parameter

The locus specific amplification efficiency parameter is calculated by ModelMaker using the same crime and reference file. Its calculation of this parameter is determined by the following equation:

15 seconds:

LSAE Variance - fitted gamma curve = exp(64.25832145393541) - mean = 0.015562186770111413
24 seconds:
LSAE Variance - fitted gamma curve = exp(56.143702296033844) - mean = 0.017811436708024916

After the settings and parameters have been established through Model Maker, the software can be validated as a tool for assisting in sample interpretation.

Validation

The validation of this software included an assessment of the Markov Chain Monte Carlo (MCMC) portion of the software as well as the likelihood ratio calculator portion of the software.

The Markov Chain Monte Carlo portion of the software was evaluated by examining single source, 2-person, 3-person, and 4-person DNA mixtures and determining whether correct genotype deconvolution, mixture ratios were obtained. In addition, the MCMC process is known to be a random process that will produce slightly different results each time it is run. The reproducibility of the process was investigated. In addition, the ability of STRmix to deconvolute mixtures into component genotypes was assessed through comparisons to known contributors and known non-contributors using the Database search tool.

The likelihood ratio portion of the software using the same single source, 2-person, 3-person, and 4-person DNA mixtures previously described that will examine the LR ratio calculations with a mix of known contributors, known non-contributors, and assumed contributors.

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

The biological model used by STRmix to build an expected DNA profile is described by the following equation:

$$T_{an}^{l} = A_{r}^{l} t_n X_{an}^{l} \times e^{d_n \times \left(mwt_a^l - offset \right)}$$

STRmix then splits total allelic product into allelic and stutter height, using the following equations:

Allele **Stutter**

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

The MCMC process involves thousands to millions of iterations as it attempts to better describe the observed data. Eventually, the mass parameters reach an equilibrium point where only a small set of variables are continually being selected to describe the data.

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.

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.

MCMC

The MCMC and Metropolis-Hastings are central processes to STRmix. The validation of this software package was done by providing 186 DNA mixtures covering a wide range of mixture samples. Each of these mixtures was examined in detail to record the known genotype weight of every contributor. 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. Overall, the MCMC process of deconvoluting DNA mixtures was very robust with correct deconvolutions obtained for even low level 4-person mixtures. In general, 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.

Reproducibility

Since the MCMC process is random, each time a mixture is run through STRmix it produces a different answer each time it is run. This variability is at its largest when the likelihood ratio is low, if the number of iterations is insufficient to solve the problem, or if the problem is complex. Variability from four different sources (replicate CE injections, replicate CE plate loads of same amplified DNA, replicate PCR amplifications, and replicate LR calculations) was measured in the developmental validation. Of all these, the MCMC process displayed the least variability.

A subset of the 2-, and 3-person mixtures created as part of the GlobalFiler validation were used for this experiment. 4-person mixtures were not selected for this study due to the length of time those mixtures required to run in STRmix. In the mixture study, different ranges of template DNA were targeted.

The individual weights assigned to each genotype possibility are derived from the proportion of time each possibility is accepted as a better answer than a separate guess. The variability in the weights is derived from the randomness of each guess in the MCMC process. As such, there are a couple of means to verify the reproducibility of the MCMC process. One method would be to check the assigned weights of each possible genotype at each marker and compare them across replicates. Another means of checking this would be done based on the examining the likelihood ratios for contributors to the mixtures. Likelihood ratios are calculated in STRmix by taking into account the various possible genotypes for each contributor including the weights assigned to each possibility. Since the weights of each possibility are taken into account, the value of the likelihood ratio is reflective of the differences in genotype weight and can be used as a measure of the reproducibility of the MCMC process.

Overall, the results of the mixture deconvolution displayed very good reproducibility. The variability in the mixture results generally increased as the number of genotype possibilities that could explain the data increased. The data suggests that as the total amount of template DNA is decreased, or the contribution level of a contributor is lowered, the variability increases. In addition, the variability increased when the contributors, even robust level contributors, had similar contribution levels to other contributors in the mixture, or when there was a large amount of allele sharing between the contributors.

Comparison to Known Contributors and Known Non-Contributors(Sensitivity and Specificity)

The goal of this study is to determine whether the MCMC deconvolutions result in correct inclusions when compared to a large number of subjects (specificity) and whether the deconvolutions remain effective as template decreases (sensitivity).

STRmix allows the user to search a deconvoluted DNA mixture against a database directly, without the need for deriving a single source component. The Database Search function can be

used as a quality assurance tool for comparison of complex mixtures. The minimum LR value can be used as a list management tool to filter out of the results file all comparisons that lead to an LR below the specified value. The results summarize the input files, deconvolution and database search run conditions, and list of individuals whose comparison to the genotype PDF has yielded an LR above the defined cut-off value. The LR threshold set to 0 will return the results for all individuals within the database.

2-person mixtures
Ten 2-person mixtures were deconvoluted using STRmix. After running the mixtures through STRmix the deconvolution results were compared to the database file of 76 individuals. Each 2-person mixture resulted in likelihood ratios favoring inclusion for the individuals known to comprise the mixtures. All other non-contributors in the database had likelihood ratios of zero (i.e., excluded).

3-peron mixtures
Seventeen 3-person mixtures were deconvoluted using STRmix. After running the mixtures through STRmix the deconvolution results were compared to the database file of 76 individuals. In eight of the mixtures, including two of the low level mixtures, all non-contributors had likelihood ratios of zero (i.e., exclusion). In the remaining nine mixtures, the non-contributors all had negative log likelihood ratios favoring exclusion.

4-person mixtures
Sixteen 4-person mixtures were deconvoluted using STRmix. These were mixtures that included high, mid, and low level mixtures with a range of contributor proportions. Seven of the mixtures had contributors with dropout. After running the mixtures through STRmix, the deconvolution results were compared to the database file of 76 individuals. All sixteen of the 4-person mixtures resulted in the inclusions of the individuals known to comprise the mixtures. In addition to the correct inclusions, one mixture (mixture ID: 4-63) also had a single non-contributing profile from the database that also resulted in a likelihood ratio that favored inclusion.

STRmix deconvolutions have been demonstrated to be very robust. There is a high degree of specificity as established by the high level of accuracy of the inclusions and exclusions.

Likelihood Ratio
Following deconvolution of evidence samples, STRmix has the capability of generating likelihood ratios (LRs) to determine the degree to which the evidence suggests that an individual, or a group of individuals, contributed DNA to the evidence sample. The likelihood ratios are calculated by finding the ratio of the probability of obtaining the evidence under an inclusionary hypothesis, H_1 (also referred to as the prosecutor's hypothesis, H_p) to the probability of obtaining

the evidence under an exclusionary hypothesis, H_2 (also referred to as the defense attorney's hypothesis, H_d).

Previously analyzed samples of known composition, prepared for the Globalfiler validation, were selected for this verification. Twenty four single source samples (15, 16, 17, 19, 20, 23, 27, 28, 29, 31, 34, 35, 36, 38, 39, 40, 49, 51, 52, 56, 61, 62, 71, and 161), two two-person mixtures (2-19 and 2-39), and one three-person mixture (3-9) were chosen.

This study demonstrated that the likelihood ratio calculations performed by STRmix can be replicated within a very small margin within Excel. This study also demonstrates that the laboratory has a very good understanding of the manner in which STRmix calculates Likelihood ratios under a variety of different scenarios.

Adjudicated Cases

Six adjudicated cases with a sexual component (sex crimes and child abuse) were selected because these cases contained samples known to have mixtures of DNA, are representative of the types of cases encountered, had a high probative value, and represented a range of previously validated DNA typing kits.

Based on the results obtained, the GlobalFiler amplification kit yielded more information overall than the Profiler Plus, COfiler, Identifiler, and Identifiler Plus amplification kits with similar amounts of input DNA.

The GlobalFiler kit displayed general concordance with the results obtained from previous analysis; however, concordance could only be ascertained for samples with previous STR typing and only at the loci contained in the typing kits previously used for these samples. Differences in relative proportions of contributors to some samples were noted. These differences were due to the fact that several samples were intentionally prepared in a manner to generate more evenly balanced mixtures than were previously obtained in an effort to make these samples amenable to mixture deconvolution by STRmix.

The Y chromosomal markers within GlobalFiler, while not very discriminating, provide additional potential to indicate that a lower level contributor to a mixture is male than previous testing kits with a single gender informative marker.

The additional discriminating loci in the GlobalFiler kit provide better assessment of the number of contributors to mixed samples and increase the chance of detecting low-level minor contributors to a sample, who may be masked by fewer and less discriminating loci contained in previously used amplification kits.

Previously reported conclusions originally generated for the evidence with regards to the inclusion of the victim and suspect still hold for the data generated with the GlobalFiler

amplification kit; however, the strength of the associations are magnified by the presence of additional discriminating loci within the GlobalFiler kit compared to previous testing kits.

Conclusions

All of the above sections describe a software package that is robust, fit for purpose of assisting in the interpretation of single source to five-person mixtures. Overall the results obtained with STRmix suggest it would improve the power of STR testing over current methods. STRmix allows for both greater strength of the evidence when an individual is included, and also allows for both inclusions and exclusions from samples that previously would have been deemed inconclusive.

Based on the data obtained from the validation of STRmix and the reasons stated above, STRmix should be implemented in casework at the SDPD.

Validation Analysts

_____	_____
Melissa Strong Criminalist	Marybeth Sciarretta Criminalist
_____	_____
Adam Dutra Criminalist	Shawn Montpetit DNA Technical Manager

For a signed copy please contact the Quality Assurance Manager

_____	_____
Supervisor Review Date	**Supervisor Review Date**

_____

Quality Assurance Review Date

SDPD Forensic Science Section – Forensic Biology Unit

STRmix® ModelMaker

Determining lab, kit and condition specific settings for STRmix interpretation

---

Introduction

STRmix™ applies a fully continuous probabilistic genotyping approach to DNA profile interpretation. The STRmix biological model uses lab specific parameters to calculate expected peak height data for alleles and N-1 stutter. The results are compared to observed data, and the goodness of fit is determined. This happens many times during the deconvolution process. The amount of stutter varies from amplification to amplification, as does the peak height balance and locus specific amplification efficiency. Peak height balance also varies with input amount; the lower the amount of starting material, the more the stochastic effects can affect peak height balance. These three things (stutter variance, peak height balance, and locus specific amplification efficiency variance) are taken into account when determining goodness of fit. They are important in comparing the expected profile to the observed profile. This is where ModelMaker fits into the process. STRmix contains a tool called ModelMaker to help determine the variance values for a specific kit on a specific instrument platform. Normally, STRmix is provided evidence and models to get genotype weights. But it can also be used initially to get the models, by providing it weights and evidence (single source samples and their references). ModelMaker is ideally used once, prior to any mixtures generated in the lab being analyzed with STRmix. If there is a change to the amplification kit used, or a major protocol change occurs (i.e., instrument, number of cycles or injection time), it should be run again to check the variance parameters. While STRmix is supplied with default values for many commercially available kits, the reason behind any internal validation is that the local lab environment (i.e. products, instruments, and analysts) can affect DNA results, especially when it comes to sensitivity, stutter and peak height balance. It is ideal to supply SDPD specific variance parameters to STRmix in order to optimally model DNA profiles that are generated in the SDPD crime lab. This validation study describes the use of ModelMaker in the SDPD lab for samples injected for the standard duration of 15 seconds, as well as for an increased injection time (24 seconds).

Purpose

Prior to the validation and implementation of STRmix™ within a laboratory there are a number of laboratory specific variables that need to be optimized. These include:

1. Stutter ratios
2. Analytical threshold (or limit of detection)
3. Capillary electrophoresis (CE) instrument's saturation limit
4. Drop-in parameters
5. Variance constants for stutter and allele
6. Locus specific amplification efficiency parameter

7. Population settings including allele frequencies and theta values

ModelMaker is a tool to help a lab determine the variance constants for stutter (k^2), allele (c^2), and locus specific amplification efficiency variance (A^I) for different STR profiling kits and protocols (i.e., longer injection protocols or increased cycles). In other words, we provide ModelMaker with #1-4 above, and it gives us #5 and #6. ModelMaker functions by using independent amplifications at different target amounts, from very high input to vey low input; low enough that alleles are dropping out. By including this range of data, ModelMaker is able to get an idea of how these parameters change with peak height (similar to previous validations in the lab where peak height balance is determined for different input amounts/peak heights). ModelMaker uses the MCMC process to determine these variance parameters. The foundational concept in ModelMaker is: if the genotypes, stutter expectations, instrument saturation, and drop-in rate are known, the dataset can be evaluated to determine the stutter, allele, and locus specific amplification variances that best fit the lab's data. This study will include a description of #1-6 above, and #7 will be covered in the Likelihood Ratio validation study.

Materials and Methods

Stutter ratios

Lab-specific stutter is provided to STRmix in two ways. The first is a *stutter file*, and the second is a *stutter exceptions file*. STRmix can only model n-1 repeat stutter. All other types of stutter are disregared (edited as artifacts) when it comes to STRmix. Stutter is observed at almost all loci in GlobalFiler. The only two loci where no stutter has been observed is the Yindel (Locus 6) and Amelogenin. The GlobalFiler Stutter Study can be referenced for details about the compilation of this data, and for graphs of allele-specific stutter at every locus. The *stutter file* was created from the n-1 repeat data plotted in these graphs in the GlobalFiler study (presented as ratio instead of percentage). Specifically, it was created by taking a linear regression of all n-1 repeat stutter.

The *stutter exceptions file* is referenced before the stutter file. If there is a 0 in the stutter exceptions file, expected stutter will be calculated using the *stutter file*. The *stutter exceptions file* provides a look-up table for the stutter ratio based on the longest uninterrupted sequence (LUS) or the lengths of a multiple core repeats within an allele (multi-sequence model). Each locus whose linear regression line did not fit the data well was examined (see results section for a list of these loci). For these loci, stutter was modeled differently (using the LUS or multi-sequence model) in order to minimize differences observed and expected stutter ratios for every allele. LUS values can be found for every locus in Appendix 1 of the Butler Methodology textbook (1). While there sometimes multiple options for LUS at several alleles, the one that fit the data best was chosen. When the LUS model still did not fit the data, the multi-sequencce model was used. This model was communicated by Dr. Buckleton, and uses the formula:

Stutter Ratio = slope*(MAX(LUS1-lag, 0) + MAX(LUS2-lag, 0) + MAX(LUS3-lag, 0) + MAX(LUS4-lag, 0)) + C

Where C and Lag are constants. Essentially, it takes into account multiple sequence stretches, and not just the longest one. This data can be examined in the validation folder under GlobalFiler stutter.

Analytical threshold

The analytical threshold was determined in the GlobalFiler baseline study. Briefly, negative controls and DNA-containing samples were amplified (peaks removed) to examine the noise peaks. Each channel was examined separately, and the analytical threshold was determined to be between 2 and 3 times the highest of the documented noise peaks during the validation study. Refer to the GlobalFiler baseline study for more details.

Capillary electrophoresis (CE) instrument's saturation limit

Saturation was determined using the samples from the GlobalFiler Sensitivity Study. A series of 10 single source samples amplified 10 times each with template amounts ranging from ~20pg to ~6ng to ensure that at least samples reached a saturation level. Capillary electrophoresis was performed on these samples using both 3500 instruments according to the manufacturer's recommended parameters. A detection threshold of 100 RFU was applied during GeneMapper ID-X (v1.4) analysis, and data was analyzed with the stutter filter off. Allele peak heights and n-1 repeat stutter peak heights were recorded.

Drop-in parameters

STRmix provides the option of considering the possibility of drop-in occurring, and taking that into account during mixture deconvolution. DNA is present at low levels in the environment, and the sensitivity of the DNA typing methods being employed make it possible to detect very small amounts of DNA. Drop-in is unexplained peaks observed within a profile. It is possible to have an extra or unexpected allele detected in a sample, even using standard amplification methods (i.e. Identifiler Plus at 28 cycles). Because drop-in is a possibility, the goal of this study was to provide STRmix with realistic drop-in parameters so that it can accurately account for it in its biological model. The SDPD does not have sufficient data for the GlobalFiler kit to accurately estimate a drop-in rate. As such, an alternative to using GlobalFiler data was needed. The validation of the GlobalFiler kit indicates that the sensitivity of the GlobalFiler kit at 29 cycles is similar to the Identifiler Plus sensitivity at 28 cycles. Given the similarity in sensitivity between the two kits; the drop-in rate for Identifiler Plus could serve as a stand-in for a GlobalFiler drop-in rate until such time as one could reasonably be calculated.

To obtain an estimate of drop-in, the parameters of number of observations of drop-in out of the total number of possibilities (loci x sample number) and the heights of those peaks are required. The height of a drop-in peak should follow a gamma distribution (2). The gamma distribution is a two-parameter family of continuous probability distributions. The gamma distribution can be parameterized in terms of a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, called

a rate parameter. Both parameters are positive real numbers. A random variable X that is gamma-distributed with shape α and rate β is denoted:

$$X \sim \Gamma(\alpha, \beta) \equiv \mathrm{Gamma}(\alpha, \beta)$$

If X_i has a Gamma(k_i, θ) distribution for i = 1, 2, ..., N (i.e., all distributions have the same scale parameter θ), then:

$$\sum_{i=1}^{N} X_i \sim \mathrm{Gamma}\left(\sum_{i=1}^{N} k_i, \theta\right)$$

provided all X_i are independent. In statistics, maximum-likelihood estimation (MLE) is a method of estimating the parameters of a statistical model. When applied to a data set and given a statistical model, maximum-likelihood estimation provides estimates for the model's parameters. For the SDPD Identifiler Plus data, the unexpected results from the years 2010 through 2014 were examined. The number of times unattributable types were detected in reagent blanks or negative amplification controls were counted. From the observed sample set, a gamma distribution is projected using a maximum likelihood estimation to establish the α and β variables. The observed results over this period of time also provided a drop in frequency. The Identifiler Plus data (collected on a 3130 Genetic analyzer) was adapted to reflect what drop-in might look at with GlobalFiler on a 3500 Genetic analyzer. The considerations were that sensitivity is about the same between the kits, and the sample peak heights are 2-3 times higher when analyzed on a 3500 instrument compared to a 3130 instrument.

Variance constants for stutter and allele

The STRmix Manual (3) recommends providing a range of samples (at least 90) of varying profile quality and peak heights. This was accomplished using the samples from the GlobalFiler Sensitivity Study (reformulation data). Briefly, a series of 13 single source samples amplified 6 or 7 times each with template amounts ranging from ~12pg to ~1ng. Capillary electrophoresis was performed on these samples using the 3500 instrument according to the manufacturer's recommended parameters. The first set of samples (101 total) was injected for 15 seconds. To determine how the varaiance constance change when the samples are injected for a longer period of time, another dilution series of single source samples (117 total samples; ~12pg to 200pg) was injected for 24 seconds on the 3500.

A detection threshold of 100 RFU was applied using GMID-X (v1.4), and data was analyzed with all n-1 repeat stutter filters off (evidence analysis method) so that all stutter peaks were called as alleles. Other types of stutter (n+1 and n-0.5 repeat) were still filtered and artifacts were labeled as such so as not to be included in the exported genotype table. In addition, these samples from each dilution series were also analyzed with all stutter filtered to produce a record of known

genotypes. Settings were created in GMID-X (v1.4) to reflect these two different analysis methods. STRmix models n-1 repeat stutter, so all evidence samples amplified with GlobalFiler to be deconvoluted by STRmix should be analyzed with the ***STRmix_GF_evidence*** analysis method, and the ***GF_evidence*** panel in GMID-X (v1.4), which filters only n+1 and n-0.5 repeat stutter. All reference samples amplified with GlobalFiler should be analyzed with the ***GF_reference*** analysis method and the ***GlobalFiler_Panel_v1*** panel, which filters all types of stutter. See Table 1 for a list of the stutter filters (see GlobalFiler Stutter Study for more details and stutter graphs of each locus).

Table 1 GMID-X v1.4 stutter filters.

Locus	N-1 Stutter (%)		Stutter Filter (%)		
	Minimum	Maximum	GMID-X *N-1*	GMID-X *N+1*	GMID-X *N-.5*
D3S1358	3.22	11.52	10.98	2.00	0
vWA	3.01	10.08	10.73	1.00	0
D16S539	2.62	8.42	9.48	2.00	0
CSF1PO	3.98	7.26	8.77	1.00	0
TPOX	1.44	3.63	5.55	0	0
D8S1179	2.94	9.53	9.60	2.00	0
D21S11	3.91	9.77	10.45	2.00	0
D18S51	4.03	15.63	12.42	2.50	0
DYS391	3.02	6.18	7.43	0	0
D2S441	2.36	8.38	8.60	2.50	2.00
D19S433	3.91	11.00	10.25	2.00	0
TH01	1.16	3.69	4.45	0	0
FGA	3.75	10.99	11.55	2.00	0
D22S1045	2.24	12.59	16.26	6.69	0
D5S818	1.95	9.95	9.39	2.00	0
D13S317	1.61	9.01	9.19	2.00	0
D7S820	2.04	7.93	8.32	2.00	0
SE33	4.43	18.60	14.49	2.90	5.00
D10S1248	5.54	11.47	11.46	2.25	0
D1S1656	4.16	13.20	12.21	2.00	2.45
D12S391	4.78	15.09	13.66	2.00	0
D2S1338	3.96	14.15	11.73	2.00	0

Two text files were created from each set of these samples. The "evidence" file with the single source samples analyzed as evidence (with stutter) contained allele, height and size information for every called peak. The "reference" file contains the reference genotype alleles (stutter removed) with size only (no height). The full genotype was entered in for the reference file if the low level samples had dropout (copied from a more robust sample). The names of the profiles in each file were identical and in the same order. While it is important to have a dataset that contains some dropout to accurately represent allele variance and the probability of dropout, any profiles with less than 10 datapoints or with peaks ablove the saturation level are not included in the final curve fitting.

These two files were entered into ModelMaker after all other settings (settings from the original kit formulation) and stutter files were set, and ModelMaker was run. The mass parameters for the contributors in the dataset, $M_1 \ldots M_c$, are optimized separately from the variance constants and from each other. Observed (O) and expected (E) peak heights were modeled by the equations below:

$$\log_{10}\left(\frac{O_a}{E_a}\right) \sim N\left(0, \frac{c^2}{E_a}\right) \text{ for allelic peak } a \text{ and}$$

$$\log_{10}\left(\frac{O_{a-1}}{E_{a-1}}\right) \sim N\left(0, \frac{k^2}{O_a}\right) \text{ for stutter peak } a\text{-}1$$

The analysis was then carried out by repeating loops of:

1. Optimizing $M_1 \ldots M_c$ while holding variance constant
2. Optimizing variance while holding $M_1 \ldots M_c$ constant

until all values within both vectors have converged.

There are three things to note:

1) The variance constants, c^2 and k^2, are modelled by gamma distributions, $\Gamma(\alpha_x \beta_x)$

2) The split of allele, a, and stutter peak, a-1, heights, to allow for separate variances for each data type

3) The inverse proportionality of the variance in the stutter model on observed peak height, O_a, rather than expected stutter peak height, E_{a-1}

ModelMaker determined the variance constants for stutter and allele using the settings described above and the supplied evidence and reference files. These results were collected from the ModelMaker output.

Locus specific amplification efficiency parameter

This parameter was estimated from the same dataset and files described in the section above (variance constants for sutter and allele). These two files from each set (15 and 24 seconds) were entered into ModelMaker after all other settings and stutter files were set, and ModelMaker was run. The ModelMaker results determined the locus specific amplification efficiency parameter using the lab generated data.

Results and Discussion

Stutter ratios

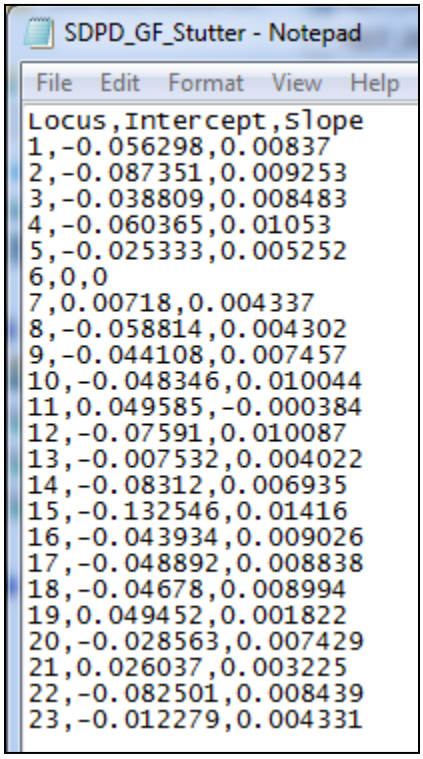

Figure 1 – Stutter file

The linear regresstion provides the slope and the y-intercept. This relationship describes an allele specific stutter ratio based on allele designation and is used to calculate the expected height of stutter and allele peaks within STRmix. See Figure 1 for the SDPD stutter file values. One important thing to note is that not all STR loci are composed of simple repeat structures where the expected stutter increases with allele number within a locus. Some loci have compound/complex stutter sequences.This creates stutter patterns that don't necessarily have very steep lines when taking a linear regression of all the data. For an example of this, see locus 11 (D2S441), which has a slope close to 0 due to two very different stutter populations. For this locus, and several other loci, the stutter file is not the best resource for estimating allele specific stutter. For this reason, a *stutter exceptions file* is also provided to STRmix.

Each locus that whose linear regression line did not fit the

data well was examined and stutter was modeled differently in able to minimize differences observed and expected stutter ratios for every allele. These loci were given stutter exceptions for the SDPD stutter exceptions file: (**LUS**) - vWA, D21, TH01, FGA, (**MSM**) - D3, D8, D2S441, D19, D22, SE33, D1, D12, D2S1338. The stutter ratio for alleles at these loci can be found in the stutter exceptions file.

Analytical threshold

The analytical threshold was determined in the GlobalFiler baseline study to be 100 0RFU for all channels. Refer to the GlobalFiler baseline study for more details.

Capillary electrophoresis (CE) instrument's saturation limit

The saturation limit was examined in two ways. The first way simply determined the maximum peak height observed for all samples in the sensitivity study (which included samples amplified with a target of ~7ng of DNA. There were 49 peaks that exceeded 30,000 RFU. The first of which was seen in a sample with a ~1ng target. The maximum peak detected (32,936 RFU) was seen in a sample with a target of ~2ng. The plateau seen around ~32,000 even though target DNA input amounts seemed to suggest a saturation limit of the 3500 instrument.

The second method of estimating saturation utilized the stutter peak heights. While allele peaks saturate at some point due to CE instrument limits, the stutter peaks do not because they are typically only ~10% of the parent peak. And because they are an amplification artifact, their height is dependent on template amount.

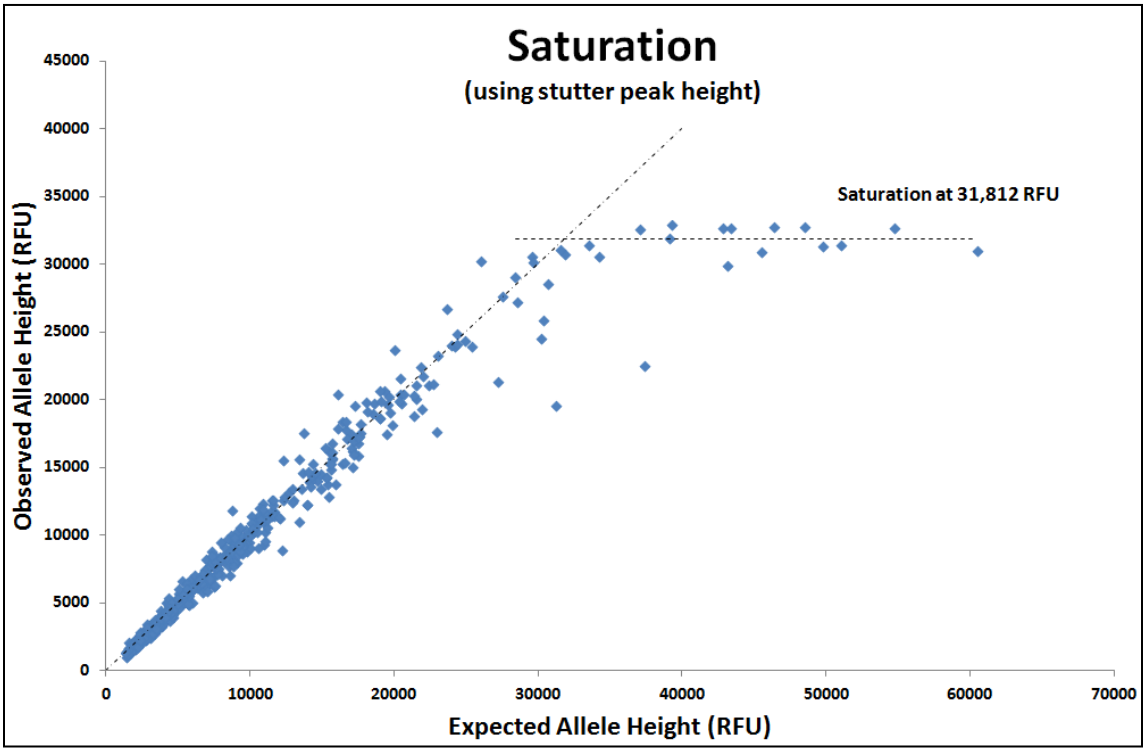

Figure 2 – Saturation for the 3500 Genetic analyzers

Stutter peak height continued to increase with template amount, unlike allele peak height. From this stutter peak, the stutter ratio described above can be used to calculate what the estimated parent allele height should be, using this equation: Expected Allele Height $=O_{a-1}/SR_a$. These values far exceeded the observed maximum allele heights. Plotting observed allele height against expected allele height gives another estimate of 3500 instrument signal limits. See Figure 2.

Drop-in parameters

The SDPD does not have sufficient data for the GlobalFiler kit to accurately estimate a drop-in rate. Data obtained with Identifiler Plus was used as a stand-in because of the similar sensitivity between kits/instruments. For the SDPD Identifiler Plus data, the unexpected results from the years 2010 through 2014 were examined. The number of times unattributable types were detected in reagent blanks or negative amplification controls were counted. Seven instances of drop-in were observed. The total number of reagents blanks was estimated by examining the average number of purification runs on the EZ1s per year and multiplying by two. The number of reagent blanks estimated in this manner was found to be approximately 1200.

The raw drop-in rate was determined to be 7/1200 x 16 ID+ loci = 7/19200 = 0.000364. This is not the true drop-in rate because the AT is at 50 and drop-in events could be occurring below this but are not detected. In order to obtain the true drop-in rate we must first calculate the percentage of drop-ins that are detected.

Since drop-ins follow a gamma distribution, we can estimate the shape of the gamma distribution based on the data above. The proportion of drop in that is detected (above 50 RFU) of the total was determined (Figure 3). For the SDPD data, the predicted proportion of detectable drop-in events was found to be 0.2953.

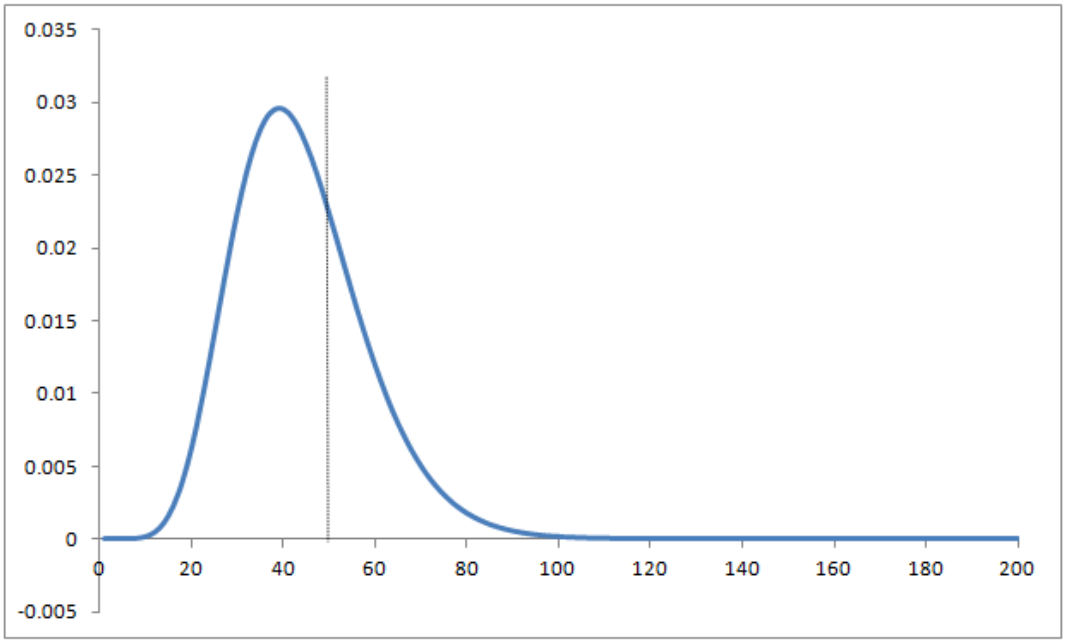

Figure 3 – The gamma distribution of drop-in events for Identifiler Plus

The raw drop-in rate divided by the percentage of detectable drop-ins gives the actual drop-in rate (0.000364/0.2953 = 0.0012). The drop-in cap was scaled up for data collected on the 3500 instruments to be 390 RFU.

Variance constants for stutter and allele

Allelic variance parameters describe variability of allele peak heights. This includes considerations of heterozygote peak height balance and what the probability of dropout is. Before entering data into ModelMaker, the heterozygote peak height balance (and how that changes with peak height) can be plotted. See Figure 4. In this figure, relationships between heterozygote peaks are plotted. The higher molecular weight (hmw) peak was divided by the lower molecular weight peak (lmw) at a locus, and a log of that ratio was calculated and plotted against the average peak height of those two peaks (plotting it in log form gives a more representative visual display of variation). One very apparent feature of this data set is the relationship between the peak height of the heterozygote alleles and how much the peak height ratio (PHR) can vary. The red lines on this graph indicate the 95% confidence interval. This is derived from the 75th quantile for the allele height variance, which is calculated by ModelMaker (see Figure 7 for variance values).

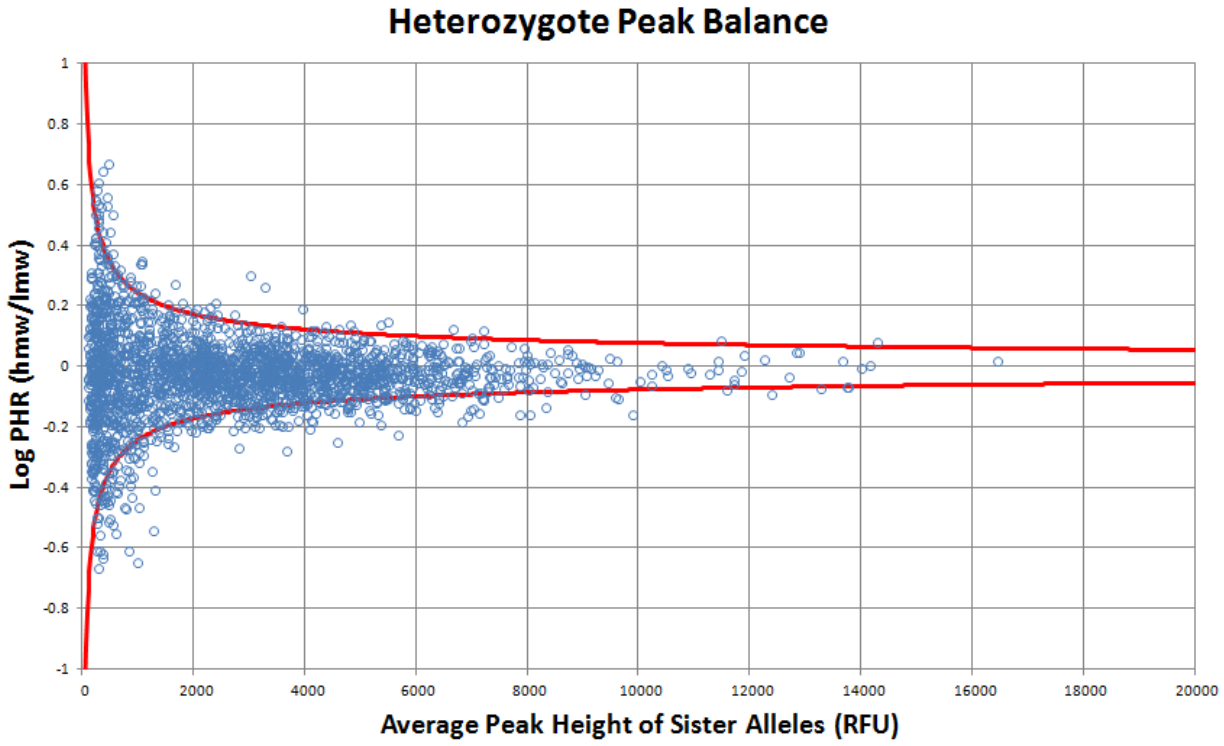

Figure 4 – Heterozygote peak height balance of samples amplified at a range of target values with 95% boundaries (red dotted line) plotted against the average height of heterozygote alleles.

We know from previous validations and published literature that peak height balance can vary a lot more as the peak heights approach the stochastic range. In the past, a stochastic threshold has

been calculated. This binary threshold has been based on the peak height of a known heterozygote allele for which its sister allele dropped below the analytical threshold. When dropout starts to occur has also examined using the GlobalFiler validation data. Instead of determining a binary threshold, however, a probablility of dropout was calculated. The possibility of dropout is peak height dependent. This is shown in Figure 5. This data is a histogram of the proportion of dropout events to the total allele observations in a given RFU range (bins = 20 RFU). This data demonstrates the more continuous method being used in probabilistic genotyping using data generated by the crime and evidence files being used in ModelMaker.

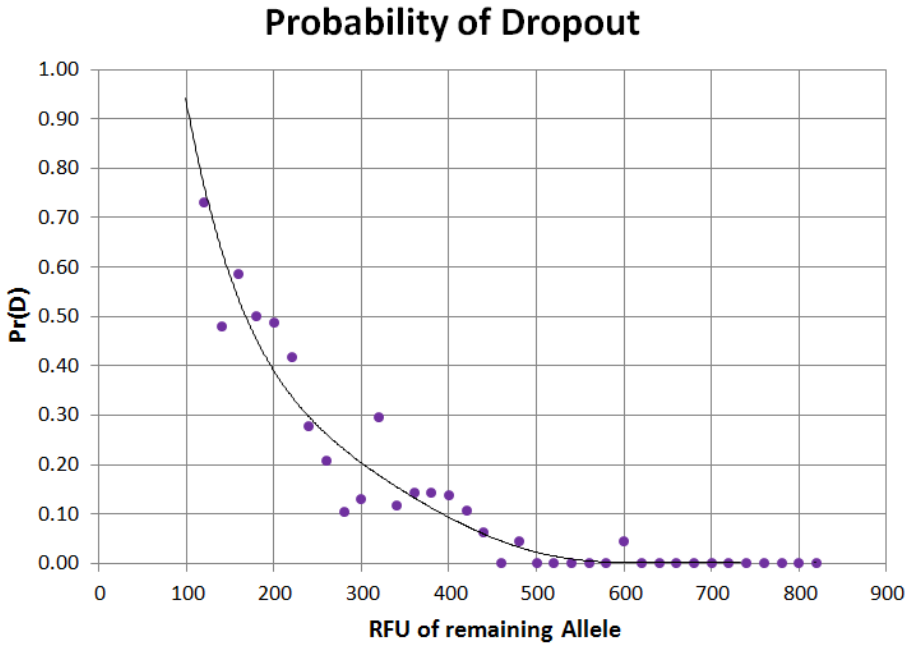

Figure 5 – Probability of dropout

ModelMaker can determine numbers that describe this variance to be used as lab specific settings based on this data. Allelic variance (c^2) is modelled using a $\Gamma(\alpha_1,\beta_1)$ prior. Stutter variance parameters describe stutter variance. Stutter variance (k^2) is modeled using a $\Gamma(\alpha_1,\beta_1)$ prior. Figures 6 and 7 show the data generated by ModelMaker in minimizing the difference between observed and expected (both for allele and stutter). See GlobalFiler Stutter Study for graphs of stutter at every locus, and how much that varied between separate amplifications.

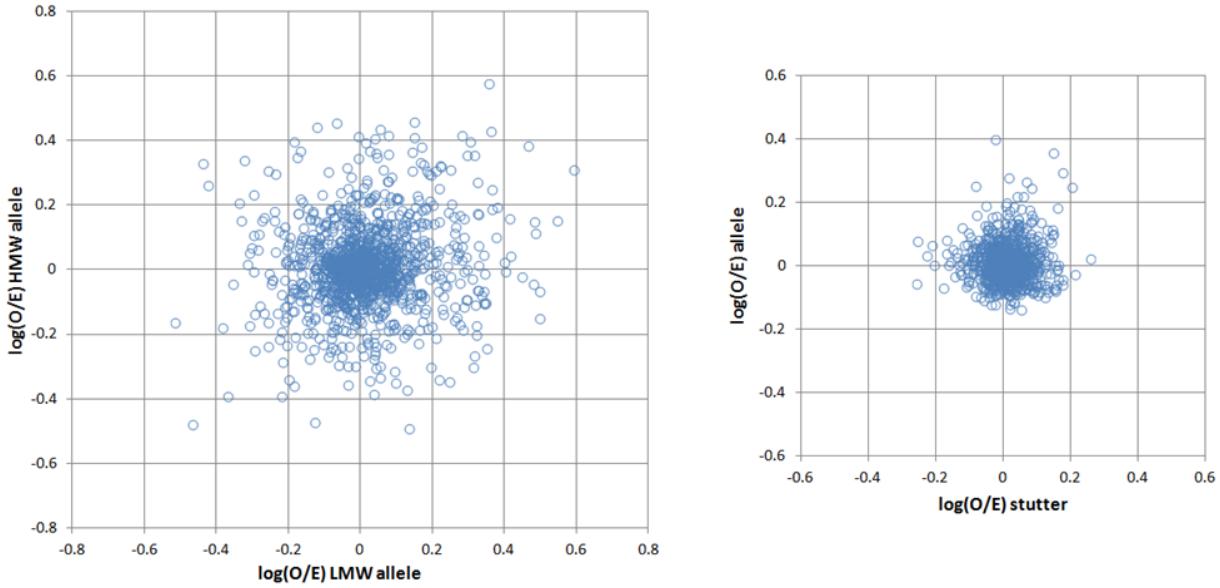

Figure 6 – STRmix ModelMaker outputs of Allele and Stutter variance for samples injected for 15 seconds. The differences between observed and expected values are minimized.

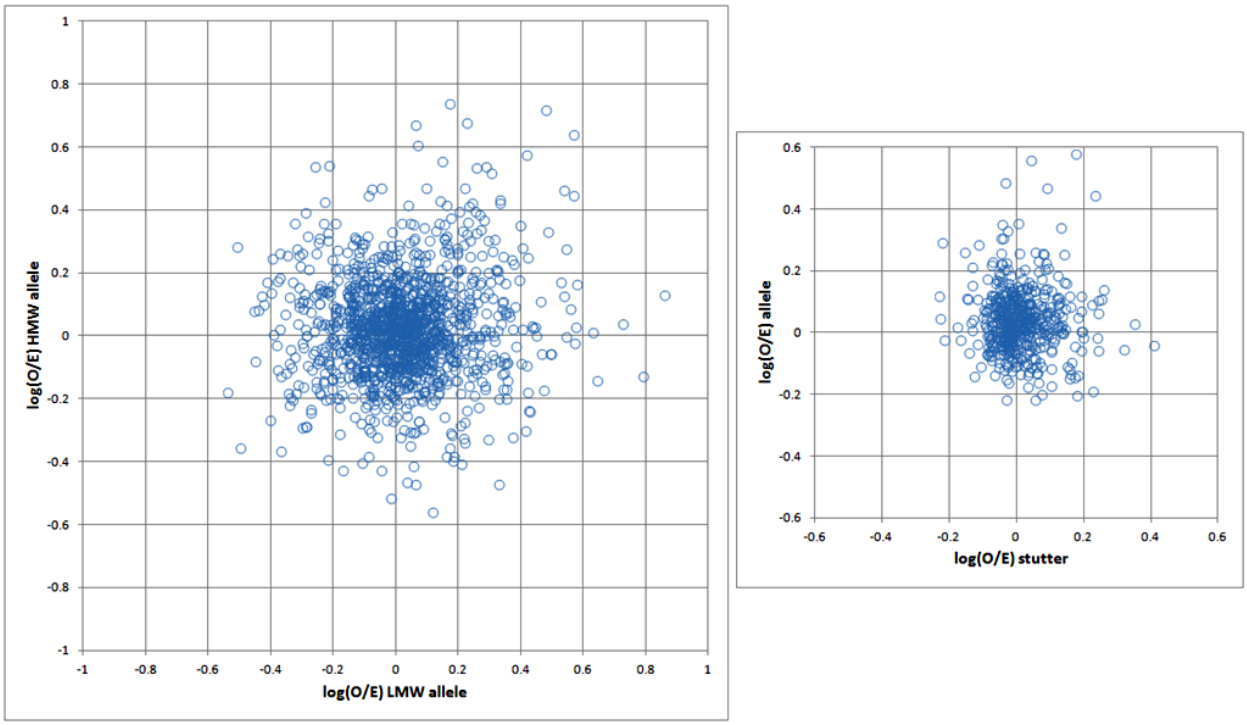

Figure 7 – STRmix ModelMaker outputs of Allele and Stutter variance for samples injected for 24 seconds. The differences between observed and expected values are minimized.

As expected, stutter variance did not change much between samples injected for 24 seconds compared to samples injected for 15 seconds. The amount of stutter is a phenomenon of amplification, and will increase proportionally with allele height when injected for a longer time. The peak height balance between heterozygotes is affected by an increased injection time. Any

imbalance is amplified when peak heights are increased ~1.6 fold. Because the this, the stochastic range is increased.
Figures 6 and 7 display the range of individual datapoints from the samples. This variance can be expressed numerically so that these variance parameters can be input as settings into STRmix. These values are obtained from a gamma disctribution (Figure 8).

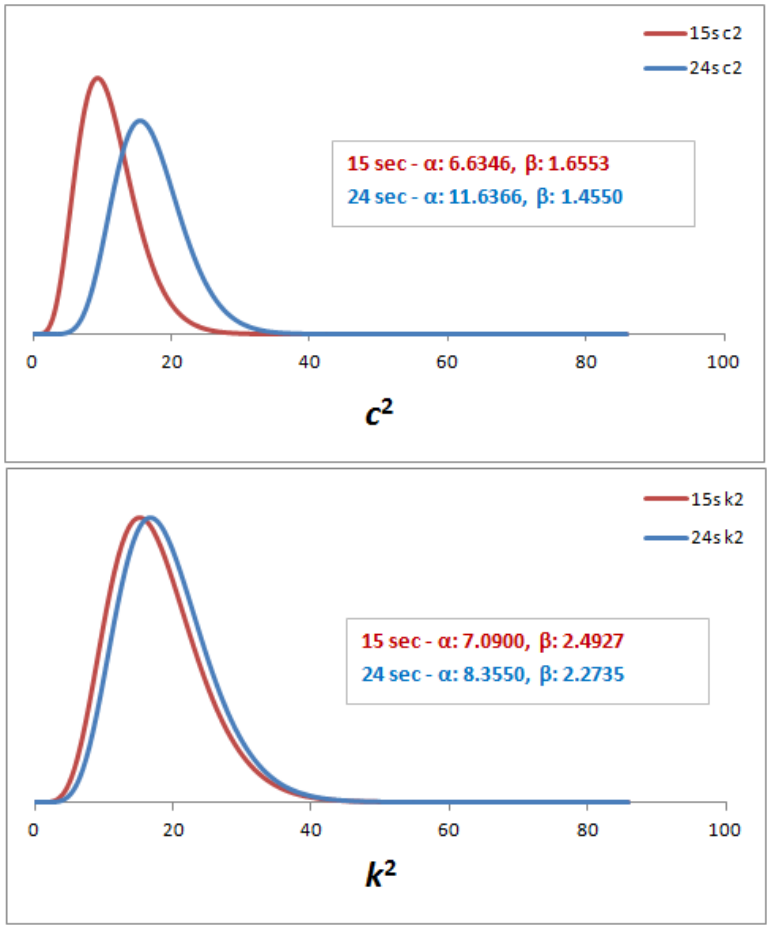

Figure 8 – The allele and stutter variance parameters.

Locus specific amplification efficiency parameter

The locus specific amplification efficiency parameter is calculated by ModelMaker using the same crime and reference file. Its calculation of this parameter is determined by the following equation:

15 seconds:

LSAE Variance - fitted gamma curve = exp(64.25832145393541) - mean = 0.015562186770111413

24 seconds:

LSAE Variance - fitted gamma curve = exp(56.143702296033844) - mean = 0.017811436708024916

Conclusions

All of the above sections describe the settings that will be used for mixture deconvolution using STRmix. There are two sets of settings to be used, depending on the amount of time the samples

are injected for. One set of settings should be applied to samples that are injected for 15 seconds (or less), and the second set of settings should be applied to samples injected for 24 seconds. These settings are summarized in Figures 9 and 10. While STRmix was supplied with default values for many commercially available kits, an internal validation of GlobalFiler on the 3500 was performed to reflect the local lab environment (i.e. products, instruments, and analysts) for two different injection times. This data was used to supply SDPD specific variance parameters to STRmix so it can optimally model DNA profiles that are generated in the SDPD crime lab. This validation study describes the use of ModelMaker in the SDPD lab, and the settings determined by it that should be used for mixture analysis in the lab.

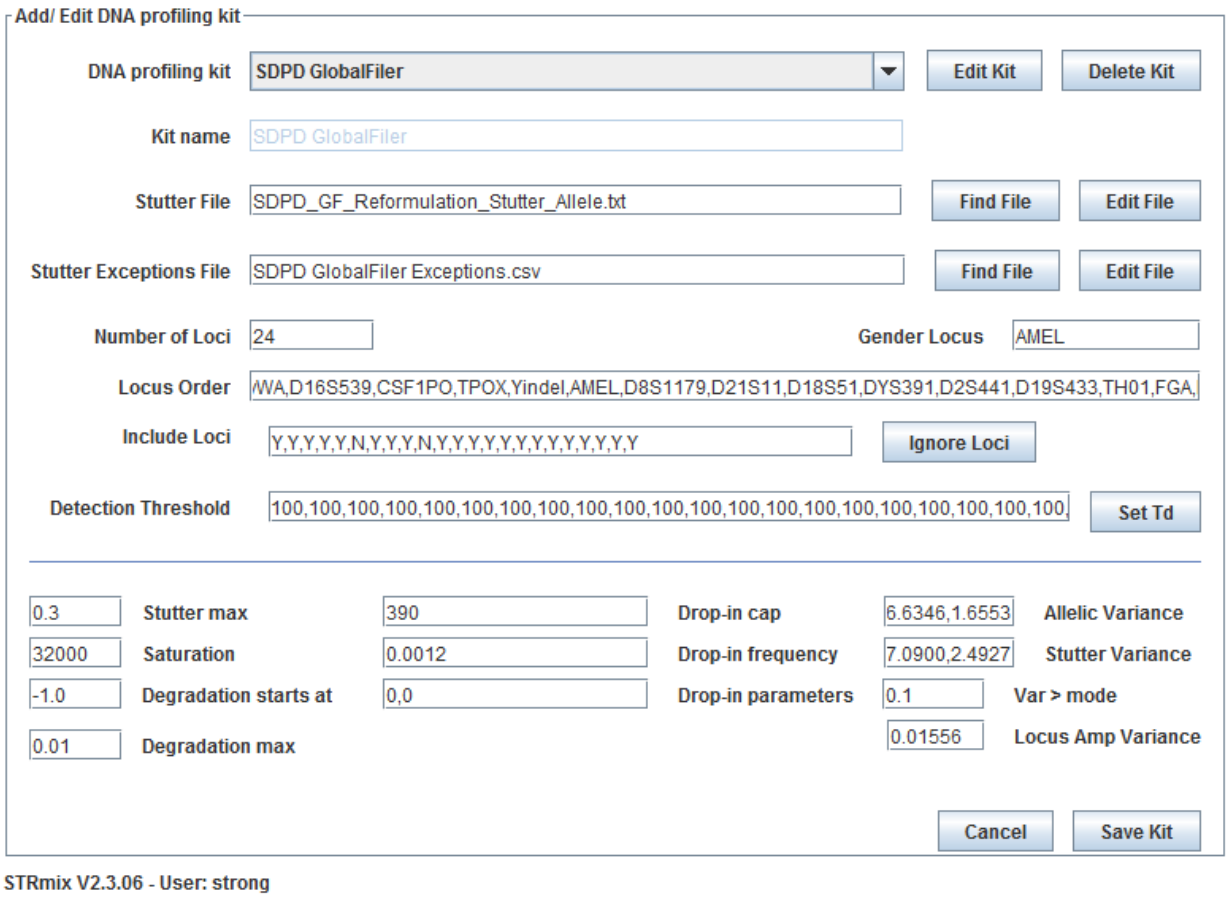

Figure 9 – SDPD ModelMaker settings (15 second injection) for v2.3.06 of STRmix.

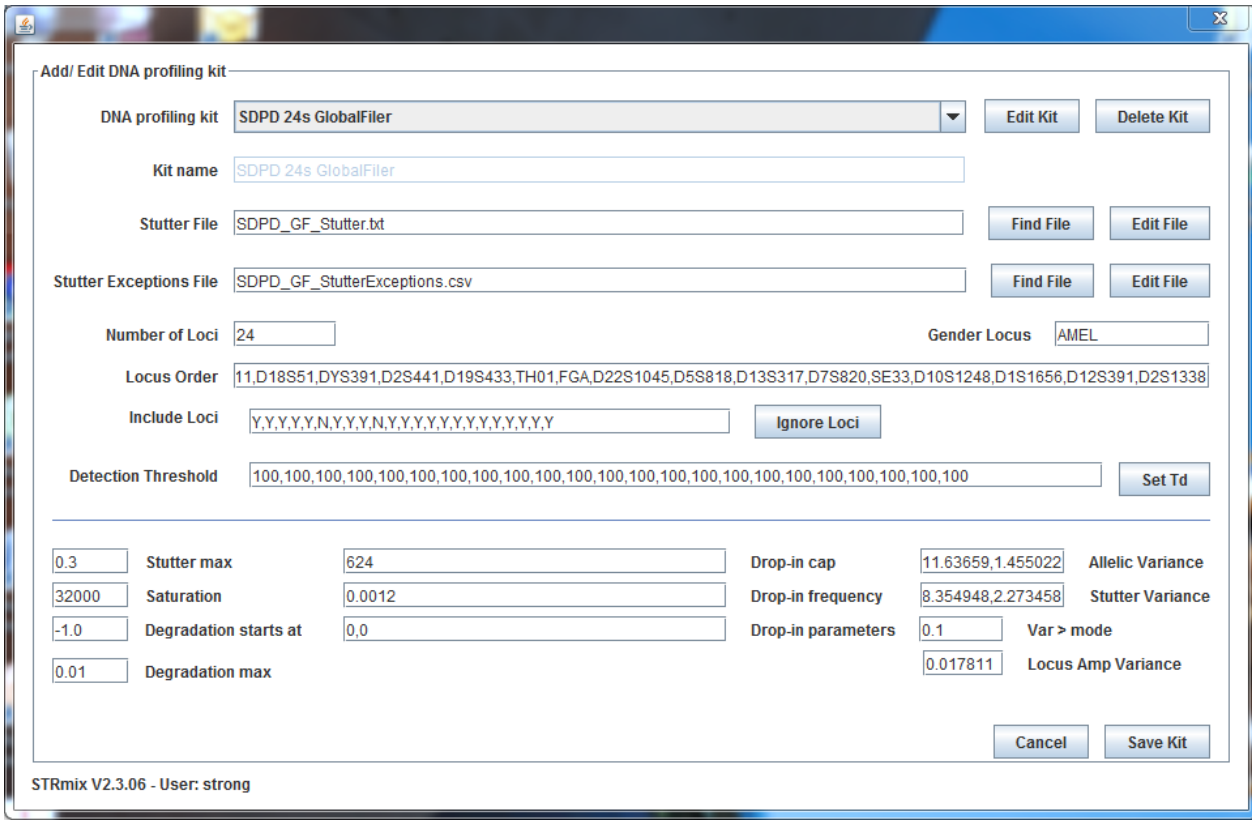

Figure 10 – SDPD ModelMaker settings (24 second injection) for v2.3.06 of STRmix.

Refernces

1. Butler, J. Advanced Topics in Forensic DNA Typing: Methodology. 2012 Academic Press.
2. Puch-Solis R. A dropin peak height model. Forensic Science International: Genetics. 2014;11:80-4.
3. STRmix v2.3 Users Manual. Issued by Institute of Environmental Science and Research Limited; Date of Issue: 20 January 2015
4. STRmix™ V2.3 Implementation and Validation Guide. Issued by Institute of Environmental Science and Research Limited; Date of Issue: 11 December 2014

SDPD Crime Laboratory – Forensic Biology Unit

Validation of the STRmix™ 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.

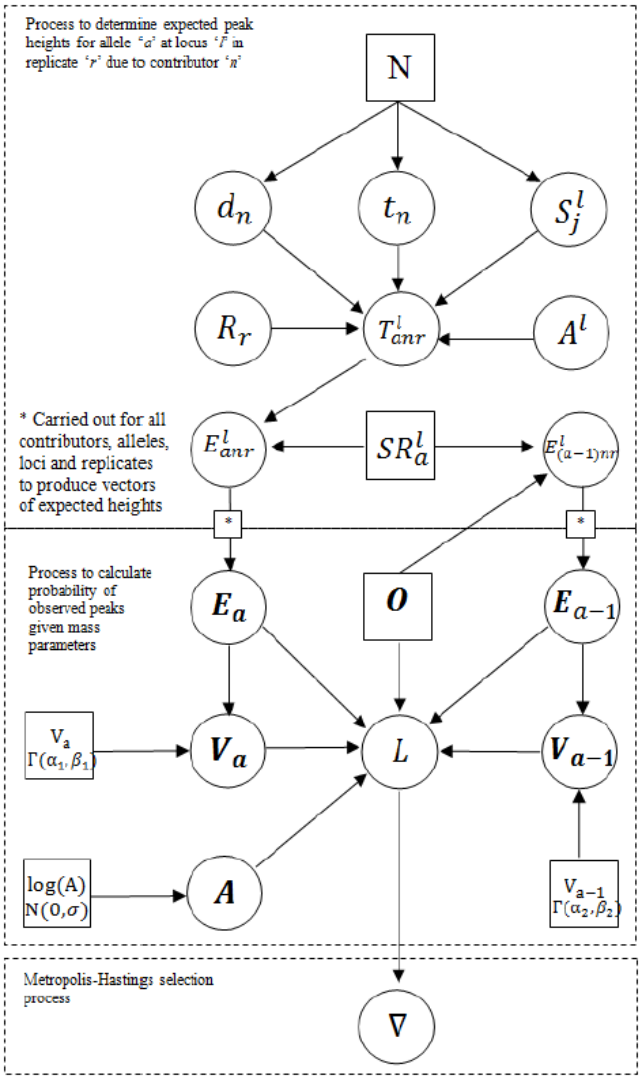

Figure 1 – A description of the MCMC process used by STRmix. Figure taken from the User's manual (1). The circles are modeled by STRmix, and the squares are input parameters by the user (See ModelMaker Validation Study for details about settings input by the user).

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.

$$T_{an}^{l} = A_{r}^{l} t_n X_{an}^{l} \times e^{d_n \times \left(mwt_a^l - offset\right)}$$

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)

$e^{d_n \times \left(mwt_a^l - offset\right)}$ 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:

Allele **Stutter**

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

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:

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

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

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

$$\Pr\left(O_a^l | E_a^l\right) \sim N\left(\log\left(\frac{O_a^l}{E_a^l}\right), 0, \frac{c^2}{E_a^l}\right)$$ (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 ~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* ≥ 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 $\leq$ 20% contributors had genotypes that did not fall into the top 99%. Likewise, 22 of these 201 contributors only contributed $\leq$ 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

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the STRmix™ Software

Sensitivity and Specificity
Comparison of deconvoluted mixtures to known contributors and known non-contributors

Introduction
STRmix uses biological models and takes the quantitative data from the electropherogram (for example peak heights) to calculate the probability of the peak heights given each of the possible genotype combinations for the individual contributors. 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 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 performs the deconvolution of for mixtures containing any number of contributors without reference to any known contributors or persons of interest (POI), unless specified by the defense hypothesis.

MCMC is based on a random number generation process. The MCMC trials numerous combinations of biological parameter values to describe the observed data and ultimately generates posterior distributions for each free parameter in the model. At each iteration of the MCMC STRmix™ builds a picture of an expected profile and ultimately compares it to the observed profile data to calculate a likelihood. In STRmix, the stepping from one MCMC guess to a better one is done using the Metropolis Hastings Algorithm (MHA). MHA compares two states, the current state and the proposed state. The algorithm considers whether to step to the proposed state or stay at the current state. If the proposed state has a higher probability density the chain always steps. If it has a lower probability density it will step some of the time. The step to a lower probability guess occurs in proportion to the ratio of the new (lower guess) to the old guess. Any proportion >0.5 means that you are more likely going to take that next step, and vice versa.

The variability in the weights produced during the MCMC process are template dependent. As the number of viable genotype options increases the weight determined for each combination spreads out across the combinations.

The performance of the MCMC output can be evaluated by examining the possible genotype combinations that make up the mixed result to determine whether the correct combination was arrived at (see the MCMC write-up). Another method of determining the effectiveness of the deconvolution is to compare the MCMC output to known contributors and non-contributors. Ideal performance would result in the inclusion of true contributors and the exclusion of non-

contributors. Examination of this can also assist in the determination of the limits of the STRmix software in obtaining accurate deconvolutions. In the establishment of STRmix into casework, false exclusions would be preferable to false inclusions.

Purpose

The goal of this study is to determine whether the MCMC deconvolutions result in correct inclusions when compared to a large number of subjects (specificity) and whether the deconvolutions remain effective as template decreases (sensitivity).

Materials and Methods

A subset of the 2-, 3-, and 4-person mixtures created as part of the GlobalFiler validation were used for this experiment. In the mixture study, different ranges of template DNA were targeted. The high level samples were prepared such that the average RFU for the highest percentage contributor was between 3K and 10K RFU. The mid-range samples were prepared such that the average RFU for the highest percentage contributor was between 1K and 3K RFU. The low level samples were prepared such that average RFU for the lowest percentage contributor had between 200 and 500 RFU (see GlobalFiler mixture study). Ten 2-person mixtures, seventeen 3-person mixtures, and sixteen 4-person mixtures were selected for use in this study. Of the ten 2-person mixtures selected for this study, three were high level, four were mid-range, and three were low level. Of the seventeen 3-person mixtures, six were high level, five were mid-range, and six were low level. Of the 4-person mixtures, five were high level, six were mid-range, and five were low level.

STRmix allows the user to search a deconvoluted DNA mixture against a database directly, without the need for deriving a single source component. The Database Search function can be used as a quality assurance tool for comparison of complex mixtures. The minimum LR value can be used as a list management tool to filter out of the results file all comparisons that lead to an LR below the specified value. The results summarize the input files, deconvolution and database search run conditions, and list of individuals whose comparison to the genotype PDF has yielded an LR above the defined cut-off value. The LR threshold set to 0 will return the results for all individuals within the database.

The calculation performed in the database search is not equivalent to when a POI is selected for comparison to a mixture. The allele frequency database used for the LR calculations can be selected from the list of available populations just as in any LR calculation, however; for the for a standard database match the calculations use a theta value = 0, returning the product rule only (p^2 for homozygotes and 2pg for heterozygotes) whereas theta is generally incorporated into LR calculations for POIs.

Also, in LR calculations concerning a POI, the propositions are variable and can be set by the user. In all database comparisons, the LR calculation are performed by considering the database individual (POI) in a 'N' person mixture by:

H_p = POI + N-1 unknowns

H_d = N unknowns.

A database file containing DNA profiles used in the validation was created based on the STRmix file requirements. The file contained 76 known DNA profiles. In general, without any additional thresholds applied, likelihood ratios greater than 1 favor inclusion to a given mixture while likelihood ratios between 0 and 1 favor exclusion (a.k.a., negative log likelihoods), and likelihood ratios of 0 indicate an exclusion.

Results and Discussion

2-person mixtures

Ten 2-person mixtures were deconvoluted using STRmix. After running the mixtures through STRmix the deconvolution results were compared to the database file of 76 individuals. Each 2-person mixture resulted in likelihood ratios favoring inclusion for the individuals known to comprise the mixtures. All other non-contributors in the database had likelihood ratios of zero (i.e., excluded).

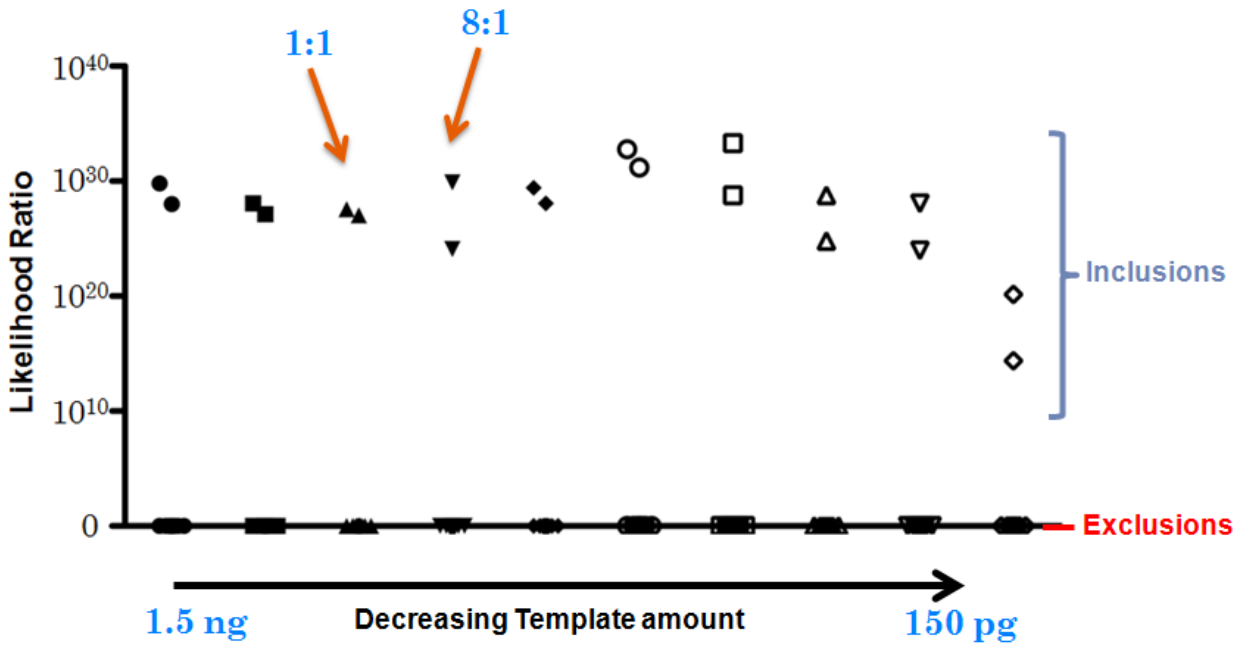

Figure 1 - 2-person mixtures

3-peron mixtures

Seventeen 3-person mixtures were deconvoluted using STRmix. After running the mixtures through STRmix the deconvolution results were compared to the database file of 76 individuals.

Several examples of 3-person mixtures are presented in Figure 2. These represent a high, mid, and low level samples, as well as a range of contributor combinations (balanced mixtures, mixtures where one person contributed a high percentage of the DNA, etc). The resultant likelihood ratios when compared to the 76 profiles in the database. Fourteen of the 3-person mixtures resulted in the correct inclusions of all individuals known to comprise the mixtures and the exclusion of all other profiles in the database. Three of the low level mixtures (mixture ID: 3-45, 3-53, and 3-61) had two of the three known contributors have likelihood ratios that favored inclusion, while the third contributor in each case had a negative log likelihood ratio (i.e., either inconclusive or suggesting exclusion).

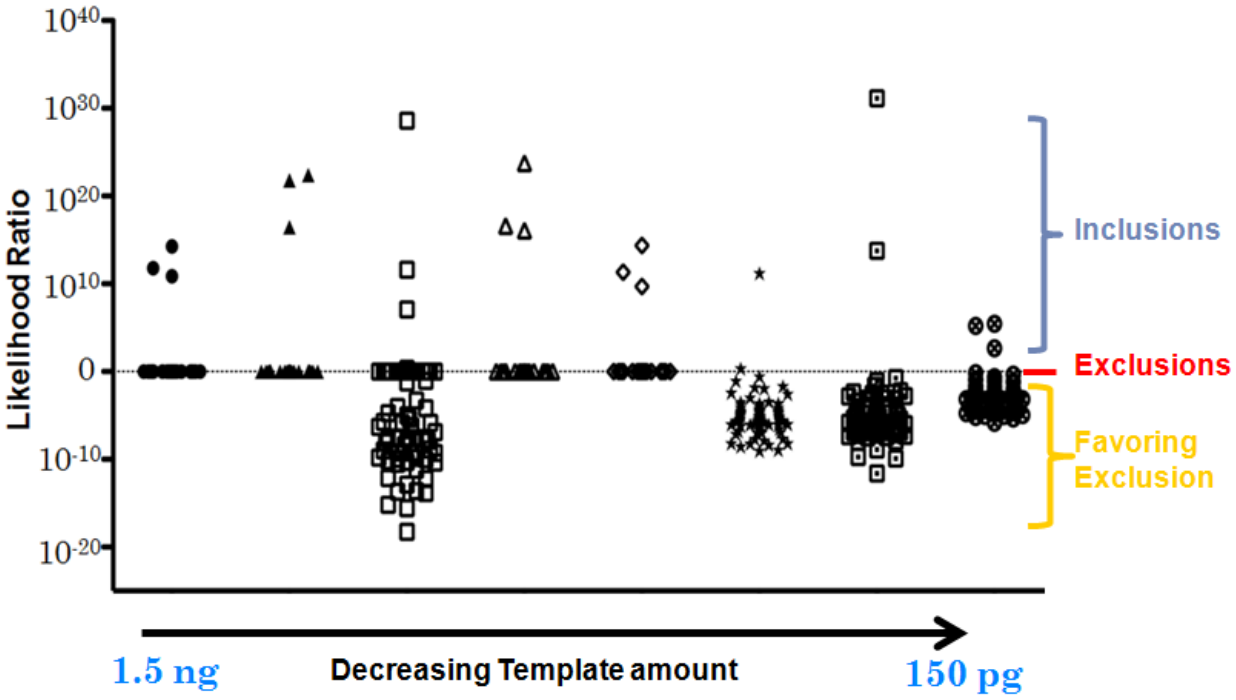

Figure 2 - 3-person mixtures

In eight of the mixtures, including two of the low level mixtures, all non-contributors had likelihood ratios of zero (i.e., exclusion). In the remaining nine mixtures, the non-contributors all had negative log likelihood ratios favoring exclusion.

One of the mixtures that excluded a known contributor (mixture ID: 3-45) is a low level 1:1:1 mixture with drop-out. Based on a review of the electropherogram, an observed contributor ratio could not be manually calculated. The mixture had extensive drop-out: more than 32 alleles were missing from the data set based on the genotypes of the known contributors. STRmix evaluated the profile as being comprised of a 40% contributor and two 30% contributors, consistent with the targeted 1:1:1 mixture. One true contributor had a likelihood ratio of 1.4×10^{11}; the second true contributor had a likelihood ratio of 2.06. The third true contributor, who had a negative log likelihood ratio suggesting exclusion, was missing 20 alleles total from the detected alleles in the

mixture, including complete genotype drop-out at several loci. Given the amount of missing information related to the third contributor, it is not surprising that a negative log likelihood ratio was obtained, or that all the other people in the database received likelihood ratios between zero and one. In fact, twelve non-contributors in the database had higher likelihood ratios than the known contributor (although still less than 1).

A second mixture that excluded a known contributor (mixture ID: 3-53) is a low level 50:30:20 mixture with drop-out. Based on a review of the electropherogram, the observed mixture ratio is approximately 48:24:24. One true contributor had a likelihood ratio of $1.9x10^{19}$; the second true contributor had a likelihood ratio of $2.2x10^{7}$. The third true contributor, who had a negative log likelihood ratio, was missing the most alleles of any contributor (11 alleles total) from the detected alleles in the mixture. Given the amount of missing information related to this contributor, it is not surprising that a negative log likelihood ratio was obtained, or that all the other people in the database had likelihood ratios between zero and one. No non-contributor in the database had a likelihood ratio higher than the true contributor with the negative log likelihood ratio.

The final three person mixture that excluded a known contributor (mixture ID: 3-61) is a low level 60:20:20 mixture with drop-out. Based on a review of the data, the observed mixture ratio is approximately 62:19:12. One true contributor had a likelihood ratio of $1.3x10^{31}$; the second true contributor had a likelihood ratio of $5.8x10^{13}$. The third true contributor who had a negative log likelihood ratio was missing the most alleles of any contributor (17 alleles total) from the detected alleles in the mixture, included six loci with complete genotype drop-out. Given the amount of missing information related to this contributor, it is not surprising that a negative log likelihood ratio was obtained, or that all the other people in the database had likelihood ratios between zero and one. Seven non-contributors in the database had a likelihood ratio higher than the true contributor with the negative log likelihood ratio.

4-person mixtures

Sixteen 4-person mixtures were deconvoluted using STRmix. These were mixtures that included high, mid, and low level mixtures with a range of contributor proportions. Seven of the mixtures had contributors with dropout. After running the mixtures through STRmix, the deconvolution results were compared to the database file of 76 individuals. Three examples of 3-person mixtures are presented in Figure 3. These represent three high level samples and the resultant likelihood ratios when compared to the 76 profiles in the database.

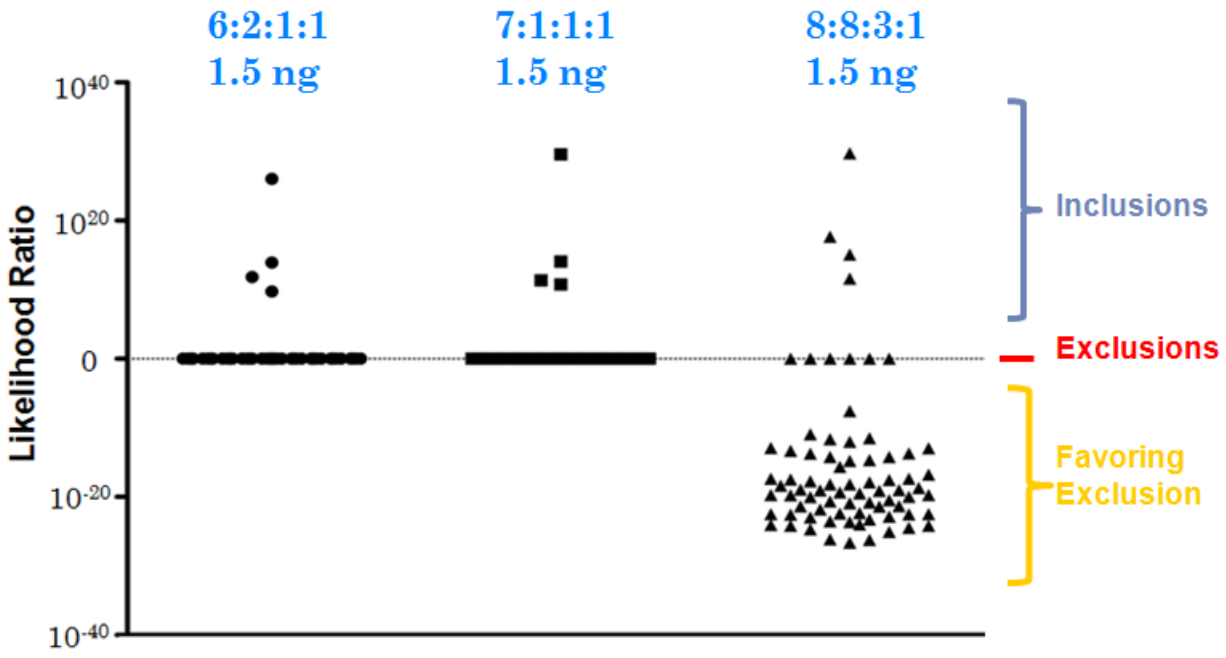

Figure 3 - 4-person mixtures

All sixteen of the 4-person mixtures resulted in the inclusions of the individuals known to comprise the mixtures. In addition to the correct inclusions, one mixture (mixture ID: 4-63) also had a single non-contributing profile from the database that also resulted in a likelihood ratio that favored inclusion. The likelihood ratio for the non-contributor was 37. In comparison, the smallest likelihood ratio for a true contributor was 530. When examined more closely, the profile for the non-contributor that received the inclusionary likelihood ratio, there were fifteen loci where the non-contributor was completely represented, five loci where only one allele from the non-contributor was present, and a single locus where the genotype was not represented at all. When mixture 4-63 was re-run conditioned with the contribution of 2 of the contributors (the two with the highest likelihood ratios), the profile that was falsely included received a likelihood ratio that strongly favored exclusion ($3x10^{-10}$). All other non-contributors in the database received likelihood ratios favoring exclusion.

Conclusions

The specificity and sensitivity of the deconvolutions produced by the MCMC process were evaluated using the database search functionality of STRmix. Comparing the deconvolution results against the seventy-six profiles in the database allowed for determining whether accurate inclusions and exclusions could be obtained from a variety of different mixture types and levels.

Correct inclusions were obtained for all 43 mixtures when compared against profiles from 76 profiles in the database. Assessing inclusions based on likelihood ratios greater than 1, only one false inclusion was obtained.

There were three instances (in mixtures with a high amount of drop-out) where one of the three contributors received likelihood ratios that favored exclusion. In each of these scenarios the amount of drop-out is likely the reason for the exclusionary likelihood ratio.

STRmix deconvolutions have been demonstrated to be very robust. There is a high degree of specificity as established by the high level of accuracy of the inclusions and exclusions. The 43 mixtures comprised of 135 contributors in the tested mixtures, 132 (or 99.7%) of the true contributors had likelihood ratios that favored inclusion when the deconvolutions were compared to the database. When all 43 mixtures were compared against the database there were 10260 comparisons performed. Considering 135 of those comparisons were to true contributors, 10124 comparisons (or 99.01%) correctly resulted in likelihood ratios that favored exclusion.

The only mixtures where false exclusions were obtained were all low level mixtures that contained extensive drop-out. These mixtures all had average peak heights below 600rfu. Mixture 3-45 had a peak height average of 255.6rfu, mixture 3-53 had a peak height average of 380.3rfu; mixture 3-61 had a peak height average of 472.8.6rfu, and mixture 4-63 had a peak height average of 512.5rfu. The peak height below which there is a concern regarding drop-out was estimated to be approximately 600rfu. Given that these mixtures were such a low level the deconvolution results are still fairly robust given that only a single false inclusion was obtained, albeit at a relatively low likelihood ratio when compared to the true contributors. In addition, the false exclusions were not surprising given the amount of drop-out in the mixtures. Running these mixtures conditioned against possible contributors with high likelihood ratios may be a mechanism for determining whether possible contributors with low likelihood ratios are actually included in the mixture.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the STRmix™ Software

Reproducibility
Evaluating the variability in the MCMC process

Introduction
STRmix uses biological models and takes the quantitative data from the electropherogram (for example peak heights) to calculate the probability of the peak heights given each of the possible genotype combinations for the individual contributors. STRmix performs the deconvolution of for mixtures containing any number of contributors without reference to any known contributors or persons of interest (POI). 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 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).

MCMC is based on a random number generation process. The MCMC trials numerous combinations of biological parameter values to describe the observed data and ultimately generates posterior distributions for each free parameter in the model. At each iteration of the MCMC STRmix™ builds a picture of an expected profile and ultimately compares it to the observed profile data to calculate a likelihood. In STRmix™, the stepping from one MCMC guess to a better one is done using the Metropolis Hastings Algorithm (MHA). MHA compares two states, the current state and the proposed state. The algorithm considers whether to step to the proposed state or stay at the current state. If the proposed state has a higher probability density the chain always steps. If it has a lower probability density it will step some of the time. The step to a lower probability guess occurs in proportion to the ratio of the new (lower guess) to the old guess. Any proportion >0.5 means that you are more likely going to take that next step, and vice versa.

The variability in the weights produced during the MCMC process is template dependent. As the number of viable genotype options increases, the weight determined for each combination spreads out across the combinations.

Several known sources of variation in the DNA analysis process are already known. These include the DNA extraction and purification process where slightly different concentrations of DNA will be obtained from replicate samples, the PCR process, and the electrokinetic injections in capillary electrophoresis. The MCMC is a new source of variability within the forensic DNA analysis process. Since the MCMC process is random, each time a mixture is run through STRmix it produces a different answer each time it is run. This variability is at its largest when

the likelihood ratio is low, if the number of iterations is insufficient to solve the problem, or if the problem is complex. Variability from four different sources (replicate CE injections, replicate CE plate loads of same amplified DNA, replicate PCR amplifications, and replicate LR calculations) was measured in the developmental validation. Of all these, the MCMC process displayed the least variability.

Purpose

The goal of this study is to determine the MCMC variability for replicate analyses of the same mixtures made in our internal validation study.

Materials and Methods

A subset of the 2-, and 3-person mixtures created as part of the GlobalFiler validation were used for this experiment. 4-person mixtures were not selected for this study due to the length of time those mixtures required to run in STRmix. In the mixture study, different ranges of template DNA were targeted. The high level samples were prepared such that the average RFU for the highest percentage contributor was between 3K and 10K RFU. The mid-range samples were prepared such that the average RFU for the highest percentage contributor was between 1K and 3K RFU. The low level samples were prepared such that average RFU for the lowest percentage contributor had between 200 and 500 RFU (see GlobalFiler mixture study). Ten 2-person mixtures and seventeen 3-person mixtures were selected that spanned the range of template amounts from robust to low level. Each of the mixtures were run through STRmix four separate times.

The following table contains the mixture ratios for the samples that were used in this study. Of the 2-person mixtures: three were robust, four were mid-level, and three were low level. In the 3-person mixture dataset: six were robust mixtures, five were mid-level, and six were low level mixtures.

Sample	Mix ratio	Sample	Mix Ratio
2-3	2 : 1	**3-1**	1 : 1 : 1
2-7	1 : 8	**3-5**	6 : 3 : 1
2-11	1 : 1	**3-9**	5 : 3 : 2
2-15	8 : 1	**3-13**	2 : 2 : 1
2-19	1 : 2	**3-17**	3 : 1 : 1
2-23	5 : 1	**3-21**	4 : 3 : 3
2-27	1 : 5	**3-25**	7 : 2 : 1
2-31	2 : 1	**3-29**	5 : 4 : 1
2-35	1 : 8	**3-33**	4.5 : 4.5 : 1
2-39	1 : 1	**3-37**	1 : 1 : 1
		3-41	2 : 1 : 1
		3-45	1 : 1 : 1

	3-49	6 : 3 : 1
	3-53	5 : 3 : 2
	3-57	2 : 2 : 1
	3-61	3 : 1 : 1
	3-65	4 : 3 : 3

The individual weights assigned to each genotype possibility are derived from the proportion of time each possibility is accepted as a better answer than a separate guess. The variability in the weights is derived from the randomness of each guess in the MCMC process. As such, there are a couple of means to verify the reproducibility of the MCMC process. One method would be to check the assigned weights of each possible genotype at each marker and compare them across replicates. Another means of checking this would be done based on the examining the likelihood ratios for contributors to the mixtures. Likelihood ratios are calculated in STRmix by taking into account the various possible genotypes for each contributor including the weights assigned to each possibility. Since the weights of each possibility are taken into account, the value of the likelihood ratio is reflective of the differences in genotype weight and can be used as a measure of the reproducibility of the MCMC process.

To assess the differences in the likelihood ratio data, the highest likelihood ratio result was divided by the lowest likelihood ratio result with the replicates. Reproducibility would be determined in this method by examining for results that larger than 1. A result of 1 in this method would indicate that there was no variation in the MCMC results between the replicates. The larger the result from this method would indicate more variability in the MCMC result and corresponding genotype weights.

Results and Discussion

2-person mixtures

Overall, the 2-person mixtures resulted likelihood ratios demonstrated good reproducibility (low variability) between replicate runs. The data indicates that as the amount of template goes down, or as the contribution level of an individual contributor goes down the variability of the MCMC goes up. Using the value of 1 to indicate no variation in the MCMC results, all values obtained were within a 1.44 fold difference. In the ten 2-person mixtures, there was no difference (highest LR – lowest LR = 1), or very little difference (highest LR – lowest LR = 1.01) for nine of the twenty total contributors. Six of the twenty total contributors had differences in LRs in the range between 1.15 and 1.30. These six contributors were to mixtures that spanned the range of template DNA used in this study and also spanned different percent contributions from 11% to 64%. The last five contributors had differences in the range between 1.31 to 1.44. In the remaining four of these, this occurred in the contributors that were donating 10% of the DNA to the sample and that were in the range where stutter and drop-out were possible. The last instance where there was a high difference occurred in a 1:1 mixture where the majority of the peaks were

below 600 RFU. A visual representation of the data is presented in Figure 1 (helpful for seeing the magnitude of change in context of the LR scale) and the actual LR replicate data itself is presented in Table 1.

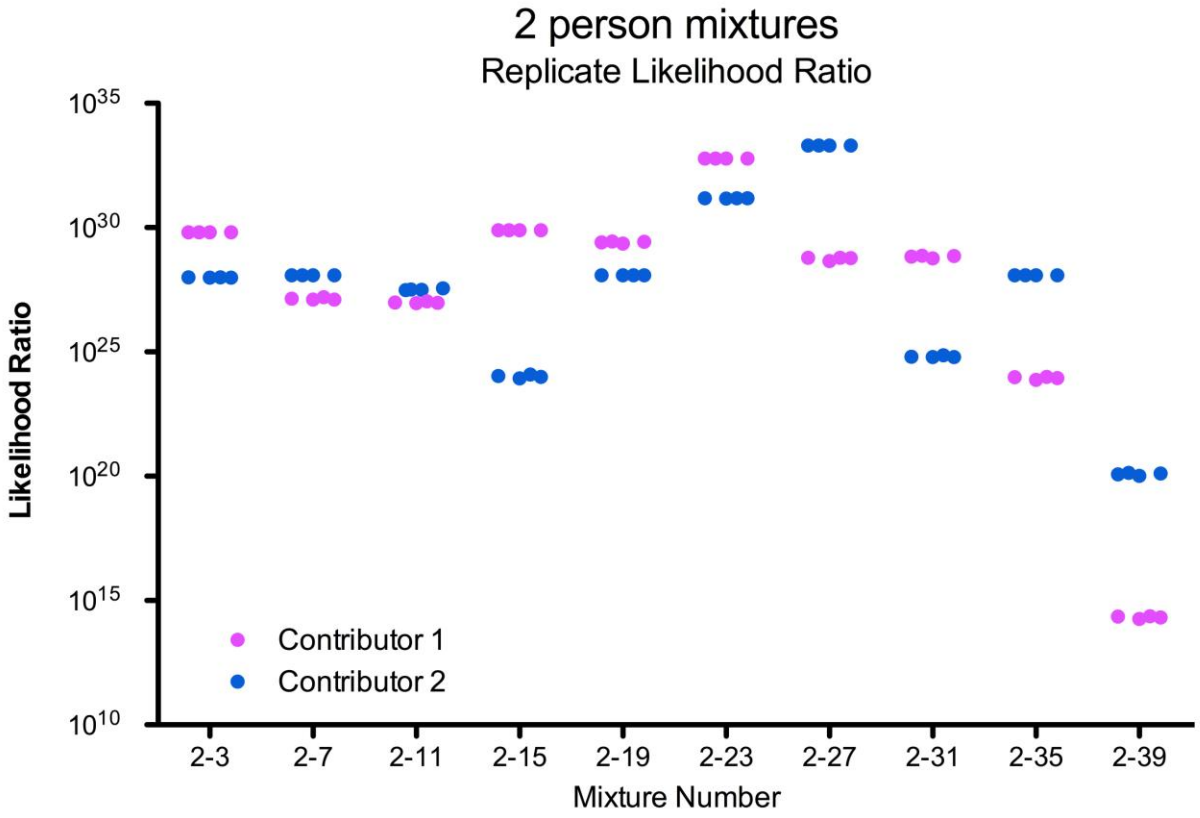

Figure 1 – Likelihood ratios for 2-person mixtures

3-peron mixtures

The 3-person mixtures displayed more variability than the 2-person mixtures. This would be expected because as the number of alleles increases the number of possible genotypes that could be contributing to the mixture also increases. Overall, the same basic data trends were observed as compared to the 2-person mixtures. The data indicates that as the amount of template goes down, or as the contribution level of an individual contributor goes down the variability of the MCMC goes up.

Additional sources of variability were observed when there was a similarity between the contributions of donors to the mixture and when donors shared alleles. For example, in mixture 3-9 the two major donors had higher variability than the minor donor. In this mixture, the two strong donors were contributing approximately 47% and 38% and had numerous loci where there was stacking of shared alleles. This sharing of alleles increased the number of possible genotypes that could have been contributing, thereby increasing the variability of the MCMC in each replicate as compared to the minor contributor who shared less alleles overall with the two other

donors to the mixture. Mixture 3-13 also displayed similar sources of variability. In this mixture the contributor with the least overlap with other contributors, which was also one of the strongest contributors, had the lowest variability. The variability observed in the other contributor genotypes was due to the low level of one of the contributors and the amount of overlap between the three contributors.

A visual representation of the data is presented in Figure 2 and the actual data itself is presented in Table 2. Note that even in samples with a large fold difference, the difference is relatively small with regard to the overall magnitude of change.

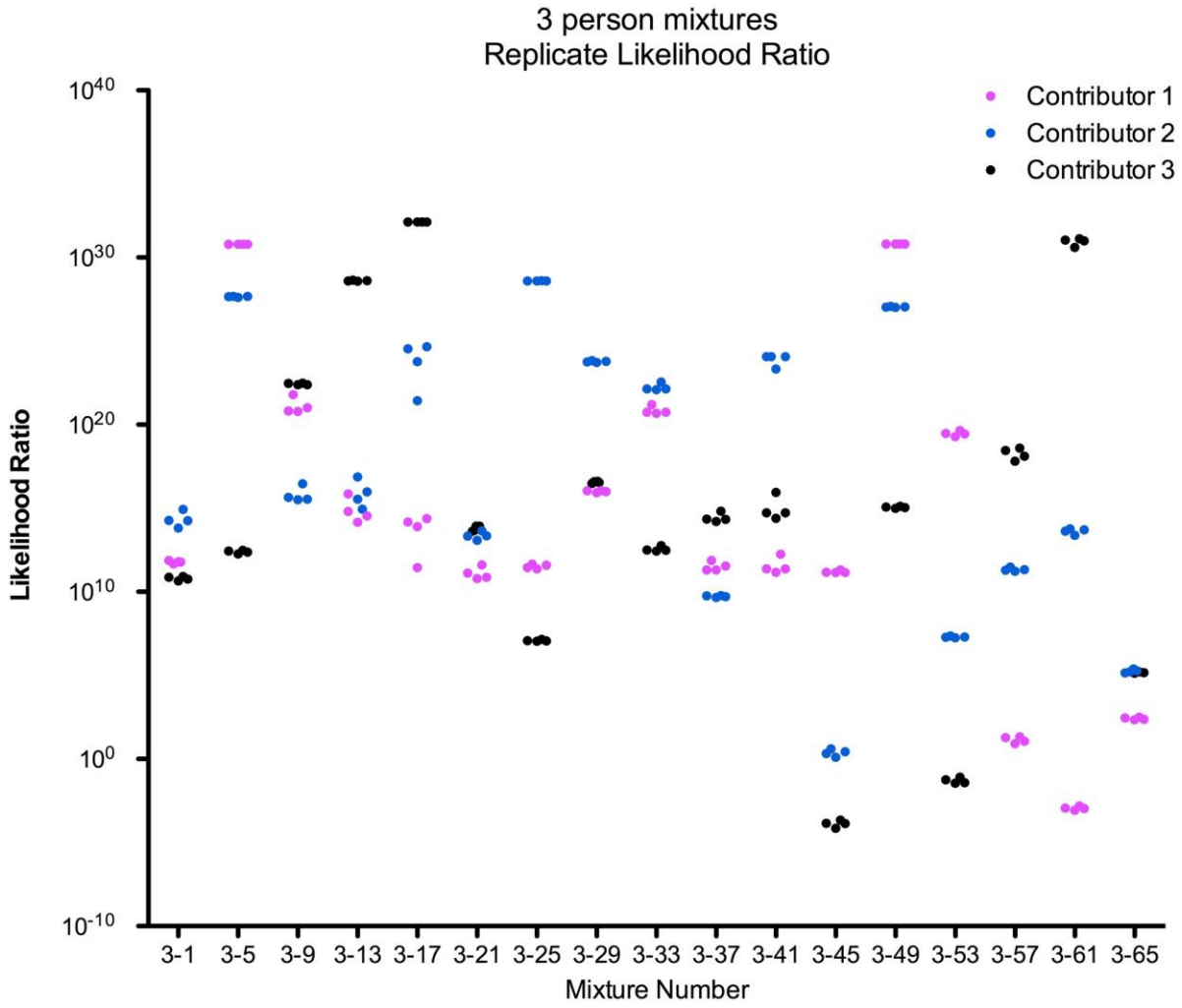

Figure 2 - 3-person mixtures

Conclusions

Overall, the results of the mixture deconvolution displayed very good reproducibility. The variability in the mixture results generally increased as the number of genotype possibilities that could explain the data increased. The data suggests that as the total amount of template DNA is

decreased, or the contribution level of a contributor is lowered, the variability increases. In addition, the variability increased when the contributors, even robust level contributors, had similar contribution levels to other contributors in the mixture, or when there was a large amount of allele sharing between the contributors.

Table 1

Mixture Contributor		Replicate							
2-3	% Contribution	A	B	C	D	Average	Max	Min	Fold Difference
150-TD	68	6.45E+29	6.40E+29	6.46E+29	6.48E+29	6.44E+29	6.48E+29	6.40E+29	1.01
156-RK	32	9.84E+27	9.76E+27	9.85E+27	9.87E+27	9.83E+27	9.87E+27	9.76E+27	1.01
2-7									
150-TD	11	1.27E+27	1.39E+27	1.59E+27	1.29E+27	1.39E+27	1.59E+27	1.27E+27	1.26
156-RK	89	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.00
2-11									
62-BS	36	9.17E+26	9.81E+26	9.43E+26	1.08E+27	9.81E+26	1.08E+27	9.17E+26	1.18
31-BB	64	3.07E+27	3.28E+27	3.15E+27	3.62E+27	3.28E+27	3.62E+27	3.07E+27	1.18
2-15									
150-TD	90	7.86E+29	7.86E+29	7.86E+29	7.86E+29	7.86E+29	7.86E+29	7.86E+29	1.00
156-RK	10	8.53E+23	1.00E+24	1.06E+24	1.23E+24	1.04E+24	1.23E+24	8.53E+23	1.44
2-19									
150-TD	23	2.30E+29	2.66E+29	2.79E+29	2.55E+29	2.58E+29	2.79E+29	2.30E+29	1.22
156-RK	77	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.00
2-23									
62-BS	77	5.95E+32	5.95E+32	5.95E+32	5.95E+32	5.95E+32	5.95E+32	5.95E+32	1.00
31-BB	23	1.49E+31	1.50E+31	1.51E+31	1.51E+31	1.50E+31	1.51E+31	1.49E+31	1.01
2-27									
62-BS	10	4.46E+28	6.16E+28	6.07E+28	5.95E+28	5.66E+28	6.16E+28	4.46E+28	1.38
31-BB	90	1.99E+33	1.99E+33	1.99E+33	1.99E+33	1.99E+33	1.99E+33	1.99E+33	1.00
2-31									
150-TD	67	7.20E+28	7.41E+28	6.74E+28	5.85E+28	6.80E+28	7.41E+28	5.85E+28	1.27
156-RK	33	6.32E+24	7.41E+24	6.27E+24	6.23E+24	6.56E+24	7.41E+24	6.23E+24	1.19
2-35									
150-TD	10	8.97E+23	9.82E+23	7.51E+23	9.51E+23	8.96E+23	9.82E+23	7.51E+23	1.31
156-RK	90	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.20E+28	1.00
2-39									
62-BS	51	1.82E+14	2.06E+14	2.26E+14	2.33E+14	2.12E+14	2.33E+14	1.82E+14	1.28
31-BB	49	1.04E+20	1.30E+20	1.19E+20	1.37E+20	1.22E+20	1.37E+20	1.04E+20	1.32

Table 2

Mixture Contributor		Replicate				Average	Max		
3-1	% Contribution	A	B	C	D			Min	Difference
131-SB	36	6.03E+11	7.57E+11	6.30E+11	4.68E+11	6.14E+11	7.57E+11	4.68E+11	1.62
142-VM	33	1.83E+14	8.58E+14	1.88E+14	6.60E+13	3.24E+14	8.58E+14	6.60E+13	13.00
40-DB	32	7.48E+10	8.32E+10	5.67E+10	4.43E+10	6.48E+10	8.32E+10	4.43E+10	1.88
3-5									
151-TR	73	6.28E+30	6.28E+30	6.28E+30	6.27E+30	6.28E+30	6.28E+30	6.27E+30	1.00
39-DC	23	4.76E+27	4.70E+27	4.60E+27	4.08E+27	4.54E+27	4.76E+27	4.08E+27	1.17
81-AR	3	1.83E+12	2.73E+12	3.05E+12	2.36E+12	2.49E+12	3.05E+12	1.83E+12	1.67
3-9									
23-JS	47	6.22E+21	1.02E+21	6.58E+20	6.24E+20	2.13E+21	6.22E+21	6.24E+20	9.97
39-DC	38	2.89E+16	4.49E+15	3.53E+15	3.19E+15	1.00E+16	2.89E+16	3.19E+15	9.05
79-PV	15	2.43E+22	2.46E+22	2.90E+22	3.06E+22	2.71E+22	3.06E+22	2.43E+22	1.26
3-13									
102-EB	15	6.29E+14	7.13E+15	3.48E+14	1.46E+14	2.06E+15	7.13E+15	1.46E+14	48.89
156-RK	44	9.42E+15	7.43E+16	3.48E+15	8.95E+14	2.20E+16	7.43E+16	8.95E+14	83.04
61-KM	41	4.33E+28	3.99E+28	3.77E+28	4.13E+28	4.05E+28	4.33E+28	3.77E+28	1.15
3-17									
103-KD	6	1.51E+14	7.84E+13	2.40E+14	2.80E+11	1.18E+14	2.40E+14	2.80E+11	856.36
125-CN	13	3.43E+24	5.96E+23	4.60E+24	2.68E+21	2.16E+24	4.60E+24	2.68E+21	1714.46
127-DF	81	1.32E+32	1.32E+32	1.32E+32	1.32E+32	1.32E+32	1.32E+32	1.32E+32	1.00
3-21									
115-RM	29	7.29E+10	6.19E+10	1.31E+11	4.04E+11	1.67E+11	4.04E+11	6.19E+10	6.53
39-DC	29	2.18E+13	1.19E+13	2.23E+13	4.40E+13	2.50E+13	4.40E+13	1.19E+13	3.69
71-LW	42	8.51E+13	8.45E+13	4.25E+13	4.96E+13	6.54E+13	8.51E+13	4.25E+13	2.00
3-25									
106-BL	11	3.87E+11	2.84E+11	2.34E+11	4.51E+11	3.39E+11	4.51E+11	2.34E+11	1.93
125-CN	77	4.03E+28	4.00E+28	3.91E+28	3.96E+28	3.98E+28	4.03E+28	3.91E+28	1.03
155-JK	12	1.15E+07	1.43E+07	1.19E+07	1.10E+07	1.22E+07	1.43E+07	1.10E+07	1.31
3-29									
71-LW	45	1.15E+16	1.09E+16	1.00E+16	8.50E+15	1.02E+16	1.15E+16	8.50E+15	1.35
81-AR	10	5.66E+23	6.01E+23	5.31E+23	6.68E+23	5.92E+23	6.68E+23	5.31E+23	1.26
93-RC	45	4.00E+16	3.85E+16	3.55E+16	3.00E+16	3.60E+16	4.00E+16	3.00E+16	1.34
3-33									
161-GZ	47	5.50E+20	5.52E+20	4.85E+20	1.58E+21	7.91E+20	1.58E+21	4.85E+20	3.26
30-SS	44	1.39E+22	1.39E+22	1.22E+22	3.53E+22	1.88E+22	3.53E+22	1.22E+22	2.90
88-KB	9	3.15E+12	2.78E+12	3.06E+12	5.78E+12	3.69E+12	5.78E+12	2.78E+12	2.08
3-37									

156-RK	33	2.06E+11	7.49E+11	2.02E+11	3.54E+11	3.78E+11	7.49E+11	2.02E+11	3.71
159-MS	34	5.04E+09	5.66E+09	4.62E+09	5.88E+09	5.30E+09	5.88E+09	4.62E+09	1.27
79-PV	33	2.23E+14	6.66E+14	1.61E+14	2.13E+14	3.16E+14	6.66E+14	1.61E+14	4.14
3-41									
115-RM	25	1.47E+11	2.33E+11	2.33E+11	1.80E+12	6.02E+11	1.80E+12	1.47E+11	12.22
125-CN	49	2.14E+23	1.17E+24	1.17E+24	1.17E+24	9.34E+23	1.17E+24	2.14E+23	5.49
79-PV	26	2.51E+14	5.30E+14	5.30E+14	8.93E+15	2.56E+15	8.93E+15	2.51E+14	35.54
3-45									
131-SB	39	1.46E+11	1.40E+11	1.42E+11	2.03E+11	1.58E+11	2.03E+11	1.40E+11	1.45
142-VM	30	2.067522	1.264069	3.968647	2.692089	2.50E+00	3.97E+00	1.26E+00	3.14
40-DB	30	0.000216	7.07E-05	0.000142	0.000137	1.41E-04	2.16E-04	7.07E-05	3.05
3-49									
151-TR	76	6.31E+30	6.31E+30	6.31E+30	6.31E+30	6.31E+30	6.31E+30	6.31E+30	1.00
39-DC	20	1.03E+27	1.12E+27	1.07E+27	1.18E+27	1.10E+27	1.18E+27	1.03E+27	1.15
81-AR	4	1.16E+15	1.11E+15	1.30E+15	9.91E+14	1.14E+15	1.30E+15	9.91E+14	1.31
3-53									
23-JS	35	1.94E+19	4.40E+19	3E+19	2.82E+19	3.04E+19	4.40E+19	1.94E+19	2.27
39-DC	32	2.25E+07	1.79E+07	19246381	19955011	1.99E+07	2.25E+07	1.79E+07	1.26
79-PV	33	0.080809	0.056229	0.037615	0.035562	5.26E-02	8.08E-02	3.56E-02	2.27
3-57									
102-EB	34	8.28E+00	21.44717	18.78794	11.54093	1.50E+01	2.14E+01	8.28E+00	2.59
156-RK	33	1.94E+11	1.69E+11	3.06E+11	2.13E+11	2.21E+11	3.06E+11	1.69E+11	1.81
61-KM	33	6.43E+17	1.27E+18	2.82E+18	3.95E+18	2.17E+18	3.95E+18	6.43E+17	6.15
3-61									
103-KD	14	0.001534	0.000877	0.001166	0.001078	1.16E-03	1.53E-03	8.77E-04	1.75
125-CN	14	5.85E+13	4.26E+13	2.36E+13	5.08E+13	4.39E+13	5.85E+13	2.36E+13	2.48
127-DF	71	1.31E+31	1.12E+31	3.99E+30	9.86E+30	9.54E+30	1.31E+31	3.99E+30	3.30
3-65									
115-RM	33	288.936	216.9714	319.9085	236.2279	2.66E+02	3.20E+02	2.17E+02	1.47
39-DC	33	160305.8	189182.8	140659.7	241323.3	1.83E+05	2.41E+05	1.41E+05	1.72
71-LW	33	144598.4	171670.6	134037.3	162322.6	1.53E+05	1.72E+05	1.34E+05	1.28

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the STRmix™ Software

Likelihood Ratio Calculation Verification

Introduction

STRmix™ uses possible genotype combinations and a biological model to create expected profiles that are compared to the quantitative data from the electropherogram (for example peak heights). STRmix™ then calculates the probability of the peak heights given the selected mass parameters values. Using Markov chain Monte Carlo (MCMC) methods with a Metropolis-Hastings algorithm, STRmix solves for genotype combinations that explain the observed data set, generating a list of genotype sets and assigning weights to each set that reflect how well they 'fit' the observed evidence data. If the proposed combination of 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 performs the deconvolution of for mixtures containing any number of contributors without reference to any known contributors or persons of interest (POI), unless specified by the defense hypothesis.

Following deconvolution of evidence samples, STRmix has the capability of generating likelihood ratios (LRs) to determine the degree to which the evidence suggests that an individual, or a group of individuals, contributed DNA to the evidence sample. The likelihood ratios are calculated by finding the ratio of the probability of obtaining the evidence under an inclusionary hypothesis, H_1 (also referred to as the prosecutor's hypothesis, H_p) to the probability of obtaining the evidence under an exclusionary hypothesis, H_2 (also referred to as the defense attorney's hypothesis, H_d). STRmix calculates these probabilities by summing the products of the weights (w_j or w_u) for each genotype set obtained during the MCMC process and the probabilities of each corresponding genotype set (S_j or S_u) under the competing hypotheses:

$$LR_C = \frac{\sum_j w_j \Pr(S_j \mid H_1)}{\sum_u w_u \Pr(S_u \mid H_2)}$$

The probabilities of the genotype sets are products of the probabilities of each genotype within the set under the given hypothesis. If Hardy Weinberg equilibrium were assumed, the probabilities for each genotype within the set would be:

- 0: for a genotype within the set, which cannot be obtained under the hypothesis
- 1: for a genotype within the set, which his assumed under the hypothesis
- $2p_ip_j$: for a genotype within the set from a randomly selected heterozygote: P_i,P_j
- p_i^2: for a genotype within the set from a randomly selected homozygote: P_i,P_i

Because population substructure is known to exist, the frequency of each allele (p_a) for the randomly selected heterozygote and randomly selected homozygote are adjusted, per Balding and Nichols, using the following formula:

$$\frac{\left[\left(m_A\theta + \left(1-\theta\right)p_a\right)\right]}{\left[m\theta + \left(1-\theta\right)\right]}$$

where m_A is the number of times that allele A has been previously observed, m is the number of times any allele has been observed, and θ is the coancestry coefficient Note that m_A and m differ for each observed allele within a genotype set, such that the adjustments for each allele in a homozygote genotype differ and the adjustment for the same allele observed in two different contributors within the genotype set also differ.

The LR calculations use posterior mean allele frequencies generated from the NIST population databases for the African American, Asian, Caucasian, and Hispanic population groups. The posterior mean frequencies are calculated using the following formula:

$$p(A) = \frac{x_a + 1/k}{N+1}$$

Where x_a is the known count of allele *'a'*, k is the number of observed allele classes at the locus, and N is the total number of alleles observed in the population data.

The LR calculated in STRmix is based on the posterior mean allele frequencies, a selected value of θ, and the weights assigned during the MCMC process. Because of uncertainty, each of these three factors could be explained by a distribution of values. In a method referred to as Highest Posterior Density (HPD), STRmix can calculate numerous LRs by selecting from a distribution of values for any or all of these factors, thereby generating a distribution of LRs likely to contain the true value of the LR. At the SDPD crime lab, we will perform HPD calculations using sampling of the weights and the posterior mean frequencies, but will select a single, conservative value for θ for each population group (0.01 for African Americans, Caucasians and Hispanics, 0.02 for Asians). We will select the 99% one sided, lower bound of the LR distribution to report for HPD calculations.

STRmix can calculate LRs for various relatives and can report a unified LR incorporating both related and unrelated individuals. STRmix can also take into account both the relative proportions of the population groups and the LR calculated for each population go give a single, stratified LR. SDPD currently does not plan to report unified or stratified LRs.

Purpose

The goal of this study is to compare LR calculations performed using Microsoft Excel to the LR calculations performed by STRmix. Verification of LR calculations was performed for single source samples and for mixed samples using various sets of hypotheses.

Because the way in which STRmix generates a distribution of weights and posterior mean allele frequencies cannot be replicated, no attempt was conducted to verify the HPD. Note that the HPD will generally be lower than the point estimate of the LR that is being verified using Excel.

Materials and Methods

Previously analyzed samples of known composition, prepared for the Globalfiler validation, were selected for this verification. Twenty four single source samples (15, 16, 17, 19, 20, 23, 27, 28, 29, 31, 34, 35, 36, 38, 39, 40, 49, 51, 52, 56, 61, 62, 71, and 161), two two-person mixtures (2-19 and 2-39), and one three-person mixture (3-9) were chosen.

STRmix analysis was conducted on the single source samples and STRmix was used to calculate LRs for the comparison of these samples to their known sources vs. unknown sources.

STRmix deconvolution of the two-person mixtures was performed three times: unconditioned and conditioned on each of the two known contributors to the mixture. STRmix deconvolution of the three-person mixture was only performed without conditioning. The following LR propositions were used for the mixed samples and their corresponding known contributors:

2-19:	Contributor 1 + Unknown	vs.	Two unknowns
	Contributor 2 + Unknown	vs.	Two unknowns
	Contributor 1 + Contributor 2	vs.	Two unknowns
	Contributor 1 + Contributor 2	vs.	Contributor 1 + Unknown
	Contributor 1 + Contributor 2	vs.	Contributor 2 + Unknown
2-39:	Contributor 1 + Unknown	vs.	Two unknowns
	Contributor 1 + Contributor 2	vs.	Contributor 1 + Unknown
3-9:	Contributor 1 + Two unknowns	vs.	Three unknowns

Prior to performing LR calculations in Microsoft Excel, posterior mean frequencies were calculated from the NIST population data for the four population groups.

Calculations of the single source LRs in Excel were performed using the posterior mean frequencies for the population group corresponding to that of the known contributor. These calculations were performed using the single source Balding and Nichols formulae described in NRC II equations 4.10a and 4.10b. Because STRmix settings were adjusted during the course of the validation, STRmix analysis with the new settings was conducted on select samples to compare LRs to those generated from their initial MCMC runs.

Calculations for the mixed sample LRs in Excel were performed using the Caucasian posterior mean frequencies, applying Balding and Nichols adjustments to each genotype within the

genotype set. Because the unconditioned MCMC results for mixtures 2-39 and 3-9 contained hundreds of genotype sets, Excel calculations for these mixtures were only performed the first three Globalfiler loci (D3S1358, vWA, and D16S539).

Results and Discussion

Single source samples

All but one LR calculated in Excel gave results that differed from STRmix by less than 1%. For one sample, the LR at a single locus differed by 3% when calculated by Excel. This appears to be due to the use of a less than ideal stutter file. This resulted in STRmix periodically accepting the stutter peak as an allele, leading to a $Pr(E|H_1)$ of less than 1. When STRmix was rerun with an improved stutter filter, the difference between the Excel and STRmix results for this sample was less than 1%

2-person mixtures

Mixture 2-19 showed differences in the LR calculations between STRmix and Excel less than a fraction of a percent for all loci except for D22S1045, which varied as much as 2.48%. It is not clear why this locus was thus affected. This difference is not substantial enough to affect the interpretation of the results. Although the STRmix LR at this locus was within a few percent of the Excel calculation, both the numerator and denominator of the STRmix LR gave values that do not appear to make sense given the data. Because of the weights assigned during MCMC, the numerator of the LR should have been greater than 0.6 at all loci irrespective of conditioning. Four of the LR calculations had numerators that were on the order of 10^{-4} and denominators that were a similar factor smaller than calculated in Excel. It is not clear what led to this inexplicable calculation by STRmix.

Mixture 2-39 was intentionally selected because it was a low level sample and alleles from both contributors had dropped out. This sample generally showed greater differences in the LR calculations than 2-19, presumably due to the consideration of dropout at all alleles except for TH01. When conditioned on Contributor 1, the difference in the LR calculated by STRmix differed from the Excel calculation by less than 1%; however the LR calculations at some loci showed differences as great as 2.2% (for D5S818). The three locus, unconditioned LR calculations for mixture 2-39 differed by less than 0.7%, with locus LR differences as great as 0.8.

3-person mixture

The three locus, unconditioned LR calculation for mixture 3-9 differed by less than 0.6%, with almost all of the difference resulting from D16S539. This locus had the greatest number of genotype sets incorporated in to the calculation: 123.

The loci with differences in the LR calculations greater than 0.1% typically had some combination of rare alleles and dropout considered in the calculation. It is probable that the

observed differences are due to rounding errors or due to a difference in how dropout was incorporated into the calculations between STRmix and Excel.

Conclusions

This study demonstrated that the likelihood ratio calculations performed by STRmix can be replicated within a very small margin within Excel. This study also demonstrates that the laboratory has a very good understanding of the manner in which STRmix calculates Likelihood ratios under a variety of different scenarios.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler and STRmix™ Software

Adjudicated Sexual Assault Cases

Introduction

Six adjudicated cases with a sexual component (sex crimes and child abuse) were selected because these cases contained samples known to have mixtures of DNA, are representative of the types of cases encountered, had a high probative value, and represented a range of previously validated DNA typing kits (Profiler Plus, COfiler, Identifiler, and Identifiler Plus) for comparison to the GlobalFiler DNA typing kit. One additional case with a sexual component (15-016405) did not have prior DNA typing of the evidence or the victim's reference, but had multiple samples for testing with a range of mixture proportions.

GlobalFiler results were to be assessed for concordance with those obtained from previous typing methods, for consistency between the conclusions drawn from previous DNA typing results, and to assess the strength of the probative value of the evidence using STRmix to interpret the DNA samples compared to the previous interpretation methods.

Materials and Methods

This study involved the evaluation of STRmix as a method to deconvolute mixtures generated from evidentiary samples. A modified differential extraction was performed on selected samples, which gave single source or predominant profiles with their previous analysis, to increase the likelihood that DNA mixtures would result. Deviations from SDPD differential extraction protocol are described below. All samples described below were quantitated using the Applied Biosystems Quantifiler Duo kit, amplified using the Applied Biosystems GlobalFiler PCR Amplification Kit, and subsequently analyzed using the Applied Biosystems 3500 Genetic Analyzer.

<u>98-083218</u>
A modified differential extraction (no post digest washes) was performed on one vaginal swab (10254970-1A). The DNA was purified using QIAGEN BioRobot® EZ1. The previously prepared DNA extract from the suspect (TE) was also tested. The evidence samples had been previously tested with the Profiler Plus and COfiler PCR Amplification kits. The reference sample had been previously tested with the Identifiler Plus PCR Amplification kit.

<u>10-000731</u>
Previously prepared DNA extracts of sperm and nonsperm fractions of a cutting from the victim's carpet (10-2) and from the suspect's reference sample (RB11-1007) were tested. The evidence samples had been previously tested with the Identifiler PCR Amplification kit. The reference sample had been previously tested with the Identifiler Plus PCR Amplification kit.

<u>10-023794</u>

A modified differential extraction (shortened nonsperm cell digestion and no post digest washes) was performed on one vaginal swab (10051843-A). A reference sample from the suspect (MJ) was also tested. The DNA was purified using QIAGEN BioRobot® EZ1. Analysis of these samples was discontinued due to a significant excess of female DNA in the evidence samples.

11-034984

A modified differential extraction (fifteen minute nonsperm cell digestion and no post digest washes) was performed on one vaginal swab (10026201-A). Reference samples from the suspect (AA), consensual partner (JC), and victim (EV) were also tested. The DNA was purified using QIAGEN BioRobot® EZ1. Analysis was discontinued on the DNA extracts from the vaginal swab due to a significant excess of female DNA.

A modified differential extraction (one hour nonsperm cell digestion and no post digest washes) was performed on an additional vaginal swab (10026201-A) and on two external genital swabs (10026201-A). The DNA was purified using QIAGEN BioRobot® EZ1. Analysis of the external genital swab was discontinued due to the low amount of recovered DNA.

The samples had been previously tested with the Identifiler Plus PCR Amplification kit.

14-018763

A previously prepared Y-screen DNA extract from the deep vaginal swab [10362116-A(Y)] and the previously prepared DNA extract from the reference samples from the suspect (10362118) and the victim (10369601) were tested. The reference samples had been previously tested with the Identifiler Plus PCR Amplification kit. DNA typing had not been previously conducted on the deep vaginal swab, but had been conducted on the sperm and nonsperm fractions of the cervical swab.

14-038266

Previously prepared DNA extracts of nonsperm fractions of two cuttings from paper towels (10408653-1 NS and 10408653-2 NS) and from reference samples from the suspect (10408728), the victim's consensual partner (10408772), and from the victim (10409928) were tested. The samples had been previously tested with the Identifiler Plus PCR Amplification kit.

15-016405

A reference sample from the victim was sampled for analysis. The DNA was purified using QIAGEN BioRobot® EZ1. Four Y-screen DNA extracts, previously prepared from cuttings from the jeans [10478817-A4(Y), 10478817-A5(Y), 10478817-A8(Y), and 10478817-A4(Y)], and a DNA extract from the suspect's reference sample (10479448) were also tested. The evidence samples and the reference sample from the victim had not been previously tested. The reference sample from the suspect had been previously tested with the Identifiler Plus PCR Amplification kit.

Results

98-083218

DNA results were obtained at all GlobalFiler loci for the sperm fraction of the vaginal swab and from the suspect reference sample. DNA results were obtained at all loci except for DYS390 for the nonsperm fraction. The sample showed no Y peak at Amelogenin and the DNA type at Yindel was very low, suggesting that the majority of the DNA in this sample was female. The victim's reference sample was not tested with GlobalFiler for comparison. As anticipated, the sperm fraction resulting from the modified differential extraction was a mixture with a greater amount of the DNA types corresponding to the nonsperm fraction than was previously obtained with the standard differential extraction. The sperm and nonsperm fractions were consistent with being mixtures of DNA from two people. The reference sample appeared to have DNA results consistent with being single source.

STRmix was able to deconvolute the mixtures from the sperm and nonsperm fractions, resulting in a 74:26 mixture ratio in the sperm fraction and a 1:99 mixture ratio in the nonsperm fraction. The original RMP reported for the sperm fraction was on the order of 10^{19}. The LR calculated by STRmix for the suspect was on the order of 10^{31} for the sperm fraction and 10^{4} for the nonsperm fraction.

10-000731

DNA results were obtained at all GlobalFiler loci for the sperm and nonsperm fractions of the cutting from the carpet and from the suspect reference sample. The victim's reference sample was not tested with GlobalFiler for comparison. The results of the sperm fraction indicated a single source sample, which previously had a single minor contributor type detected in a stutter position at D8S1179. The previous Identifiler testing of the nonsperm fraction indicated a mixture of DNA, having two minor contributor peaks detected in stutter positions. The GlobalFiler results for the nonsperm fraction were consistent with being a mixture of DNA from two people and showed unequivocal indications of a second contributor at several loci in addition to indications in stutter positions. The reference sample appeared to have DNA results consistent with being single source.

STRmix was able to deconvolute the mixture from the nonsperm fraction, resulting in a 98:2 mixture ratio. The original RMP reported for the predominant profile from the sperm and nonsperm fractions was on the order of 10^{20}. The LR calculated by STRmix for the suspect was on the order of 10^{28} for the sperm fraction and 10^{27} for the nonsperm fraction. Although no conclusions were made for the minor contributors to the original analysis, the STRmix results do appear to be amenable to comparisons for possible minor contributors; however, no additional reference samples were tested for comparisons.

11-034984

DNA results were obtained at all GlobalFiler loci for the sperm fraction of the vaginal swab and from the reference samples from the suspect and consensual partner. DNA results were obtained at all GlobalFiler loci except for Yindel and DYS391 from the nonsperm fraction of the vaginal swab and from the victim reference sample. The results of the sperm fraction indicated a mixture of DNA from two people, and the nonsperm fraction appeared to be single source. These results are similar to the results from the previous Identifiler Plus testing of these samples;

however the modified differential appeared to result in a greater proportion of female DNA in the sperm fractions. The reference samples appeared to have DNA results consistent with being single source samples.

STRmix was able to deconvolute the mixture from the sperm fraction, resulting in a 67:33 EV:AA mixture ratio. The original RMP reported for the predominant profile from the sperm and nonsperm fractions was on the order of 10^{22}. The LR calculated by STRmix for the suspect (conditioned on EV) was on the order of 10^{29} for the sperm fraction. LRs for comparisons to the victim and consensual partner were not performed; however the consensual partner could be excluded based on visual comparison to the evidence profile.

14-018763

DNA results were obtained at all GlobalFiler loci for the Y-screen DNA extract of the deep vaginal swab and from the reference sample from the suspect. DNA results were obtained at all GlobalFiler loci except for Yindel and DYS391 from the victim's reference sample. The results of the Y-screen DNA extract indicated a mixture of DNA from two people. These results are similar to the results from the previous Identifiler Plus testing of these samples; however the modified differential appeared to result in a greater proportion of female DNA in the sperm fractions. The reference samples appeared to have DNA results consistent with being single source samples.

STRmix was able to deconvolute the mixture from the deep vaginal swab, resulting in a 68:32 suspect:victim mixture ratio. The original RMP reported for the single source profile from the sperm fraction of the cervical swab was on the order of 10^{22}. The LR calculated by STRmix for the suspect (conditioned on the victim) was on the order of 10^{27} for the Y-screen DNA extract of the vaginal swab. LRs for comparisons to the victim were not performed.

14-038266

DNA results were obtained at all GlobalFiler loci for the nonsperm fractions of the cuttings from the paper towels and from the reference samples from the suspect and the victim's consensual partner. DNA results were obtained at all GlobalFiler loci except for Yindel and DYS391 from the victim's reference sample. The results from 10408653-1 NS indicated a mixture of DNA from two individuals. The results from 10408653-2 NS indicated a mixture of DNA from three individuals. The reference samples appeared to have DNA results consistent with being single source.

STRmix was able to deconvolute the mixtures from the nonsperm fractions, resulting in a 65:35 mixture ratio for 10408653-1 NS and 54:44:2 mixture ratios for 10408653-2 NS. These results are similar to the previous Identifiler Plus testing of both samples, which indicated mixtures of DNA from three people, with two major contributors and one low-level minor contributor. The original RMP reported for the victim's inclusion (the major contributor foreign to Roberto Blanco) for both nonsperm fractions was on the order of 10^{19}. The LR calculated by STRmix for the victim's comparison, conditioned on the suspect, was on the order of 10^{28} for both mixtures. Because this case involved "neutral" evidence, STRmix could also calculate other LR combinations, such as suspect and victim vs. two unknowns, which gave and LR on the order of 10^{51} for 10408653-1 NS. Although no conclusions were made for the minor contributors to the

original analysis, the STRmix results do appear to be amenable to comparisons for possible minor contributors; however, the consensual partner was excluded from both mixtures and no additional reference samples were tested for comparisons.

15-016405

DNA results were obtained at all GlobalFiler loci for the Y-screen DNA extracts of the cuttings from the jeans and from the reference sample from the suspect. DNA results were obtained at all GlobalFiler loci except for Yindel and DYS391 from reference sample from the victim. The results from 10478817-A4(Y), 10478817-A5(Y), and 10478817-A4(Y) indicated a mixture of DNA from four individuals. The results from 10478817-A8(Y) indicated a mixture of DNA from three individuals. The reference samples appeared to have DNA results consistent with being single source.

STRmix was able to deconvolute the mixtures from the jeans. MCMC was performed conditioned on the victim, who appeared to be the major contributor upon visual inspections. The deconvolution resulted in 77:11:10:2 mixture ratios for 10478817-A4(Y), 66:15:10:9 mixture ratios for 10478817-A5(Y), 92:4:4 mixture ratios for 10478817-A8(Y), and 67:26:4:4 mixture ratios for 10478817-A9(Y). These samples had not been previously tested, nor were comparisons and statistical assessments performed previously. The LR calculated by STRmix for the suspect's comparison, conditioned on the victim, was on the order of 10^9 for 10478817-A4(Y), 3 for 10478817-A5(Y), 10^{-7} for 10478817-A8(Y), and 10^4 for 10478817-A9(Y).

Conclusions

Based on the results obtained, the GlobalFiler amplification kit yielded more information overall than the Profiler Plus, COfiler, Identifiler, and Identifiler Plus amplification kits with similar amounts of input DNA.

The GlobalFiler kit displayed general concordance with the results obtained from previous analysis; however, concordance could only be ascertained for samples with previous STR typing and only at the loci contained in the typing kits previously used for these samples. Differences in relative proportions of contributors to some samples were noted. These differences were due to the fact that several samples were intentionally prepared in a manner to generate more evenly balanced mixtures than were previously obtained in an effort to make these samples amenable to mixture deconvolution by STRmix.

Samples expected to contain a significant portion of male DNA gave results at all loci, including Yindel and DYS390. Samples expected to contain exclusively female DNA gave no results at the markers on the Y chromosome. For samples with a very small portion of male DNA, results were obtained in samples sample at one of the other male markers (Yindel or DYS390) without results for a Y peak at Amelogenin. Although the Y chromosomal markers are not very discriminating, these provide additional potential to indicate that a lower level contributor to a mixture is male than previous testing kits with a single gender informative marker.

When tested with GlobalFiler, several mixed samples showed unequivocal evidence of additional contributors who were masked by other alleles or by stutter in the results obtained from the

previous typing kits. The additional discriminating loci in the GlobalFiler kit provide better assessment of the number of contributors to mixed samples and increase the chance of detecting low-level minor contributors to a sample, who may be masked by fewer and less discriminating loci contained in previously used amplification kits.

Previously reported conclusions originally generated for the evidence with regards to the inclusion of the victim and suspect still hold for the data generated with the GlobalFiler amplification kit; however, the strength of the associations are magnified by the presence of additional discriminating loci within the GlobalFiler kit compared to previous testing kits. The strengths of statistical associations are also bolstered by fact that STRmix generates likelihood ratios that take into account peak height and assumptions regarding the number of contributors. These factors were not considered in statistical assessment of the majority of the previous associations.

Variation in contribution across four samples in one case led to comparisons that strongly supported inclusion, were inconclusive, or strongly supported exclusion. Due to the number of minor contributors, it is unlikely that any of these samples would have been deemed suitable for comparisons if analyzed using Identifiler Plus under our currently validated methods. This shows that GlobalFiler in combination with STRmix allows for both greater strength of the evidence when an individual is included, but also allows for both inclusions and exclusions from samples that previously would have been deemed inconclusive.

Overall, the results obtained with the GlobalFiler kit suggest that this kit would improve the power of STR testing over current methods. Additionally, the information obtained from GlobalFiler testing could be compared to previous work performed on a case with the Profiler Plus, COfiler, Identifiler, or Identifiler Plus amplification kits.