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

Validation of the Minifiler PCR Amplification Kit on the 3500 using STRmix for interpretation

Validation Summary

Introduction

The AmpF*l*STR MiniFiler™ Amplification Kit is designed for the genotyping of degraded and/or inhibited DNA samples when a complete DNA profile is unable to be obtained with the AmpF*l*STR Identifiler™ Amplification Kit alone. In a single PCR reaction MiniFiler amplifies eight autosomal STR loci: D7S820, D13S317, D2S1338, D21S11, D16S539, D18S51, CSF1PO, FGA, and the sex determining locus Amelogenin. This kit is already in use in the laboratory for casework, however, after amplification, the sample will now be run on a 3500 genetic analyzer instead of a 3130. Furthermore, interpretation of this data will now be done with STRmix probabilistic genotyping software.

The focus of this set of validation studies was to not to assess the performance of this multiplex kit as it relates to amplification. The goal assess what an appropriate analytical threshold will be on the 3500, and to determine the peak height variance, stutter variance and Locus specific amplification efficiency variance so that samples can be interpreted with STRmix.

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.

The following validation studies were performed to ultimately assess the use of STRmix for interpretation of Minifiler results: Baseline, STRmix ModelMaker for Minifiler, MCMC, and comparison to known contributors.

Baseline Study

The baseline study established the analytical threshold on the 3500 genetic analyzer. Negative controls and samples containing DNA were amplified for the previously validated 30 cycles and capillary electrophoresis was performed on both 3500 instruments. The noise was evaluated for each dye channel and across the entire profile. Three methods were considered for determining analytical threshold. In addition to calculating based the mean and standard deviation of the noise peaks, some weight was given to the highest noise peaks over the course of the validation. With the goal of maximizing detection of alleles while minimizing artifacts, a threshold of 100 RFU was chosen for all channels and was assessed throughout the validation. Spectral artifacts were occasionally observed above this threshold, so some care should be taken when assessing the shape and location of possible artifact peaks.

STRmix ModelMaker

Several parameters are provided to STRmix ModelMaker, and the output is the models for allelic, stutter and locus specific amplification efficiency variance. These variance parameters are determined prior to any STRmix run, and are used during mixture deconvolution. 150 single source samples amplified with a range of target amounts (resulting in profiles with dropout to very robust results) were collected for use in ModelMaker. Stutter was not filtered in these samples.

Expected stutter for every locus was also provided to STRmix ModelMaker. First, stutter from samples run on a 3500 was compared to stutter previously collected in the lab. Data collected on the 3500 was used to compile stutter files. The stutter file was created from the n-1 repeat data

plotted in these graphs in the ModleMaker 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**) - D21S11 and FGA, (**MSM**) - D2S1338.

Three other parameter were also provided to STRmix ModelMaker and are also used as STRmix Minifiler kit settings. The analytical threshold was determined in the Minifiler baseline study to be 100 RFU for all channels. A saturation limit of the 3500 was observed to be ~32,000 RFU. All previous event of drop-in were compiled from the lab to estimate drop-in. The drop-in cap was scaled up for data collected on the 3500 instruments to be 612 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).

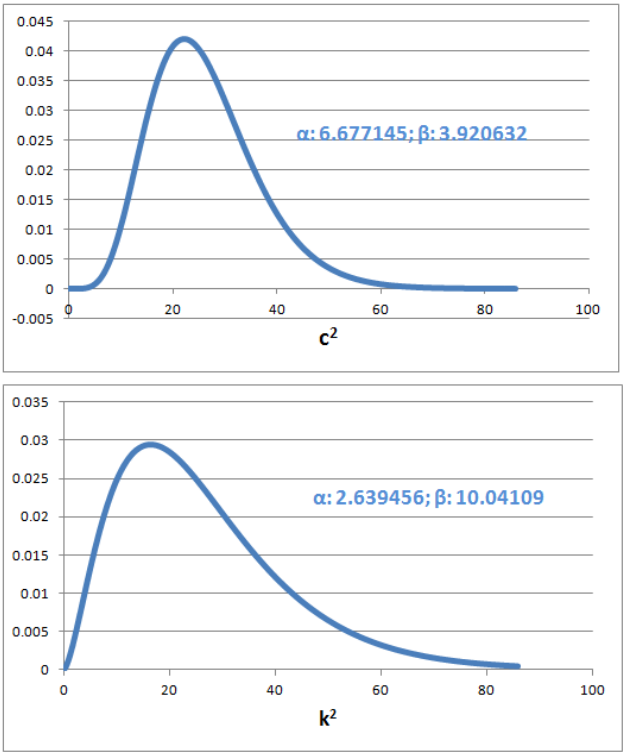

Figure 1 – The allele and stutter variance parameters.

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:
LSAE Variance - fitted gamma curve = exp(50.10322560345891) - mean = 0.01995879482719301

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

STRmix MCMC
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.

Two, three, four, and five-person 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. A subset of these mixture extracts were chosen and amplified with Minifiler.

The MCMC and Metropolis-Hastings are central processes to STRmix. The validation of this software package was done by providing 58 DNA mixtures amplified with Minifiler 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.

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. All mixtures were deconvoluted with STRmix, and results were compared to the database file of 76 individuals.

15 two person mixtures were deconvoluted using STRmix. 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).

16 three-person mixtures were deconvoluted using STRmix. All of the 3-person mixtures resulted in the correct inclusions of all individuals known to comprise the mixtures. Only one mixture (3-63) included a known non-contributors with an LR above 1. This LR was 4.35 the Known contributors to the mixture had much higher LRs, with the lowest of the three being 15201.84.

7 four-person mixtures were deconvoluted using STRmix. All 7 of the 4-person mixtures resulted in the inclusions of the individuals known to comprise the mixtures. In addition to the correct inclusions, three mixtures 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.

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 amplified with Minifiler. 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	Jasmin Kaeser Criminalist
Roxanne Kotzebue 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

Validation of the Minifiler® PCR Amplification Kit from Life Technologies Using a 3500 Genetic Analyzer

Baseline (signal-to-noise) - Analytical Threshold Study

Purpose

The goal of this study is to empirically determine the peak amplitude threshold (i.e. detection threshold) that allows for reliable interpretation of Minifiler® PCR Amplification Kit data. An analytical threshold defines the minimum height requirement at (and above) which detected peaks can be reliably distinguished from background noise. Because the analytical threshold is based upon a distribution of noise values, it is expected that occasional, non-reproducible noise peaks may be detected above the analytical threshold. An analytical threshold should be sufficiently high to filter out noise peaks. On the other hand, usage of an exceedingly high analytical threshold increases the risk of allelic data loss which is of potential exclusionary value. This study aimed to determine an ideal Analytical Threshold for the Minifiler® kit on 3500 Genetic Analyzers.

Materials and Methods

Amplification blanks and DNA-containing samples were amplified using the amount of input DNA and cycling parameters recommended by the manufacturer. The recommended parameters were: 10uL of TE for the blanks, 0.3-0.4ng of DNA for amplification and 30 cycles with the Minifiler® kit. The amplification product from the TE blanks and the DNA-containing samples were injected for capillary electrophoresis on one 3500 instrument (3500A) according to the manufacturer's recommended parameters. Six DNA containing samples and six TE blanks were used to analyze the analytical threshold. A detection threshold of 1 RFU was applied to all samples and blanks during GeneMapper ID-X (v1.4) analysis. All known allele peaks, PCR artifacts (i.e. stutter peaks and minus-A peaks), and peaks that spectrally overlapped with known peaks were removed prior to the assessment. The data was evaluated (from 65-450 bps) to determine the average baseline peak heights as well as the standard deviation of baseline peaks for each dye channel. Additionally, the RFU of the highest peak was determined for each dye channel that could not be identified as a DNA related peak or spectral artifact. The highest peak was evaluated from the panel ranges (from GeneMapper ID-X) for each dye channel (90.0-193.5 for blue, 99.3-250.6 for green, 70.0-210.4 for yellow, and 8406-296.4 for red).

Results and Discussion

The baseline study evaluated the signal-to-noise ratio for the Applied Biosystems Minifiler® kit on the 3500 Genetic Analyzer. The instrumental noise (baseline) was examined from amplified TE blanks as well as from amplified single source samples (allele peaks, PCR artifacts, and

peaks that spectrally overlapped with known peaks were removed). A summary of the average, standard deviation, and maximum peak heights can be seen in Figure 1.

Amplified Samples (DNA Peaks Removed)			
Channel	Average	StDev	Max PH
B	3.45	2.13	39
G	6.94	7.44	41
Y	13.75	4.65	49
R	20.57	6.57	58

TE Blanks			
Channel	Average	StDev	Max PH
B	3.74	2.49	39
G	7.25	4.91	24
Y	15.07	8.89	51
R	21.91	6.67	49

Figure 1 – Peak Heights from two different types of samples amplified with Minifiler® and injected on 3500 Genetic analyzers

Average noise peak height and max peak height were similar from both types of samples (Amplified samples with peaks removed and TE blanks). Occasionally, noise peaks over ten standard deviations above the average were observed. In one of the amplified samples two peaks of 277 and 101 RFU (83.12 and 84.06 bp, respectively) were observed in the green channel. One TE blank had a peak of 196 RFU at 84.22 bp in the green channel. A second TE blank had a peak of 390 RFU at 84.5 bp in the yellow channel and another peak of 100 RFU at 133.04 bp. These peaks were included in the calculations for the channel averages and standard deviations, but were not included in evaluating the maximum peak height.

The peak heights for the detection range of each dye channel were evaluated and separated according to size. The majority of the noise peak heights were below 10 RFU for the blue and green channels, below 20 RFU for the yellow channel, and below 25 for the red channel, as shown in Figure 2.

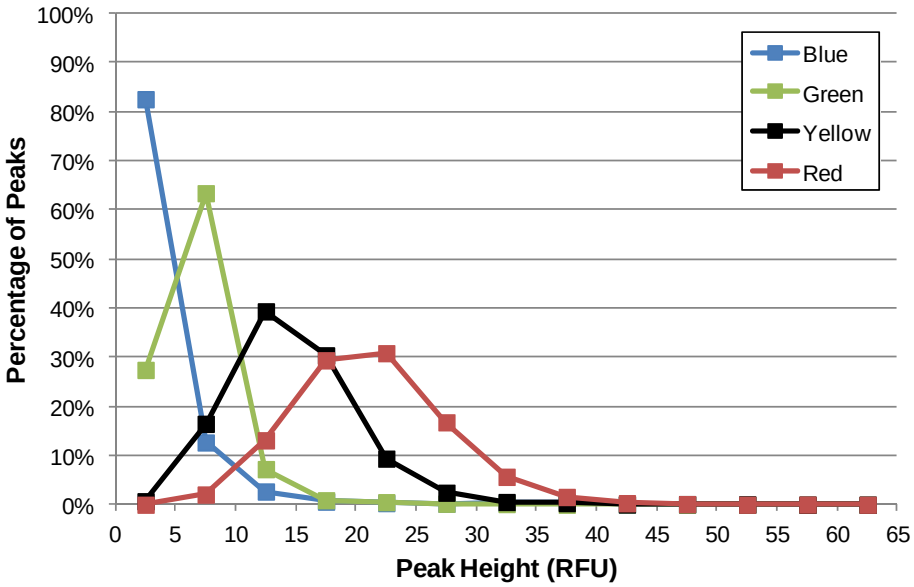

Figure 2 – Distribution graph evaluating the range of noise peaks

There are multiple ways of calculating an empirical analytical threshold from baseline data. Analytical thresholds are generally calculated by examining the baseline data and calculating the mean and standard deviation, then determining the threshold based on two or three times the standard deviation, or ten times the standard deviation. Michael Coble has previously given an example for determining the analytical threshold using two times the intensity difference between the highest peak in the baseline and the lowest trough (1). GeneMapper ID-X data generally omits negative baseline data (i.e. data below zero) in its collection of data. In order to compensate for this, one option in determining the analytical threshold would be to use a calculation of two or three times the highest peak in the baseline. Figure 3 shows three possible methods to help guide the analytical threshold value.

Method	Channel	Analytical Threshold (RFU)
Mean + 10 standard deviations	Blue	27.06
	Green	68.59
	Yellow	88.29
	Red	85.91
2x the highest baseline peak	Blue	78
	Green	82
	Yellow	102
	Red	116
3x the highest baseline peak	Blue	117
	Green	123
	Yellow	153
	Red	174

Figure 3 – Multiple methods to determine the analytical threshold

Conclusions

The data from the "Baseline Study" suggests that a peak amplitude threshold of 100 RFU is an appropriate analytical threshold for DNA casework.

References

Michael D. Coble, PhD. "Design and Execution of Validation Studies for Establishing DNA Mixture Interpretation Procedures". NFSTC DNA Mixture Interpretation Workshop. Largo, FL (March 15-17, 2011).

SDPD Forensic Science Section – Forensic Biology Unit

STRmix® ModelMaker for Minifiler

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 closeness 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 closeness 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 within a specific laboratory. 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. 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 effect 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.

Purpose

Prior to the validation and implementation of STRmix for Minifiler, several things need to be determined. The saturation limit of the 3500 was previously determined with GlobalFiler.

1. Stutter ratios
2. Analytical threshold (or limit of detection)
3. Drop-in parameters
4. Variance constants for stutter and allele
5. Locus specific amplification efficiency parameter

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^l) for different STR profiling kits and protocols (i.e. longer injection protocols or increased cycles). In other words, we provide ModelMaker with #1-3 above, and it gives us #4 and #5. 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 parameter 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 of the model. 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.

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 Minifiler. Data was previously collected and analyzed using data amplified with Minifiler on the 310 and 3130 instruments. Samples collected for evidence/reference were analyzed for stutter to determine if the current stutter ratios needed to be updated. See Figure 1 for stutter data.

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 from the 3500.

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 that 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.

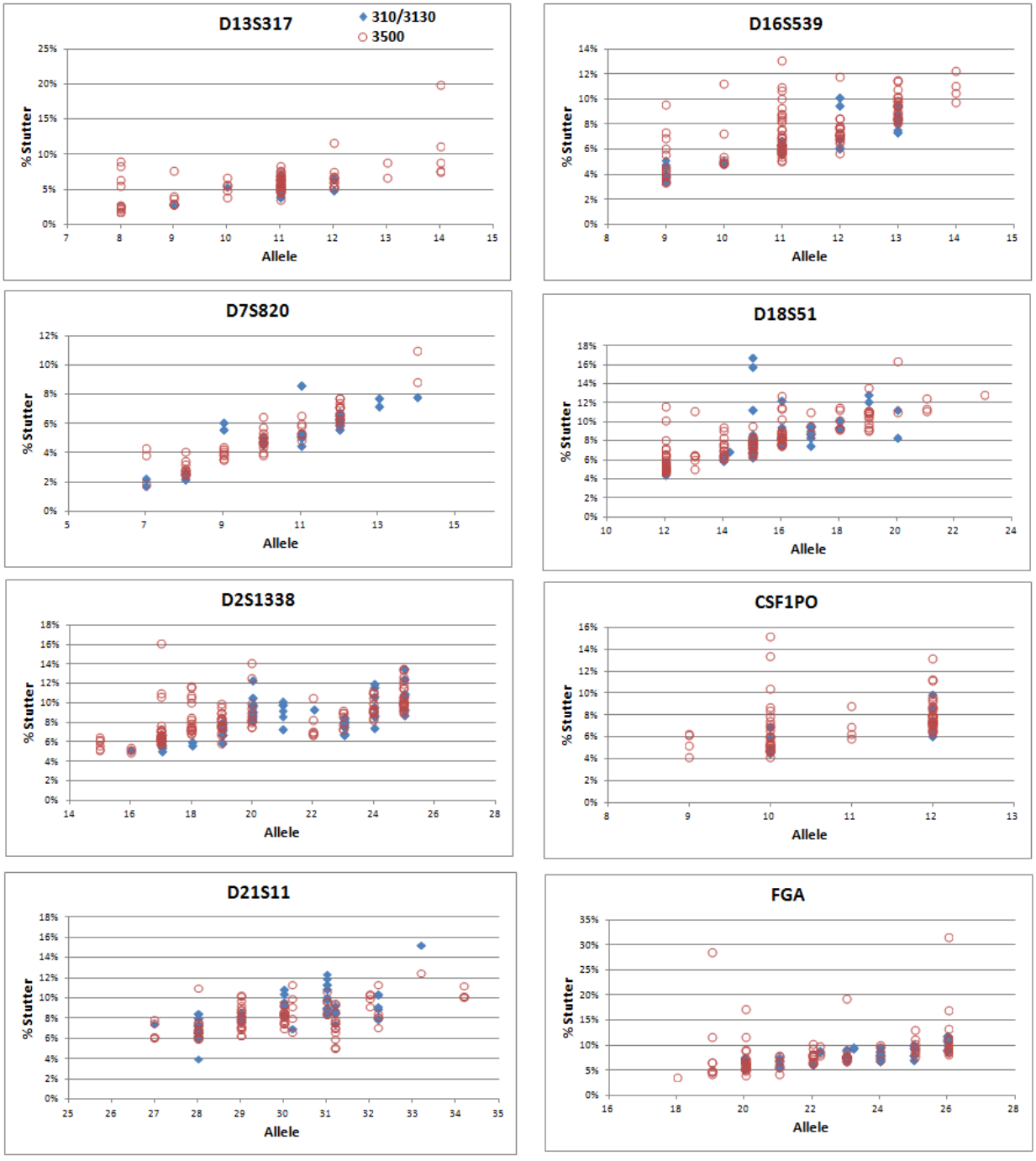

Figure 1 – Minifiler stutter on 310/3130 compared to 3500

Analytical threshold

The analytical threshold was determined in the Minifiler baseline study. Briefly, 6 negative controls and 6 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 Minifiler baseline study for more details.

Drop-in parameters

STRmix provides the option of allowing for drop-in to occur, 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 non-reproducible, 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.

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:

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

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:

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

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 Minifiler data, the unexpected results from the years 2009 through 2015 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 Minifiler data (collected on a 3130 Genetic analyzer) was adapted to reflect what drop-in might look like 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 24 single source samples amplified anywhere from 5 to 8 times each with template amounts ranging from ~10pg to ~800pg. Capillary electrophoresis was performed on these samples using the 3500 instrument according to the manufacturer's recommended parameters. The samples (150 total) was injected for 15 seconds.
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 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 genotype. 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 Minifiler to be deconvoluted by STRmix should be analyzed with the ***STRmix_MF_evidence*** analysis method. All reference samples amplified with Minifiler should be analyzed with the ***MF_reference*** analysis method.

Two text files were created from 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 samples with stutter removed and create a .txt file with allele and 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 furve 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 modelled 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 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

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. 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 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 MF stutter exceptions file: (**LUS**) - D21S11 and FGA, (**MSM**) - 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 Minifiler baseline study to be 100 RFU for all channels. Refer to the Minifiler baseline study for more details.

Drop-in parameters

The number of times unattributable types were detected in reagent blanks or negative amplification controls were counted. Nine 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 205.
The raw drop-in rate was determined to be 9/205 x 9 Minifiler loci = 7/1841 = 0.00488. This is not the true drop-in rate because the AT was 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. For the SDPD data, the predicted proportion of detectable drop-in events was found to be 0.8089. The raw drop-in rate divided by the percentage of detectable drop-ins gives the actual drop-in rate (0.00488/0.8089 = 0.006). The drop-in cap was scaled up for data collected on the 3500 instruments to be 612.36 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 2. 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 give 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.

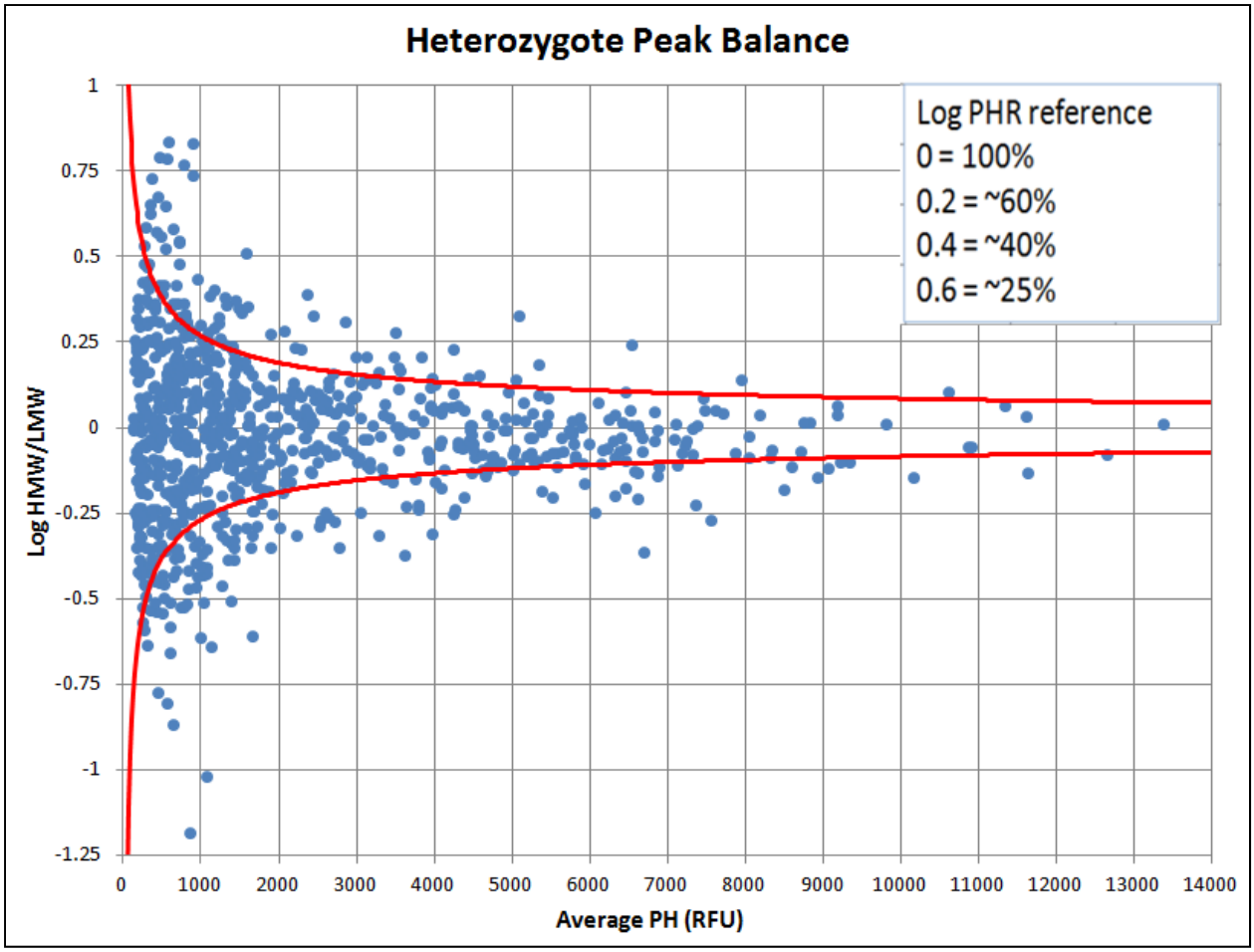

Figure 2 – Heterozygote peak height balance of samples amplified at a range of target values with 95% boundaries (red dotted line) plotted against the average height of heterozygote alleles.

We know from previous validations and published literature that peak height balance can vary a lot more as the peak heights approach a 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 Minifiler ModelMaker data. Instead of determining a binary threshold, however, a probablility of dropout was calculated. The possibility of dropout is very peak height dependent. This is shown in Figure 3. 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.

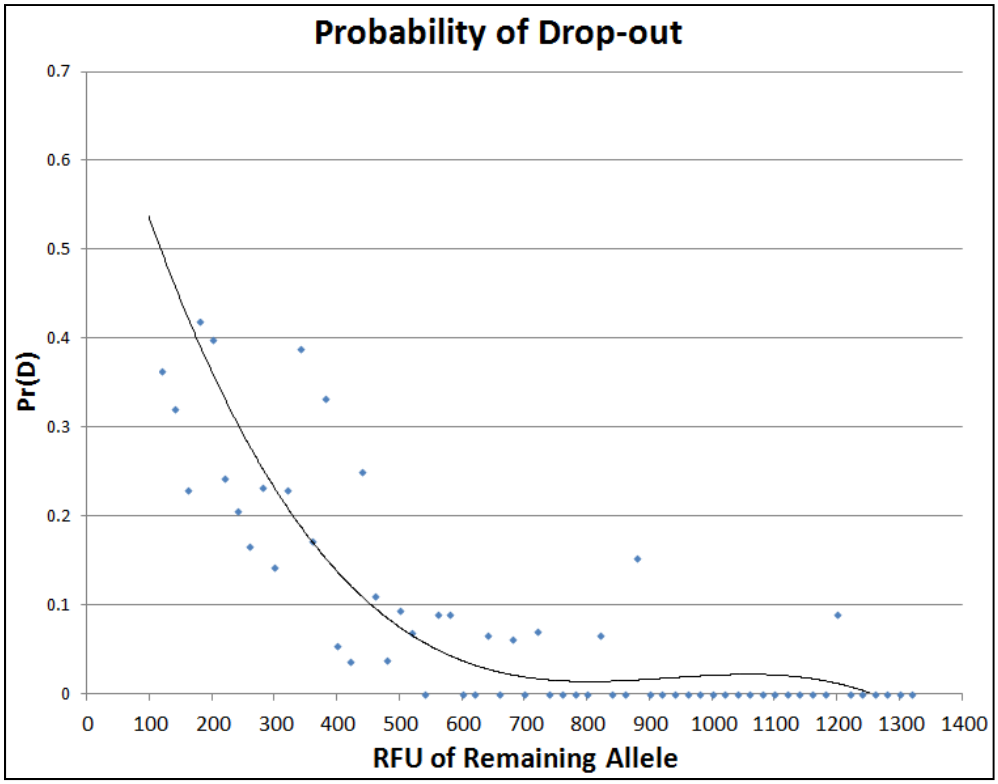

Figure 3 – Probability of dropout

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. Figure 4 shows the data generated by ModelMaker in minimizing the difference between observed and expected (both for allele and stutter).

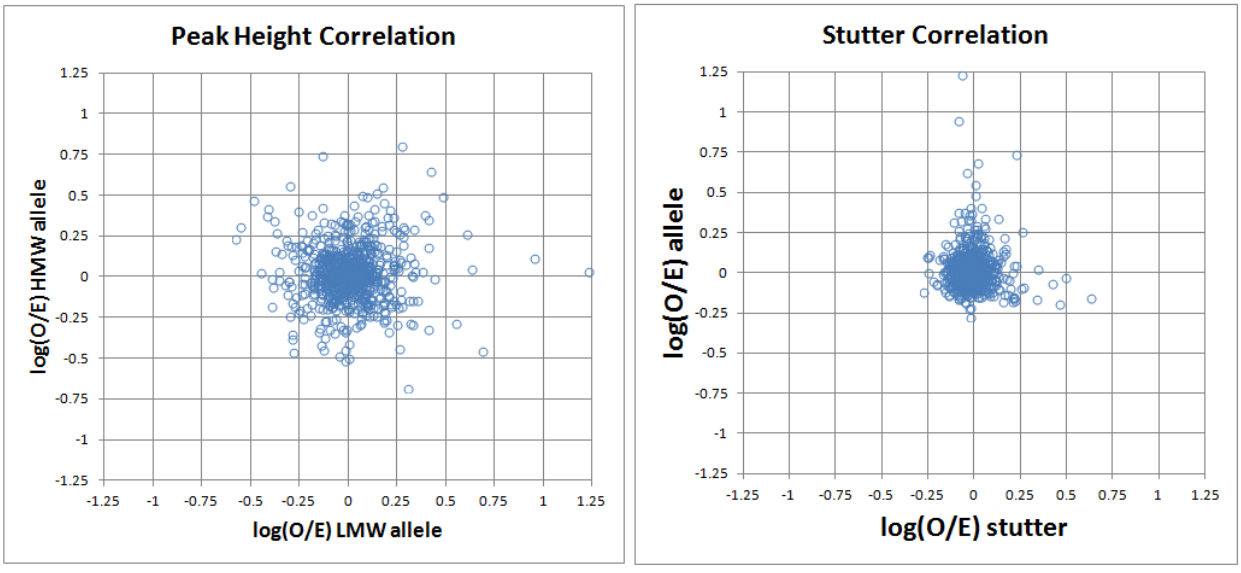

Figure 4 – STRmix ModelMaker outputs of Allele and Stutter variance. The differences between observed and expected values are minimized.

Figure 4 displays 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 5).

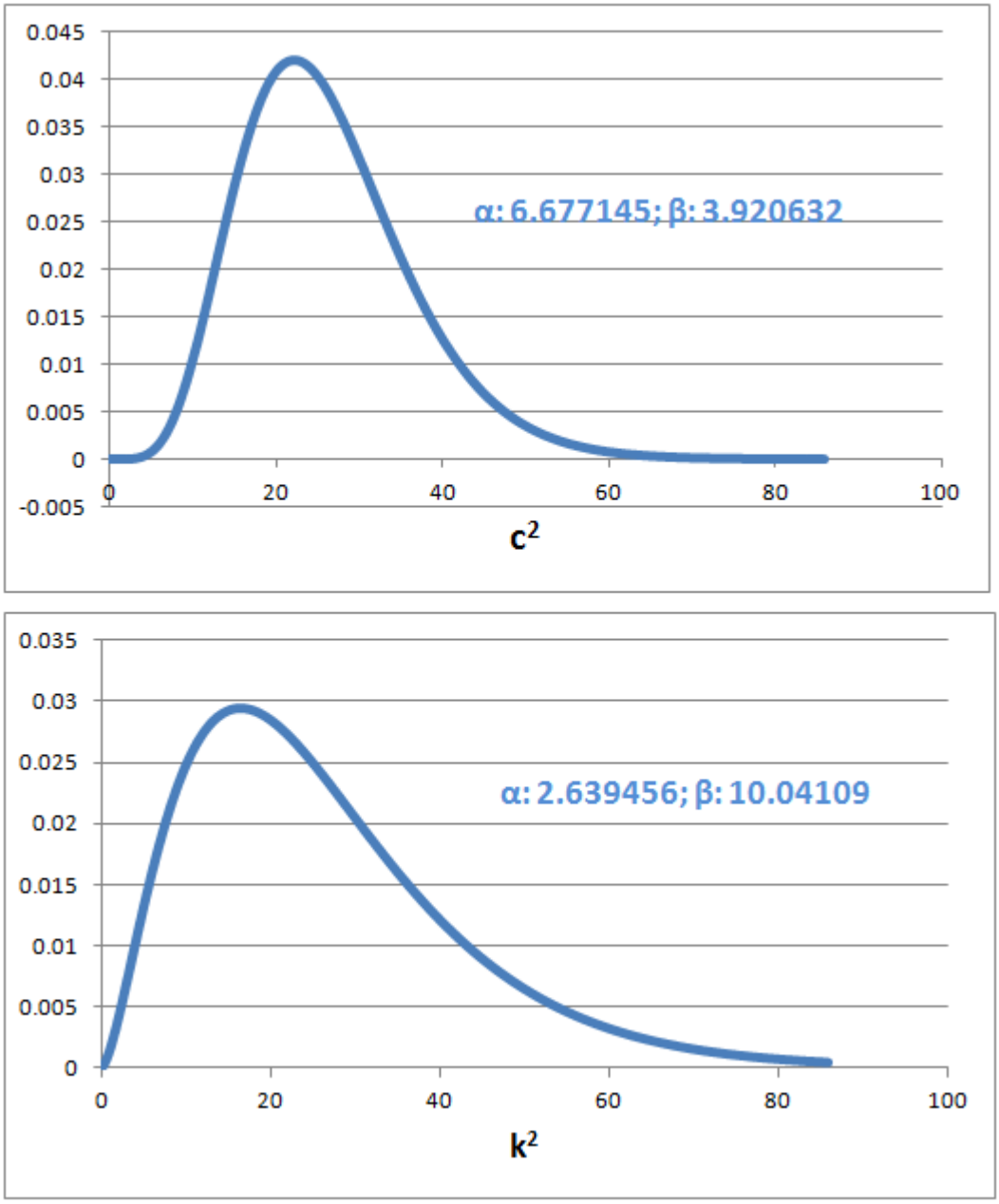

Figure 5 – The allele and stutter variance parameters.

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:

LSAE Variance - fitted gamma curve = exp(50.10322560345891) - mean = 0.01995879482719301

Conclusions

All of the above sections describe the settings that will be used for mixture deconvolution of samples amplified with Minifiler using STRmix. These settings are summarized in Figure 6. While STRmix was supplied with default values for many commercially available kits, an internal validation of Minifiler on the 3500 was performed to reflect the local lab environment (i.e. products, instruments, and analysts). 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.

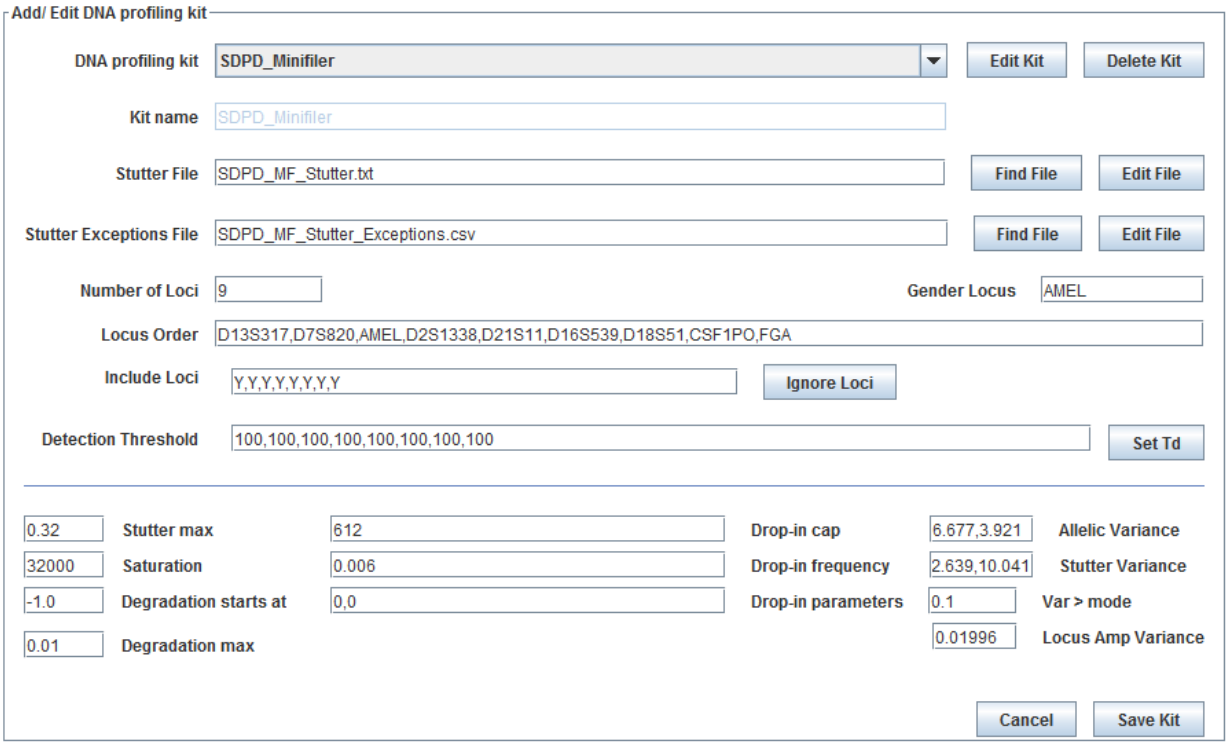

Figure 6 – SDPD ModelMaker settings for Minifiler using v2.3.06 of STRmix.

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 for Minifiler
Markov Chain Monte Carlo

Introduction
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. An overview of the MCMC process was given in the STRmix validation, and this study is to assess the deconvolution of Minifiler mixtures using STRmix.

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. This was accomplished for Minifiler utilizing 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 comparing the STRmix results to known genotype sets of the ground truth mixtures.

Materials and Methods
Two, three, four, and five-person 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. A subset of these mixture extracts were chosen and amplified with Minifiler. The samples that were overamplifed were excluded from STRmix analysis and some of them were re-amplified at lower target amounts.

All of these mixtures were evaluated assuming the number of contributors the mixture was designed to have. All of the two, three and four person mixtures were evaluated extensively. The three 5 person mixtures haven't been run due to a limit in java. 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 was compared against STRmix calculations for percent contributor.

Results and Discussion

Results for every two, three, and four person mixture were carefully scrutinized (58 total mixtures).

2 person mixtures:
For this study, 21 two-person mixtures were deconvoluted with the number of contributors set at 2. Of the 21 two-person mixtures, only one of these mixtures (2-22) had a diagnostic value that warranted a closer look. The Gelman-Rubin Convergence number was 1.24. 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. 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. See Table 1 for a summary of results.

3 person mixtures:
Of the 26 three-person mixtures, 9 of them had alleles from at least one contributor dropping out. One of the 26 mixtures had a diagnostic value that warranted a closer look, with a Gelman-Rubin Convergence number of 1.24 and an effective sample size that could not be calculated. In looking closer, it had no problem with the deconvolution into the known contributors.
Five of the three person mixtures had at least one contributor whose genotypes weights did not fall into the top 99%. In mixture 3-25, the 9% contributor's genotype did not fall into the top 99% at one locus (D13). In mixture 3-42, extreme peak imbalance and imbalance between contributors is preventing the robust deconvolution of the 16% contributor. The other two have been deconvoluted so that the genotypes fall in the top 99%. Mixture 3-50 was run while excluding data from the locus D21 because a known and documented null allele of one of the contributors is present. After excluding D21, the locus D2 is showing extreme contributor imbalance, and that is preventing the correct combination of genotypes for each contributor. All known genotypes are in the top 99%, but not in the correct combination for the contributor. In mixtures 3-52 and 3-61, dropout is not being sufficiently accounted for in the 15% and 10% contributor, respectively, at one locus.
In summary of the three person mixtures, the majority of mixtures with dropout and peak imbalance were deconvoluted in a way such that the known contributors genotypes were weighted in the top 99%. See Table 2 for a summary of results.

4 person mixtures:
Of the 11 three-person mixtures, 2 of them had alleles from at least one contributor dropping out. In this wide variety of mixtures, only 2 had a diagnostic value that warranted a closer look, with Gelman-Rubin Convergence numbers of 1.31 and 1.32. Full assessment of these samples did not indicate any other problem. All four person 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. See Table 3 for a summary of results.

Table 1 – Two person mixture deconvolution results

People	Mixture #	GR	ESS	Dropout	Reference IDs	Target % contribution	STRmix % contribution	STRmix Contributor #	Correct combination in 99%?	Genotype possibilities intuitive?
2	17	1.04	371		150-TD	67	85	2	Y	Y
					156-RK	33	15	1	Y	Y
2	21	1.11	708		150-TD	11	89	2	Y	Y
					156-RK	89	11	1	Y	Y
2	22	1.24	897		62-BS	89	86	1	Y	Y
					31-BB	11	14	2	Y	Y
2	22b	1.02	1493		62-BS	89	90	1	Y	Y
					31-BB	11	10	2	Y	Y
2	24	1.01	880		62-BS	67	74	2	Y	Y
					31-BB	33	26	1	Y	Y
2	28	1.02	695		62-BS	11	7	1	Y	Y
					31-BB	89	93	2	Y	Y
2	29	1.04	528440		150-TD	89	86	2	Y	Y
					156-RK	11	14	1	Y	Y
2	30	1.09	495		150-TD	83	84	1	Y	Y
					156-RK	17	16	2	Y	Y
2	31	1.03	1293		150-TD	67	66	1	Y	Y
					156-RK	33	34	2	Y	Y
2	32	1.05	2664		150-TD	50	62	1,2	Y	Y
					156-RK	50	38	2,1	Y	Y
2	33	1.01	334		150-TD	33	41	2	Y	Y
					156-RK	67	59	1	Y	Y
2	33b	1.02	785		150-TD	33	27	2	Y	Y
					156-RK	67	73	1	Y	Y
2	34	1.06	413		150-TD	17	16	1	Y	Y
					156-RK	83	84	2	Y	Y
2	36	1.06	223		62-BS	89	86	2	Y	Y
					31-BB	11	14	1	Y	Y
2	37	1.01	568		62-BS	83	79	1	Y	Y
					31-BB	17	21	2	Y	Y
2	38	1.07	NaN		62-BS	67	52	1,2	Y	Y
					31-BB	33	48	2,1	Y	Y
2	39	1.02	668		62-BS	50	52	1,2	Y	Y
					31-BB	50	48	2,1	Y	Y
2	40	1.01	333		62-BS	33	29	2	Y	Y
					31-BB	67	71	1	Y	Y
2	40b	1.02	581		62-BS	33	32	1	Y	Y
					31-BB	67	68	2	Y	Y
2	41	1.05	405		62-BS	17	9	2	Y	Y
					31-BB	83	91	1	Y	Y
2	42	1.01	1452		62-BS	11	7	1	Y	Y
					31-BB	89	93	2	Y	Y

Table 2 – Three person mixture deconvolution results

People	Mixture #	GR	ESS	Dropout	Reference IDs	Target % contribution	STRmix % contribution	STRmix Contributor #	Correct combination in 99%?	Genotype possibilities intuitive?
3	25	1.02	3880		125-CN	70	79	1	Y	Y
					106-BL	20	13	2,3	Y	Y
					155-JK	10	9	3,2	N	Y
3	26	1.11	943		34-WD	70	75	3	Y	Y
					28-GL	20	17	1,2	Y	Y
					35-JD	10	8	2,1	Y	Y
3	28	1.24	NaN		41-AD	60	66	2	Y	Y
					49-JB	30	24	1	Y	Y
					47-ER	10	10	3	Y	Y
3	30	1.03	452060		56-TB	50	47	3	Y	Y
					39-DC	40	44	1	Y	Y
					81-AR	10	9	2	Y	Y
3	31	1.07	4201		23-JS	50	49	2	Y	Y
					39-DC	30	33	1,2,3	Y	Y
					79-PV	20	18	3,1	Y	Y
3	32	1.11	2315		30-SS	50	34	3	Y	Y
					52-KK	30	34	2	Y	Y
					23-JS	20	32	1	Y	Y
3	33	1.06	1475		30-SS	45	48	2	Y	Y
					161-GZ	45	42	1	Y	Y
					88-KB	10	10	3	Y	Y
3	35	1.08	6966		61-KM	40	55	3,1	Y	Y
					102-EB	40	9	2	Y	Y
					156-RK	20	36	1,3	Y	Y
3	36	1.06	2559		51-HH	40	35	2,1,3	Y	Y
					62-BS	40	33	1,2,3	Y	Y
					94-RL	20	32	3,2,1	Y	Y
3	38	1.02	4939		39-DC	35	47	1,3	Y	Y
					93-RC	35	47	3,1	Y	Y
					142-VM	30	6	2	Y	Y
3	40	1.03	1754		131-SB	60	76	2	Y	Y
					106-BL	20	9	1	Y	Y
					152-LM	20	16	3	Y	Y
3	41	1.01	1416		125-CN	50	57	1	Y	Y
					115-RM	25	13	3	Y	Y
					79-PV	25	30	2	Y	Y
3	42	1.18	4027		61-KM	50	48	3,1	Y	Y
					30-SS	25	36	3,1	Y	Y
					69-CL	25	16		N	Y
3	43	1.18	3797		71-LW	40	25	1,3	Y	Y
					115-RM	30	25	3,1	Y	Y
					39-DC	30	50	2	Y	Y

Table 2 (con't) – Three person mixture deconvolution results

People	Mixture #	GR	ESS	Dropout	Reference IDs	Target % contribution	STRmix % contribution	STRmix Contributor #	Correct combination in 99%?	Genotype possibilities intuitive?
3	46	1.04	2415	14	79-PV	33	33	1,2,3	Y	Y
					71-LW	33	30	1,2,4	Y	Y
					127-DF	33	36	1,2,5	Y	Y
3	47	1.02	1184	3	125-CN	70	78	3	Y	Y
					106-BL	20	11	1,2,3	Y	Y
					155-JK	10	11	2,1,3	Y	Y
3	48	1.06	2717		34-WD	70	76	1	Y	Y
					28-GL	20	12	2,3	Y	Y
					35-JD	10	12	3,2	Y	Y
3	50	1.09	3469	1	41-AD	60	71	2	N	Y
					49-JB	30	15	3	N	Y
					47-ER	10	14	1	N	Y
3	52	1.04	2582	1	56-TB	50	57	1,2	Y	Y
					39-DC	40	28	2,1	Y	Y
					81-AR	10	15	3,2	N	Y
3	54	1.07	3767	2	30-SS	50	33	1,2	Y	Y
					52-KK	30	33	1	Y	Y
					23-JS	20	34	3,2,1	Y	Y
3	56	1.02	10080	1	150-TD	45	50	1,3	Y	Y
					127-DF	45	40	3,1	Y	Y
					155-JK	10	10	2	Y	Y
3	60	1.08	4059	3	39-DC	35	45	1	Y	Y
					93-RC	35	30	2,1	Y	Y
					142-VM	30	25	3,2,1	Y	Y
3	61	1.01	4258	1	127-DF	60	67	3	Y	Y
					125-CN	20	23	1	Y	Y
					103-KD	20	10	2	N	N
3	63	1.04	769		125-CN	50	42	1,2,3	Y	Y
					115-RM	25	28	1,2,3	Y	Y
					79-PV	25	29	1,2,3	Y	Y
3	64	1.05	5843		61-KM	50	38	2,3	Y	Y
					30-SS	25	31	1,2,3	Y	Y
					69-CL	25	30	3,2,1	Y	Y
3	66	1.12	2349	10	150-TD	40	34	1,2,3	Y	Y
					74-LM	30	33	1,2,3	Y	Y
					132-JF	30	33	1,2,3	Y	Y

Table 3 – Four person mixture deconvolution results

People	Mixture #	GR	ESS	Dropout	Reference IDs	Target % contribution	STRmix % contribution	STRmix Contributor #	Correct combination in 99%?	Genotype possibilities intuitive?
4	24	1.31	3432		138-SE	25	24	1	Y	Y
					134-DG	25	25	2	Y	Y
					113-MS	25	25	3	Y	Y
					49-JB	25	25	4	Y	Y
4	25	1.13	406816		17-PTO	60	59	3	Y	Y
					71-LW	20	17	4,1	Y	Y
					150-TD	10	12	1,2,4	Y	Y
					79-PV	10	12	2,1,4	Y	Y
4	33	1.32	5780		133-AL	40	45	1,3	Y	Y
					51-HH	40	43	3,1	Y	Y
					155-JK	15	8	2,4	Y	Y
					131-SB	5	5	4,2	Y	Y
4	47	1.07	10397		17-PTO	60	50	2	Y	Y
					71-LW	20	26	4,2	Y	Y
					150-TD	10	15	1,4,3	Y	Y
					79-PV	10	9	3,1,4	Y	Y
4	48	1.19	NaN		93-RC	60	63	1	Y	Y
					69-CL	20	24	4	Y	Y
					109-JV	10	7	2,3	Y	Y
					20-JG	10	6	3,2	Y	Y
4	50	1.01	1689		38-SC	50	30	3	Y	Y
					94-RL	20	28	2,3,1	Y	Y
					23-JS	20	29	1,3,4	Y	Y
					47-ER	10	13	4,3,2,1	Y	Y
4	53	1.14	4622		159-MS	40	43	3	Y	Y
					74-LM	20	16	4,3,2,1	Y	Y
					36-TH	20	30	2,3,1	Y	Y
					132-JF	20	11	1,4	Y	Y
4	54	1	3849	3	99-GC	40	32	1,2,3,4	Y	Y
					19-DS	20	25	4,3,2,1	Y	Y
					41-AD	20	22	2,3,4,1	Y	Y
					91-KM	20	21	3,4,1,2	Y	Y
4	58	1.11	8947	1	139-GF	35	44	4,3	Y	Y
					30-SS	35	33	3,4	Y	Y
					106-BL	20	10	1,2,3,4	Y	Y
					156-RK	10	14	1,3,4	Y	Y
4	59	1.09	11810		156-RK	40	44	4	Y	Y
					103-KD	40	40	3,4	Y	Y
					56-TB	10	9	2,1,3,4	Y	Y
					61-KM	10	7	1,2,3,4	Y	Y
4	62	1.18	5362		17-PTO	35	25	4,2,3	Y	Y
					71-LW	35	42	2,3	Y	Y
					112-KM	25	22	3,2,4	Y	Y
					27-MJF	5	10	1	Y	Y

Conclusion

The MCMC and Metropolis-Hastings are central processes to STRmix. This study was designed to test STRmix in deconvolution of ground truth mixtures amplified with Minifiler. While it

would have been optimal to have mostly low level samples with a lot of dropout, the robust performance of the kit on the few samples that were like that suggests that STRmix can be used on compromised samples. 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 all the contributors making up these 2, 3 and 4 person mixtures, all 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. 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), the more ambiguity in their exact genotype combination. Finally, the more dropout there is associated with a mixture, the more ambiguous the results can be. A large 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 for Minifiler

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 for Minifiler 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 Minifiler validation were used for this experiment. See the Minifiler MCMC for a list and description of these mixtures. 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 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

15 two 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), see Figures 1 and 2.

3-peron mixtures

16 three-person mixtures were deconvoluted using STRmix. After running the mixtures through STRmix the deconvolution results were compared to the database file of 76 individuals. See Figures 1 and 2 for results. 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. All of the 3-person mixtures resulted in the correct inclusions of all individuals known to comprise the mixtures. Only one mixture (3-63) included a known non-contributors with an LR above 1. This LR was 4.35 the Known contributors to the mixture had much higher LRs, with the lowest of the three being 15201.84.

4-person mixtures

7 four-person mixtures were deconvoluted using STRmix. These were mixtures that included high, mid, and low level mixtures with a range of contributor proportions. After running the mixtures through STRmix, the deconvolution results were compared to the database file of 76 individuals. See Figures 1 and 2 for results.

All 7 of the 4-person mixtures resulted in the inclusions of the individuals known to comprise the mixtures. In addition to the correct inclusions, three mixtures also had a single non-contributing profile from the database that also resulted in a likelihood ratio that favored inclusion.

In mixture 4-33, the LR of the non-contributor was 19.49. The lowest LR of the four contributors for this mixture was 43196.51.

In mixture 4-48, the LR of the non contributor was 3.78. The lowest LR of the four known contributors was 773.37.

Mixture 4-53 resulted in more ambiguity. There were two known non-contributors with an LR above 1: 1.24 and 222.35. The lowest of the four known contributors was 3985.33.

All other non-contributors in the database received likelihood ratios favoring exclusion.

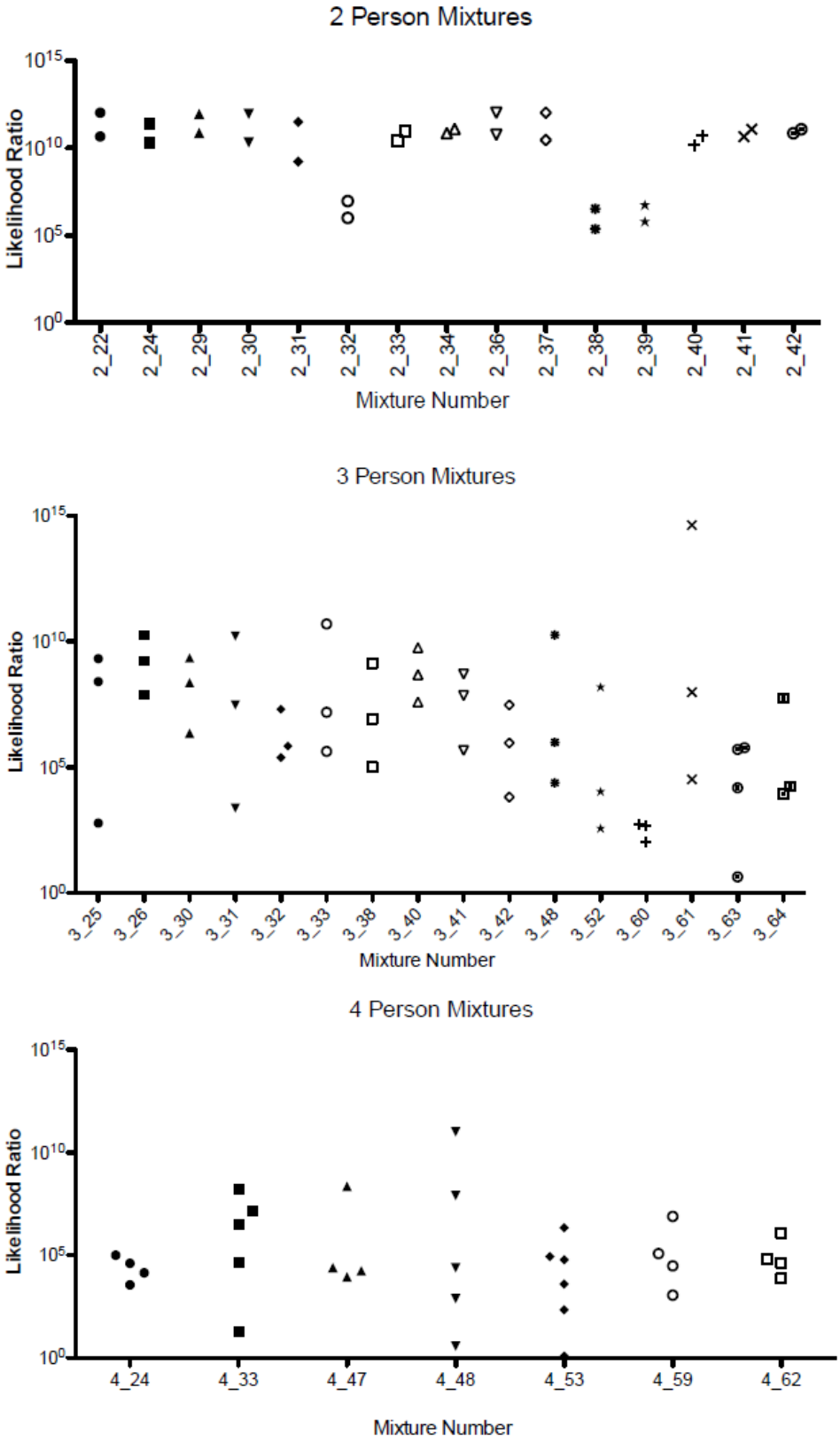

Figure 1 – the Likelihood Ratios of known contributors

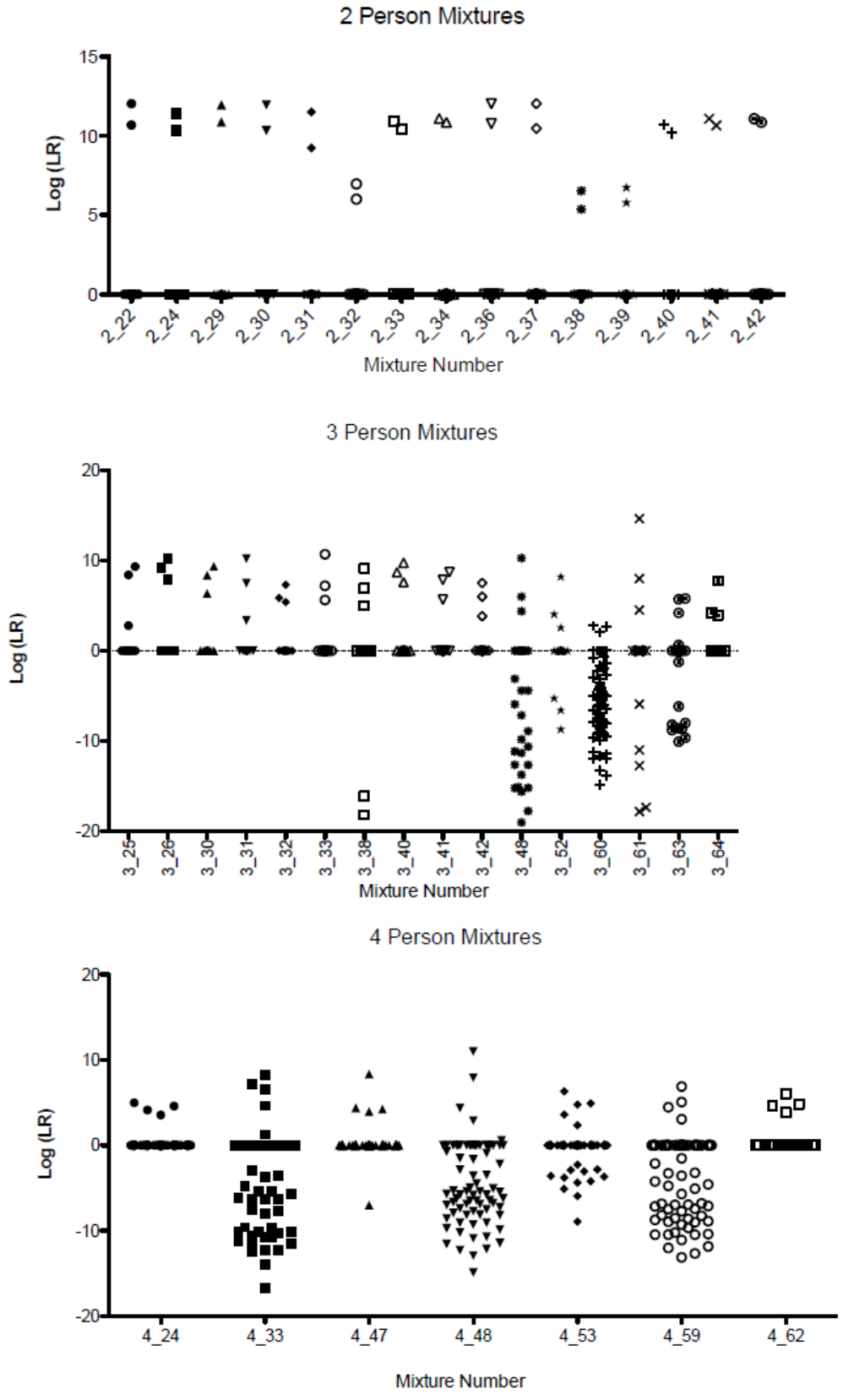

Figure 2 – the LOG Likelihood Ratios of all contributors in the database

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 76 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 the mixtures when compared against profiles from 76 profiles in the database. Assessing inclusions based on likelihood ratios greater than 1, four false inclusions were obtained. No known contributors were falsely excluded.

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.