LOS ANGELES COUNTY SHERIFF'S DEPARTMENT
SCIENTIFIC SERVICES BUREAU
BIOLOGY SECTION

VALIDATION OF STRmix™ v. 2.5.11 using the POWERPLEX FUSION 6C KIT

Interpreting Powerplex Fusion 6C DNA profiles using the probabilistic software STRmix™ v. 2.5.11 produced acceptable and valid results and is deemed suitable for use on casework.

Validation Team:

Signature Redacted 10-26-17

Gregory Hadinoto, Senior Criminalist **Date**

Signature Redacted 10-26-17

Cindy Carroll, DNA Technical Leader **Date**

Signature Redacted 10-26-17

Learden Matthies, Supervising Criminalist **Date**

Approved: Signature Redacted 10-26-17

Cindy Carroll, DNA Technical Lead **Date**

Reviewed: Signature Redacted 11/6/2017

Robert Taylor, Assistant Director **Date**

Validation approved: Signature Redacted 12/19/2017

Karla Taylor, Quality Manager **Date**

The interpretation of Powerplex Fusion 6C profiles with STRmix™ v. 2.5.11 is approved for casework.

EFFECTIVE DATE: 12/20/17

Validation of STRmix™ Software

Introduction

Currently the laboratory uses a binary approach, along with the random match probability (RMP) statistical method, for DNA interpretation. The binary approach is a laborious process that is done manually by an analyst and is often complicated by allele drop out, peak height imbalances, stutter, and the presence of more than one contributor. Binary methods are being superseded by probabilistic genotyping methods. Fully continuous probabilistic genotyping methods utilize peak heights and the Markov Chain Monte Carlo (MCMC) process to calculate the probability of the possible genotypes for individual contributors in the form of a likelihood ratio (LR). The likelihood ratio considers the probability of obtaining the evidence DNA profile given two competing hypotheses. Typically these are:

- H_p: The person of interest is the source or is a contributor of DNA to the sample
- H_d: The person of interest is not the source or is not a contributor of DNA to the sample

The LR value increases if H_p is true and decreases if H_d is true. Continuous models make substantially better use of the DNA profile data, whereas binary systems, by necessity, discard and simplify the available information. They also offer a less laborious and more consistent approach to DNA interpretation. For these reasons, the laboratory validated STRmix™ (v2.5.11), which is a fully continuous probabilistic genotyping software used to interpret DNA profiles. STRmix™ has previously been subjected to developmental validation[1]. An internal validation was performed following the FBI Quality Assurance Standards (QAS)[2] validation requirements defined in standard 8.3.1 and the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems[3]. The validation is comprised of two parts: Part 1: Establishing STRmix™ mass parameters, and Part 2: Internal validation experiments defined as Sections A through N.

This validation describes the experiments, results, and conclusions for the following (listed in order of discussion):

- **Section A: Single Source Profiles and Accuracy**
- **Section B: Use of Peak Heights**
- **Section C: Weights**
- **Section D: Sensitivity, Specificity and Mixtures**
- **Section E: Alternate Hypotheses**
- **Section F: Assigning Number of Contributors**
- **Section G: Drop In**
- **Section H: Forward and Reverse Stutter**
- **Section I: Intra-Locus Peak Heights**
- **Section J: Inter-Locus Peak Heights**
- **Section K: Challenge Testing**
- **Section L: Known Mock or Non-Probative Casework Samples**
- **Section M: Precision**
- **Section N: NIST or NIST Traceable Sample**

This software was validated using the laboratory's current typing kit (Promega PowerPlex® Fusion 6C at 29 cycles), CE instrumentation (3500 at 3 kV and 10 second injections unless otherwise noted), and profile analysis software (GMID-X v. 1.4). The data was analyzed using the analytical threshold of 75 RFU. Stutter filters were not utilized during analysis of the non-reference profiles though all other identifiable artifacts were deleted from the profiles to be used with STRmix™. The samples used for this validation were generated during the laboratory's Fusion 6C internal validation, unless otherwise noted.

Various LRs are calculated throughout the validation. LRs that do not include a theta ($\theta = 0.01$) correction or any correction for MCMC or allele probability uncertainty will be referred to as database LRs for the purposes of this validation. Point estimate LRs include a theta correction and allele frequency uncertainty, but do not include MCMC uncertainty calculations. The 1-sided 99% lower highest posterior density (HPD), factor of N! LRs include a theta correction and both allele frequency and MCMC uncertainty calculations. They will be referred to as HPD LRs for the purposes of this validation. All three versions of LRs are used for various experiments in this validation. Based on the recommendations of the software developers, the laboratory has chosen to use the HPD LR for reporting results. All LR calculations were performed using the original 2013 National Institute of Standards and Technology (NIST) database[4]. Since the completion of the validation the database was updated to the revised NIST database[5]. The most common number from the four racial groups is reported unless otherwise noted.

Based on the results of this internal validation, the DNA results produced from the PowerPlex® Fusion 6C STR kit analyzed on a 3500 and interpreted with STRmix™ produces reliable and reproducible results; and thus, are deemed suitable for use in forensic DNA casework. These studies were used to establish the laboratory's DNA analysis and interpretation methods using the STRmix™ software.

All raw and analytical electronic data are saved in the following location: *S:\Section_Files\Forensic Biology\12. Validations\STRmix\01. LASD Validation.*

References

[1] Bright, J. et al., Forensic Science International: Genetics 23 (2016); Developmental validation of STRmix™ expert software for the interpretation of forensic DNA profiles.

[2] FBI Quality Assurance Standards for Forensic DNA Testing Laboratories (2011)

[3] SWGDAM Guidelines for Validation of Probabilistic Genotyping Systems (2015)

[4] Hill, C. et al., Forensic Science International: Genetics 7 (2013); U.S. Population Data for 29 Autosomal STR Loci

[5] Steffen, C.R., Forensic Science International: Genetics 7 (2017); http://dx.doi.org/10.1016/j.fsigen.2017.08.011. Corrigendum to U.S. Population Data for 29 Autosomal STR Loci.

Part I: Establishing STRmix™ Mass Parameters *(SWGDAM 3.2.4)*

There are a number of parameters that are not optimized by the MCMC process in a STRmix™ analysis. These parameters are set by the user and are either determined by analysis of empirical data or modeled using the Model Maker software within STRmix™. The parameters that were defined by the laboratory prior to use of STRmix™ are:

1. Stutter Ratios
2. Drop-in parameters
3. Saturation
4. Allelic and stutter peak height variances
5. Locus specific amplification efficiency (LSAE) variance.

N-1 and N+1 repeat stutter ratios:

Single source samples from 118 donors were analyzed with an analytical threshold of 15 RFU. Locus specific N-1 and N+1 repeat stutter ratios (SR) were calculated for all alleles detected with one exception. If a locus had two alleles one repeat apart, then the stutter for the lower molecular weight peak was not calculated in the overall mean since the parent peak would be artificially increased by the stutter of the higher molecular weight allele. Loci in which the alleles were two repeats apart were used for the study. N-1 repeat stutter ratios were analyzed using regression analysis against the allele number or the longest uninterrupted sequence (LUS) of repeats. Some loci with complicated internal structures were addressed by using average stutter rates observed for each detected allele or a multi-sequence analysis approach. These approaches to modeling N-1 stutter are described in more detail below. N+1 stutter ratios for all loci except D22S104 are dependent on the allele height, not allele number or LUS. For these loci the average observed forward stutter ratio was used. D22S1045 N+1 repeat stutter was determined using regression analysis against LUS.

The STRmix™ 'per allele' back stutter model is either based on the allele designation itself or the longest uninterrupted stretch (LUS) of common repeats in the allele. For the allele designation method, per allele stutter ratios were calculated using a linear equation regressing stutter ratio against allele. Within STRmix™ stutter is estimated using the model $SR = m \times Allele + c$ where the intercept (c) and the slope (m) are determined using regression. These equations are stored in an allele designation file which for this discussion is described as the default stutter file. The slope and intercept values for these equations are listed in Table 1.

For some loci, a better explanatory variable for stutter ratio is the longest uninterrupted stretch (LUS) of common repeats within the allele rather than the designation itself. Values for LUS can be determined by sequencing alleles. These values were taken from the STRmix™ support site.

STRmix™ utilizes a Stutter Exception File that is an allele by allele listing of expected stutter ratios at each locus and is accessed by the software first during a run. If there is no value in this file the software will determine the expected stutter ratio using the default stutter file. LUS, average values, and multi-sequence values are listed in the stutter exceptions file for alleles that were detected in the validation. Rare allele variants (or any allele not part of the validation) will have stutter modeled on the regression equation listed in the default stutter file.

Structural variations at some alleles in some loci can have a large impact on the expected amount of stuttering. The 14 allele at the vWA locus serves as an example. The different structural variants of the vWA locus influences the LUS value[1] and therefore the amount of stutter. Based on population data

from the Novroski paper the majority of variants have a LUS of four with different variants having a LUS of up to 11. Both short and long LUS structural variants were observed in the LASD validation data (see Appendix A). The expected LUS for a 14 allele based on the validation data is 7.2 and the associated stutter amount is 2.5%. If an unknown casework sample with a 14 allele has the longer LUS value then the stutter peak may be much higher than what is modeled, forcing the software to assign it as an allele when it may indeed only be stutter. Until more sequencing data can be obtained by the forensic community, examples of this may arise at other loci. Analysts should be aware of this phenomenon when analyzing STRmix™ results. If the deconvolution does not meet qualitative expectations at a particular locus and this stands out from the rest of the profile then the sample may need to be re-run, re-run with more iterations, or the locus may need to be ignored.

STRmix™ provides an Excel file that categorizes STR loci based on the best fit for modeling stutter ratios. They are divided into three categories: 1. Simple repeat loci that can be addressed with the default stutter file 2. LUS loci and 3. Average stutter rates observed for each allele. The results of the internal study were compared to this Excel file. Their file was created before the multi-sequence approach was included in discussions for modeling stutter. The multi-sequence approach was determined to be the best fit for one locus (SE33) with this validation data, and therefore, the LASD stutter model has one extra category. For SE33, it is known that there are some alleles with two or three different length polymorphisms that contribute to stuttering once they are over a certain number of repeats. A stretch of repeats that does not contribute to stuttering is called the lag. Every stretch of repeats over the lag contributes to the overall stutter rate, in proportion to its length. The equation representing the multi-sequence approach is as follows:

$$SR = m \sum \max (l_i - x, 0) + c$$

where m is the slope constant, l_i is the LUS for each stretch of repeats longer than the lag, x is the lag constant (the number of repeats that do not contribute to stuttering) and c is the y-intercept constant. This equation is similar to the classic line formula $y=mx+b$, except that "x" is a sum instead of a single number. A summary of the method of the predicted SR for each locus is provided in Table 1. The regression plots for all loci including those based on both the allele designation and LUS (if applicable) are shown in Appendix A of this validation.

Locus #	Marker	Allele	LUS	Mean	Multi-Sequence	Allele Slope	Allele Intercept
1	D3S1358		x			0.010164685	-0.071957608
2	D1S1656		x			0.005151832	0.013230829
3	D2S441			x		-0.000423204	0.055138699
4	D10S1248	x				0.009589379	-0.052535711
5	D13S317	x				0.010253717	-0.059391143
6	Penta E	x				0.004031982	-0.017300465
7	D16S539	x				0.011701995	-0.061603349
8	D18S51	x				0.008144953	-0.04096002
9	D2S1338			x		0.004495675	-0.009950519
10	CSF1PO	x				0.012045344	-0.068285041
11	Penta D			x		0.003105275	-0.015870101
12	TH01		x			0.002774586	1.64221E-05
13	vWA		x			0.014870925	-0.172276613
14	D21S11			x		0.005371824	-0.077479658
15	D7S820	x				0.01062843	-0.056148243
16	D5S818	x				0.011098031	-0.061899686
17	TPOX	x				0.006362334	-0.031863971
18	D8S1179			x		0.00444825	0.017697242
19	D12S391	x				0.010004303	-0.098930137
20	D19S433		x			0.010463576	-0.072561107
21	SE33				x	0.00243043	0.049510235
22	D22S1045	x				0.014142586	-0.124366793
23	FGA	x				0.007474273	-0.089294496

Table 1. Type of stutter model applied per locus. Slope and intercept listed using allele model.

STRmix™ also allows the user to input a maximum allowable stutter ratio. The maximum allowable stutter ratio reduces run time by only permitting peaks in a stutter position below a certain percentage to be considered as stutter. This parameter has been set at 0.3 (30%) for N-4 stutter and 0.15 (15%) for N+4 stutter based on inspection of laboratory stutter ratio data.

Peak Saturation:

It is known that the 3500 instrument camera saturation level is approximately 30,000 RFU[(2)]. This also has been observed in our lab with over-amplified validation and casework samples. The laboratory does not routinely interpret profiles with saturated data. These samples are either reinjected at a lower injection time or diluted and reinjected. Therefore, the saturation point was set to 30,000 RFU.

Allele and Stutter peak height variance and LSAE:

Within STRmix™ the variability of peaks within profiles is described using a model containing a variance constant. Setting variance parameters involves measuring the variability in a number of single source profiles using the Model Maker function within STRmix™. The profiles should encompass the range of profile quality encountered in casework from low level partial profiles (minimum 10 peaks) to full profiles approaching the camera's saturation threshold. The allele variance is described as c^2 and the stutter variance is k^2. These variables are modeled by gamma distributions and determined through the MCMC process. The starting position for these values within the MCMC is the mode of the gamma distribution based on empirical values.

That empirical data came from ten laboratory donors amplified at 10 different input amounts covering a range from 100pg to 2ng and two Fusion 6C positive control samples. Each sample was injected on the 3500 using the laboratory's current injection times of 10 and 5 seconds. The Model Maker function within the software then determined the allele, stutter and LSAE variance parameters. Variances were initially calculated using software version 2.4.06 separately for the 10 and 5 second injections and then combined. A summary of the allele and stutter variances for each injection time and the combined data are shown in Figure 1.

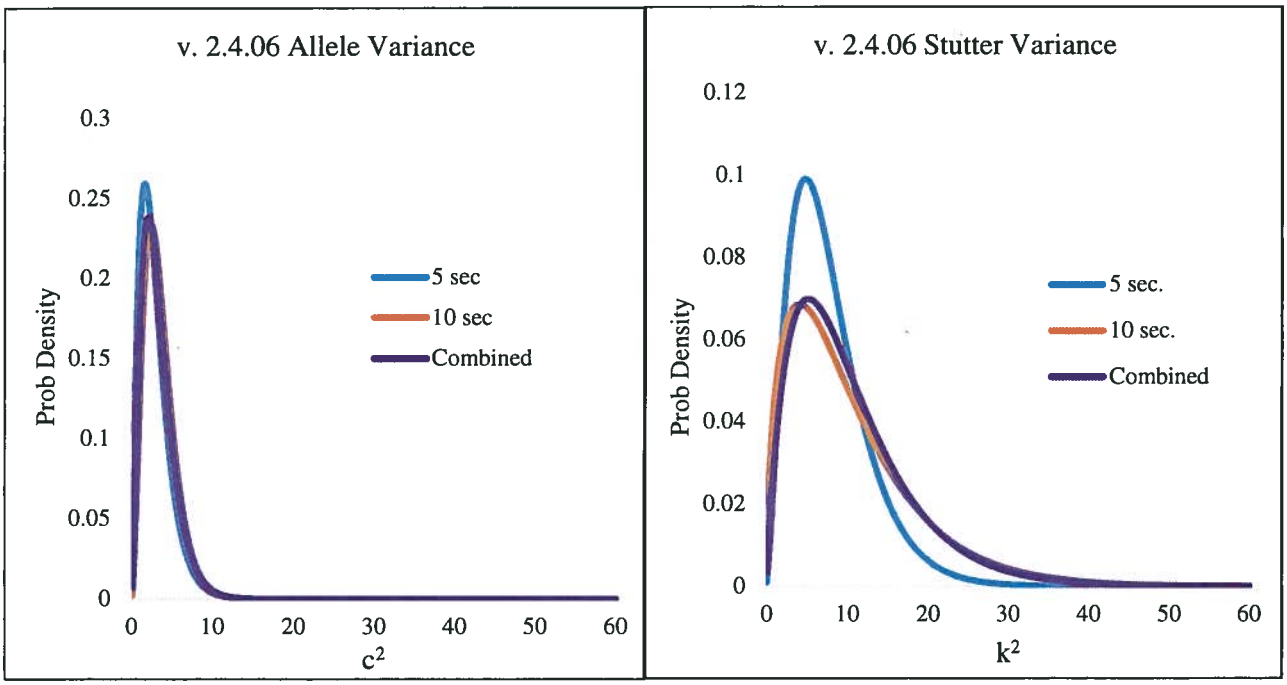


Figure 1. Allele and stutter variance gamma distributions for 10 and 5 second injections and combined data for v. 2.4.06

Heterozygote balance was calculated for all heterozygote loci for the Model Maker profiles. Heterozygote balance (Hb) was calculated as:

$$Hb = \frac{O_{HMW}}{O_{LMW}}$$

where O_{HMW} refers to the observed height of the high molecular weight allele and O_{LMW} refers to the observed height of the low molecular weight allele. In single source samples variability in Hb reduces as the average peak height (APH) at a locus increases. The variance of Hb is expected to be twice the variance of the individual allelic peaks assuming the variance of each peak is the same. This allows an approximate comparison between the variance from the STRmix™ MCMC approach and a readily determined variable from the empirical data. The plot of log(Hb) versus APH for the data and the expected 95% bounds (plotted as dotted lines) calculated as:

$$\pm\sqrt{2} \times 1.96 \times \sqrt{\frac{c^2}{APH}}$$

where $c^2 = 2.88$, which is the 50th percentile from the gamma distribution from the combined data set. The 95% bounds encapsulate sufficient data as demonstrated in Figure 2 (coverage = 96.6%) demonstrating that the values for variance are sufficiently optimized. It serves as an approximate check of Model Maker.

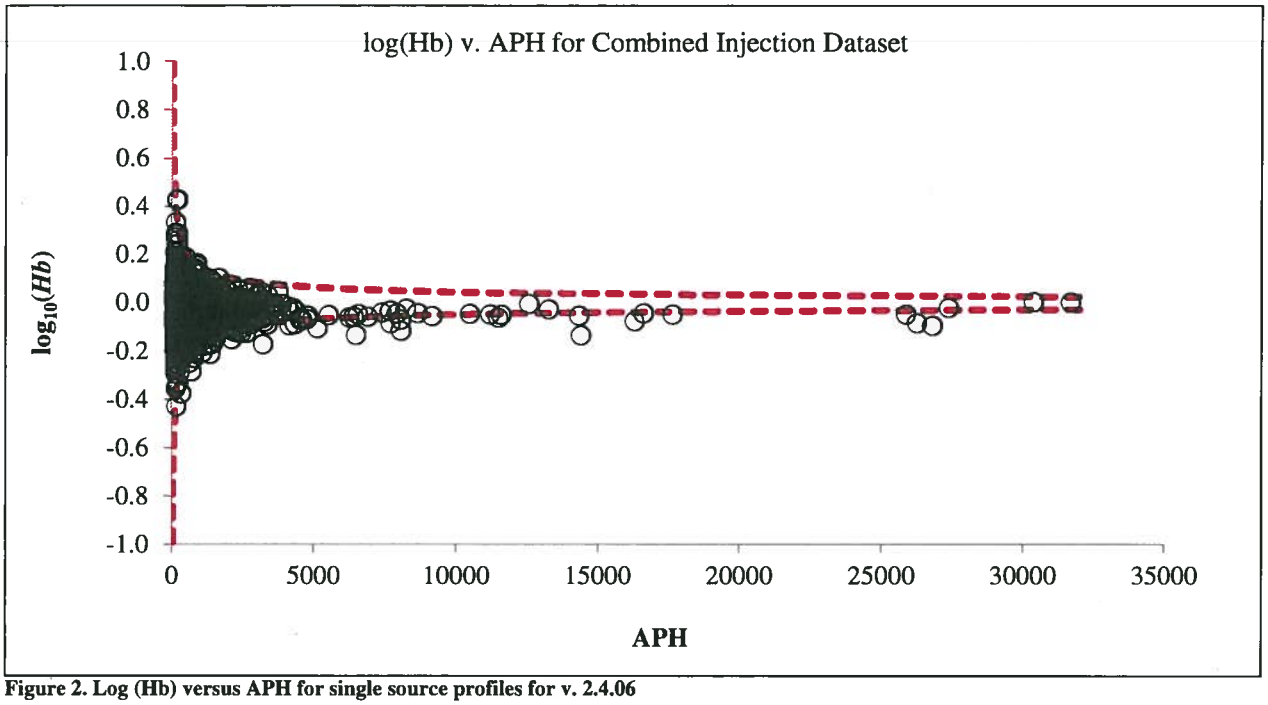


Figure 2. Log (Hb) versus APH for single source profiles for v. 2.4.06

The correlation plots for LMW versus HMW allele and allele versus stutter peaks for the combined dataset for version 2.4.06 are shown in Figure 3. The correlation plots for the 10 and 5 second data are similar and the data is not shown. The distribution of the points within the figures is as expected, with no observed correlation. The points being evenly distributed among the 4 quadrants is demonstrative of a lack of correlation. There are two outliers observed in the stutter correlation plot. These are larger than expected stutter peaks that were labeled at analysis, however, they do not affect the results. The appropriateness of these values was also tested by interpreting a range of mixed DNA profiles conducting the specificity and sensitivity testing as described in Section D.

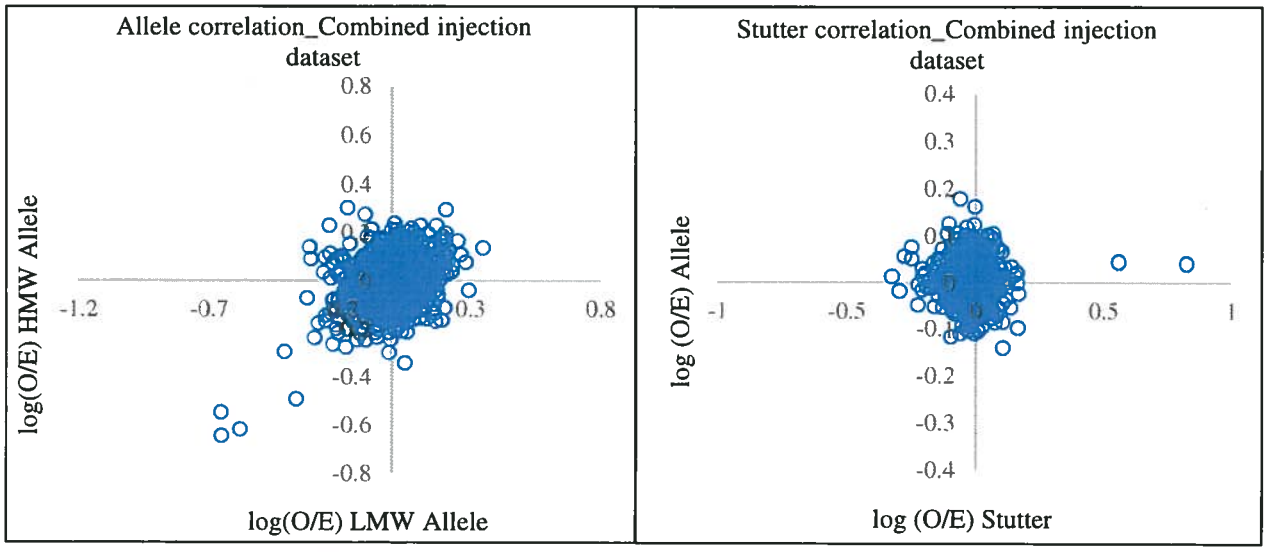


Figure 3. Correlation plots for v. 2.4.06

Based on the combined data compiled and provided by scientists at the Environmental Science & Research (ESR) Laboratory it was determined that a single set of variance values could be used for both the 10 and 5 second injections. All samples from sections C, D and G and some of the samples from section A and L of this validation were run using STRmix™ version 2.4.06 under the combined injection time parameters listed in Table 2. During the validation an update to STRmix™ 2.5.11 became available. A performance check was then conducted on a subset of single source and mixture samples analyzed thus far to compare the results from each version of the software using the original variance values. Model Maker was then rerun within version 2.5.11 using the combined 5 and 10 second data and new variance values were generated (see Table 2). These new values were applied to the performance check samples. The results from the 3 different runs were similar and the new version of the software was installed and used for the remainder of the validation studies. According to the STRmix™ upgrade bulletin[3], point estimate LRs generated for single source samples with a weight =1 should be identical if the same theta value and allele frequency database are used and mixture profiles should yield different but similar LRs due to the expected variability within the MCMC. The expected results were obtained. A summary of the LRs and the diagnostic data including the average log likelihood, Gelman-Rubin (GR), the allele variance (c^2), and the stutter variance (k^2) generated from the three-way comparison for single source profiles and mixtures is shown in Tables 3-6. The full reports from the performance check samples are stored electronically.

Finally, the allele and stutter variance plots, correlation plots, and log(Hb) versus APH plot for version 2.5.11 are shown in Figures 4 through 6. The 95% bounds encapsulate 96.7% of the data when applying the 50th percentile c^2 value from the gamma distribution (c^2 = 2.84) for version 2.5.11. The remaining sections of this validation were run using version 2.5.11.

STRmix™ Version	Number of Profiles Analyzed	Allele Variance Parameters (Mode)	Stutter Variance Parameters (Mode)	Mean LSAE Variance
2.4.06	102	Gamma 2.643, 1.244 (2.044)	Gamma 1.984, 5.303 (5.218)	0.005
2.5.11	102	Gamma 2.743, 1.208 (2.105)	Gamma 2.119, 4.137 (4.630)	0.004

Table 2. Allele and stutter variance values for v. 2.4.06 and 2.5.11

Sample ID	STRmix™ Version	Ave Log likelihood	GR	c^2	k^2	HPD LR	Point estimate LR (AA)	Point estimate LR (CAUC)	Point estimate LR (Asian)	Point estimate LR (HISP)
1ng PQ183	2.4.06	66.24	1.02	1.800	6.500	3.92E+29	1.62E+33	1.80E+32	1.05E+30	1.97E+31
	2.5.11 old MM	66.33	1.01	1.839	6.447	1.59E+29	1.62E+33	1.80E+32	1.05E+30	1.97E+31
	2.5.11 new MM	66.46	1.01	1.867	6.334	1.46E+29	1.62E+33	1.80E+32	1.05E+30	1.97E+31

Table 3. Comparison of STRmix™ versions 2.4.06 and 2.5.11 for a single source sample (PQ183). Old MM = Model Maker variance values from 2.4.06 and new MM = Model Maker variance values from 2.5.11

		STRmix™ Mixture proportions						HPD LR	
Sample ID	STRmix™ Version	PQ183 (M)	PQ212 (F)	Ave Log likelihood	GR	c^2	k^2	PQ183 (M)	PQ212 (F)
FM 1:4 _1ng-a	2.4.06	83%	17%	89.90	1.01	1.800	6.600	2.29E+29	6.70E+28
	2.5.11 old MM	83%	17%	89.91	1.00	1.804	6.618	1.49E+29	
	2.5.11 new MM	83%	17%	90.07	1.00	1.839	6.367	1.21E+29	
MF 1:1 200pg-a	2.4.06	55%	45%	59.24	1.01	1.500	8.000	6.81E+15	1.66E+16
	2.5.11 old MM	55%	45%	59.62	1.01	1.628	7.892	3.73E+15	
	2.5.11 new MM	55%	45%	58.58	1.01	1.681	7.493	6.64E+15	

Table 4. Comparison of STRmix™ versions 2.4.06 and 2.5.11 for 2 person mixtures. Old MM = Model Maker variance values from 2.4.06 and new MM = Model Maker variance values from 2.5.11

		STRmix™ Mixture proportions							HPD LR		
Sample ID	STRmix™ Version	PQ243 (M)	PQ183 (M)	PQ212 (F)	Ave Log likelihood	GR	c^2	k^2	PQ243 (M)	PQ183 (M)	PQ212 (F)
MMF 1:1:1 300pg-a	2.4.06	26%	40%	34%	68.50	1.01	1.700	6.700	3.51E+10	1.46E+09	1.35E+10
	2.5.11 old MM		40%		72.18	1.01	1.703	6.357		5.89E+08	
	2.5.11 new MM		40%		69.40	1.01	1.747	6.002		6.65E+08	
MMF 1:5:1 300pg-b	2.4.06	12%	72%	16%	60.12	1.02	1.400	5.800	1.90E+00	1.66E+29	3.71E+08
	2.5.11 old MM		72%		60.20	1.02	1.423	5.844		8.97E+28	
	2.5.11 new MM		72%		59.38	1.05	1.475	6.464		9.87E+28	
MMF 1:5:1 1ng-b	2.4.06	13%	76%	11%	98.80	1.01	1.900	3.800	6.75E+16	1.62E+29	9.73E+16
	2.5.11 old MM		76%		98.47	1.01	1.937	3.755		1.07E+29	
	2.5.11 new MM		76.00%		98.48	1.01	2.012	3.645		1.07E+29	
MMF 1:1:20 1ng-b	2.4.06	3%	4%	93%	83.96	1.06	2.100	7.400	4.38E+09	1.83E+05	1.53E+29
	2.5.11 old MM		4%		83.48	1.04	2.185	7.855		3.76E+05	
	2.5.11 new MM		4%		82.07	1.01	2.277	7.247		4.28E+05	

Table 5. Comparison of STRmix™ versions 2.4.06 and 2.5.11 for 3 person mixtures. Old MM = Model Maker variance values from 2.4.06 and new MM = Model Maker variance values from 2.5.11

		STRmix™ Mixture proportions								HPD LR			
Sample ID	STRmix™ Version	PQ212 (F)	PQ183 (M)	PQ243 (M)	PQ94 (M)	Ave Log likelihood	GR	c^2	k^2	PQ212 (F)	PQ183 (M)	PQ243 (M)	PQ94 (M)
FMMM 1:1:1:1 300pg-a	2.4.06	26%	29%	24%	21%	81.02	1.00	1.300	5.400	7.35E+07	7.76E+05	5.96E+09	1.77E+10
	2.5.11 old MM		31%			73.60	1.01	1.609	6.932		3.64E+05		
	2.5.11 new MM		31%			71.20	1.02	1.714	5.940		3.67E+05		
FMMM 1:1:1:10 1ng-a	2.4.06	7%	8%	7%	78%	113.13	1.02	2.200	18.000	1.87E+09	1.57E+09	1.99E+09	2.96E+29
	2.5.11 old MM		8%			112.80	1.01	2.310	17.740		1.17E+09		
	2.5.11 new MM		8%			112.04	1.00	2.215	16.780		1.09E+09		

Table 6. Comparison of STRmix™ versions 2.4.06 and 2.5.11 for 4 person mixtures. Old MM = Model Maker variance values from 2.4.06 and new MM = Model Maker variance values from 2.5.11

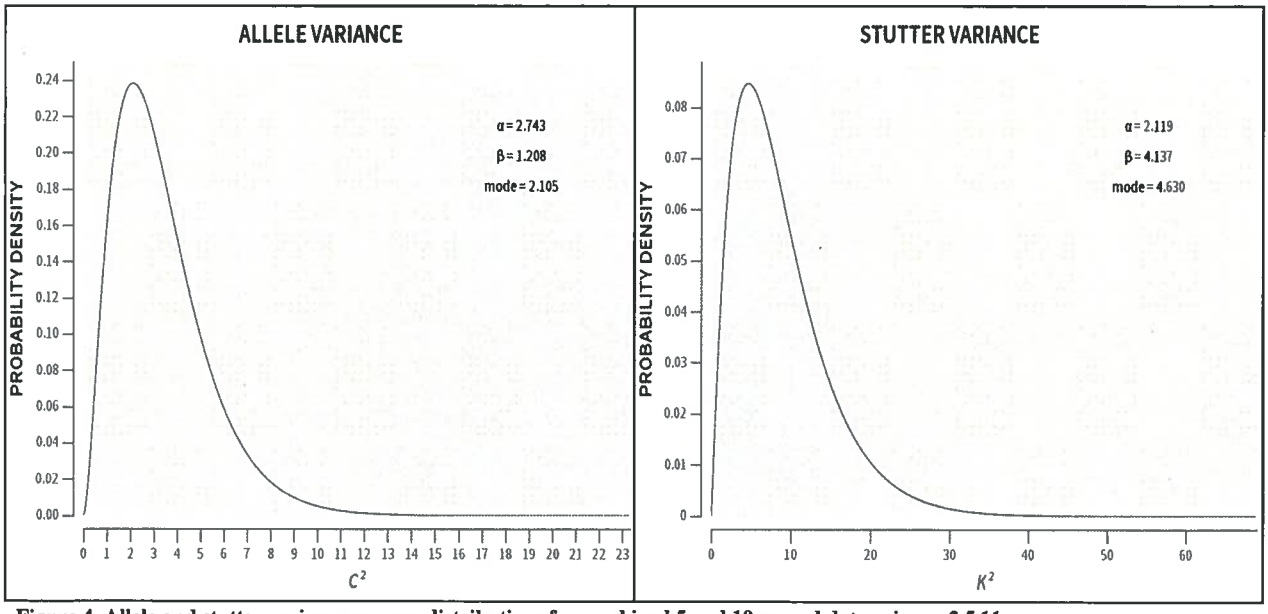


Figure 4. Allele and stutter variance gamma distributions for combined 5 and 10 second data using v. 2.5.11

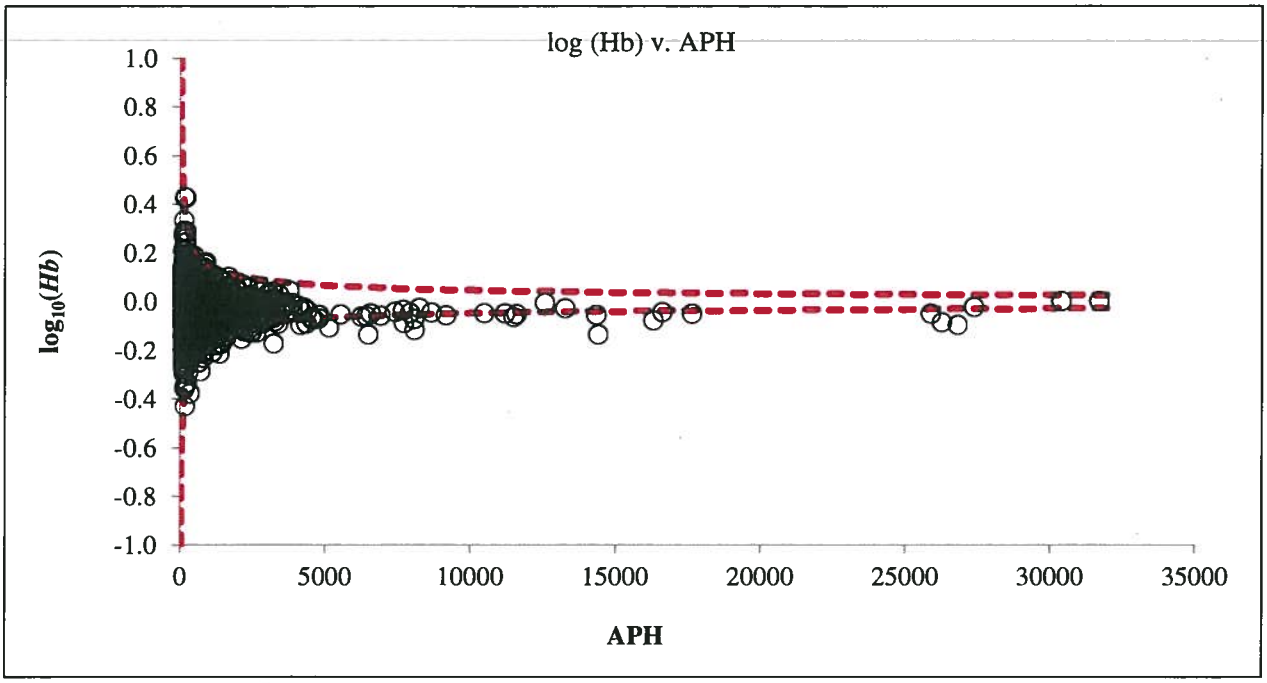


Figure 5. Log (Hb) vs. APH for single source profiles for combined 5 and 10 second data using v. 2.5.11

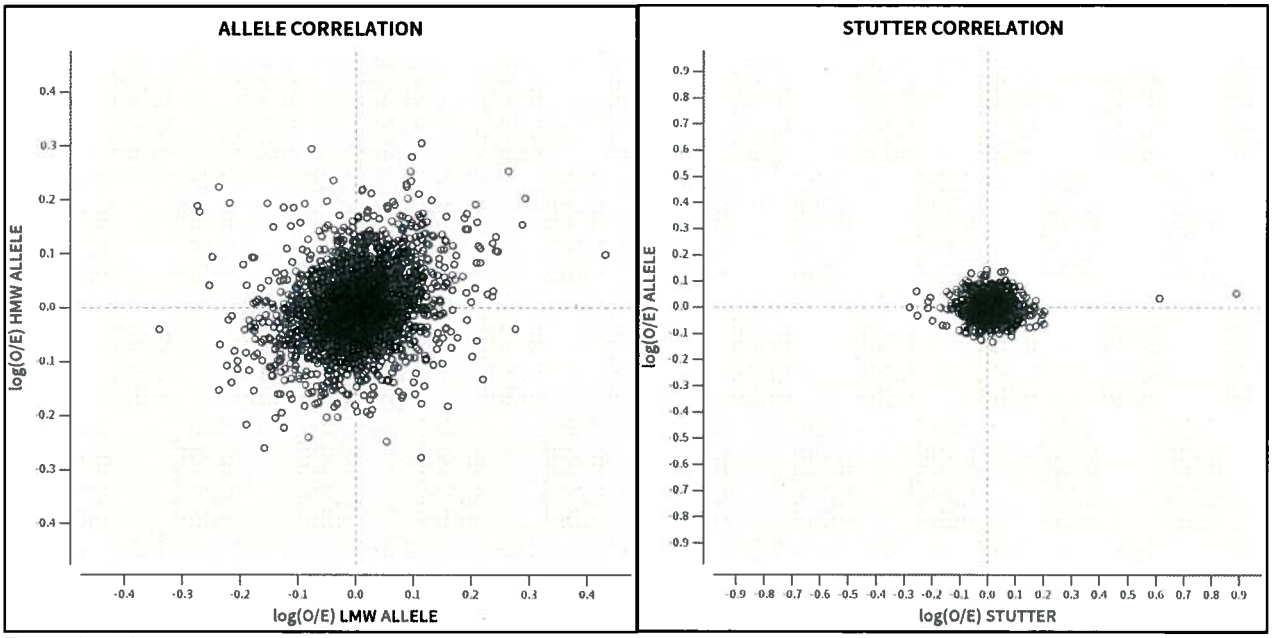


Figure 6. Allele and stutter correlation plots for combined 5 and 10 second data using v. 2.5.11

Drop-in:

The STRmix™ validation plan stated that the drop-in rate would be set to zero since no drop-in events were detected during the Fusion 6C internal validation. However, it was decided to revisit the drop-in rate since the Fusion 6C kit has now been implemented in casework. There are four parameters used for the modeling of drop-in by STRmix™. These include the following:

1. Analytical threshold
2. A cap on the maximum allowed height for a peak to be modeled as drop-in
3. Drop-in rate or frequency
4. Two parameters (α and β) for the gamma distribution model

Drop-in rates are determined by recording the counts and corresponding heights of drop-in peaks observed in negative control samples. A total of 473 negative amplification controls (NAC) and extraction reagent blanks were evaluated. Ninety-three of these samples were NACs from the contamination study performed during the Fusion 6C internal validation. These NACs were injected for 10 or 15 seconds. The remaining 380 samples were 10 second injection data compiled from 63 casework batching plates (CB, CX and PC runs) that spanned six months of laboratory casework with the Fusion 6C kit. The samples were analyzed in GMID-X using the laboratory's analytical threshold of 75 RFU. For each of the 473 profiles, the 23 autosomal loci in the Fusion 6C kit were assessed for the presence of a drop-in allele detected above 75 RFU. This resulted in the examination of 10,879 loci.

Out of 10,879 loci only one locus had a drop-in allele, resulting in a calculated drop-in frequency per locus of 1 / 10,879 = 0.0000919. The peak height of the allele was 82 RFU. Based on these results, the drop-in frequency was set to 0.0001 and the drop-in cap was set to 100 RFU. Since only one drop-in observation was detected, there was insufficient data to determine the parameters (α and β) needed to inform a gamma distribution. Therefore, they were set to zero at this time. This results in a uniform distribution model, which applies the empirically derived drop-in frequency. The final parameters are summarized in Table 7. The lab will continue to monitor drop-in and may have to adjust these values as more data is collected.

Parameter	Values used in STRmix™
Drop-in frequency	0.0001
Drop-in cap	100 RFU*
Drop-in α and β	0,0

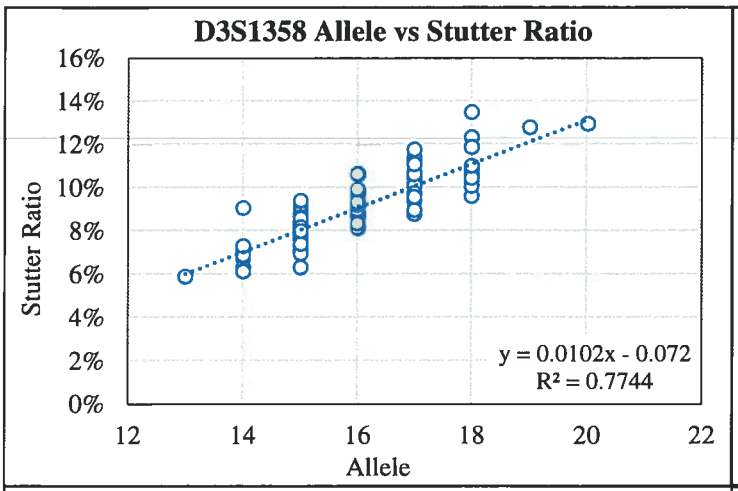
Table 7. Drop-in settings *Includes alleles with peaks heights = 100 RFU.

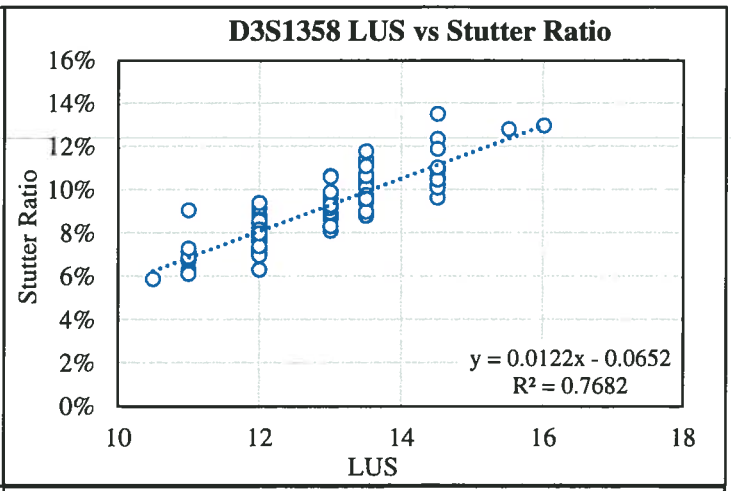
References:

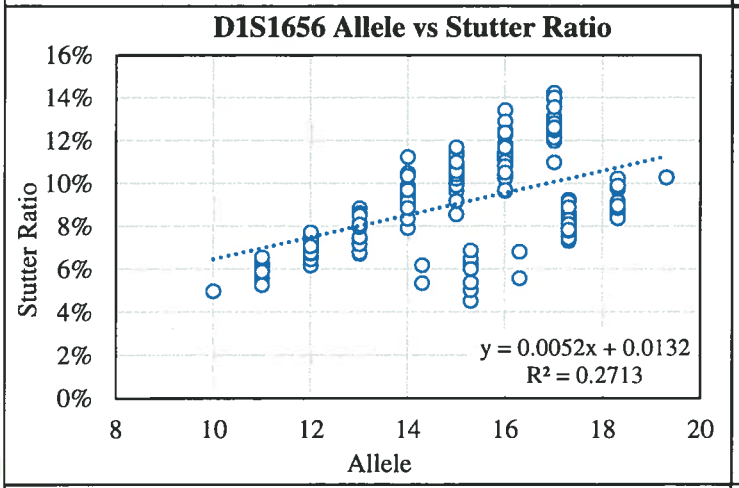
(1) Novroski, N. et al., Forensic Science International: Genetics 25 (2016); Characterization of genetic sequence variation of 58 STR loci in four major population groups.

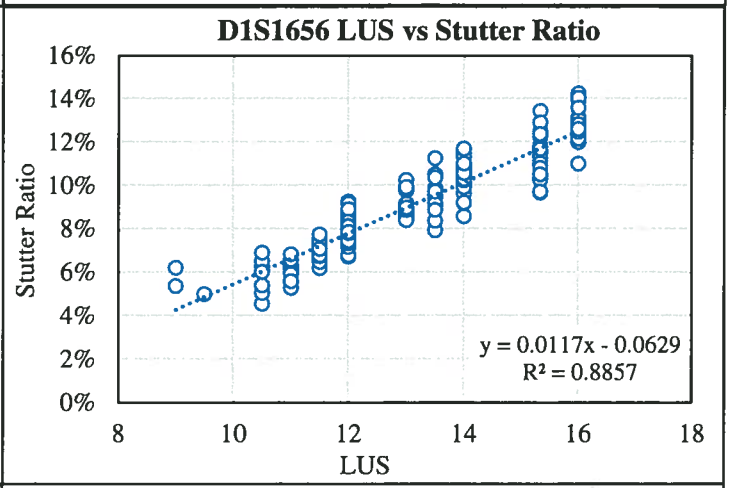
(2) Butler, J. Forensic DNA Typing: Interpretation Chapter 2 (2015), page 32

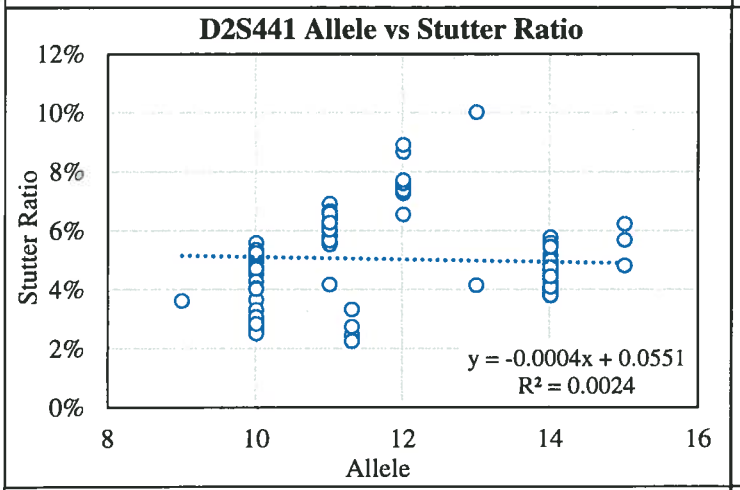
(3) STRmix v2.5.11 Release and Testing Report July 2017

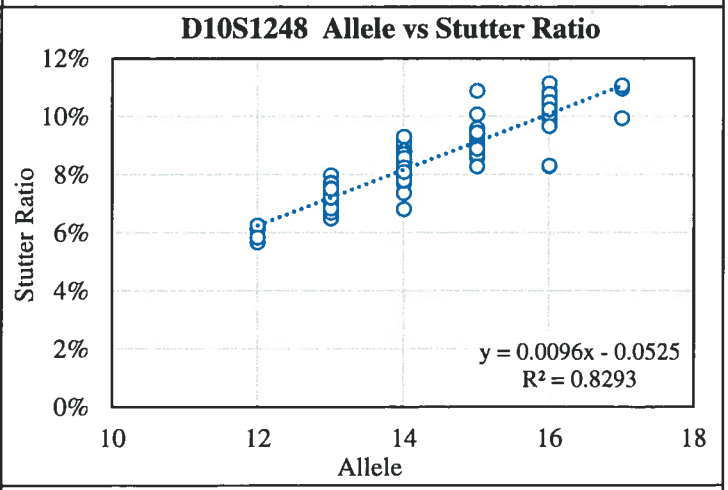

D3S1358 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0102x - 0.072
R² = 0.7744

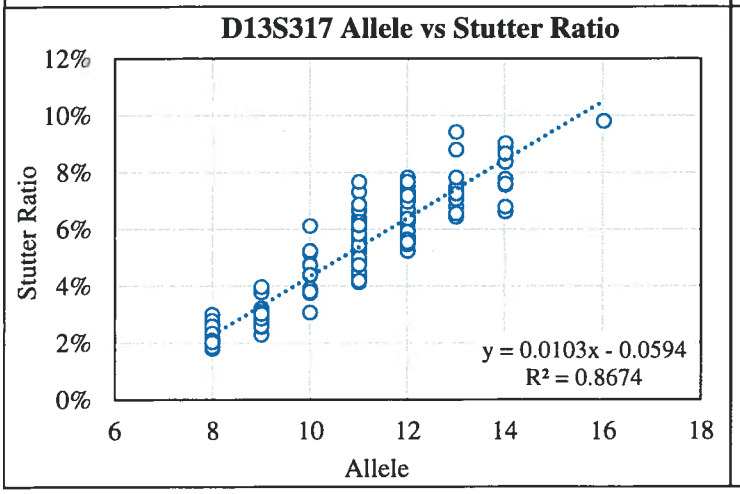

D3S1358 LUS vs Stutter Ratio
Stutter Ratio
LUS
y = 0.0122x - 0.0652
R² = 0.7682

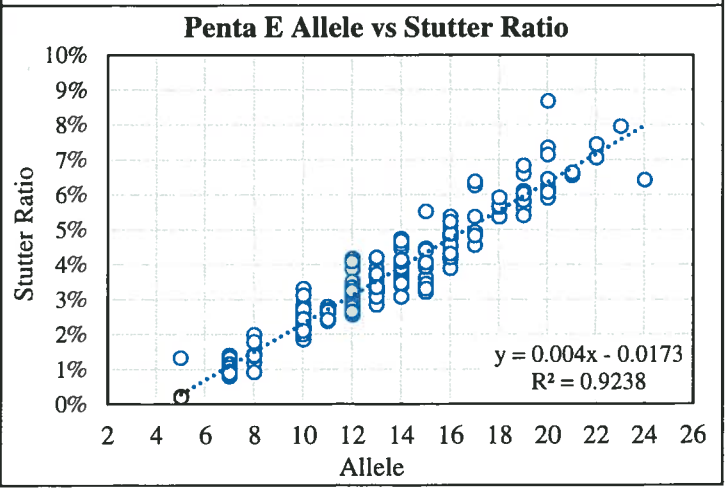

D1S1656 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0052x + 0.0132
R² = 0.2713

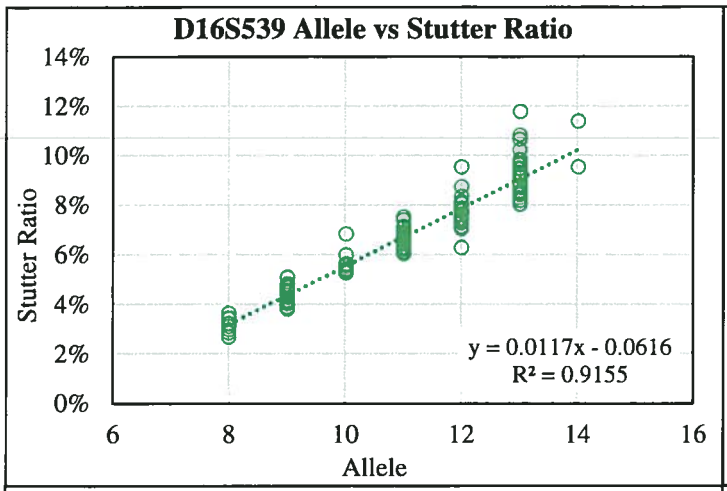

D1S1656 LUS vs Stutter Ratio
Stutter Ratio
LUS
y = 0.0117x - 0.0629
R² = 0.8857

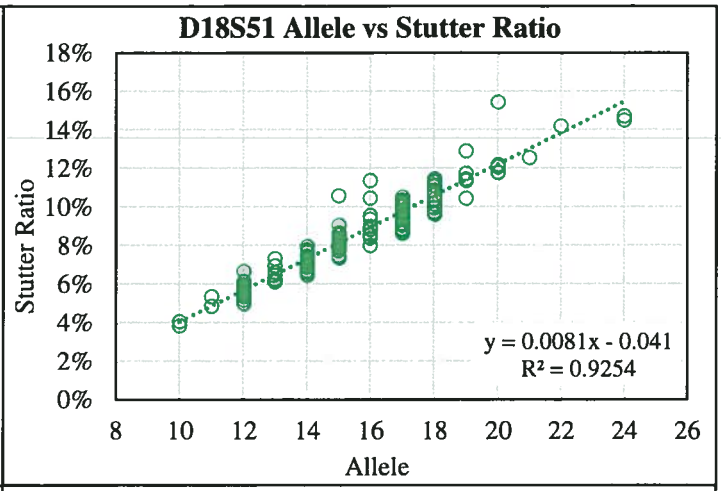

D2S441 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = -0.0004x + 0.0551
R² = 0.0024

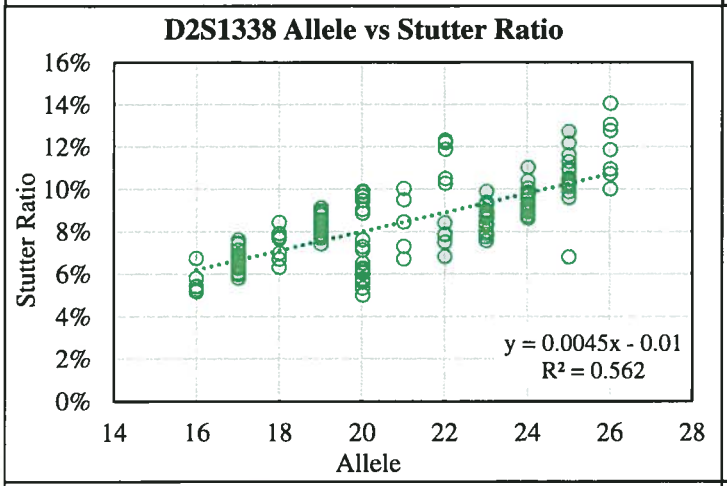

D10S1248 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0096x - 0.0525
R² = 0.8293

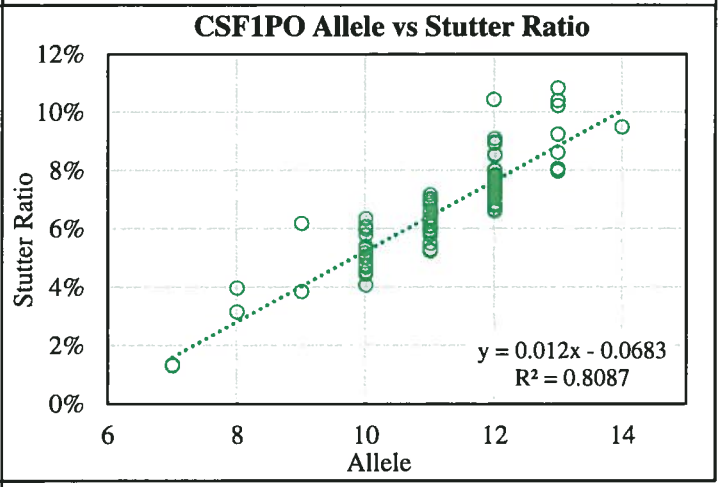

D13S317 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0103x - 0.0594
R² = 0.8674

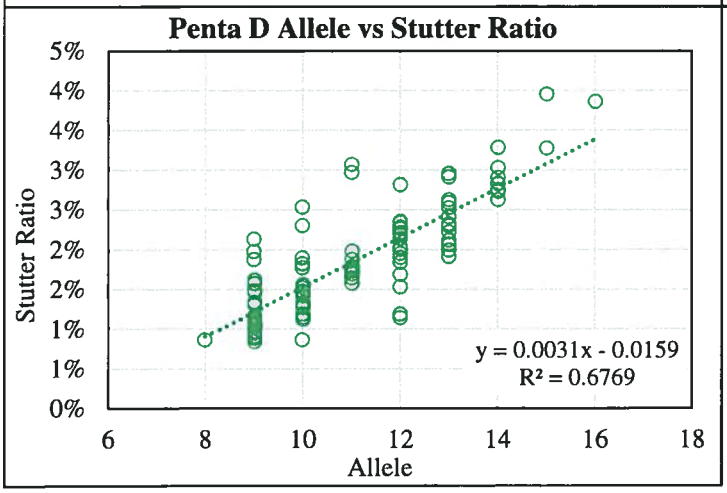

Penta E Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.004x - 0.0173
R² = 0.9238

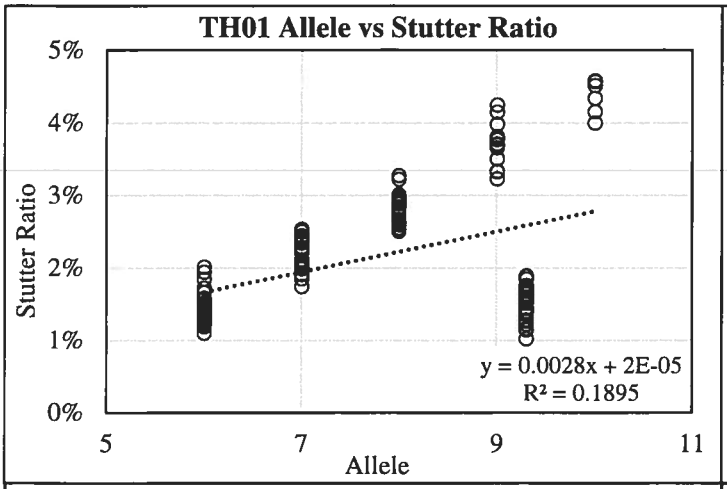

D16S539 Allele vs Stutter Ratio
Stutter Ratio
14%
12%
10%
8%
6%
4%
2%
0%
6
8
10
12
14
16
Allele
y = 0.0117x - 0.0616
R² = 0.9155

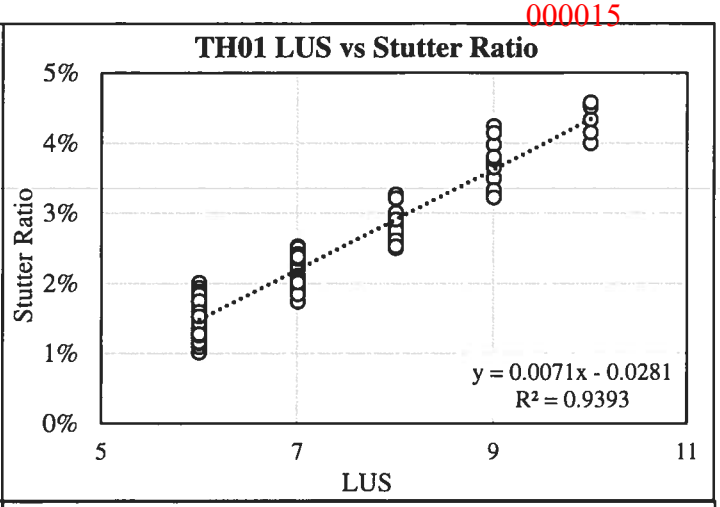

D18S51 Allele vs Stutter Ratio
Stutter Ratio
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
8
10
12
14
16
18
20
22
24
26
Allele
y = 0.0081x - 0.041
R² = 0.9254

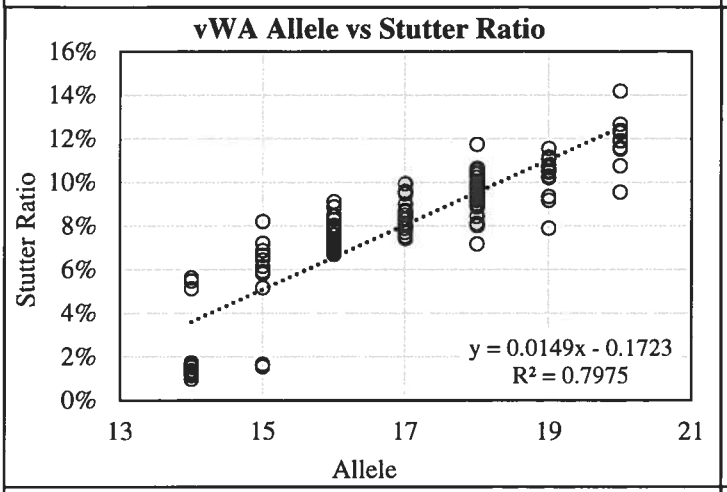

D2S1338 Allele vs Stutter Ratio
Stutter Ratio
16%
14%
12%
10%
8%
6%
4%
2%
0%
14
16
18
20
22
24
26
28
Allele
y = 0.0045x - 0.01
R² = 0.562

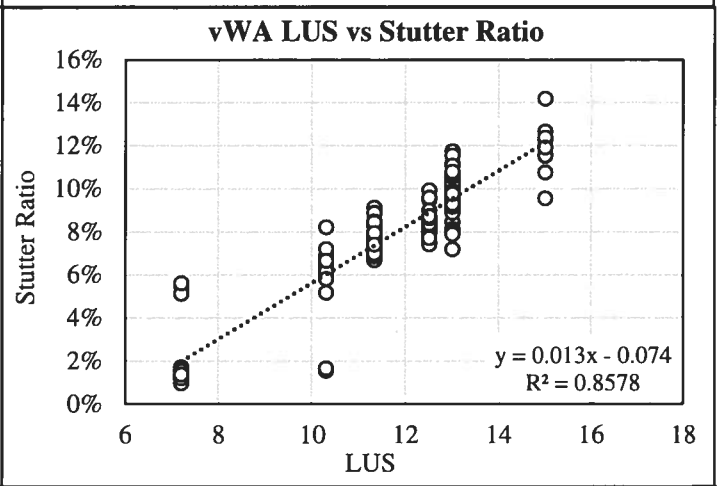

CSF1PO Allele vs Stutter Ratio
Stutter Ratio
12%
10%
8%
6%
4%
2%
0%
6
8
10
12
14
Allele
y = 0.012x - 0.0683
R² = 0.8087

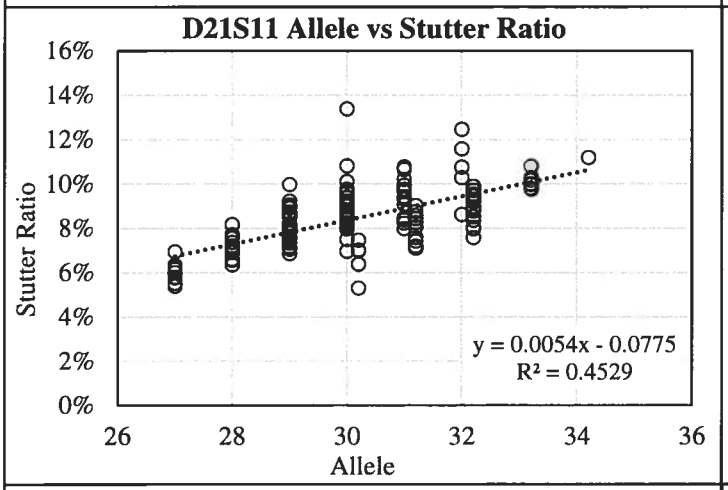

Penta D Allele vs Stutter Ratio
Stutter Ratio
5%
4%
4%
3%
3%
2%
2%
1%
1%
0%
6
8
10
12
14
16
18
Allele
y = 0.0031x - 0.0159
R² = 0.6769

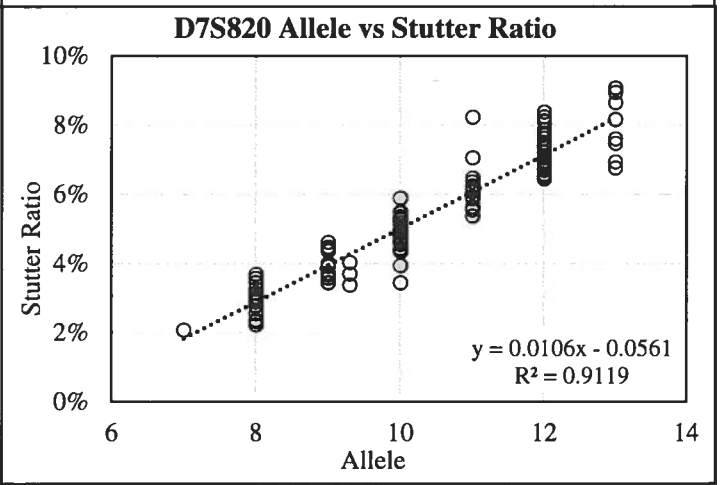

TH01 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0028x + 2E-05
R² = 0.1895

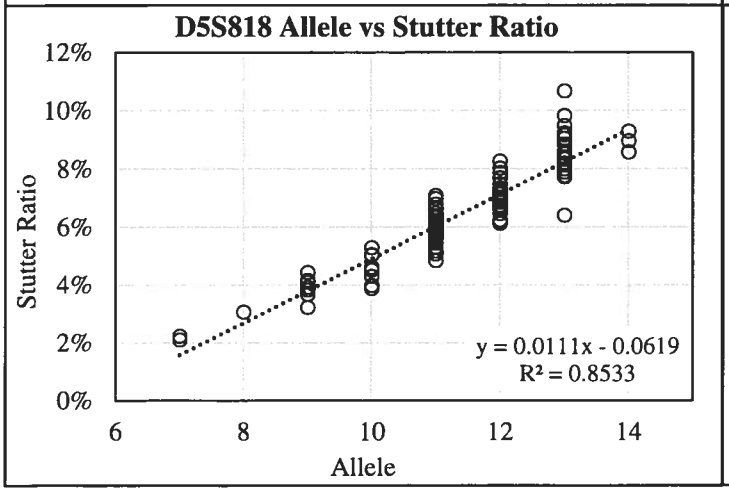

TH01 LUS vs Stutter Ratio
Stutter Ratio
LUS
y = 0.0071x - 0.0281
R² = 0.9393

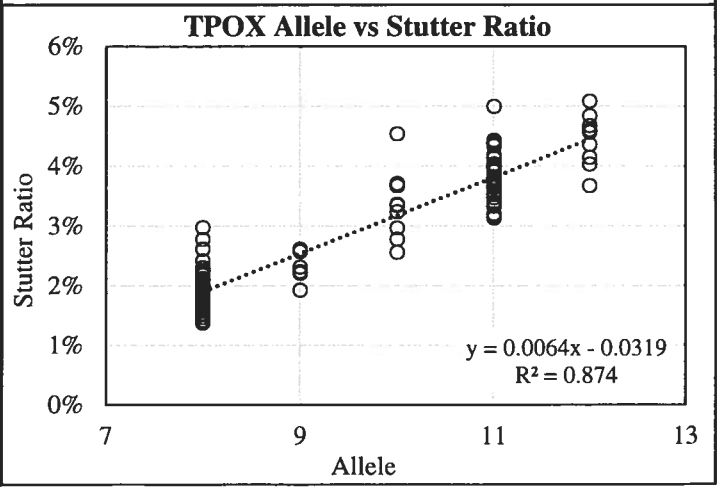

vWA Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0149x - 0.1723
R² = 0.7975

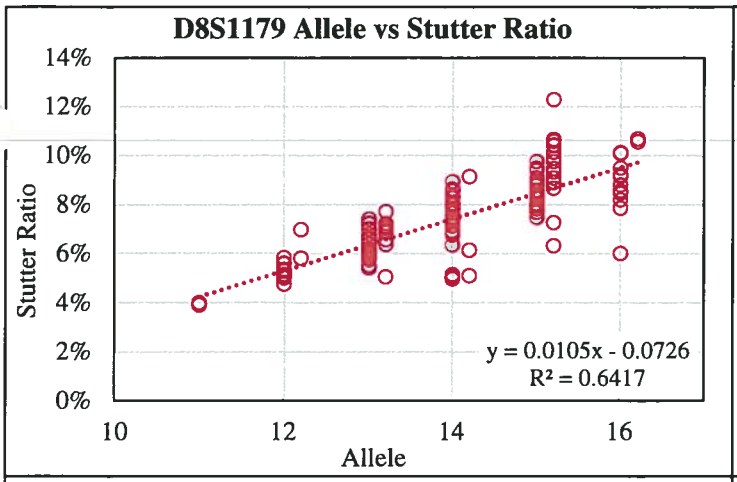

vWA LUS vs Stutter Ratio
Stutter Ratio
LUS
y = 0.013x - 0.074
R² = 0.8578

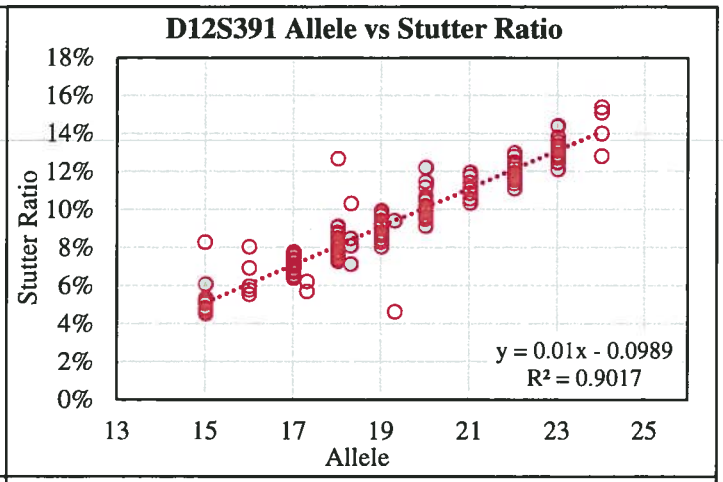

D21S11 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0054x - 0.0775
R² = 0.4529

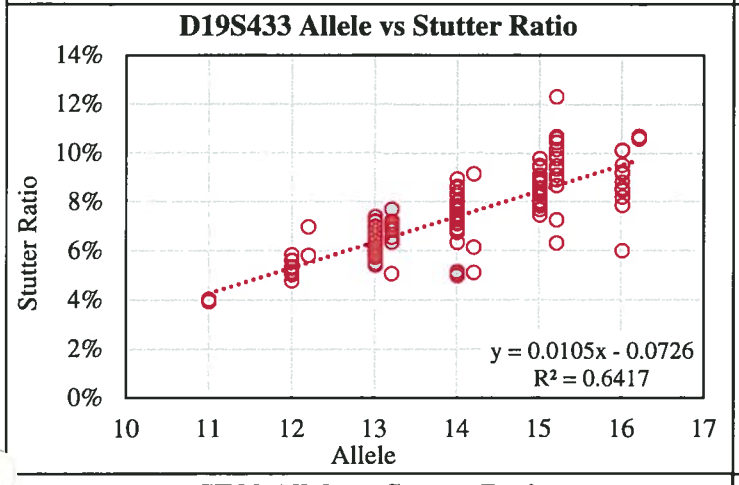

D7S820 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0106x - 0.0561
R² = 0.9119

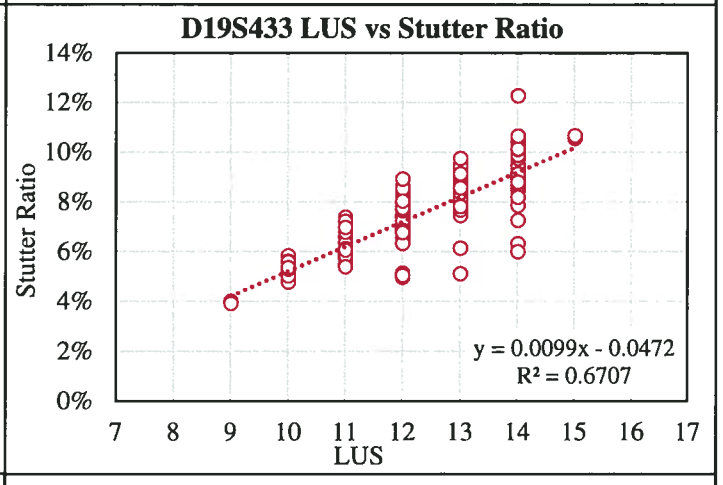

D5S818 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0111x - 0.0619
R² = 0.8533

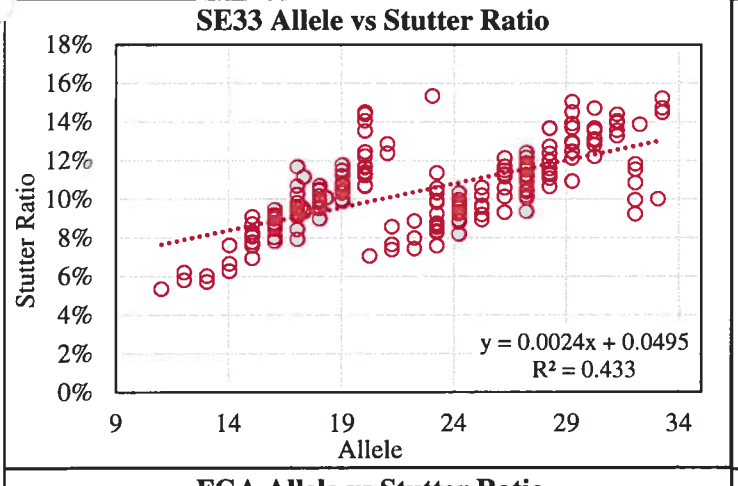

TPOX Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0064x - 0.0319
R² = 0.874

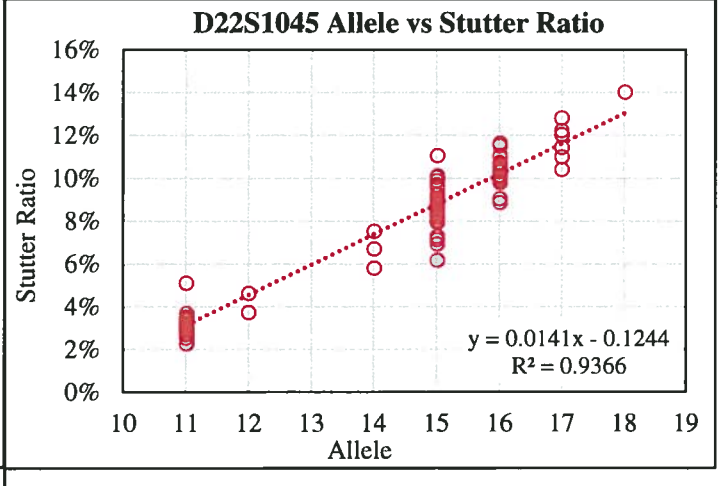

D8S1179 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0105x - 0.0726
R² = 0.6417

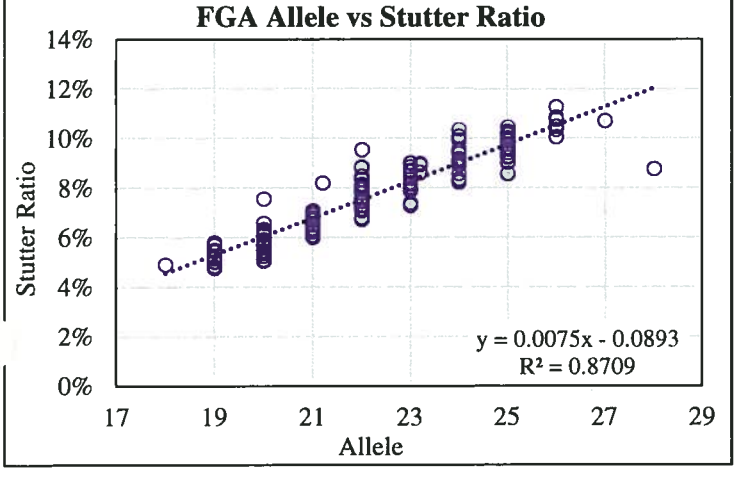

D12S391 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.01x - 0.0989
R² = 0.9017

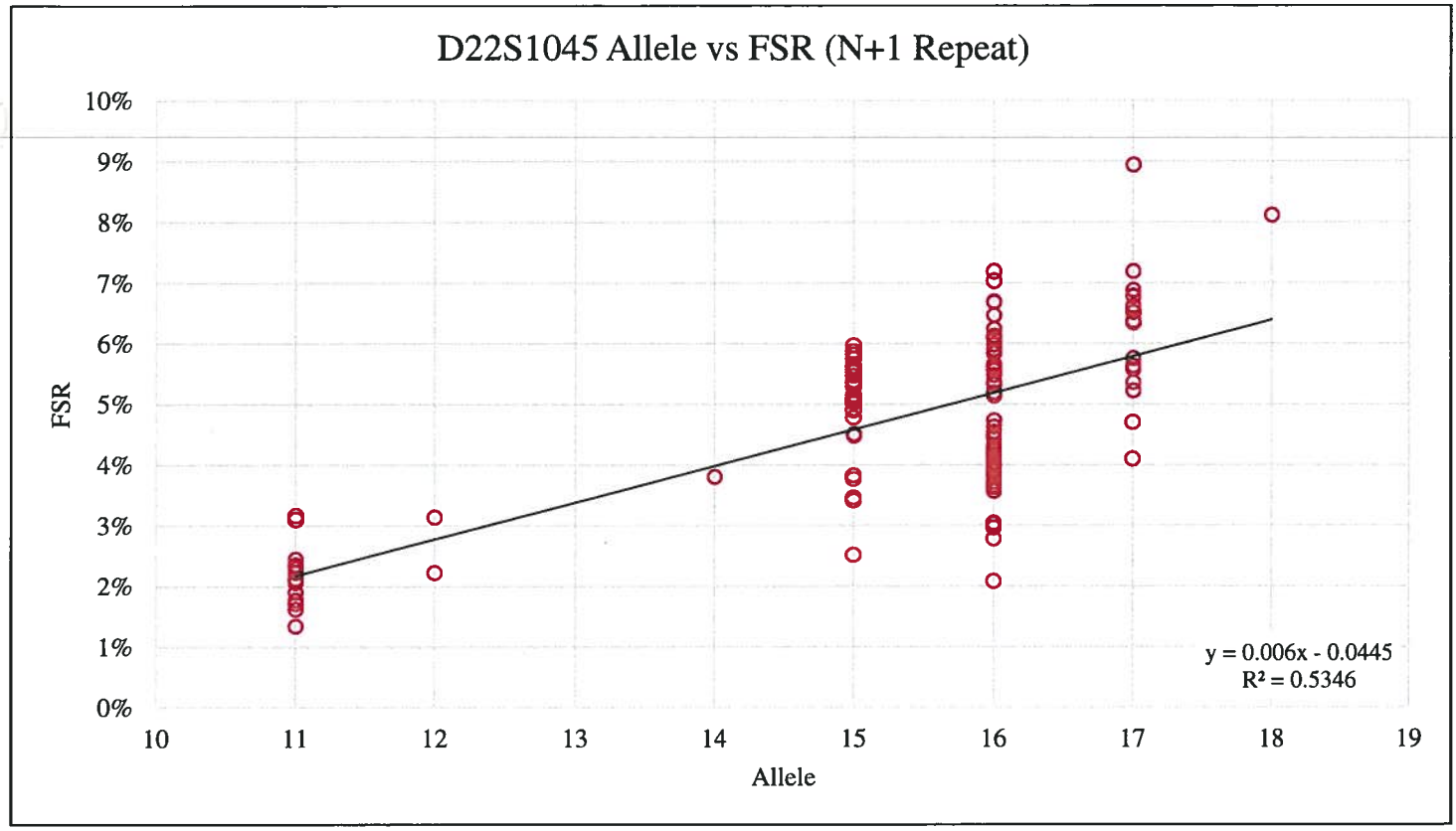

D19S433 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0105x - 0.0726
R² = 0.6417


D19S433 LUS vs Stutter Ratio
Stutter Ratio
LUS
y = 0.0099x - 0.0472
R² = 0.6707


SE33 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0024x + 0.0495
R² = 0.433


D22S1045 Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0141x - 0.1244
R² = 0.9366


FGA Allele vs Stutter Ratio
Stutter Ratio
Allele
y = 0.0075x - 0.0893
R² = 0.8709


D22S1045 Allele vs FSR (N+1 Repeat)
10%
9%
8%
7%
6%
5%
4%
3%
2%
1%
0%
FSR
10
11
12
13
14
15
16
17
18
19
Allele
y = 0.006x - 0.0445
R² = 0.5346

Section A: Single Source Profiles and Accuracy

This section covers the following standards:

4.1.5. Single-source specimens

4.2.1.2. For single-source specimens with high quality results, genotypes derived from non-probabilistic analyses of profiles above the stochastic threshold should be in complete concordance with the results of probabilistic methods.

Genotype weights produced by STRmix™ should be intuitively correct such that the most supported genotypes have the highest weights. STRmix™ should return a single genotype at each locus with a weight = 1.0 (or 100%) for a complete single source profile with optimal RFU values. As the DNA input is decreased more drop out and/or peaks in the stochastic zone will be observed. For these types of samples genotypes that consider drop out should be weighted more heavily by the software. In order to verify whether the weights assigned to different genotype combinations are appropriate, a dilution series of two single source laboratory donor samples (PQ207 and PQ183) generated during the Fusion 6C internal validation sensitivity study were used. PQ207 was amplified at DNA input amounts of 1.0ng, 0.500ng, 0.250ng, 0.100ng, 0.050ng, 0.025ng and 0.013ng. PQ183 was amplified at DNA input amounts of 1.0ng, 0.750ng, 0.114ng, 0.060ng, 0.048ng, 0.036ng, 0.027ng, and 0.025ng. The subsequent profiles had peak heights ranging from full profiles with optimal RFU values (i.e. ~1000 to 3000 RFU) to very partial profiles where drop out was observed. The profiles were interpreted with STRmix™ using the following propositions:

- H_p = The DNA originated from the person of interest (POI)
- H_d = The DNA originated from an unknown individual

The data was evaluated to ensure the HPD LR and template amounts from the STRmix™ output decreased as the DNA input dropped. The template value (in RFU) within the STRmix™ report is the average of the per chain modes of the template amounts of DNA (in RFU) generated for each contributor accepted during post burn-in. The results are plotted in Figures 1 and 2.

Figure 1 shows the log(HPD LR) progressing from a high value (log(LR) > 30 or LR > 10^{30}) for a full single source profile towards a log(HPD LR) = 0 (or LR = 1) as the DNA input decreased. This demonstrates that the weights for genotypes considering drop out increased as the template dropped.

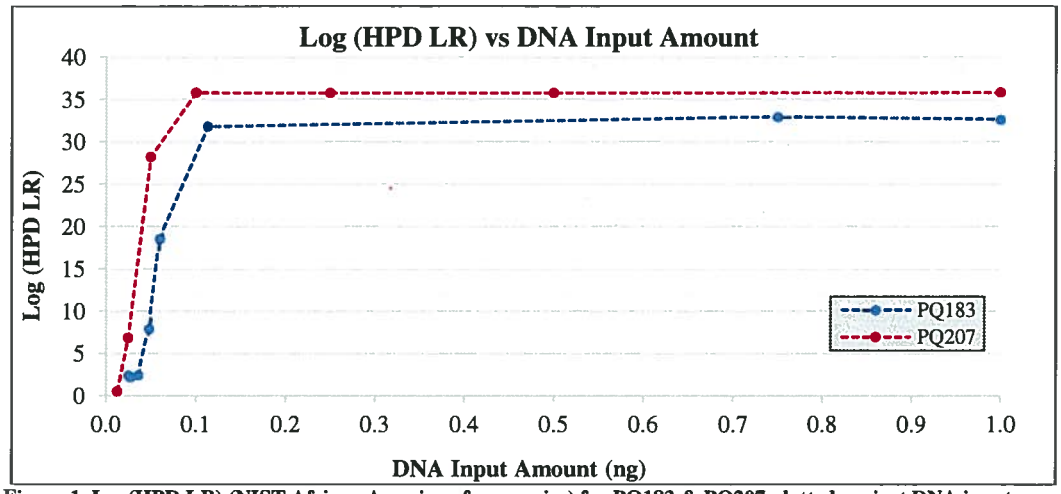


Figure 1. Log(HPD LR) (NIST African American frequencies) for PQ183 & PQ207 plotted against DNA input amounts (ng)

Figure 2 shows that the STRmix™ derived template amounts decreased as the DNA input was reduced.

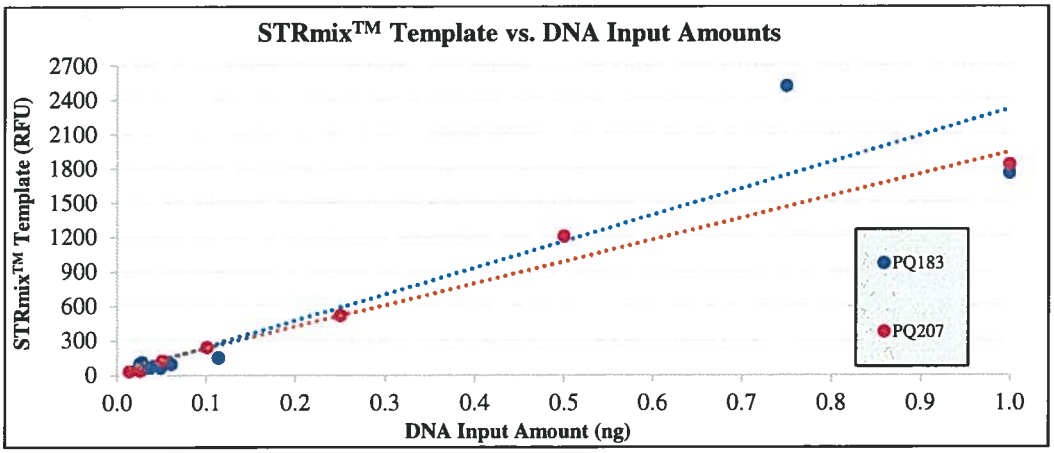


Figure 2. STRmix™ derived template amounts (RFU) change as DNA input amounts for PQ183 and PQ207 change

The Fusion 6C internal validation data used to determine the stochastic threshold was used to assess the probability of drop out (P_D) given peak height. The laboratory's stochastic threshold for the 5 and 10 second injection times were established to be 200 and 400 RFU, respectively. The plots are shown in Figure 3. The P_D increased substantially for peaks between 75 RFU (analytical threshold) and 150 RFU. For peaks above 150 RFU, the P_D was below 20% and progressed towards zero. No drop out events were detected beyond approximately 325 RFU.

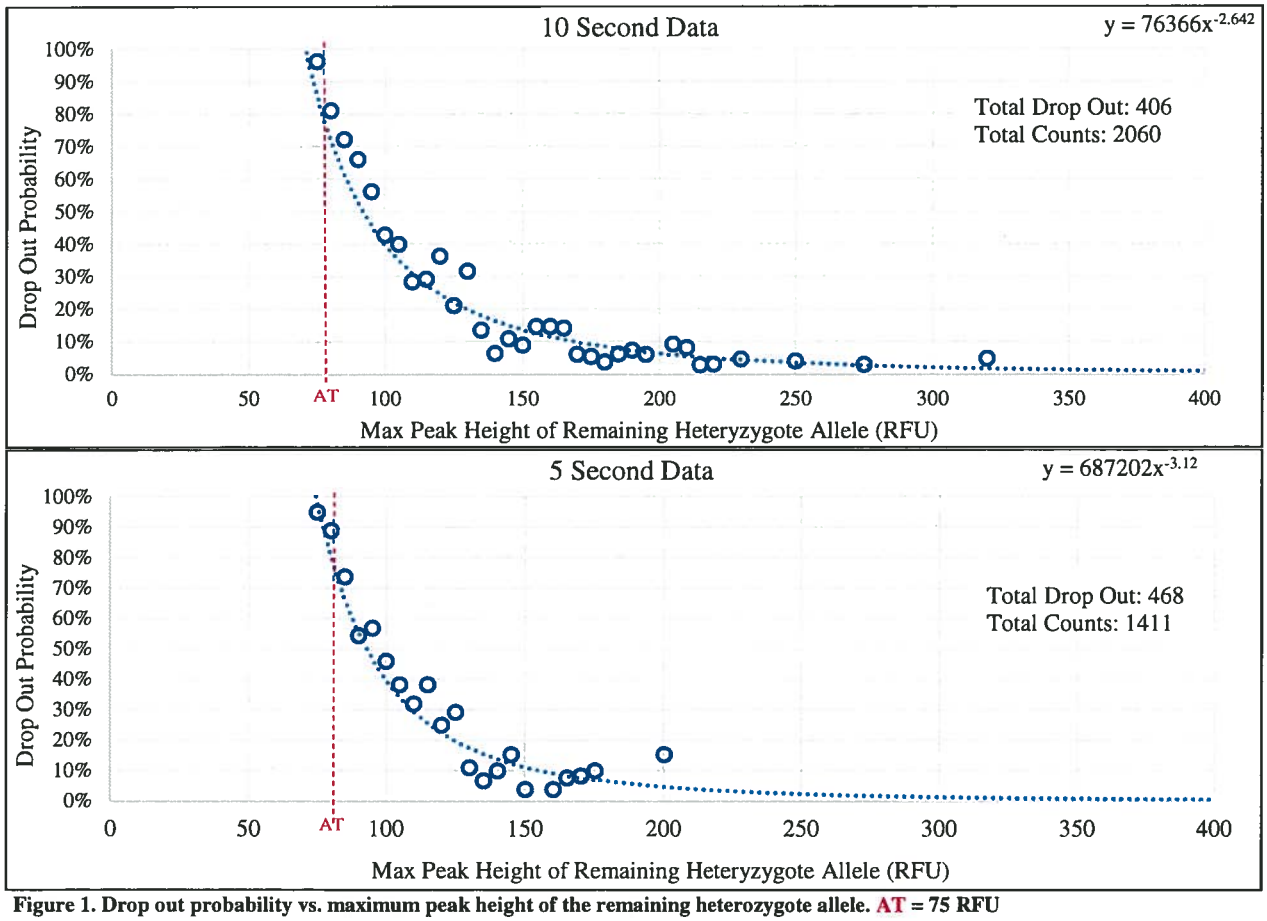


Figure 1. Drop out probability vs. maximum peak height of the remaining heterozygote allele. AT = 75 RFU

Based on these results, it was expected that STRmix™ would more routinely consider genotype combinations that consider drop out once the peak heights were less than approximately 325 RFU. To evaluate this, the STRmix™ template amounts were compared to the point estimate LRs for the PQ207 and PQ183 serial dilution samples. The results are summarized in Table 1. The point estimate LR was used rather than the HPD LR in order to alleviate the effects of allele frequency and MCMC uncertainty calculations. The point estimate LR value is expected to remain constant once a full single source profile with an individual genotype at each locus is achieved (weight = 1.0). However, once drop out combinations are also considered, the point estimate LR is expected to decrease. As shown in Table 1, the point estimate LR decreased at a template amount of 240 RFU for PQ207 and 150 RFU for PQ183. As the template decreased further resulting, in more drop out combinations, so did the point estimate LR. These results are consistent with the probability of drop out expectations described above.

Sample PQ207 DNA Input (ng)	**STRmix™ Template (RFU)**	**Point Estimate LR**
0.013	30	6.35E+00
0.025	34	1.75E+07
0.050	120	5.24E+28
0.100	240	2.13E+36
0.250	518	2.28E+36
0.500	1208	2.28E+36
1.000	1830	2.28E+36
Sample PQ183 DNA Input (ng)	**STRmix™ Template (RFU)**	**Point Estimate LR**
0.025	93	4.47E+02
0.027	108	2.21E+02
0.036	60	5.65E+02
0.048	60	1.34E+08
0.060	90	1.29E+19
0.114	150	2.25E+32
0.750	2516	1.62E+33
1.000	1755	1.62E+33

Table 1. Comparison of STRmix™ template amounts to point estimate LR values using the NIST African American allele frequencies. The red text indicates the samples where the LR changed.

In order to confirm that STRmix™ yields accurate LR values, the LR for the profile and at each locus for two of the single source profiles in Table 1 (PQ183-1ng and PQ207-0.25ng) were calculated by-hand using Microsoft Excel and compared to the point estimate LR derived from STRmix™. The Balding and Nichols formulae[1] (or equation 4.10 from NRCII[2]) and the posterior mean allele frequencies were applied to mimic the calculations performed by STRmix™. This was performed using a $\theta = 0.01$.

When setting θ to 0.01, the Balding and Nichols equations for a single source profile are as followed:

$$\frac{2[\theta + (1 - \theta)p_i][\theta + (1 - \theta)p_j]}{(1 + \theta)(1 + 2\theta)} \quad \text{for heterozygous loci}$$

$$\frac{[3\theta + (1 - \theta)p_i][2\theta + (1 - \theta)p_j]}{(1 + \theta)(1 + 2\theta)} \quad \text{for homozygous loci}$$

where p_i is the allele frequency for allele I, p_j is the allele frequency for allele j, and θ is the F_{st} value.

The posterior mean allele frequencies are calculated using the following equation:

$$\frac{x_i + 1/k}{N_a + 1}$$

where for a given locus, x_i is the number of observations of allele i in a database, N_a is the total number of alleles in that database, and k is the number of allele designations with non-zero observations in the database at that locus. The Excel-calculated and STRmix™ results for the NIST African American allele frequency dataset with $\theta = 0.01$ are given in Table 2. Concordant values were obtained for the locus LRs as well as the profile LR. Small variations in the locus LRs are attributable to differences in rounding.

Locus	PQ207 (θ = 0.01)		PQ183 (θ = 0.01)	
	Excel LR	STRmix™ Point Estimate LR	Excel LR	STRmix™ Point Estimate LR
D3S1358	7.440E+00	7.440E+00	2.457E+01	2.460E+01
D1S1656	1.314E+02	1.314E+02	5.186E+01	5.190E+01
D2S441	7.004E+00	7.004E+00	1.488E+01	1.490E+01
D10S1248	2.204E+01	2.204E+01	1.158E+01	1.160E+01
D13S317	3.148E+02	3.148E+02	4.319E+01	4.320E+01
Penta E	3.710E+02	3.710E+02	1.016E+02	1.020E+02
D16S539	2.051E+01	2.051E+01	1.220E+01	1.220E+01
D18S51	3.970E+01	3.970E+01	1.260E+02	1.260E+02
D2S1338	5.409E+01	5.409E+01	3.100E+01	3.100E+01
CSF1PO	2.619E+01	2.619E+01	3.465E+01	3.470E+01
Penta D	2.696E+01	2.696E+01	2.496E+01	2.500E+01
TH01	1.506E+01	1.506E+01	8.882E+00	8.880E+00
vWA	1.064E+01	1.064E+01	3.648E+01	3.650E+01
D21S11	2.895E+01	2.895E+01	3.998E+01	4.000E+01
D7S820	6.109E+01	6.109E+01	1.552E+01	1.550E+01
D5S818	2.589E+01	2.589E+01	2.702E+01	2.700E+01
TPOX	6.164E+00	6.164E+00	6.776E+00	6.780E+00
D8S1179	3.374E+02	3.374E+02	8.653E+00	8.650E+00
D12S391	3.999E+01	3.999E+01	2.688E+01	2.690E+01
D19S433	1.895E+01	1.895E+01	2.877E+01	2.880E+01
SE33	6.242E+02	6.242E+02	8.111E+02	8.110E+02
D22S1045	1.376E+01	1.376E+01	9.982E+00	9.980E+00
FGA	4.206E+01	4.206E+01	2.047E+01	2.050E+01
Total LR	**2.287E+36**	**2.287E+36**	**1.621E+33**	**1.620E+33**

Table 2. By-hand (Excel) vs STRmix™ point estimate LR calculations

In order to ensure the database search function in STRmix™ generates accurate LR results, the calculation was repeated to obtain the profile LR for each sample using a $\theta = 0$. When setting θ to zero, the Balding and Nichols equations condense to the product rule such that $2p_ip_j$ is used for heterozygous loci and p_i^2 is used for homozygous loci. This is the calculation applied during the database search function. The results using the NIST African American allele frequency dataset with $\theta = 0$ are given in Table 3.

PQ207 ($\theta = 0$)		PQ183 ($\theta = 0$)	
Excel LR	STRmix™ Database LR	Excel LR	STRmix™ Database LR
3.22E+39	**3.22E+39**	**8.75E+34**	**8.75E+34**

Table 3. By-hand (Excel) vs STRmix™ LR database LR calculations

The results in Tables 2 and 3 demonstrate that STRmix™ yields accurate LR values based on the population genetic model being applied.

References

(1) Balding DJ, Nichols RA., Forensic Science International 64 (1994); DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands.

(2) National Research Council (NRC II) (1996); The Evaluation of Forensic DNA Evidence.

Section B: Use of Peak Heights *(SWGDAM 4.1.4, 4.1.7, 4.1.7.2)*

This section covers the following standard:

4.1.4. Allelic peak height, to include off-scale peaks
4.1.7. Partial profiles, to include the following:

4.1.7.2. DNA degradation

STRmix™ uses peak heights to inform the genotype combinations of contributors to a profile. This was demonstrated in Section A for optimal and sub-optimal DNA input amounts. The genotype weights and LR values decreased as the template amount and peak heights decreased, and drop out began to be considered. Allelic peaks may reach the saturation level of the 3500 camera when high DNA input amounts are amplified. This means that the allele peak height is not accurately captured, and therefore, the observed stutter peak height may be higher than what is anticipated when applying the stutter ratio for that allele. As described in Part I of this validation, the saturation threshold was set to 30,000 RFU. The laboratory does not routinely interpret saturated profiles. However, degraded samples are typically amplified at a higher target amount than 1ng (i.e. up to ~6ng) and this may result in saturated peaks at Amelogenin and/or other smaller loci. These samples are generally corrected with either dilution and/or reinjection at a reduced time of 5 seconds. In order to assess the impact of saturated data in these types of samples on profile interpretation, a single source degraded sample (PQ-86-1999BS) was used and amplified at optimal (1ng) and above optimal DNA input amounts (2ng, 3ng, 4ng, 5ng, 6ng, and 6.84ng). The seven resulting profiles were interpreted in STRmix™ using the following propositions and the weights were reviewed:

- H_p= The DNA originated from the person of interest (POI)
- H_d= The DNA originated from an unknown individual

All profiles resulted in intuitive genotypes where the weight = 1.0 for the expected genotype combinations. This is represented in Table 1 by identical point estimate LR values for each profile regardless if saturated peaks were present. The HPD LR values were also similar and varied slightly as expected due to allele frequency and MCMC uncertainty.

PQ86 1999BS* Amp Input amount (ng)	STRmix™ derived template (RFU)	STRmix™ derived degradation (RFU / base pair)	Point Estimate LR	HPD LR
1.0	6799	26.85	5.54E+29	1.76E+29
2.0	9806	39.43	5.54E+29	1.49E+29
3.0	14325	57.76	5.54E+29	1.82E+29
4.0	17119	68.51	5.54E+29	1.76E+29
5.0	22361	89.94	5.54E+29	1.71E+29
6.0	31935	126.24	5.54E+29	1.56E+29
6.84	38660	157.64	5.54E+29	1.73E+29
*D16S51 omitted due to tri-allele				

Table 1. LR results for 10 second injection data.

Stutter (k^2) and allele (c^2) variances were evaluated to determine if they were affected by saturated data. The stutter and allele variances were plotted against the STRmix™ derived template amount and are given in Figure 1. The stutter variance increased relative to the mode of the prior distribution with higher template. This was not unforeseen since stutter peaks may be larger than expected when the parent peak

height is above the camera's saturation limit (~30,000 RFU). For peaks above the saturation threshold, STRmix™ calculates the expected height of the stutter peak using the proposed *expected* allele height and not the observed height. This leads to a slightly higher variance between the observed and expected stutter peaks.

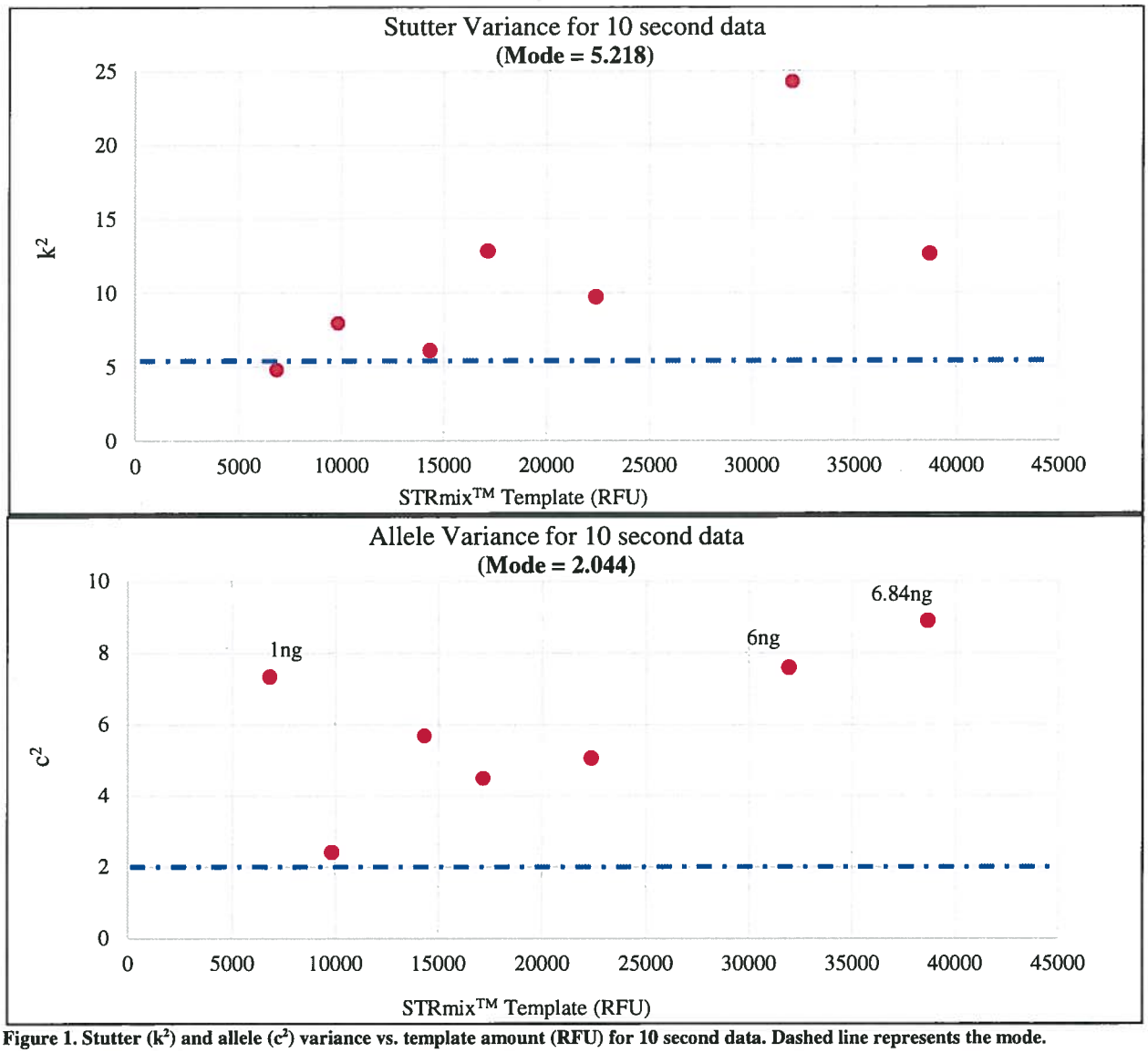


Figure 1. Stutter (k^2) and allele (c^2) variance vs. template amount (RFU) for 10 second data. Dashed line represents the mode.

The allele variance also increased as compared to the mode of the prior distribution with higher template, with one exception (Figure 1). The 1.0ng sample had an elevated allele variance that was similar to the 6.0ng sample. A high allele variance indicates that a greater divergence between the observed and expected was needed to explain the data. The D22S1045 locus for the 1.0ng sample had a heterozygous genotype with a peak height ratio of 35%. Since this sample was interpreted as a single source, STRmix™ had to force pair these two largely imbalanced alleles. This imbalance was not observed in the profiles amplified at target amounts 2ng through 5ng, and thus, the allele variance values were lower. The 6ng and 6.84ng samples both had peak heights above the saturation level. When STRmix™ models an observed saturated peak, it uses the laboratory's preset saturation threshold (i.e. 30,000 RFU) as the maximum expected allele height (described in the STRmix v.2.5.11 User Manual). The further the saturated peak is

above 30,000 RFU the greater the divergence between the observed and expected peak, which may result in a higher allele variance in order to explain the data.

Since the laboratory's protocol is to reinject these types of samples at a lower injection time, this data set was also interpreted using the 5 second injection time. The results are summarized in Table 2. The lower injection time reduced the peaks heights below the saturation threshold for all of the samples.

PQ86 1999BS* Amp Input amount (ng) 5 second injection	STRmix™ derived template (RFU)	STRmix™ derived degradation (RFU / base pair)	Point Estimate LR	HPD LR
1.0	2955	11.61	5.54E+29	1.59E+29
2.0	4309	17.18	5.54E+29	1.53E+29
3.0	6641	26.563	5.54E+29	1.71E+29
4.0	7658	30.602	5.54E+29	1.75E+29
5.0	11194	44.885	5.54E+29	1.80E+29
6.0	13793	54.823	5.54E+29	1.59E+29
6.84	18221	73.886	5.54E+29	1.80E+29
*D16S51 omitted due to tri-allele				

Table 2. LR results for 5 second injection data.

The current allele and stutter variance prior distribution modes were generated during the model maker analysis. Both 5 and 10 second injection data was used. As discussed in part I of this validation, a combined 5 and 10 second allele variance and stutter variance were selected as the laboratory's values rather than individual values for each injection time. To ensure that 5 second injection time data was modeled appropriately, these samples were compared to the results obtained for the 10 second data. All profiles resulted in intuitive genotypes with weights = 1.0 for the expected genotype combinations. Table 2 shows that the same point estimate LR values and similar HPD LR values were obtained as for the 10 second injection data.

Both the stutter and allele variances decreased for the 5 second data as shown in the plots of Figure 2, indicating there was less divergence between the observed and expected peak heights during the MCMC modeling. This was expected since the stutter peaks could be modeled more accurately with allelic peaks below 30,000 RFU. The 1ng sample had the highest allele variance for the 5 second data. Upon further inspection of the profile, the peak height imbalance observed for the 1.0ng 10 second injection at D22S1045 was still present in the 5 second injection (peak height ratio = 32%).

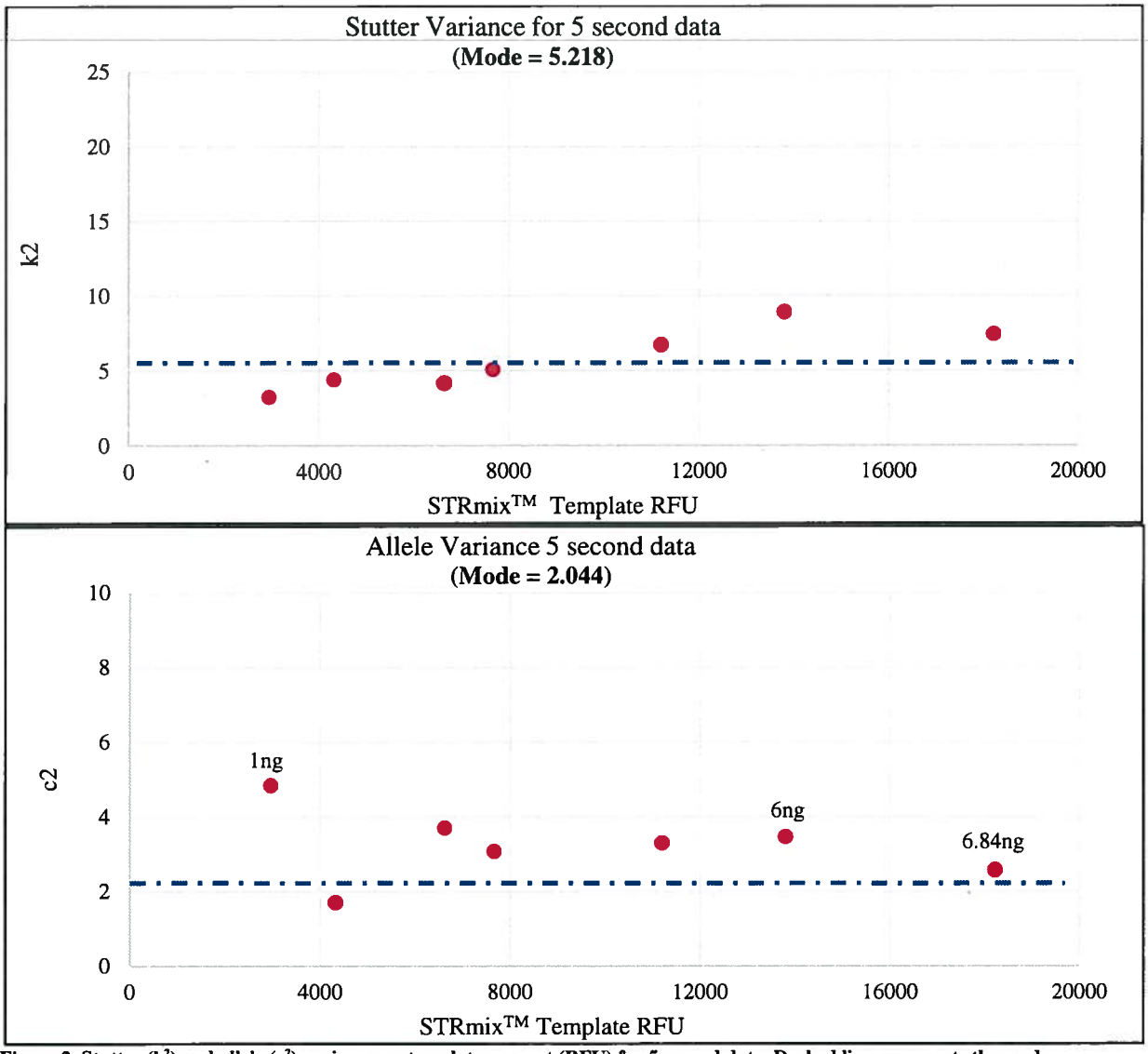


Figure 2. Stutter (k^2) and allele (c^2) variance vs. template amount (RFU) for 5 second data. Dashed line represents the mode.

Although the stutter and allele variances increased for saturated peaks, STRmix™ was able to appropriately model the data resulting in genotype weights = 1.0 for all profiles. While saturated data will not be prohibited from entry into STRmix™, analysts will be cautioned to carefully scrutinize the resulting deconvolutions to ensure that the genotype weights assigned are consistent with the qualitative expectations.

Section C: Inspection of Weights

This section covers the following standard:

> *4.2.1.3. Generally, as the analyst's ability to deconvolute a complex mixture decreases, so do the weightings of individual genotypes within a set determined by the software.*

The weights are considered as the primary output from STRmix™. They are used as a diagnostic of the deconvolution process and should be intuitively correct where the most supported genotypes have the highest weights. Poorly intuitive weights are an indication of poor biological modeling or an incorrect variance constant. In order to test whether the weights assigned to different genotype combinations are appropriate given the profile, 32 two-contributor mixture samples created using laboratory donors PQ183 and PQ212 for the Fusion 6C internal validation study were used. Male to female (M:F) mixtures and female to male (F:M) mixtures comprised of various ratios and amplified at 1ng and 200pg were interpreted with STRmix™. The mixture ratios consisted of 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, 1:20, 1:40, 1:60, and 1:80 for the 1ng samples and 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, 1:20, and 1:40 for the 200pg samples. Separate LRs for the major contributor and the minor contributor were calculated using the following propositions:

- H_p: The DNA originated from the POI (major or minor contributor) and an unknown individual
- H_d: The DNA originated from two unknown individuals

A plot of log(HPD LR) for each mixture is shown in Figures 1 through 4. For both contributors, the log(HPD LR) decreased as the alleles from the major and minor contributors were no longer distinguishable at all loci. Since multiple genotypes were possible at a locus once the mixture was indistinguishable, the locus genotype weightings were no longer equal to1.0. These values decreased by nearly half for the 1:1 mixtures when compared to the fully deconvoluted major and minor contributors. The log(HPD LR) for the major contributor increased until it was a fully distinguishable profile (genotype weights = 1.0) while the log(HPD LR) for the minor contributor decreased as its DNA template amount dropped.

The laboratory's current definition for deeming a major contributor in a profile is that at least 80% of the detected loci have a distinguishable genotype. The data was evaluated to determine at which mixture proportions a major contributor was deduced by STRmix™ as defined by the laboratory. This occurred when the major contributor was at approximately 66% and greater for the 1ng samples and approximately 75% and greater for the 200pg samples.

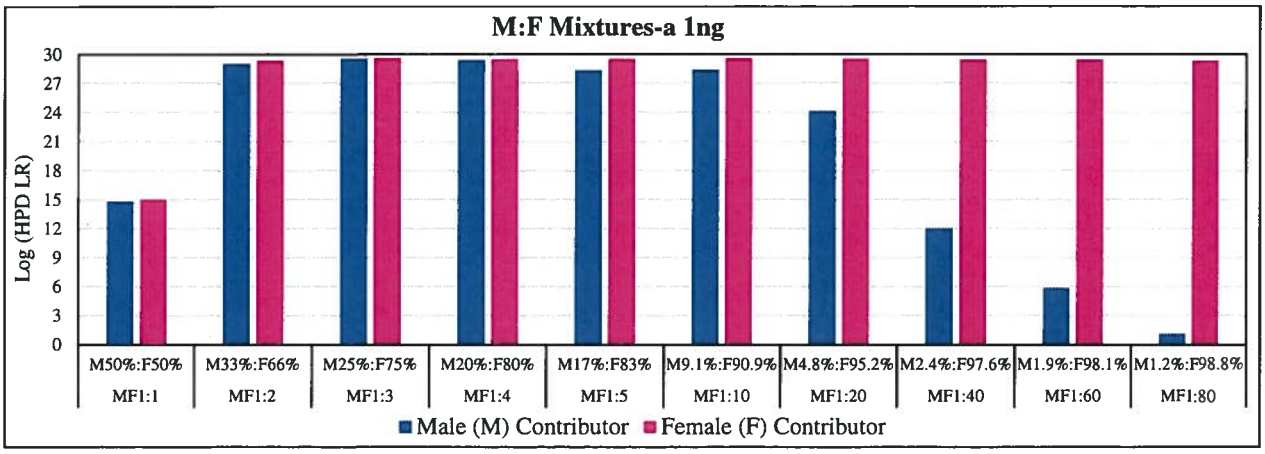


Figure 1. Log(HPD LR) versus mixture proportions for M:F mixtures with target amount 1ng

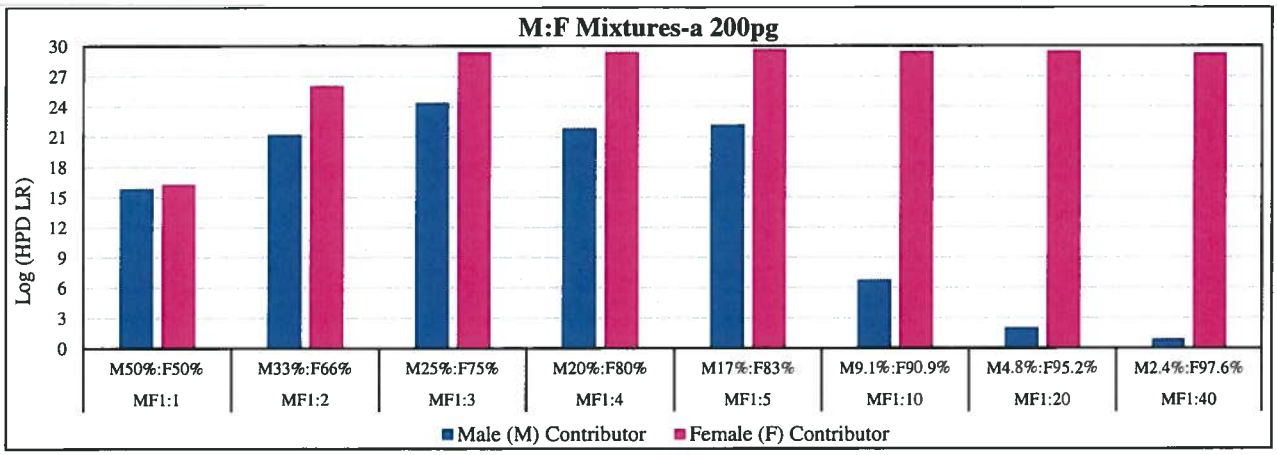


Figure 2. Log(HPD LR) versus mixture proportions for M:F mixtures with target amount 200pg. Note: MF1:10 0.2ng-a and MF1:40 0.2ng-a were analyzed with informed mixture proportion priors (IMPP)

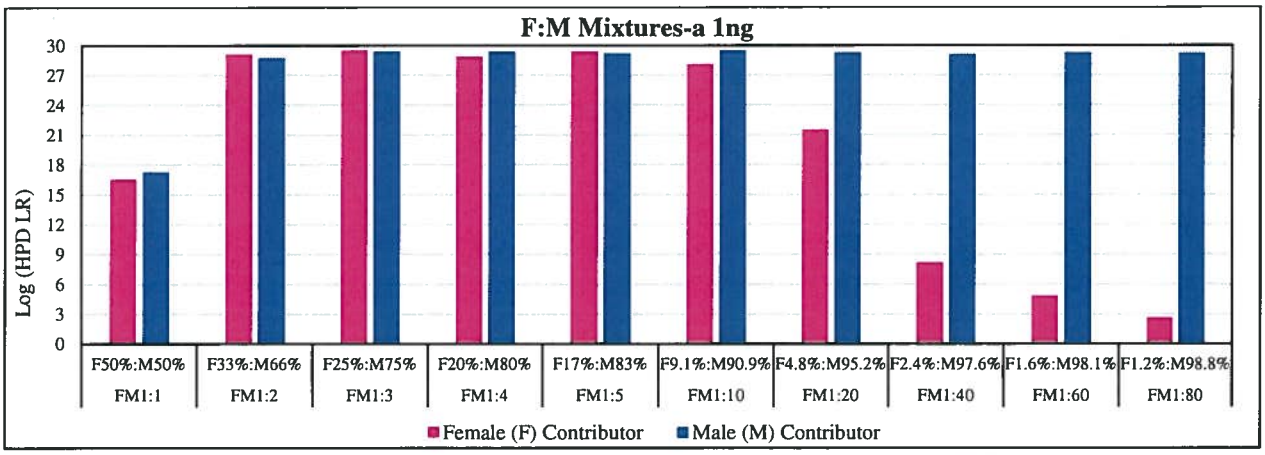


Figure 3. Log(HPD LR) versus mixture proportions for F:M mixtures with target amount 1ng. Note: FM1:80 1ng-a was analyzed with IMPP.

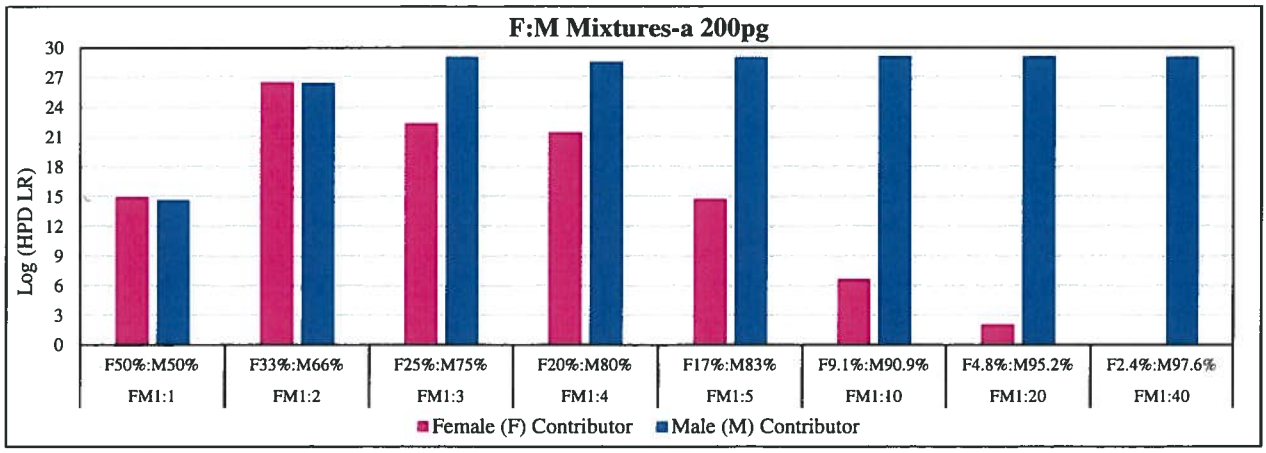


Figure 4. Log(HPD LR) versus mixture proportions for F:M mixtures with target amount 200pg. Note: FM1:40 0.2ng was analyzed with IMPP.

As with genotype weights, the mixture proportions are used as a diagnostic of the interpretation. The STRmix™ calculated mixture proportions should be similar to expected proportions based on the qualitative review of the profile. The mixture proportions calculated in STRmix™ for these two-contributor mixtures were evaluated to determine if the values changed as the mixture ratios varied and if they were consistent with the expected proportions.

The STRmix™ derived mixture proportions were intuitive based on the profile and within approximately 10% of the expected values for all samples except four. The data for these four samples are summarized in Table 1. The four mixtures were distinguishable profiles, and therefore STRmix™ was expected to deconvolve a single major genotype at each locus resulting in a genotype weight = 1.0. A single major genotype was appropriately deconvolved at all loci for sample MF1:10 0.2ng-a, but not for samples MF1:40_0.2ng-a, FM1:80_1ng-a, and FM1:40_0.2ng-a. This is possibly due to the greater disparity observed between the known mixture proportions and the STRmix™ derived proportions for the latter three samples. The minor component for each problematic mixture had less than 20pg of input DNA indicating that the mixture proportions may be less accurate as template amount decreases. Since the STRmix™ mixture proportions did not meet qualitative expectations based on the review of the profiles, the samples were re-interpreted using informed mixture proportion priors (IMPP) (STRmix™ user guide v. 2.5.11 recommendation). This resulted in intuitive mixture proportions for all four samples as shown in Table 1. Furthermore, the major contributor genotypes yielded weights = 1.0 for all loci. The IMPP data was used in Figures 2 through 4 above.

Two-Contributor Mixtures	STRmix™ % Contributor 1	STRmix™ % Contributor 2	Expected % Contributor 1	Expected % Contributor 2	Minor DNA Amount (pg)
MF1:10 0.2ng-a	15%	85%	9%	91%	18
MF1:10 0.2ng-a with IMPP (0.91, 0.09 variance = 0.00097656)	10%	90%			
MF1:40 0.2ng-a	18%	82%	2%	98%	5
MF1:40 0.2ng-a with IMPP (0.98, 0.02 variance = 0.015625)	3%	97%			
FM1:80 1ng-a	13%	87%	1%	99%	12
FM1:80 1ng-a with IMPP (0.99, 0.01 variance = 0.00390625)	1%	99%			
FM1:40 0.2ng	20.0%	80.0%	2%	98%	5
FM1:40 0.2ng with IMPP (0.98, 0.02 variance = 0.00390625)	3.0%	97.0%			

Table 1. STRmix™ calculated versus expected mixture proportions (%) for two-contributor mixtures: with and without IMPP

As shown in Figure 5, the calculated STRmix™ mixture proportions for the major and minor contributors increase and decreased, respectively, as their proportions changed. The STRmix™ mixture proportions differed by 10% or less from the expected proportions for all samples once IMPP were applied to the four affected samples listed in Table 1.

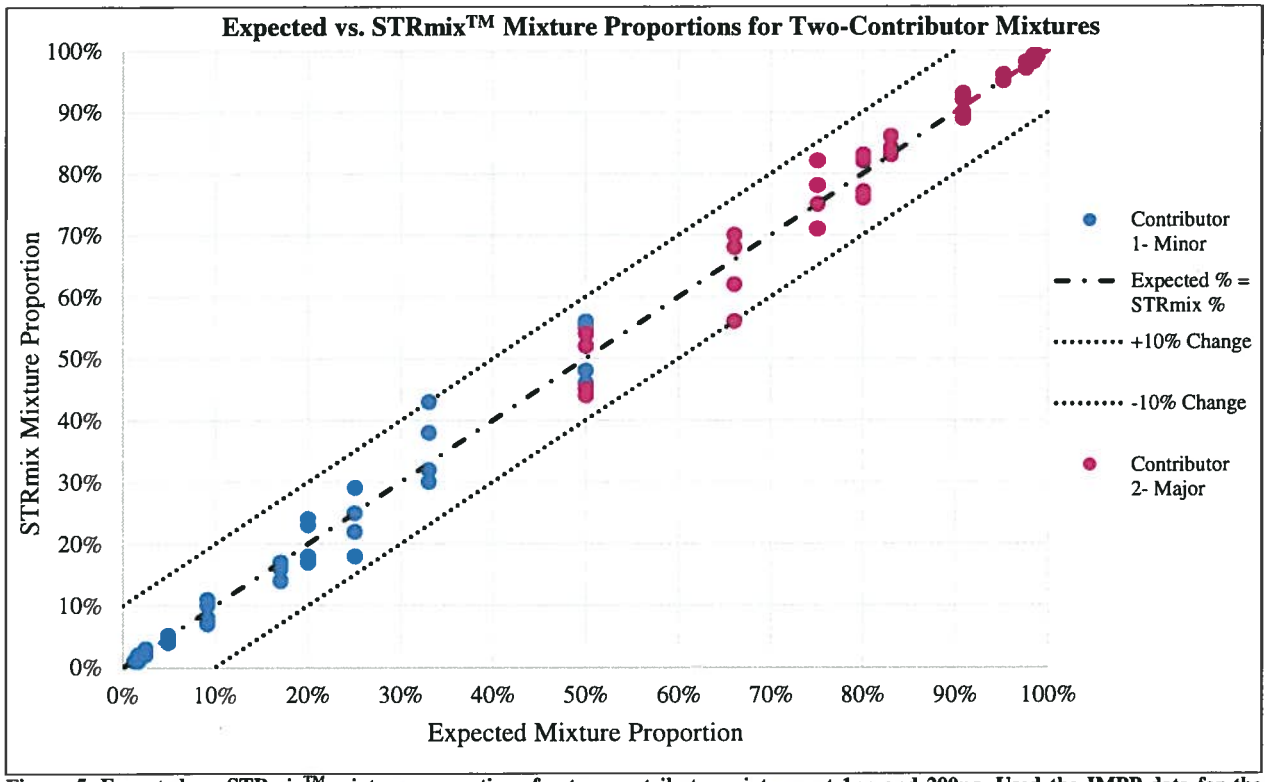


Figure 5. Expected vs. STRmix™ mixture proportions for two contributor mixtures at 1ng and 200pg. Used the IMPP data for the samples in Table 1.

The mixture proportions calculated in STRmix™ for the three (34 total) and four (34 total) contributor mixture samples listed in Section D were also compared to the expected proportions. Out of 68 mixture samples, 65 resulted in STRmix™ derived mixture proportions that were intuitive based on the qualitative expectation of the profiles. The remaining three samples that did not yield intuitive mixture proportions were three-contributor mixtures amplified at 300pg. The minor component had 12pg or less of template DNA with an expected proportion of 4% or less. The data for these samples is shown in Table 2. Samples MMF1:10:20_0.3ng a and b returned a false exclusionary LR value of 0 and MMF1:5:20 returned an LR value of 0.00164 for the minor contributor. The three mixtures were re-analyzed using IMPP and in all instances the STRmix™ mixture proportions improved (data shown in Table 2). The LR values for MMF1:10:20_0.3ng- a and b were no longer exclusionary (0.0883 and 0.330, respectively), and the MMF1:5:20 value increased to 0.560. These LR values of less than one are consistent with the low amount of DNA present for the minor contributor (designated as "Contributor 1" in Table 2) and the qualitative assessment of the profiles.

Three-Contributor Mixtures	STRmix™ % Contributor 1	STRmix™ % Contributor 2	STRmix™ % Contributor 3	Expected % Contributor 1	Expected % Contributor 2	Expected % Contributor 3	Minor ~DNA Amount (pg)
MMF1:5:20_0.3ng a	8%	14%	77%	4%	19%	77%	12
MMF1:5:20_0.3ng a with IMPP (0.77, 0.19, 0.04 variance = 9.7656E-4 ,4.8828E-4, 4.8828E-4)	4%	18%	78%				
MMF1:10:20_0.3ng a	27%	34%	39%	3%	32%	65%	9
MMF1:10:20_0.3ng a with IMPP (0.65, 0.32, 0.03, variance = 0.0019531)	1%	33%	66%				
MMF1:10:20_0.3ng b	28%	34%	38%	3%	32%	65%	9
MMF1:10:20_0.3ng b with IMPP (0.65, 0.32, 0.03, variance = 0.0019531)	4%	33%	63%				

Table 2. STRmix™ calculated versus expected mixture proportions (%) for three contributor mixtures: with and without IMPP

Similar to the two-contributor samples, the STRmix™ calculated mixture proportions of the major and minor contributors for the three and four contributor mixtures increase and decrease, respectively, as their proportions changed. As shown in Figures 6 and 7, the STRmix™ mixture proportions differed by 10% or less from the expected proportions for all samples once IMPP were used, except for one sample in the three-contributor mixtures (MMF1:1:1_0.075ng b) and three samples in the four-contributor mixtures (FMMM1:1:1:1_0.1ng a, FMMM1:1:1:1_0.1ng b, and FMMM1:5:5:10_0.3ng b,). Each of these samples presented as either an N-1 or N-2 contributor mixture based on the maximum allele count, but were analyzed using the known number of contributors (N).

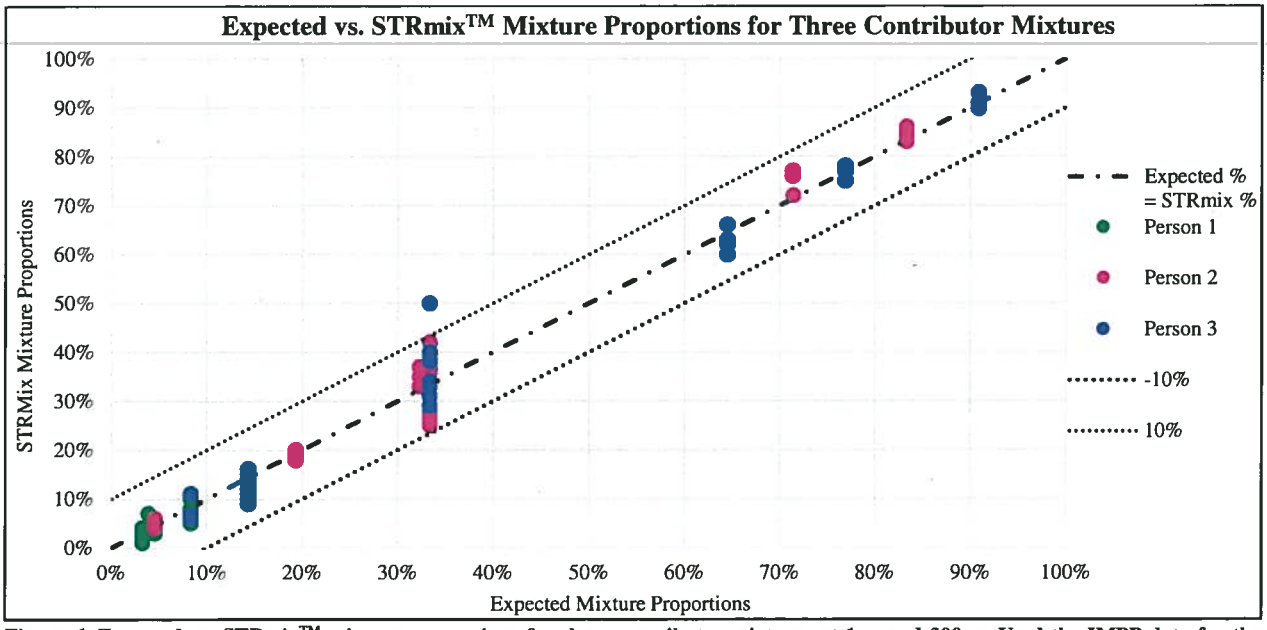


Figure 6. Expected vs. STRmix™ mixture proportions for three contributor mixtures at 1ng and 300pg. Used the IMPP data for the sample in Table 2.

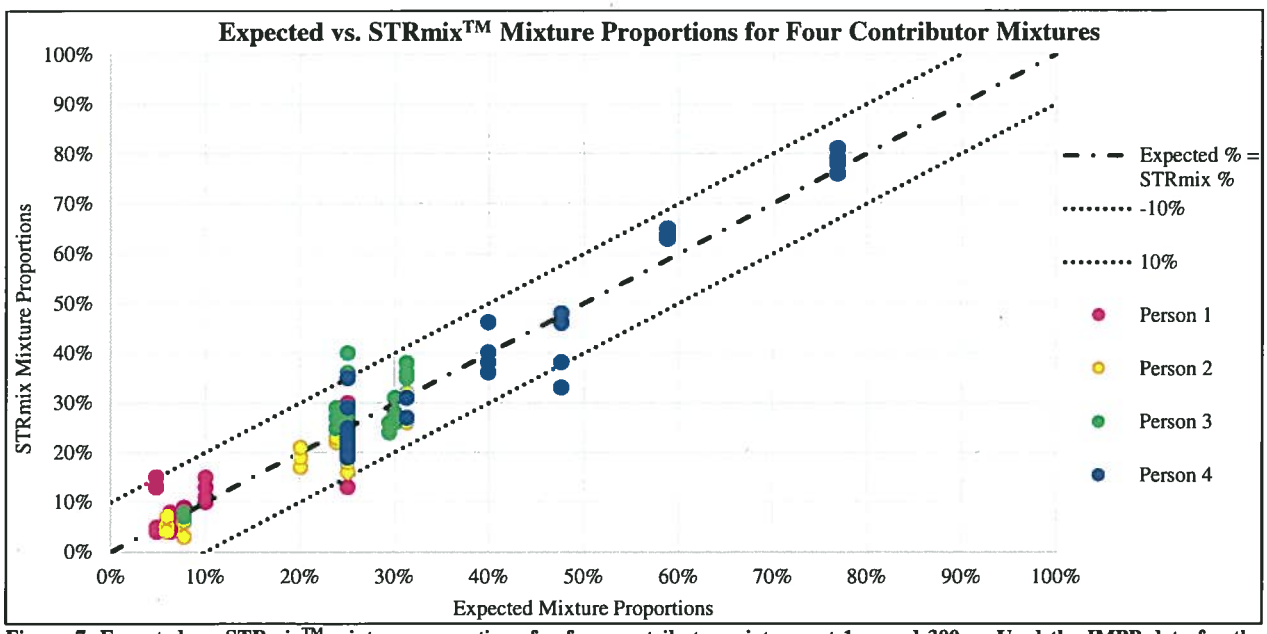


Figure 7. Expected vs. STRmix™ mixture proportions for four contributor mixtures at 1ng and 300pg. Used the IMPP data for the samples in Table 2.

In general, the weights and mixture proportions generated by STRmix™ produced reliable and accurate results for two, three and four person mixtures and were consistent with the qualitative expectations observed in the profile. However, samples with trace level contributors containing less than approximately 20pg may result in poor STRmix™ mixture proportions, which can affect the major contributor genotype weightings and could lead to a false exclusion (LR = 0) of the minor contributor. For these samples, applying IMPP better resolved the trace level contributors while improving the genotype weights of the higher DNA template contributors. It is recommended that analysts use IMPP when mixture proportions and/or weights are not intuitive based on the qualitative assessment of the profile in an attempt to improve the deconvolution.

Section D: Sensitivity, Specificity, and Mixtures

This section covers the following standards:

4.1.2. Hypothesis testing with contributors and non-contributors
4.1.6. Mixed specimens
4.1.6.1. Various contributor ratios
4.1.6.2. Various total DNA template quantities
4.1.6.3. Various numbers of contributors. The number of contributors evaluated should be based on the laboratory's intended use of the software. A range of contributor numbers should be evaluated in order to define the limitations of the software.
4.1.6.5. Sharing of alleles among contributors
4.1.7. Partial profiles, to include the following:
4.1.7.1. Allele and locus drop-out
4.1.13. Sensitivity and specificity as described for Developmental Validation (3.2.1 & 3.2.2)
3.2.1 Sensitivity- Studies should assess the ability of the system to reliably determine the presence of a contributor's DNA over a broad variety of evidentiary typing results (to include mixtures and low-level DNA quantities). This should be evaluated using various sample types (e.g. different number of contributors, mixture proportions and template quantities).
3.2.1.1 Sensitivity studies should demonstrate the potential for Type I errors (i.e. incorrect rejection of a true hypothesis), in which, for example a contributor fails to yield a LR greater than 1, and thus, his/her presence in the mixture is not supported.
3.2.1.2 Sensitivity studies should demonstrate the range of LR values that can be expect for contributors.
3.2.2 Specificity- Studies should assess the ability of the system to provide reliable results for non-contributors over a broad variety of evidentiary typing results (to include mixtures and low-level DNA quantities). This should be evaluated using various sample types (e.g. different number of contributors, mixture proportions and template quantities).
3.2.2 Specificity- Studies should assess the ability of the system to provide reliable results for non-contributors over a broad variety of evidentiary typing results (to include mixtures and low-level DNA quantities). This should be evaluated using various sample types (e.g. different number of contributors, mixture proportions and template quantities).
3.2.2.1 Specificity studies should demonstrate the potential for Type II errors (i.e. failure to reject a false hypothesis), in which, for example a non-contributor yields a LR greater than 1, and thus, his/her presence in the mixture is supported.
3.2.2.2 Specificity studies should demonstrate the range of LR values that can be expect for non-contributors.

Sensitivity is defined as the ability of the software to reliably resolve the DNA profile of known contributors (H_p true) and specificity is defined as the ability of the software to reliably exclude non-contributors (H_d true) within a DNA profile for a range of starting DNA templates. Sensitivity and specificity was tested by calculating the LR for 107 samples consisting of one, two, three and four persons varying in DNA quantities and mixture proportions (4.1.6.1, 4.1.6.2, 4.1.6.3), including samples with low amounts of DNA that demonstrate drop out (4.1.7.1). The mixture samples generated for the Fusion 6C mixture validation studies were used since they were designed to the extent in which casework samples would be interpreted (4.1.6.3). These samples have homozygous and heterozygous alleles along with varying amounts of allele sharing across the loci (4.1.6.5). Additional three and four person mixtures were created using different donor combinations to supplement the current mixture set. As previously outlined in section C of the validation, several mixture samples required analysis utilizing informed mixture proportion priors (IMPP) in order to allow for a more accurate interpretation. For this section of the validation, those samples include: 1:40FM_200pg-a, 1:40MF_200pg-a, 1:10:20MMFa_300pg, 1:10:20MMFb_300pg, and 1:5:20MMFa_300pg; which are included in the figures and tables of this

summary. Each of the 107 profiles was compared to the known contributors resulting in a total of 309 comparisons.

A database of 207 non-contributors comprised of LASD staff and family members were also compared to the same profiles resulting in a total of 21,840 comparisons using the Database search function within STRmix™ in order to determine the database LR value. The LRs from the database search do not include a θ correction nor any correction for MCMC or allele probability uncertainty. The samples used in the study are shown in Table 1.

Single source	2 Person Mixtures	3 Person Mixtures	4 Person Mixtures
0.025ng	MF1:1_1ng a	Set 1	Set 1
0.027ng	MF1:2_1ng a	MMF1:1:1_1ng a	FMMM1:1:1:1_1ng a
0.036ng	MF1:3_1ng a	MMF1:1:1_1ng b	FMMM1:1:1:1_1ng b
0.048ng	MF1:5_1ng a	MMF1:1:1_0.3ng a	FMMM1:1:1:1_0.3ng a
0.060ng	MF1:10_1ng a	MMF1:1:1_0.3ng b	FMMM1:1:1:1_0.3ng b
0.750ng	MF1:20_1ng a	MMF1:5:1_1ng a	FMMM1:1:5:10_1ng a
1.0ng	MF1:40_1ng a	MMF1:5:1_1ng b	FMMM1:1:5:10_1ng b
	MF1:60_1ng a	MMF1:5:1_0.3ng a	FMMM1:1:5:10_0.3ng a
	MF1:80_1ng a	MMF1:5:1_0.3ng b	FMMM1:1:5:10_0.3ng b
	FM1:1_1ng a	MMF1:10:1_1ng a	FMMM1:1:1:10_1ng a
	FM1:2_1ng a	MMF1:10:1_1ng b	FMMM1:1:1:10_1ng b
	FM1:3_1ng a	MMF1:10:1_0.3ng a	FMMM1:1:1:10_0.3ng a
	FM1:5_1ng a	MMF1:10:1_0.3ng b	FMMM1:1:1:10_0.3ng b
	FM1:10_1ng a	MMF1:1:20_1ng a	Set 2
	FM1:20_1ng a	MMF1:1:20_1ng b	FMMM1:2:3:4_1ng a
	FM1:40_1ng a	MMF1:1:20_0.3ng a	FMMM1:2:3:4_1ng b
	FM1:60_1ng a	MMF1:1:20_0.3ng b	FMMM1:2:3:4_0.3ng a
	FM1:80_1ng a	MMF1:5:20_1ng a	FMMM1:2:3:4_0.3ng b
	MF1:1_0.2ng a	MMF1:5:20_1ng b	FMMM1:5:5:10_1ng a
	MF1:2_0.2ng a	MMF1:5:20_0.3ng a	FMMM1:5:5:10_1ng b
	MF1:3_0.2ng a	MMF1:5:20_0.3ng b	FMMM1:5:5:10_0.3ng a
	MF1:5_0.2ng a	MMF1:10:20_1ng a	FMMM1:5:5:10_0.3ng b
	MF1:10_0.2ng a	MMF1:10:20_1ng b	FMMM1:5:5:5_1ng a
	MF1:20_0.2ng a	MMF1:10:20_0.3ng a	FMMM1:5:5:5_1ng b
	MF1:40_0.2ng a	MMF1:10:20_0.3ng b	FMMM1:5:5:5_0.3ng a
	FM1:1_0.2ng a	Set 2	FMMM1:5:5:5_0.3ng b
	FM1:2_0.2ng a	MMF1:1:1_0.9ng a	FMMM1:1:1:1_1.2ng a
	FM1:3_0.2ng a	MMF1:1:1_0.9ng b	FMMM1:1:1:1_1.2ng b
	FM1:5_0.2ng a	MMF1:1:1_0.6ng a	FMMM1:1:1:1_0.8ng a
	FM1:10_0.2ng a	MMF1:1:1_0.6ng b	FMMM1:1:1:1_0.8ng b
	FM1:20_0.2ng a	MMF1:1:1_0.3ng a	FMMM1:1:1:1_0.4ng a
	FM1:40_0.2ng a	MMF1:1:1_0.3ng b	FMMM1:1:1:1_0.4ng b
		MMF1:1:1_0.15ng a	FMMM1:1:1:1_0.2ng a
		MMF1:1:1_0.15ng b	FMMM1:1:1:1_0.2ng b
		MMF1:1:1_0.075ng a	FMMM1:1:1:1_0.1ng a
		MMF1:1:1_0.075ng b	FMMM1:1:1:1_0.1ng b

Table 1. Section D samples

The following propositions were considered when comparing the profiles to known contributors and non-contributors:

- H_p: The DNA originated from the POI (known contributor or non-contributor) and N-1 unknown individuals
- H_d: The DNA originated from N unknown individuals

The range of LR values expected for known contributors and non-contributors was investigated, along with Type I and Type II errors. A Type I error, which is defined as an incorrect rejection of a true hypothesis, is a false exclusion of a known contributor (log(LR) < 0 or LR < 1)[1]. A Type II error, which is defined as a failure to reject a false hypothesis, is a false inclusion of a non-contributor (log(LR) > 0 or LR > 1)[1]. These studies were used to determine the uninformative or inconclusive range of LR results. The data was also assessed to determine if the addition of more information such as increased DNA template improves the performance of STRmix™. For sensitivity, the LR for known contributors should be high and trend to zero as template decreases, whereas for specificity, the LR should trend upwards to zero as less information is present within the profile.

The log(LR) values were plotted against the input DNA amount per contributor for the one, two, three, and four contributor samples. Exclusions (LR = 0) are plotted as log(LR) = -30. The DNA amount was calculated for the known contributors using the theoretical mixture proportions and amplification target. The input amount used for non-contributors was the DNA amount associated with the minor component of the profile. The results of all comparisons are provided in Figures 1 through 4.

The plots in Figures 1 through 4 show that as DNA template increases the log(LR) increases. As the number of contributors increases and the template decreases the two distributions of known and non-contributors converged on log(LR) = 0. At high template amount STRmix™ correctly and reliably gave a high LR for the true contributors and a low LR for non-contributors. At low template amount or high contributor numbers STRmix™ correctly and reliably reported that the analysis of the sample trends towards uninformative or inconclusive (LR = 1 or log(LR) = 0). The plots in this section can help inform the limits of STRmix™, particularly when false negatives (Type I) and false positives (Type II) may arise. Type I errors are identified as the blue points below the horizontal line of log(LR) = 0. Type II errors are depicted as the orange points above the horizontal line of log(LR) = 0.

The one contributor samples had a false positive rate of 1.11% and a false negative rate of 0%. As depicted in Figure 1, the maximum database LR for a false positive was 655 (log(LR) = 2.82) and it was associated with a template amount of 27pg.

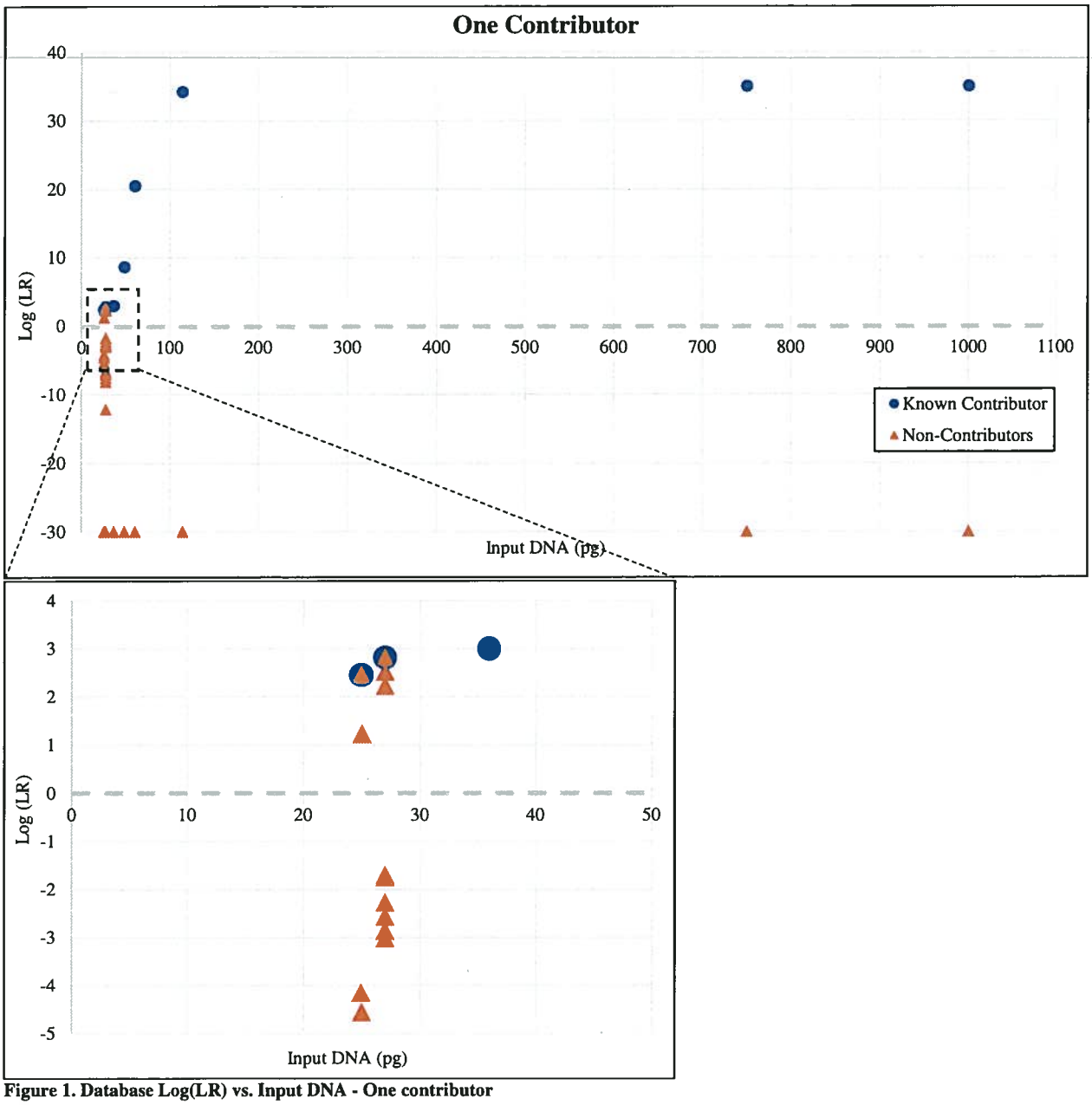


Figure 1. Database Log(LR) vs. Input DNA - One contributor

The two contributor samples had a false positive rate of 1.95%. As shown in Figure 2, the maximum database LR for a false positive was 6,750 (log(LR) = 3.83) and it was associated with a template amount of 16 pg. The false negative rate was 0% out of 64 known contributor comparisons when using the database LR. Known contributors that yielded a database LR between 1 and 1,000 (0 < log(LR) < 3) were recalculated to obtain the HPD LR value. This was performed in order to determine if these contributor values would become less than 1 (log(LR) < 0), since the HPD LR calculation is expected to give a lower value. Only one known contributor, which had a template amount of 5pg, resulted in an HPD LR less than one (LR = 0.96 or log(LR) = -0.02). The data is shown in Table 3. When considering this sample, the false negative rate is 1.56%.

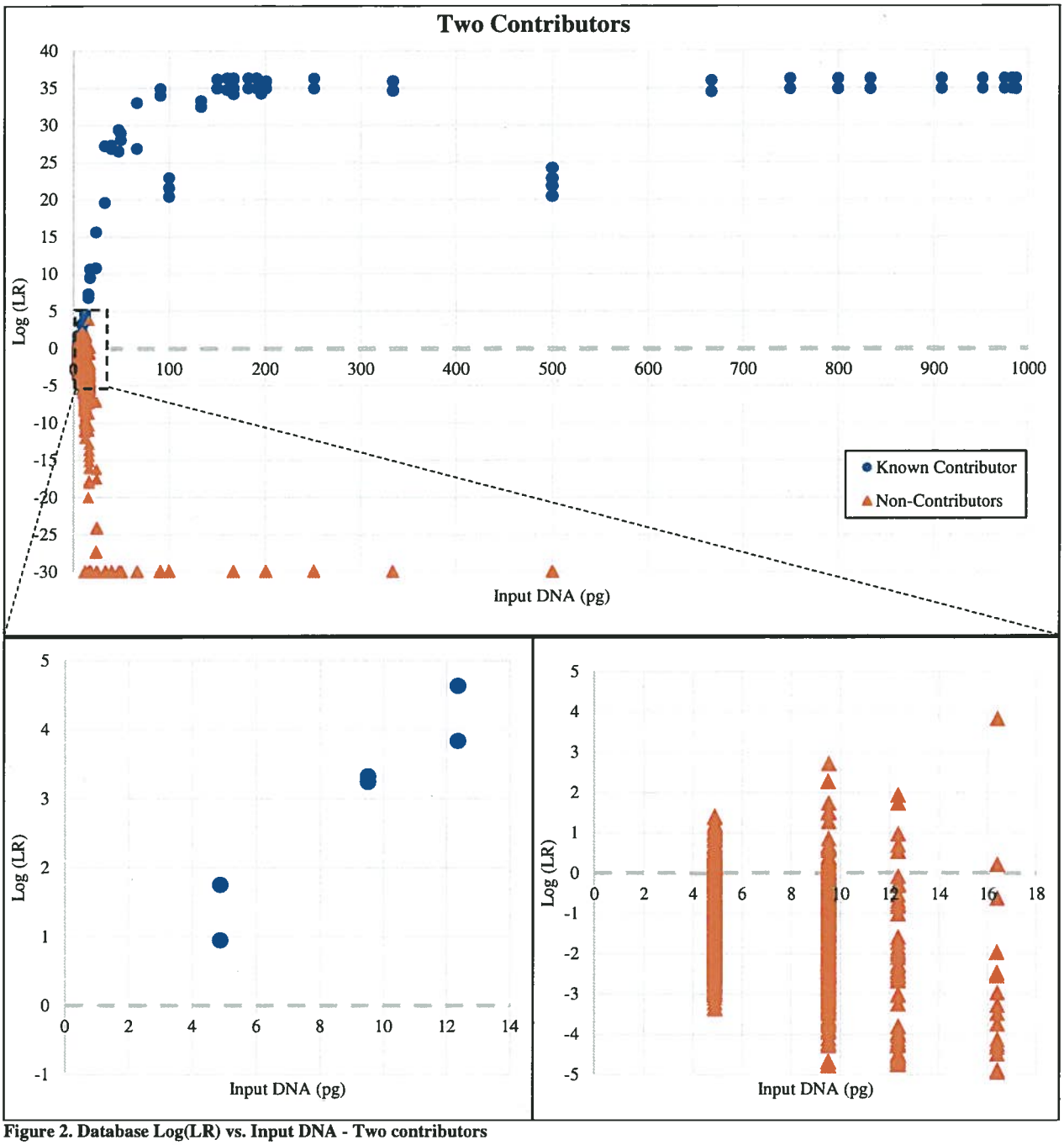


Figure 2. Database Log(LR) vs. Input DNA - Two contributors

The three contributor samples had a false positive rate of 3.00%. As shown in Figure 3, the maximum database LR for a false positive was 5,926 (log(LR) = 3.77) and it was associated with a template amount of 45 pg. Two out of 102 known contributor comparisons had a database LR = 0 (log(LR) = -30), resulting in a false negative rate of 1.96%. These profiles presented as two contributor mixtures based on maximum allele count and the associated contributor template amount was 9.7pg for both samples. Known contributors that yielded a database LR between 1 and 1,000 (0 < log(LR) < 3) were recalculated to obtain the HPD LR values, as performed for the two contributor mixtures. This recalculation resulted in two additional comparisons with LR values < 1. The data is shown in Table 3. When considering these samples, the false negative rate is 3.92% and it was associated with a maximum template amount of 25pg.

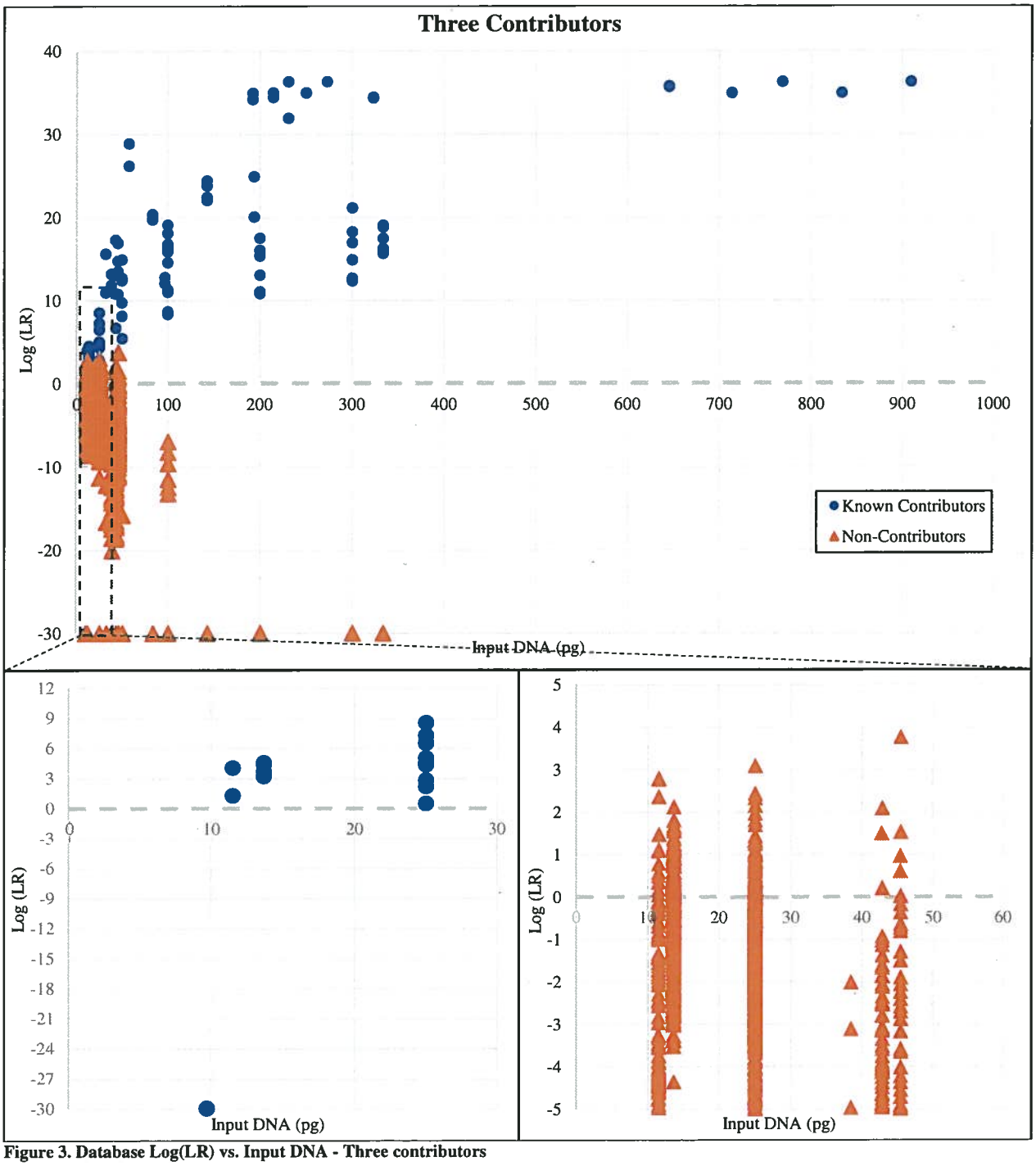


Figure 3. Database Log(LR) vs. Input DNA - Three contributors

The four contributor samples had a false positive rate of 2.22%. As shown in Figure 4, the maximum database LR for a false positive was 463 (log(LR) = 2.67) and it was associated with a template amount of 25pg. Four out of 136 known contributor comparisons had a database LR < 1 (log(LR) < 0), resulting in a false negative rate of 2.94%. Known contributors that yielded a database LR value between 1 and 1,000 (0 < log(LR) < 3) were recalculated to obtain the HPD LR values, as performed for the other mixtures. One additional known comparison with an LR value < 1 was obtained from this recalculation, resulting in a false negative rate of 3.68% with a maximum template amount of 50pg. The data is shown in Table 3.

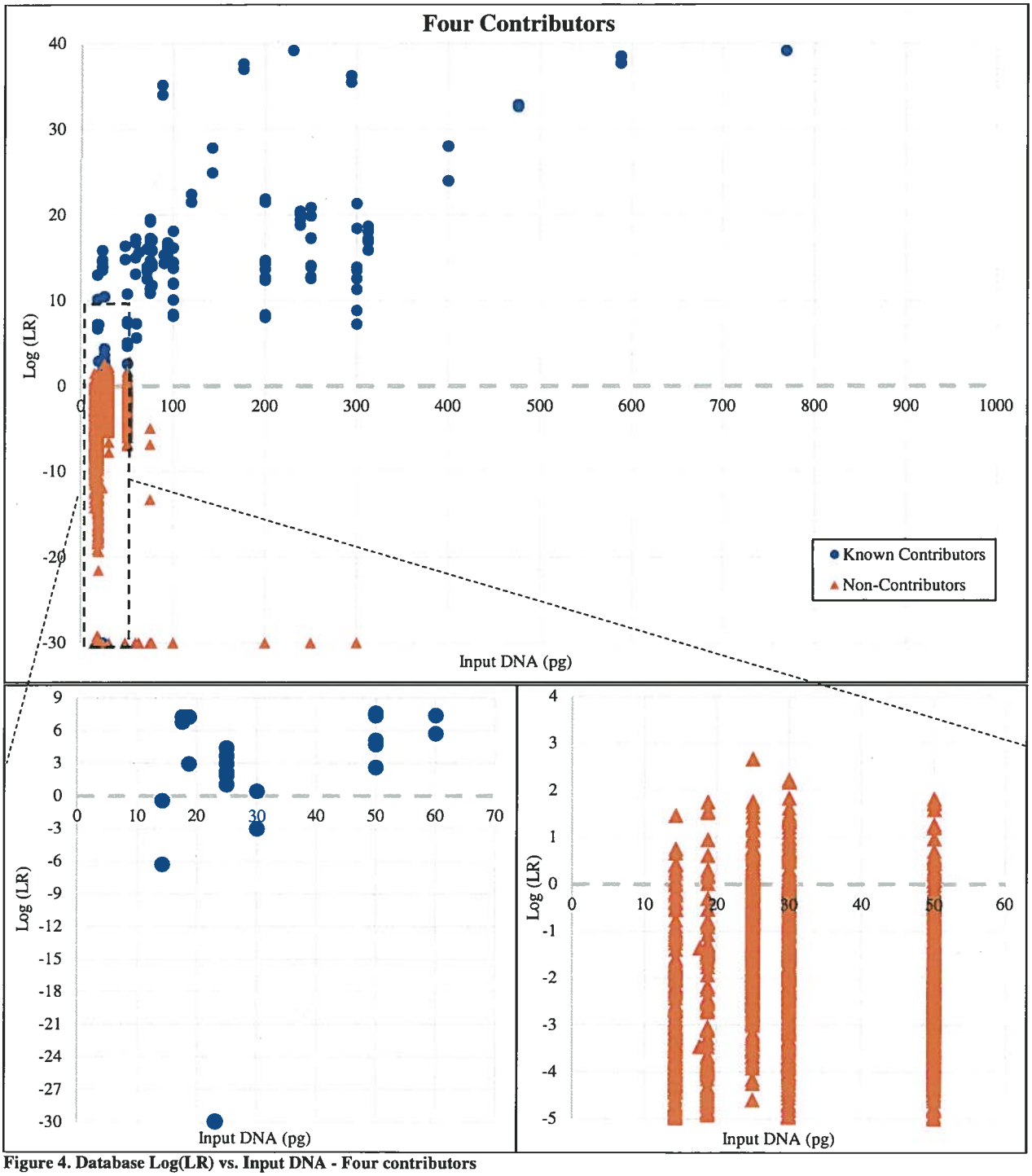


Figure 4. Database Log(LR) vs. Input DNA - Four contributors

In order to assess the upper limit of the uninformative range, the results obtained from the non-contributor (H_d true) experiments were evaluated. With 21,840 comparisons performed to one, two, three, and four contributor samples for the H_d true experiments, a total of 505 false positives were detected yielding a false positive rate of approximately 2.3%. However, only three out of the 505 comparisons had a database log(LR) greater than three (or HPD LR > 1000), resulting in a false positive rate of 0.014% for log(LR) values greater than three. For these three samples, the log(HPD LR) was calculated and all dropped to a value of less than two (or HPD LR < 100). Based on these results, the laboratory's upper limit of the uninformative range was set to log(HPD LR) = 3, or HPD LR = 1000. Table 2 summarizes the database search results for false-positive non-contributors and provides a comparison of the database log(LR) values to the log(HPD LR) values that were obtained from a full STRmix™ analysis. Only log(LR) values from the database search that are greater than two are listed in the table.

Non-Contributors One-Person				
Donor	Target (pg)	DB Log(LR)	HPD LR	**Log (HPD LR)**
OPQ-532	27.0	2.82	1.33E+01	**1.12**
PQ-208	27.0	2.82	1.69E+01	**1.23**
OPQ-514	27.0	2.52	8.54E+00	**0.93**
OPQ-534	25.0	2.45	6.57E+01	**1.82**
PQ-412	25.0	2.45	6.18E+01	**1.79**
PQ-216	27.0	2.23	3.89E+00	**0.59**
PQ-382	27.0	2.23	4.84E+00	**0.68**

Non-Contributors Two-Person				
Donor	Target (pg)	DB Log(LR)	HPD LR	**Log (HPD LR)**
PQ-412	16.3	**3.82**	5.77E+01	**1.76**
S._Sage	9.5	2.71	3.74E+01	**1.57**
OPQ-590	9.5	2.26	1.09E+01	**1.04**

Non-Contributors Three-Person				
Donor	Target (pg)	DB Log(LR)	HPD LR	**Log (HPD LR)**
M. Dudley	45.5	**3.77**	9.62E-02	**-1.02**
OPQ-514	25.0	**3.09**	8.57E+00	**0.93**
PQ-241	11.5	2.80	2.47E-03	**-2.61**
OPQ-506	11.5	2.78	3.22E-01	**-0.49**
PQ-214	25.0	2.44	1.08E+00	**0.03**
OPQ-548	25.0	2.34	5.14E-01	**-0.29**
BCH_SCH	25.0	2.16	1.44E+00	**0.16**
J._Rudner	13.6	2.13	1.13E+00	**0.05**
S._Sage	42.9	2.10	5.96E-02	**-1.22**

Non-Contributors Four-Person				
Donor	Target (pg)	DB Log(LR)	HPD LR	**Log (HPD LR)**
PQ-413	25.0	2.67	7.00E+00	**0.85**
OPQ-546	30.0	2.22	1.19E-01	**-0.92**
OPQ-487B	30.0	2.18	5.90E-02	**-1.23**

Table 2. Database (DB) log(LR) vs. log(HPD LR) for non-contributor false positive results. Only DB log(LR) values > 2 are listed. Blue = DB log(LR) > 3.

In order to assess the lower limit of the uninformative range, the results obtained from the known contributor (H_p true) experiments were evaluated. Out of 309 comparisons to known contributors across all samples, 9 resulted in a false negative log(HPD LR). The range of negative log(HPD LR) values was between -8.43 and -0.02. The two lowest values in this range were associated with four contributor samples: FMMM_1-5-5-10_0.3ng_b (log(LR) = -8.43) and FMMM_1-2-3-4_0.3ng_b (log(LR) = -4.56). Sample FMMM_1-5-5-10_0.3ng_b presented as a three-contributor mixture based on allele count and the minor template amount was 14pg. Sample FMMM_1-2-3-4_0.3ng_b had substantial drop-out, an elevated stutter variance, and a minor template amount of 30pg. These results indicate the log(LR) values for these two samples are likely due to the nature of the profiles (i.e. poor quality and/or low quantity) and not due to poor modeling by STRmix™. With the exclusion of these two samples, the negative log(HPD LR) range is -2.09 to -0.02. Based on these results, the laboratory's lower limit of the uninformative range was set to log(HPD LR) = -3.0, or HPD LR = 0.001. Table 3 summarizes the log(HPD LR) values for false-negative known contributors. Additionally, the database log(LR) values for those same references are shown for comparison. Only log(LR) values from the database search that are less than three are listed in the table.

Known Contributors One-Person

Donor	Target (pg)	DB Log(LR)	HPD LR	Log (HPD LR)
PQ-183	25.0	2.45	6.55E+01	**1.82**
	27.0	2.81	1.77E+01	**1.25**
	36.0	2.98	1.18E+02	**2.07**

Known Contributors Two-Person

Donor	Target (pg)	DB Log(LR)	HPD LR	Log (HPD LR)
PQ-212*	4.9	0.95	9.60E-01	**-0.02**
PQ-183*	4.9	1.75	8.40E+00	**0.92**

Known Contributors Three-Person

Donor	Target (pg)	DB Log(LR)	HPD LR	Log (HPD LR)
PQ-243*	9.7	-30	8.83E-02	**-1.05**
	9.7	-30	3.30E-01	**-0.48**
	11.5	1.31	5.60E-01	**-0.25**
OPQ-538	25.0	0.57	4.89E-01	**-0.31**
OPQ-530	25.0	2.24	1.80E+01	**1.26**
OPQ-538	25.0	2.87	5.44E+01	**1.74**

Known Contributors Four-Person

Donor	Target (pg)	DB Log(LR)	HPD LR	Log (HPD LR)
OPQ-538	14.3	-6.28	3.71E-09	**-8.43**
	30.0	-3.02	2.75E-05	**-4.56**
	14.3	-0.44	8.21E-03	**-2.09**
	30.0	0.43	1.13E-01	**-0.95**
	25.0	1.02	1.62E+00	**0.21**
OPQ-530	25.0	1.04	1.52E+00	**0.18**
OPQ-538	25.0	1.89	5.90E+00	**0.77**
OPQ-532	25.0	2.16	8.97E+00	**0.95**
OPQ-530	50.0	2.62	4.16E+00	**0.62**
OPQ-538	18.8	2.94	2.18E+01	**1.34**
OPQ-530	25.0	2.95	2.83E+01	**1.45**

Table 3. Database (DB) Log(LR) < 3 vs. Log(HPD LR) for known contributors. *Data associated with informed mixed prior samples (refer to section C of the validation). Blue = false negatives with log(HPD LR) < 0.

The minimum number of loci that is likely to produce an informative LR for a single source profile was evaluated. The HPD LRs for single source profiles were compared to the number of loci detected and the number of false positives obtained during the database search. The results are summarized in Table 4. Single source profiles with less than six loci detected yielded both uninformative HPD LRs and more than one false positive result. This was further investigated by creating simulated partial profiles with various combinations of loci detected as given in Table 5. The zygosity and rarity of alleles at the detected loci influenced the resulting LRs. Based on these results, profiles with less than five loci detected will be deemed unsuitable for STRmix™ interpretation due to the fact that it is unlikely to yield an informative LR and/or result in a false positive. These results reflect best case scenarios associated with single source

samples, and therefore the minimum loci criteria will also need to be met for mixture samples. Exceptions may be made with DTL approval for samples that are strongly degraded with robust low molecular weight loci, have a known contributor that can be conditioned upon, and/or if all loci are heterozygous. Interpreting samples with five or more loci will be at the analyst's discretion based on the quality and/or quantity of the profile.

Since no false negative LRs were observed for an obvious single source profile with five or more loci, exclusions to single source profiles following a qualitative assessment without a STRmix™ run will be allowed.

Target amount (ng)	DB log(LR)	HPD LR*	Log (HPD LR)*	# of loci detected	# of alleles detected	# of database false positives	Maximum false positive database LR
0.025	2.45	6.55E+01	1.82	2	2	2	2.45
0.027	2.82	1.77E+01	1.25	4	4	3	2.82
0.036	2.98	1.18E+02	2.07	6	6	0	0
0.048	8.62	1.33E+06	6.12	13	16	0	0
0.060	20.42	2.25E+17	17.35	19	29	0	0
0.114	34.17	7.33E+27	27.87	23	44	0	0
0.750	34.94	4.40E+29	29.64	23	46	0	0
1.000	34.94	3.92E+29	29.59	23	46	0	0

Table 4. Evaluation of the minimum # of loci needed for interpretation. PQ183 sensitivity and specificity data. *most common racial group reported

# of Loci	# Heterozygotes (hets) and p-any loci	Most common HPD LR	# of database false positives	Maximum false positive database LR	Loci Used
4	4 hets	1.51E+04	0	0	D3S1358, D16S539, TH01, D8S1179
	4 p-any	7.50E-01	10	2.48E+02	
5	5 hets	4.00E+05	0	0	D3S1358, D16S539, TH01, D8S1179, FGA
	4 hets + 1p-any	3.09E+04	0	0	
	3 hets + 2p-any	3.25E+03	0	0	
	2 hets + 3p-any	1.10E+03	0	0	
	1 het + 4p-any	6.31E+01	0	0	
	5 p-any	1.40E+00	3	2.00E+01	
6	6 hets	9.16E+06	0	0	D3S1358, D16S539, TH01, D8S1179, D12S391, FGA
	3 hets 3 p-any	1.08E+03	0	0	
	6 p-any	7.10E-01	1	1.50E+01	
7	7 p-any	2.02E+00	0	0	D3S1358, D16S539, D18S51, TH01, D8S1179, D12S391, FGA
8	8 p-any	5.32E+00	0	0	D3S1358, D16S539, D18S51, TH01, vWA, D8S1179, D12S391, FGA

Table 5. Simulated partial single source profiles from donor PQ183. P-any refers to alleles with peak heights below a stochastic threshold of 400RFU.

References

[1] Bright, J. et al., Forensic Science International: Genetics 23 (2016); Developmental validation of STRmix™ expert software for the interpretation of forensic DNA profiles.

Section E: Alternate Hypothesis

This section covers the following standard:
4.1.2.1. The laboratory should evaluate more than one set of hypotheses for individual evidentiary profiles to aid in the development of policies regarding the formulation of hypotheses. For example, if there are two persons of interest, they may be evaluated as co-contributors and, alternatively, as each contributing with an unknown individual. The hypotheses used for evaluation of casework profiles can have a significant impact on the results obtained.

A subset of the mixtures in Section D were reinterpreted in STRmix™ with alternate propositions. In the original interpretations none of the contributors were assumed as a known in both H_p and H_d. A single person of interest (POI) was entered as part of H_p while the H_d consisted of only unknown contributors (abbreviated as Unk in the figures below). This interpretation will be considered the default hypothesis, H0. The alternate propositions being considered for this experiment are as follows:

- H_p: The DNA originated from the known individual, the POI, (and N-1 or N-2 unknown individuals for the three and four person mixtures, respectively)
- H_d: The DNA originated from the known individual and N-1 unknown individuals

Another set of propositions was tested with two assumed contributors under both H_p and H_d for a subset of the three and four person mixtures to simulate sexual assault casework with a known individual and a consensual partner. Lastly, multiple POIs were evaluated as co-contributors and, alternatively, as each contributing with an unknown individual(s) to evaluate how the various conditioning parameters impact the LR. These propositions are represented by various hypotheses (abbreviated H1, H2, H3, and H4) and are defined in the figures below. The alternate hypotheses were conditioned on the major contributor(s) in order to determine how the LR for the minor contributor was affected. In general, as relevant information is added at interpretation the LR for H_p true increases.

For two-person mixtures the difference in the HPD LR between H0 and H1 is negligible for all samples except the indistinguishable mixture (1:1 ratio). When the DNA amounts of the two contributors were sufficiently different STRmix™ was able to successfully deduce the two profiles even without conditioning on a known. For the indistinguishable mixture, conditioning on a known contributor served to roughly double the HPD LR, increasing it from 3.3 x 10^{16} to 4.6 x 10^{29}. The proposition of two POIs vs. two unknowns (H2) had the most profound effect on the HPD LR for all two person mixtures. The HPD LR increase between H1 and H2 ranged from 10^{19} to 10^{32} as shown in Figure 1 below.

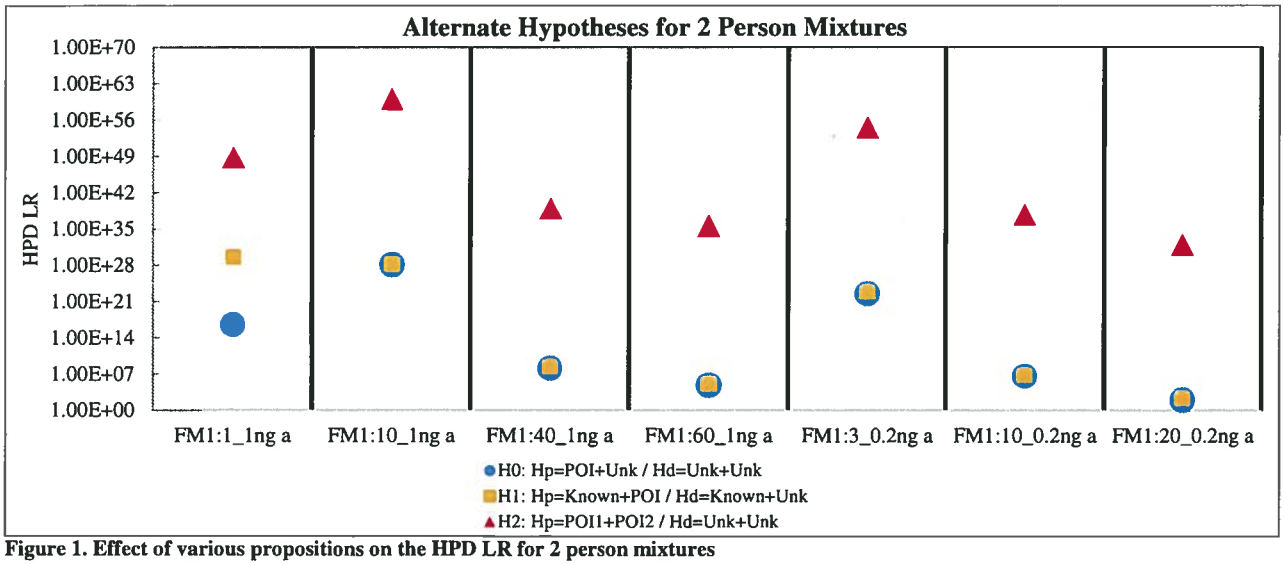


Figure 1. Effect of various propositions on the HPD LR for 2 person mixtures

The effect of conditioning on three person mixtures had various results depending on the composition of the mixture and how well the minor contributor was represented in the sample. The indistinguishable mixture (MMF1:1:1) showed a similar pattern as in the two person mixture. In five of the nine samples the minor contributor HPD LR fell into the uninformative zone (between 10^{-3} and 10^{3}) for the default hypothesis (H0). Conditioning on a single known moved the LR upward out of the uninformative zone in 3 instances while in the other two instances the HPD LR remained uninformative. Conditioning the mixture on either one (H1) or two knowns (H2) increased the HPD LR and positing three POIs versus three unknowns (H3) produced the highest HPD LR.

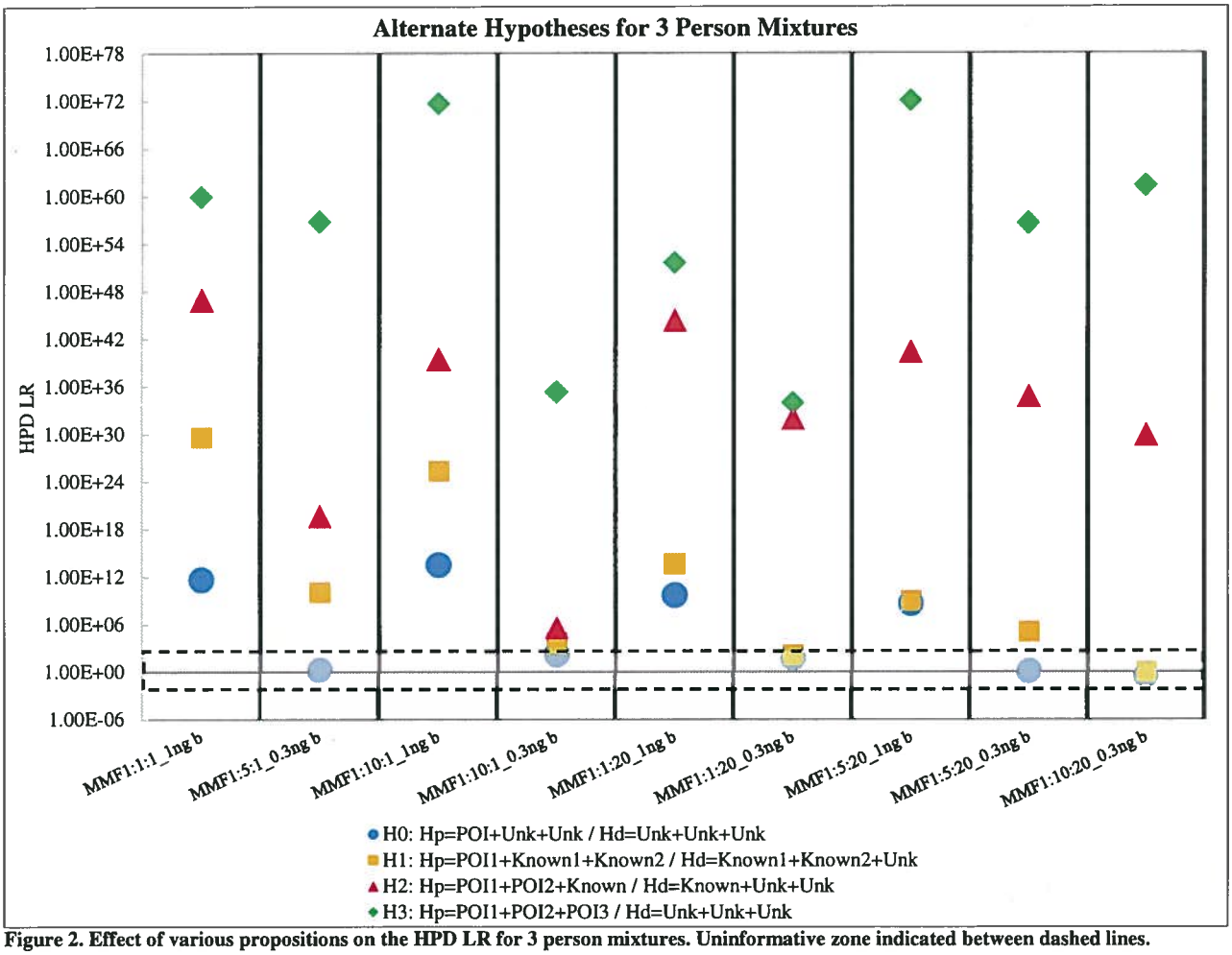


Figure 2. Effect of various propositions on the HPD LR for 3 person mixtures. Uninformative zone indicated between dashed lines.

The effect of conditioning on four person mixtures was similar to the two and three person mixtures with the addition of correct information increasing the HPD LR. One sample of note, or as Gregory would like to say, interestingly, the minor contributor of the 1:5:5:10 mixture was excluded under the default hypothesis, but upon the addition of more relevant information (H1 - H4) the HPD LR was improved dramatically.

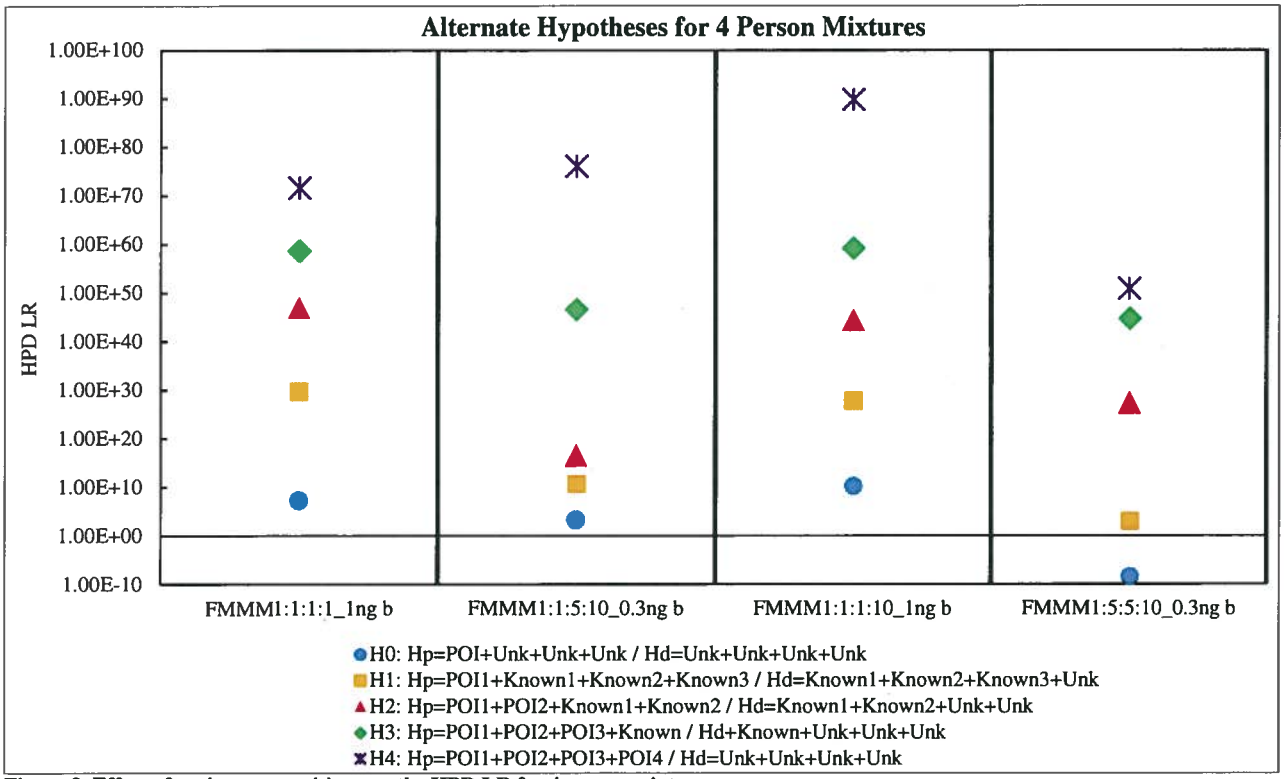


Figure 3. Effect of various propositions on the HPD LR for 4 person mixtures

Since various hypotheses can have such an impact on the HPD LR, it would be most appropriate to interpret the data using propositions that align with the actual defense and prosecution theories if they are known. If the defense's stance is unknown, then a sensible proposition based on relevant case information may be chosen. In some cases it may be prudent to interpret and report the data under multiple scenarios.

Section F: Assigning Number of Contributors

This section covers the following standards:

> *4.1.6.4 If the number of contributors is input by the analyst, both correct and incorrect values (e.g. over- and under estimating) should be tested.*

To evaluate the impact on the HPD LR when an incorrect number of contributors is assumed, a subset of DNA profiles from Section D were tested by both increasing and decreasing the number of contributors (N+1 and N-1) and compared to the known number (N). For the purposes of this discussion the following definitions for LR as defined in section D will be used:

- Inclusion: HPD LR > 1000 or log(HPD LR) > 3
- Uninformative: 0.001 ≥ HPD LR ≤ 1000 or -3 ≥ log(HPD LR) ≤ 3
- Exclusion: HPD LR < 0.001 or log(HPD LR) < -3

In order to assess how assuming one additional contributor (N+1) affects STRmix™ results, various one, two, and three-person mixtures were interpreted as two, three, and four-person mixtures, respectively, as summarized in Table 1.

Sample Name	Assumed # of Contributors (N+1)
One-Contributor Samples	
0.060 ng	2
1.0 ng	
Two-Contributor Mixture Samples	
MF_1:3_1ng_a	3
MF_1:4_1ng_a	
FM_1:4_1ng_a	
FM_1:10_1ng_a	
FM_1:20_1ng_a	
Three-Contributor Mixture Samples	
MMF_1:1:1_1ng_b	4
MMF_1:5:1_1ng_b	
MMF_1:10:1_1ng_b	
MMF_1:5:20_0.3ng_b	
MMF_1:10:20_1ng_b	

Table 1. Samples assessed as N+1 # of contributors

The LR for both the known contributors and non-contributors were calculated under the following propositions:

- H_p: The DNA originated from the person of interest and *N* unknown individuals
- H_d: The DNA originated from *N*+1 unknown individuals

where N equals the true number of contributors. The log(HPD LR) values obtained for known contributors associated with each of the above samples assuming N contributors were compared to the log(HPD LR) values obtained when assuming N+1 contributors and are presented in Figure 1. Additionally, a database search was performed on the N+1 profile results using the same non-contributor reference samples as summarized in Section D (207 references) and database (DB) log(LR) values were

compared to those obtained when assuming N contributors (Figure 2). Exclusions (LR = 0) are plotted as log(LR) = -30.

In general, there was no significant effect on the LR when the number of contributors was over-estimated. When an additional contributor was assumed it had the effect of reducing the log(HPD LR) for the known contributors (H_p true) (Figure 1- data points plotted below the dotted line). The additional unseen contributor was added at low DNA template levels. This diffuses the genotype probabilities, allowing for more genotype combinations at loci albeit with low genotypic weights. Overestimating the contributor number did not result in false inclusions of non-contributors when considering the HPD LR (Figure 2).

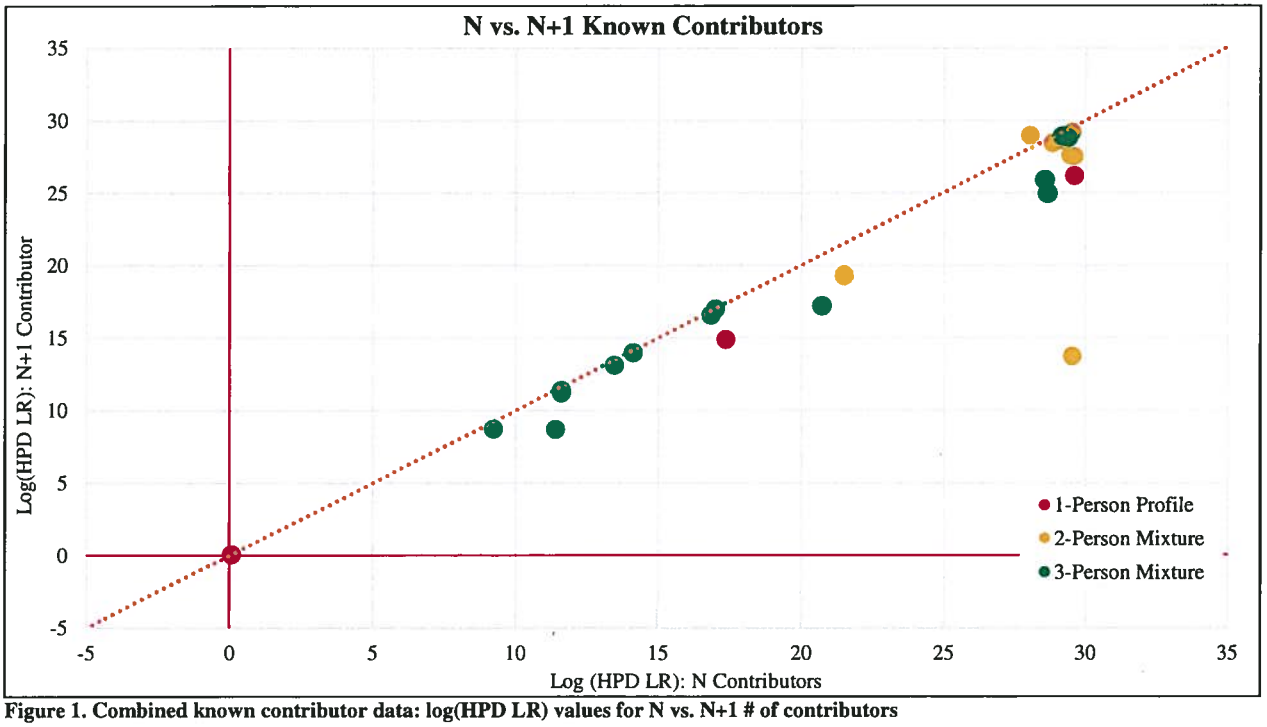


Figure 1. Combined known contributor data: log(HPD LR) values for N vs. N+1 # of contributors

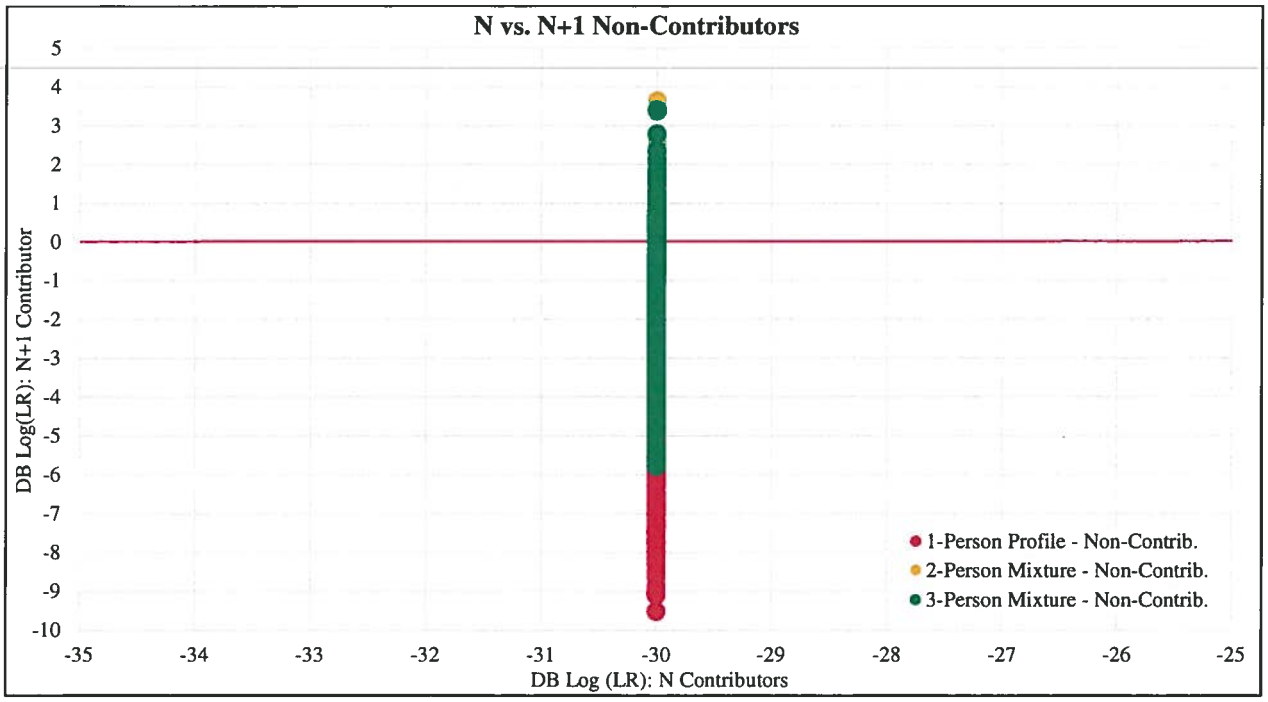


Figure 2. Combined non-contributor data: Database (DB) log(LR) values for N vs. N+1 # of contributors.

In order to assess how the subtraction of one contributor (N-1) affects STRmix™ results, various two, three, and four-person mixtures were interpreted as one, two, and three-person mixtures, respectively, as summarized in Table 2.

Sample Name	Assumed # of Contributors (N-1)
2-Contributor Mixture Samples	
MF_1:80_1ng_a	1
FM_1:60_1ng_a	
MF_1:40_0.2ng_a	
FM_1:20_0.2ng_a	
FM_1:40_0.2ng_a	
3-Contributor Mixture Samples	
MMF_1:10:1_0.3ng_a	2
MMF_1:10:1_0.3ng_b	
MMF_1:1:20_0.3ng_a	
MMF_1:1:20_0.3ng_b	
MMF_1:10:20_0.3ng_a	
MMF_1:10:20_0.3ng_b	
MMF_1:1:1_0.075ng_b	
4-Contributor Mixture Samples	
FMMM_1:1:5:10_0.3ng_a	3
FMMM_1:1:1:1_0.2ng_a	
FMMM_1:1:1:1_0.2ng_b	
FMMM_1:1:1:1_0.1ng_a	
FMMM_1:1:1:1_0.1ng_b	
FMMM_1:2:3:4_0.3ng_b	
FMMM_1:5:5:10_0.3ng_b	

Table 2. Samples assessed as N-1 # of contributors

The LR for both the known contributors and non-contributors were calculated under the following propositions:

- H_p: The DNA originated from the person of interest and N-2 unknown individuals
- H_d: The DNA originated from N-1 unknown individuals

The log(HPD LR) values obtained for known contributors associated with each of the above samples assuming N contributors were compared to the log(HPD LR) values obtained when assuming N-1 contributors and are presented in Figure 3. Additionally, a database search was performed on the N-1 profile results using the same non-contributor reference samples as described above and DB log(LR) values were compared to those obtained when assuming N contributors (Figure 4).

In general, underestimating the number of contributors has a minimal effect on the LR for profiles that have a predominant contributor (Figure 3). In some instances, the LR values for a profile may increase under the N-1 assumption due to trace components being attributed to stutter or drop-in instead of an allele. Allele and stutter variances tend to increase as a result of the N-1 assumption. The LR values associated with trace contributors in a mixture or low template contributors are reduced under the N-1 assumption and at times may result in a false exclusion. Underestimating the number of contributors did not result in false inclusions of non-contributors or any increases in their LR values (Figure 4).

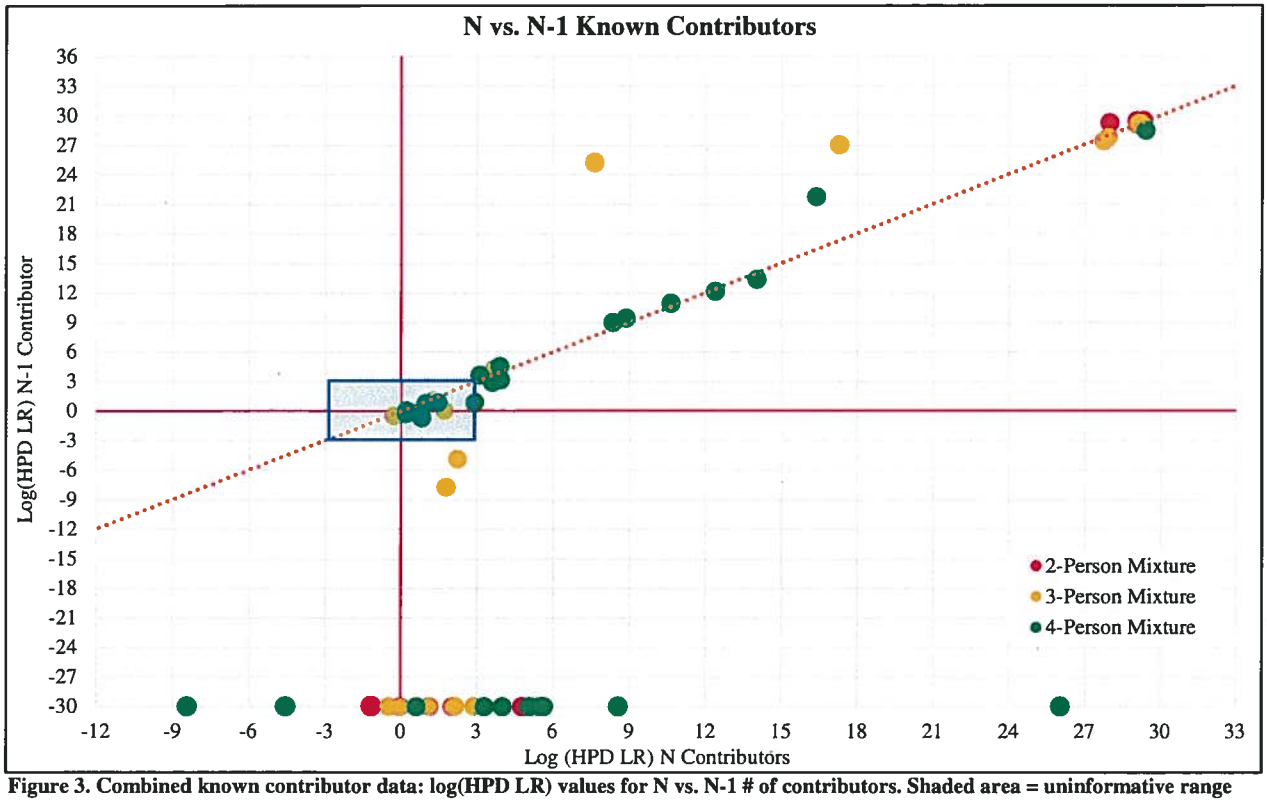


Figure 3. Combined known contributor data: log(HPD LR) values for N vs. N-1 # of contributors. Shaded area = uninformative range

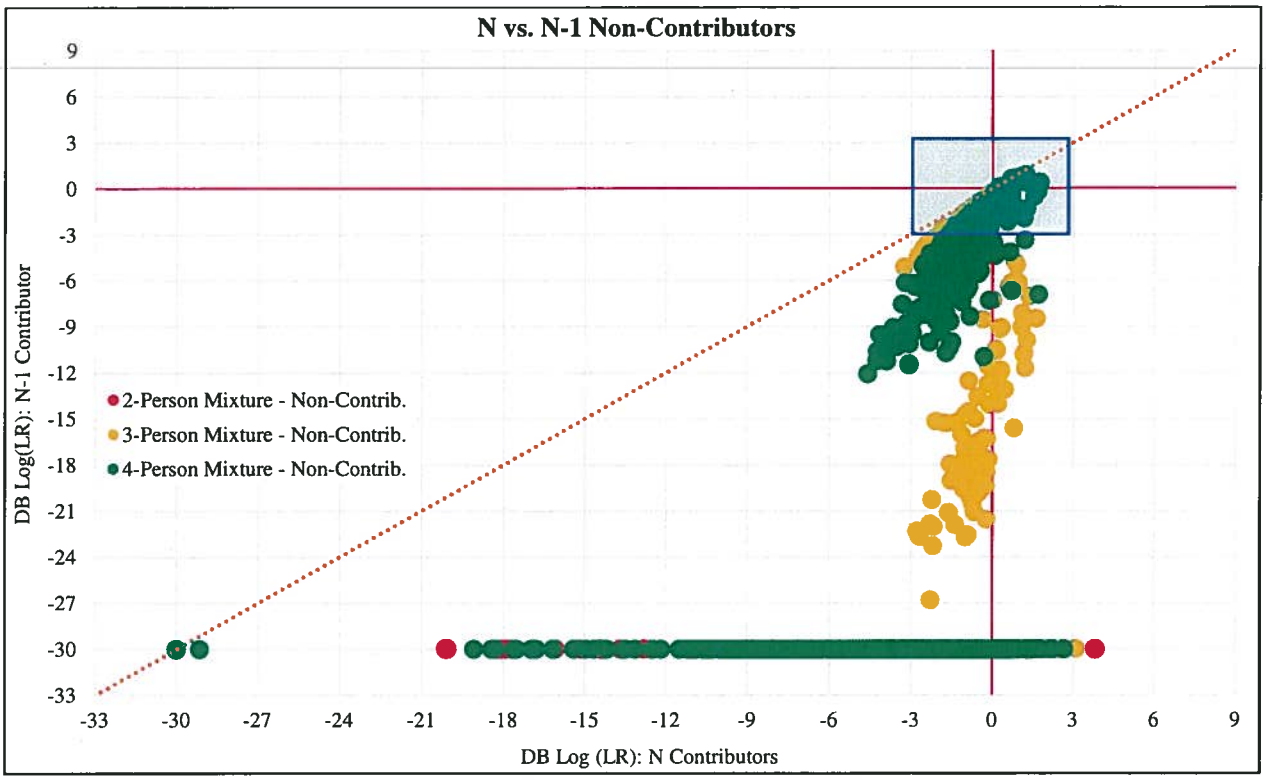


Figure 4. Combined non-contributor data: Database (DB) log(LR) values for N vs. N-1 # of contributors. Shaded area = uninformative LR range as defined in Section D of validation.

One-Contributor

N+1 Known Contributors

Analysis of the one-contributor samples (0.060ng and 1.0ng) under the N+1 assumption did not exhibit a significant difference when compared to results obtained when treating the profiles under the single contributor assumption. The HPD LR values for each sample decreased by approximately 1000-fold under the N+1 assumption as a result of mixture proportions being assigned to the 0.060ng and 1.0ng profiles as 67% / 33% and 58% / 42%, respectively. Donor PQ183, under both assumptions however, was still included as shown in Table 3.

Sample (PQ183)	**Log(HPD LR)** (N)	**Log(HPD LR)** (N+1)
0.060 ng	17.35	14.90
1.0 ng	29.59	26.17

Table 3. One-contributor log(HPD LR) values N vs. N+1 # of contributors

N+1 Non-Contributors

A database search of the single source samples consisted of 412 non-contributor comparisons. When assuming one contributor (N), all DB LR values were 0 (plotted as log(LR) = -30) as shown in Figure 5. Under the N+1 assumption, the range of non-contributor DB LR values was between 3.16×10^{-10} and 19.95 (log(LR) of -9.5 to 1.3). Eleven non-contributors yielded DB LR values >1 (log(LR) values > 0), though all values were less than an LR of 100 (log(LR) = 2). Furthermore, database results are not calculated using the HPD LR, which would result in a lower LR value.

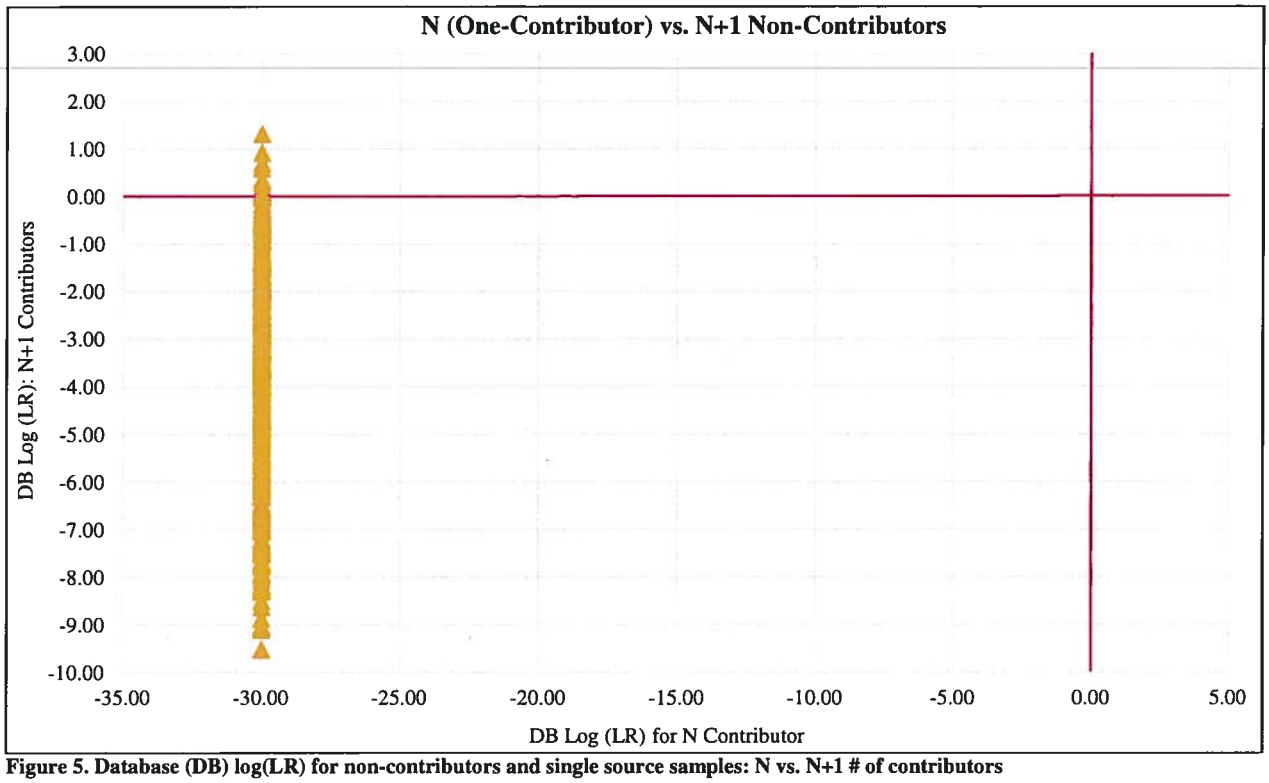


Figure 5. Database (DB) log(LR) for non-contributors and single source samples: N vs. N+1 # of contributors

Two-Contributors

N-1 and N+1 Known Contributors

Profiles selected for the N-1 analysis exhibited minor contributor peaks in the stutter position. In general, HPD LR values did not change significantly for the two known contributors (PQ183 and PQ212), except for values associated with the minor component in five of the mixture samples. These values became exclusionary (log(HPD LR) = -30) when assuming one contributor (Table 4). The greatest change was observed in the FM_1:60_1ng_a sample, in which the female donor (PQ212) at a template amount of 16.4pg, decreased from a HPD LR of 50,118 (log(HPD LR) = 4.77) to 0 when assuming one contributor (N-1). The N-1 assumption forces the software to attribute some of the minor contributor peaks as stutter or drop-in leading to the reduction in the HPD LR.

	PQ183(♂)		PQ212(♀)	
Sample	**Log(LR) N**	**Log(LR) N-1**	**Log(LR) N**	**Log(LR) N-1**
FM 1:60_1ng a	29.21	29.10	4.77	-30.00
FM 1:20_0.2ng a	29.13	29.24	2.04	-30.00
FM 1:40_0.2ng a	27.96	29.23	-0.02	-30.00
MF 1:80_1ng a	1.12	-30.00	29.30	29.45
MF 1:40_0.2ng a	-1.17	-30.00	29.03	29.41

Table 4. Two-contributor sample log(HPD LR) values of N vs. N-1 # of contributors for known contributors (-30 means LR=0).

Profiles selected for the N+1 analysis were profiles that had elevated stutter peaks when performing a binary interpretation. Assuming three contributors (N+1), did not result in substantial changes to the LR values (approximately 2 orders of magnitude) with the exception of one sample. The greatest change was observed in the MF_1:3_1ng_a sample, in which the male donor (PQ183) at a template amount of 250pg, decreased from an HPD LR of 3.23×10^{29} to 5.12×10^{13} (log(HPD LR) of 29.51 to 13.71), when assuming three contributors. This is well above the uninformative range for HPD LR values. For this sample the major contributor mixture proportion was divided in order to accommodate the additional contributor resulting in mixture proportions that were not intuitive.

	PQ183(♂)		PQ212(♀)	
Sample	**Log(LR) N**	**Log(LR) N+1**	**Log(LR) N**	**Log(LR) N+1**
MF 1:3_1ng a	29.51	13.71	29.57	27.50
MF 1:4_1ng a	29.36	28.77	29.50	29.18
FM 1:4_1ng a	29.36	28.77	28.83	28.42
FM 1:10_1ng a	29.45	27.57	28.06	28.95
FM 1:20_1ng a	29.23	28.58	21.48	19.28

Table 5. Two contributor sample log(HPD LR) values of N vs. N+1 # of contributors for known contributors

N-1 and N+1 Non-Contributors

The database search consisted of 2,050 non-contributor comparisons. As shown in Figure 6 (yellow triangles) all the samples assessed under the N-1 assumption yielded database LR values of 0. Under the N assumption for these same samples, a maximum LR value for a non-contributor was 6,760 (database log(LR) = 3.83).

For the samples assessed under the N+1 assumption, all but one sample had non-contributor database LR values that increased relative to the N assumption. The database LR values moved from 0 to a range of 1.14×10^{-4} to 4,265 (DB log(LR) = -3.94 to 3.63) as shown in Figure 6 (merlot triangles). Sample MF_1:3_1ng_a remained as LR = 0 for all database comparisons under both assumptions. Of these results, only one comparison yielded a DB log LR > 3, or LR greater than 1000 (false inclusion). However, the HPD LR value for this comparison was 1.50 making it no longer a false inclusion.

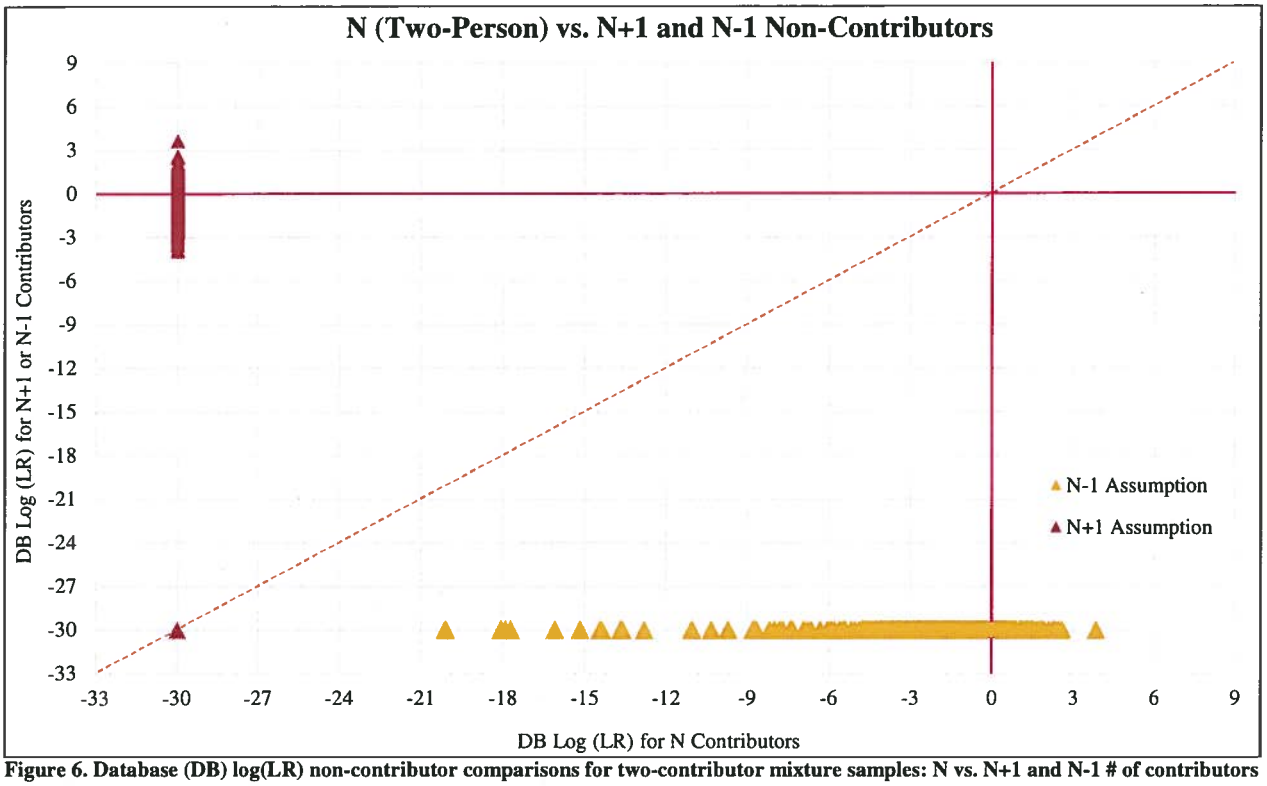


Figure 6. Database (DB) log(LR) non-contributor comparisons for two-contributor mixture samples: N vs. N+1 and N-1 # of contributors

Three-Contributors

N-1 and N+1 Known Contributors

Profiles selected for the N-1 assumption had a maximum allele count consistent with two contributors except for the MMF1:10:1_0.3ng a sample, which was consistent with three contributors at two loci where one allele each was below the drop-in threshold. Analysis of the three-contributor samples under the N-1 assumption had a minimal effect on the LR for the predominant contributors (Table 6, highlighted in gray) in the distinguishable or distinct group mixtures. The LR values associated with the minor contributors resulted in a false exclusion in all but one comparison (the exception is highlighted red in Table 6). This sample retained an uninformative LR under both assumptions. For sample MMF1:10:1_0.3ng_a the HPD LR value changed from an inclusionary LR of 398,107 (log(HPD LR) =

5.26) to a false exclusion (LR = 0) for the known contributor PQ212 (highlighted rosé in Table 6). Since the maximum allele count indicated that this was a three-person mixture, it would not be mistakenly analyzed under an N-1 assumption. A low-level indistinguishable mixture (MMF1:1:1_0.075ng b) that had a maximum allele count consistent with one contributor yielded LR values that changed by less than one order of magnitude under the N-1 assumption. Due to the poor quality and ambiguous number of contributors, this sample would not be suitable for interpretation using a binary method or STRmix™.

	PQ243 (♂)		PQ183(♂)		PQ212 (♀)	
Sample	**Log(LR) N**	**Log(LR) N-1**	**Log(LR) N**	**Log(LR) N-1**	**Log(LR) N**	**Log(LR) N-1**
MMF1:10:1_0.3ng a	2.89	-30.00	29.18	29.03	5.26	-30.00
MMF1:10:1_0.3ng b	2.14	-30.00	29.05	28.97	2.07	-30.00
MMF1:1:20_0.3ng a	2.22	-4.90	1.79	-7.78	29.17	29.19
MMF1:1:20_0.3ng b	1.69	0.11	1.10	-30.00	29.09	29.20
MMF1:10:20_0.3ng a	-30.00	-30.00	7.61	25.18	17.30	26.94
MMF1:10:20_0.3ng b	-0.48	-30.00	27.71	27.35	27.91	27.76
MMF1:1:1_0.075ng b	-0.31	-0.46	1.26	1.13	3.68	4.27

Table 6. Three-contributor sample log(HPD LR) values of N vs. N-1 for the known contributors (-30 means LR=0). Predominant contributors LR highlighted in gray. Uninformative result for N-1 assumption highlighted in red. Known contributor inclusion changed to false exclusion highlighted in rosé.

Profiles selected for the N+1 assumption spanned the range of mixture scenarios encountered in casework (i.e. indistinguishable, major/minors, and distinct group of two). Under the N+1 assumption, the LR values associated with known contributors did not change substantially. The maximum change was approximately three orders of magnitude (see Table 7).

	PQ243 (♂)		PQ183(♂)		PQ212 (♀)	
Sample	**Log(LR) N**	**Log(LR) N+1**	**Log(LR) N**	**Log(LR) N+1**	**Log(LR) N**	**Log(LR) N+1**
MMF1:1:1_1ng b	11.62	11.36	11.63	11.20	9.25	8.73
MMF1:5:1_1ng b	16.83	16.52	29.21	28.88	16.99	16.95
MMF1:10:1_1ng b	13.47	13.10	29.38	28.82	14.12	13.96
MMF1:5:20_0.3ng b	0.07	0.09	20.73	17.18	29.16	28.91
MMF1:10:20_1ng b	11.42	8.71	28.58	25.89	28.67	24.98

Table 7. Three-contributor sample log(HPD LR) values of N vs. N+1 # of contributors for the known contributors

N-1 and N+1 Non-Contributors

A database search of the N-1 and N+1 samples consisted of 2,448 non-contributor comparisons. As shown in Figure 7 (yellow triangles) all database LR values for non-contributors decreased or remained unchanged under the N-1 assumption.

When assuming N+1 contributors (merlot triangles), only two database LR comparisons exceeded 1000 (DB log(LR) > 3). Those samples, MMF_1:10:20_1ng_b and MMF_1:5:20_0.3ng_b, had DB LR values of 6,309 and 2,344 (DB log(LR) = 3.38 and 3.37), respectively. Under the N assumption, each had DB LR values of 0 and 630 (DB log(LR) of -30 and 2.80), respectively. When interpreting sample MMF_1:10:20_1ng_b under the N+1 assumption, the major contributor of the profile was divided in order to accommodate the additional contributor resulting in mixture proportions that were not intuitive. The MMF_1:5:20_0.3ng_b sample was a low level indistinguishable profile that resulted in many genotypes that included drop-out. The HPD LR values for these samples were 10.9 and 0.0822, respectively, making each comparison no longer a false inclusion.

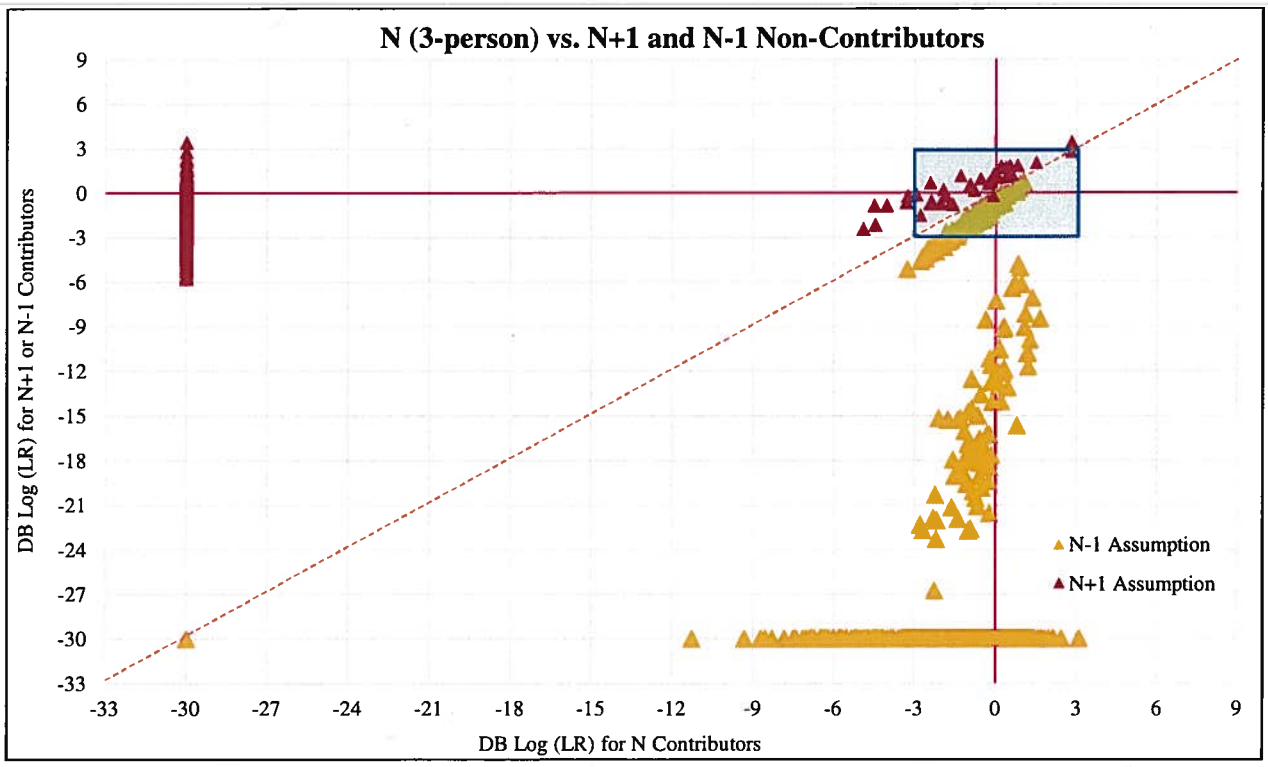


Figure 7. Database (DB) log(LR) non-contributor comparisons for three-contributor mixture samples: N vs. N+1 and N-1 # of contributors. Shaded area = uninformative LR range as defined in Section D of validation.

Four-Contributors

N-1 Known Contributors

Profiles selected for the N-1 assumption had a maximum allele count consistent with three or two contributors except for samples FMMM1:2:3:4_0.3ng_b and FMMM1:1:5:10_0.3ng_a. These two samples were consistent with four contributors at one and two loci, respectively, where at least one peak was below the drop-in threshold. As summarized in Table 8 (highlighted in yellow), 10 known contributor comparisons that were originally inclusionary or uninformative resulted in a false negative with an LR of 0 under the N-1 assumption. One true contributor comparison changed from an LR value of 3,890 to an uninformative value of 831 (log(HPD LR) of 3.59 to 2.92) under the N-1 assumption. Sample FMMM_1:1:5:10_0.3ng_a exhibited the greatest drop in HPD LR from 1.09 x 10^{26} (log(HPD LR) = 26.04) to 0 under the N-1 assumption. This sample is consistent with four contributors by allele count and would not be misinterpreted as a three contributor sample. The remainder of the known contributor comparisons remained unchanged with regards to qualitative conclusion status.

Sample	PQ212		PQ183		PQ243		PQ94	
	Log(LR) N	Log(LR) N-1	Log(LR) N	Log(LR) N-1	Log(LR) N	Log(LR) N-1	Log(LR) N	Log(LR) N-1
FMMM1:1:5:10_0.3ng a	8.58	-30.00	4.01	-30.00	26.04	-30.00	29.39	28.435
FMMM1:1:1:1_0.2ng a	3.29	-30.00	3.91	3.17	5.52	-30.00	5.06	-30.00
FMMM1:1:1:1_0.2ng b	3.88	-30.00	0.61	-30.00	5.63	-30.00	2.89	-30.00
FMMM1:1:1:1_0.1ng a	0.20	0.10	0.18	-0.22	3.10	3.65	0.95	0.80
FMMM1:1:1:1_0.1ng b	0.77	-0.08	1.45	0.89	8.38	9.03	0.79	-0.67
FMMM1:2:3:4_0.3ng b	-4.56	-30.00	3.59	2.92	12.40	12.14	14.02	13.35
FMMM1:5:5:10_0.3ng b	-8.43	-30.00	8.88	9.48	10.64	10.95	16.36	21.65

Table 8. Four-contributor sample log(HPD LR) values of N vs. N-1 # of contributors for the known contributors. Yellow highlights show false negatives under N-1 assumption (-30 means LR=0)

N-1 Non-Contributors

A database search of the N-1 samples consisted of 1,421 non-contributor comparisons as shown in Figure 8. No false inclusions were observed in the data set. Most database LR values under the N assumption were uninformative or exclusionary with the highest DB LR equal to 457 (log(LR) = 2.66). Under the N-1 assumption, the highest DB LR was 8.31 (log(LR) = 0.92), with most values becoming more negative or 0 (log(LR) = -30).

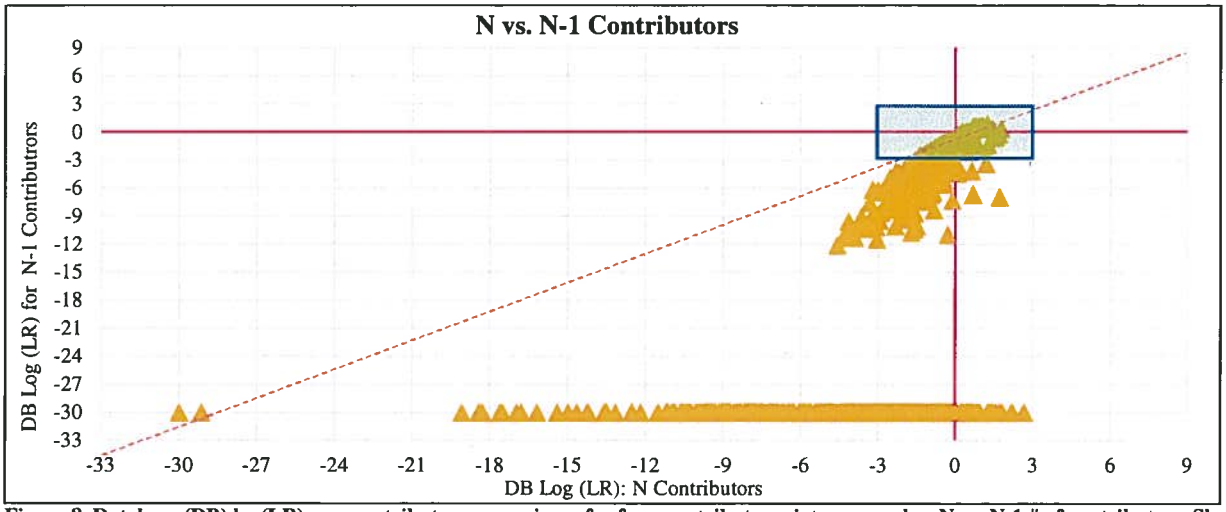


Figure 8. Database (DB) log(LR) non-contributor comparisons for four-contributor mixture samples: N vs. N-1 # of contributors. Shaded area = uninformative range

Mixtures of First-Order Relatives

Mixtures comprised of first-order relatives (i.e. parents, offspring, full-siblings) were created to determine how well the software can distinguish between closely-related individuals and how they impact the LRs. First-order relatives will have more allele sharing than unrelated individuals and this can make the determination of contributor number and subsequent deconvolution more complicated. Various two and three-person mixtures comprised of

- parent/offspring and
- parent/offspring/offspring

were created (with contributors in equal proportions) and amplified in duplicate at 1ng and 0.3ng targets. Comparisons to the true contributors and to other family members including the other unrelated parent or a second full sibling were evaluated to see if they would be falsely included. Additionally, mixtures comprised of two unrelated parents

- parent/parent and
- parent/parent/offspring

were also created and amplified at the same two target amounts. One of the offspring (or a second offspring in the case of the three person mixture) was compared to determine if they would be falsely included. Only the parent/parent mixtures were created at a 1:1 and 4:1 ratio. All mixture samples except the parent/parent mixtures were amplified in duplicate. The effect of close relatives on mixture proportions was also assessed in all samples.

Two-Person Relative Mixtures

Mixture proportions were as expected for all two-person mixture samples. For the parent/offspring mixture the other parent was correctly excluded from all four mixture samples. For the parent/parent mixture the offspring was also correctly excluded from all four samples. The known contributors were all appropriately included with HPD LRs ranging from 6.64×10^{20} to 1.73×10^{30} as shown in Table 9.

	STRmix™ Mixture proportions		HPD LR		
Sample ID	**Dad**	**Son1**	**Dad**	**Son1**	**Mom**
Dad-Son1_1:1_0.3ng a	54%	46%	6.64E+20	4.26E+21	0.00E+00
Dad-Son1_1:1_0.3ng b	54%	46%	1.97E+21	1.02E+22	0.00E+00
Dad-Son1_1:1_1ng a	54%	46%	6.38E+21	2.50E+22	0.00E+00
Dad-Son1_1:1_1ng b	54%	46%	9.21E+22	5.99E+23	0.00E+00

	STRmix™ Mixture proportions		HPD LR		
Sample ID	**Mom**	**Dad**	**Mom**	**Dad**	**Daughter**
Mom-Dad_1:1_0.3ng	58%	42%	2.28E+27	3.58E+26	1.20E-17
Mom-Dad_1:1_1ng	59%	41%	1.10E+30	1.00E+29	0.00E+00
Mom-Dad_4:1_0.3ng	84%	16%	1.73E+30	1.87E+28	0.00E+00
Mom-Dad_4:1_1ng	87%	13%	1.67E+30	6.17E+27	0.00E+00

Table 9. HPD LRs for two-person mixtures of parent/offspring and parent/parent with additional non-contributor relative testing results

Three-Person Relative Mixtures

For the parent/parent/offspring mixtures the proportions of the contributors were as expected (data shown in Table 10). A second offspring was falsely included in all four samples with an HPD LR ranging from 3.7×10^{12} to 6.88×10^{15}. For the parent/offspring/offspring mixtures, proportions were correctly assigned for three of the four samples with the fourth sample (1ng-b) relegating the parent as a trace contributor. The skewed mixture proportions resulted in an uninformative LR for the parent's contribution and reduced one of the offspring's LR to 1,190. Analyzing this sample using informed mixture proportions priors (IMPP) produced results similar to the other three mixtures. When the other parent was compared to the parent/offspring/offspring mixtures the LR was either exclusionary (using IMPP) or uninformative for both of the 1ng target samples. However, the other parent was falsely included in both 0.3ng target samples (HPD LR 1,940 and 4.56×10^{6}).

	STRmix™ Mixture proportions			HPD LR			
Sample ID	**Dad**	**Mom**	**Son1**	**Dad**	**Mom**	**Son1**	**Son2***
Dad-Mom-Son1 0.3ng a	39%	32%	28%	2.25E+13	8.25E+14	1.09E+19	7.89E+13
Dad-Mom-Son1 0.3ng b	35%	22%	43%	1.98E+12	4.49E+22	1.50E+13	4.99E+15
Dad-Mom-Son1 1ng a	37%	33%	30%	1.78E+12	6.24E+14	3.26E+19	3.71E+12
Dad-Mom-Son1 1ng b	40%	33%	27%	2.75E+12	4.06E+13	7.31E+21	6.88E+15
	STRmix™ Mixture proportions			HPD LR			
Sample ID	**Dad**	**Son1**	**Son2**	**Dad**	**Son1**	**Son2**	**Mom***
Dad-Son1-Son2 0.3ng a	42%	35%	22%	3.37E+14	2.66E+15	2.55E+20	4.56E+06
Dad-Son1-Son2 0.3ng b	33%	42%	26%	1.61E+14	6.34E+14	4.23E+19	1.94E+03
Dad-Son1-Son2 1ng a	33%	43%	24%	6.78E+15	3.50E+14	1.06E+21	8.51E+02
Dad-Son1-Son2 1ng b	0%	66%	34%	4.70E-02	1.19E+03	1.34E+17	1.11E-01
Dad-Son1-Son2 1ng b (IMPP)	33%	32%	35%	3.53E+16	1.05E+19	1.22E+18	0.00E+00

Table 10. HPD LRs for three-person mixtures of parent/parent/offspring and parent/offspring/offspring with additional non-contributor relative testing results. *falsely included relative

A subset of the two and three-person relative mixture samples was analyzed under alternate conditioning hypotheses. The two-person relative samples were tested under the hypotheses listed in Table 11.

Hypothesis	H_p	H_d
H_p0	Dad + Unk	Unk1 +Unk2
H_p1	Dad + Son1	Dad + Unk
H_p2	Dad + Son1	Unk1 + Unk2

Table 11. Alternate hypotheses for 2-person relative mixtures. Unk=unknown

Similar to alternate conditioning hypotheses for unrelated individuals, adding relevant information at interpretation increases the LR for H_p true. This is demonstrated in the simple parent/offspring two-person mixture as shown in Figure 9.

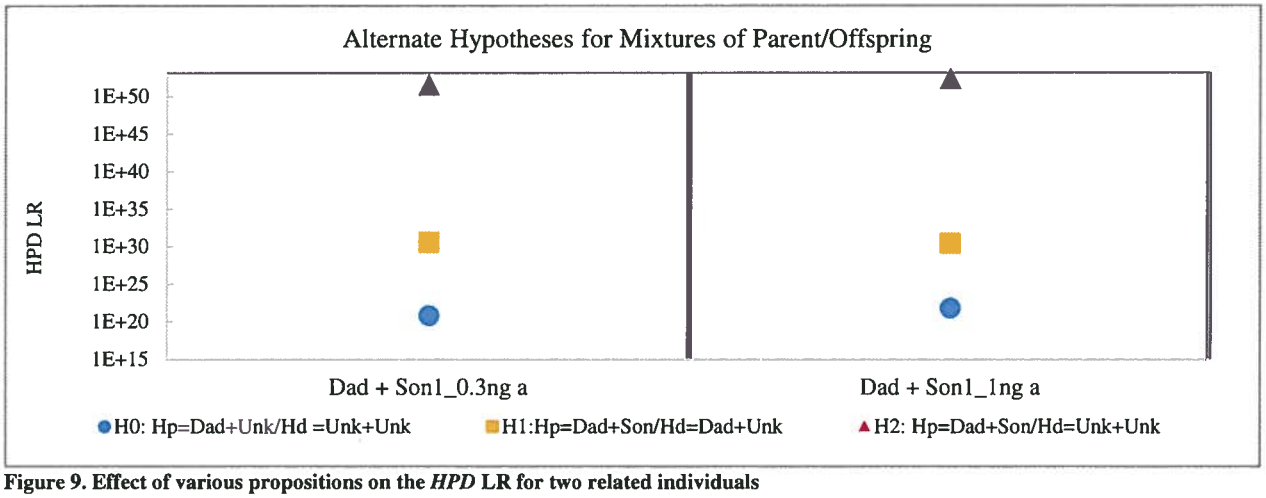


Figure 9. Effect of various propositions on the *HPD* LR for two related individuals

The three-person relative samples were tested under the hypotheses listed in Table 12.

Hypothesis	H_p	H_d
H_p0	Dad + Unk1 + Unk2	Unk1 +Unk2 + Unk3
H_p1	Dad + Mom + Son1	Dad + Mom + Unk
H_p2	Dad + Mom +Son1	Dad + Unk1 + Unk2
H_p3	Dad + Mom + Son1	Unk1 + Unk2 + Unk3

Table 12. Alternate hypotheses for 3-person relative mixtures. Unk=unknown

The results are summarized in Table 13. Similar to the two-person relative mixtures, adding relevant information at interpretation increases the LR for H_p true. A high allele variance was observed for the 0.3ng_a sample under H_p1, which suggests greater peak height imbalances. This was confirmed upon a qualitative review of the profile using the known references. The mixture proportions for this sample were not consistent with the expected proportions. The tested individual (offspring) under this hypothesis was modeled incorrectly as a trace contributor resulting in an uninformative LR. Given these results, the sample was reanalyzed using the IMPP. The mixture proportions, allele variance and HPD LR improved and were consistent with qualitative expectations. Similar results were observed with the 1ng_a sample under H_p1 and the allele variance was even higher due to greater peak height imbalances. The results improved when analyzing this sample using IMPP.

	STRmix™ Mixture proportions						
Sample ID	Dad	Son1	Mom	c^2 (Mode 2.106)	k^2 (Mode 4.629)	HPD LR	Hypothesis
Dad-Mom-Son1 0.3ng a	39%	32%	28%	2.141	9.609	2.25E+13	Hp0
Dad-Mom-Son1 0.3ng a	*53%*	*0%*	*47%*	*8.88*	*7.612*	*5.04E+01*	*Hp1*
Dad-Mom-Son1 0.3ng a (IMPP)	35%	29%	36%	7.21	8.94	6.65E+12	Hp1-IMPP
Dad-Mom-Son1 0.3ng a	41%	31%	27%	1.952	7.819	7.47E+50	Hp2
Dad-Mom-Son1 0.3ng a	43%	32%	25%	2.175	6.803	1.76E+61	Hp3
Dad-Mom-Son1 1ng a	37%	33%	30%	2.355	6.77	1.77E+12	Hp0
Dad-Mom-Son1 1ng a	*51%*	*4%*	*45%*	*20.013*	*5.24*	*1.64E+13*	*Hp1*
Dad-Mom-Son1 1ng a (IMPP)	37%	33%	30%	2.757	6.517	4.22E+29	Hp1 IMPP
Dad-Mom-Son1 1ng a	37%	33%	30%	2.771	6.479	1.09E+52	Hp2
Dad-Mom-Son1 1ng a	45%	33%	22%	2.523	5.995	1.93E+61	Hp3

Table 13. Summary of the alternate hypotheses for three-person relative mixtures

The three-person samples of parent/offspring/offspring and parent/parent/offspring appeared to be mixtures of only two individuals by maximum allele count (MAC), and therefore were also evaluated under an N-1 assumption. The HPD LRs for the known contributors were calculated under the following propositions:

- H_p: The DNA originated from the person of interest and N-2 unknown individuals
- H_d: The DNA originated from N-1 unknown individuals

The N-1 assumption produced varied results for the parent/offspring/offspring samples as shown in Table 14. In one instance (0.3ng_a), the HPD LR values for all 3 known contributors were similar under both N and N-1 assumptions. However, in the second 0.3 ng_b sample, the parent was falsely excluded and one of the offspring returned an uninformative LR. The second offspring's (son2) LR remained largely unchanged in all four samples (both 0.3ng and 1ng targets) under both assumptions. The N-1 assumption yielded HPD LR = 0 for the parent and one of the offspring for both the 1ng targets. The mother was falsely included in one of the four parent/offspring/offspring mixtures under the N-1 assumption. For the parent/parent/offspring both parents were either excluded or their LR values were greatly reduced as shown in Table 15. The known offspring was included in all samples. The second offspring (non-contributor) was falsely included in one sample (0.3ng_b).

		STRmix™ Mixture proportions			HPD LR			
Sample ID	# of contributors analyzed	Dad	Son1	Son2	Dad	Son1	Son2	Mom
Dad-Son1-Son2 0.3ng a	N	42%	35%	22%	3.37E+14	2.66E+15	2.55E+20	4.56E+06
Dad-Son1-Son2 0.3ng a	N-1	59%*	41%*	N/A	2.13E+14	1.57E+16	1.27E+22	2.39E+05
Dad-Son1-Son2 0.3ng b	N	33%	42%	26%	1.61E+14	6.34E+14	4.23E+19	1.94E+03
Dad-Son1-Son2 0.3ng b	N-1	70%*	70%*	N/A	**0.00E+00**	5.31E-01	9.12E+17	**0.00E+00**
Dad-Son1-Son2 1ng a	N	33%	43%	24%	6.78E+15	3.50E+14	1.06E+21	8.51E+02
Dad-Son1-Son2 1ng a	N-1	66%*	34%*	N/A	**0.00E+00**	**0.00E+00**	4.84E+17	**0.00E+00**
Dad-Son1-Son2 1ng b (IMPP)	N	33%	32%	35%	3.53E+16	1.05E+19	1.22E+18	**0.00E+00**
Dad-Son1-Son2 1ng b	N-1	66%	34%	N/A	**0.00E+00**	**0.00E+00**	4.15E+17	**0.00E+00**

Table 14. Summary of N-1 assumption for parent/offspring/offspring

		STRmix™ Mixture proportions			HPD LR			
Sample ID	# of contributors analyzed	Dad	Mom	Son1	Dad	Mom	Son1	Son2
Dad-Mom-Son1 0.3ng a	N	39%	32%	28%	2.25E+13	8.25E+14	1.09E+19	7.89E+13
Dad-Mom-Son1 0.3ng a	N-1	36%	64%	N/A	**0.00E+00**	**0.00E+00**	2.57E+29	**0.00E+00**
Dad-Mom-Son1 0.3ng b	N	35%	22%	43%	1.98E+12	4.49E+22	1.50E+13	4.99E+15
Dad-Mom-Son1 0.3ng b	N-1	39%	61%	N/A	1.68E+05	8.06E+06	6.43E+27	3.84E+10
Dad-Mom-Son1 1ng a	N	37%	33%	30%	1.78E+12	6.24E+14	3.26E+19	3.71E+12
Dad-Mom-Son1 1ng a	N-1	32%	68%	N/A	**0.00E+00**	**0.00E+00**	4.27E+29	**0.00E+00**
Dad-Mom-Son1 1ng b	N	40%	33%	27%	2.75E+12	4.06E+13	7.31E+21	6.88E+15
Dad-Mom-Son1 1ng b	N-1	34%	66%	N/A	**0.00E+00**	**0.00E+00**	5.31E+29	**0.00E+00**

Table 15. Summary of N-1 assumption for parent/parent/offspring

In casework, if it is suspected that a mixture may be comprised of first-order relatives then assessing the number of contributors will require special consideration, especially in mixtures of three or more. It is most common to underestimate the number of contributors due to the increased amount of allele sharing and maximum allele count may not be the best guide. When the number of contributors is underestimated, the mixture proportions are no longer intuitive. The weighted genotypes may no longer be concordant with the true contributors resulting in a false exclusion and/or inclusion. In order to help determine the number of contributors analysts should inspect the profile for peak height imbalances and consider an assumed known reference if available. The software's diagnostics (i.e. variance values and mixture proportions) may indicate if the assumed number of contributors needs to be revised. Analysts should use caution when interpreting mixtures believed to be comprised of first-order relatives.

Section G: Drop In (QAS 8.3.1 & SWGDAM 4.1.8)

This section covers the following standard:

4.1.8 Allele drop-in

The drop-in parameters were established during the Part I implementation segment of the validation. The final parameters applied to STRmix™ are summarized in Table 1.

Parameter	Values used in STRmix™
Drop-in frequency	0.0001
Drop-in cap	100 RFU
Drop-in α and β	0,0

Table 1. Drop-in settings

To test these settings, three experiments were performed using a single source profile (PQ207) that had a drop-in peak artificially added to one heterozygous locus. The heterozygous locus, D5S818 (genotype 10,11), was used for each experiment. The drop-in peak was intentionally added to a non-stutter position (allele 13) in order to avoid any affects from stutter modeling. The template amount of the single source profile and/or the height of the drop-in peak (i.e. above or below the 100 RFU cap) were varied for each experiment. The modified profile, along with the associated unmodified profile, were interpreted with STRmix™ assuming one contributor using the following propositions:

- H_p = The DNA originated from the person of interest (POI)
- H_d = The DNA originated from an unknown individual

The aim of the first experiment was to check that STRmix™ would properly model the additional peak as drop-in if it were below the defined cap of 100 RFU and unable to pair with the remaining alleles at the locus due to extreme peak height imbalance. A drop-in peak of 99 RFU was added to a single source profile (PQ207_0.500ng) with optimal peak heights ranging from approximately 700 to 2100 RFU. The peak heights at D5S818 for alleles 10 and 11 were 883 and 876 RFU, respectively. STRmix™ appropriately modeled the additional peak as drop-in and returned a single genotype at the locus with a weight = 1. The resulting D5S818 locus LR was identical to the locus LR for the unmodified single source profile.

The goal of the second experiment was to confirm that STRmix™ would model the additional peak as both drop-in and an allele when its peak height was both less than the cap and of similar height to the remaining alleles such that they could pair. A drop-in peak of 99 RFU was added to D5S818 in a single source profile (PQ207_0.050ng) with peak heights ranging from drop out to less than 250 RFU. The heterozygous alleles at D5S818 were also changed to 99 RFU so that all three alleles were suitable for pairing. STRmix™ appropriately modeled the additional peak as both drop-in and as a true allele. This weighted the true genotype slightly less than 1 (weight = 0.997633) since more than one genotype was now considered. Consequently, the D5S818 locus LR for the modified profile was marginally less than the locus LR for the unmodified profile that only had a single genotype with a weight = 1.

The purpose of the third experiment was to verify that STRmix™ would not model the additional peak as drop-in if it was above the 100 RFU cap and the interpretation assumed one contributor. With the presence of an additional allele above 100 RFU that is not in a stutter position, STRmix™ should view the profile as a mixture of two contributors, and thus, be unable to model the data when assuming one contributor. A drop-in peak of 101 RFU was added to the single source profile (PQ207_0.500ng) described in experiment 1. As expected, an error occurred and the interpretation was halted as the

profile could no longer be explained by one contributor. Figure 1 shows the error window generated by the software.

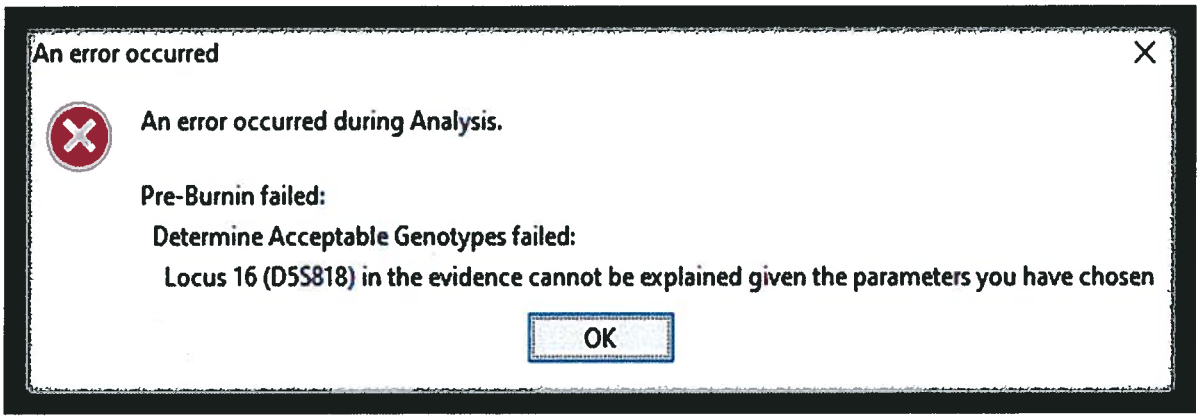


Figure 1. Error message for an input profile that cannot be explained by the assumed number of contributors

In all three experiments, STRmix™ correctly modeled each modified profile given the height of the added drop-in peak, the height of the remaining profile alleles, and the assumption of the number of contributors used for the interpretation.

Section H: Stutter (SWGDAM 4.1.9)

This section covers the following standard:

4.1.9. Forward and reverse stutter

STRmix™ models both back stutter (N-1 repeat) and forward stutter (N+1 repeat). The stutter models (i.e. allele, LUS, multi-sequence, average) applied for each locus by the laboratory were described in the Part I implementation segment of the validation. Furthermore, the stutter variance (k^2) was determined during the Model Maker analysis as described in Part I. The modeling of stutter peaks can be seen in the interpretation of single source profiles where stutter peaks are retained at interpretation. As part of the MCMC burn-in process they are considered as alleles in the genotype but those combinations are not accepted. Therefore, they receive no weight and are not listed in the component interpretation output of the STRmix™ report. In mixed DNA profiles, where the minor contributor is of a similar height as the stutter peaks, they start to be considered as minor alleles. To verify the established stutter values and stutter variance are performing as described above, two single source profiles and one mixture profile interpreted in STRmix™ were evaluated for the known N-1 and N+1 stutter peaks.

The two single source samples used were selected to mimic a best case scenario profile (PQ207_1000pg) and a more challenging profile (PAC2). The PQ207_1000pg profile has peak heights in the optimal RFU range (i.e. ~1000 – 3000 RFU) with no elevated stutter ratios when compared to the laboratory's current stutter filters applied during a binary interpretation. The PAC2 profile has high RFU peaks with some above the saturation threshold of 30,000 RFU. It is known that when a peak reaches the saturation level of the camera, the resulting stutter ratio may be artificially enhanced since there is no longer a linear response in signal with higher amounts of template. Both profiles were interpreted in STRmix™ assuming one contributor. The software correctly modeled all known N-1 and N+1 stutter peaks as stutter. This is exhibited by the return of the correct genotype with a weight = 1.0 for all loci as shown in Figures 1 and 2.

COMPONENT INTERPRETATION

CONTRIBUTOR 1 (100%)

Questioned contributor

LOCUS	GENOTYPE	WEIGHT
D3S1358	15, 17	100.00%
D1S1656	12, 12	100.00%
D2S441	11, 11	100.00%
D10S1248	13, 16	100.00%
D13S317	8, 9	100.00%
Penta E	17, 18	100.00%
D16S539	9, 13	100.00%
D18S51	14, 17	100.00%
D2S1338	17, 25	100.00%
CSF1PO	8, 12	100.00%
Penta D	9, 10	100.00%
TH01	8, 9	100.00%
vWA	15, 17	100.00%
D21S11	27, 29	100.00%
D7S820	10, 13	100.00%
D5S818	10, 11	100.00%
TPOX	8, 11	100.00%
D8S1179	10, 10	100.00%
D12S391	17, 21	100.00%
D19S433	14, 14	100.00%
SE33	30.2, 30.2	100.00%
D22S1045	15, 15	100.00%
DYS391		
FGA	24, 24	100.00%
DYS576		
DYS570		

Figure 1. STRmix™ component interpretation for PQ207_1000pg

COMPONENT INTERPRETATION

CONTRIBUTOR 1 (100%)

Questioned contributor

LOCUS	GENOTYPE	WEIGHT
D3S1358	17, 18	100.00%
D1S1656	12, 13	100.00%
D2S441	10, 14	100.00%
D10S1248	13, 15	100.00%
D13S317	9, 11	100.00%
Penta E	7, 14	100.00%
D16S539	9, 13	100.00%
D18S51	16, 18	100.00%
D2S1338	22, 25	100.00%
CSF1PO	12, 12	100.00%
Penta D	12, 13	100.00%
TH01	6, 9.3	100.00%
vWA	16, 19	100.00%
D21S11	29, 31.2	100.00%
D7S820	8, 11	100.00%
D5S818	12, 12	100.00%
TPOX	11, 11	100.00%
D8S1179	14, 15	100.00%
D12S391	18, 23	100.00%
D19S433	13, 14	100.00%
SE33	15, 16	100.00%
D22S1045	16, 16	100.00%
DYS391		
FGA	20, 23	100.00%
DYS576		
DYS570		

Figure 2. STRmix™ component interpretation for PAC2

The stutter variance values for both samples were also evaluated. The PQ207_1000pg profile yielded a stutter variance of 3.142, which is near the mode (4.629) derived from the prior distribution during the Model Maker analysis. This was expected given the optimal RFU peak heights of the profile and the lack of any elevated stutter peaks. The stutter variance for the PAC2 profile was 56.404, which is substantially greater than the mode. The elevated stutter variance observed in this profile was not unexpected. As discussed in Section B of this validation, STRmix™ calculates the expected height of the stutter peak using the proposed expected allele height and not the observed height when the parent peak is above the saturation threshold. This may result in greater disparity between the observed and expected stutter peaks, and thus, require a higher stutter variance value in order to model the data.

A two person mixture (MF_1ng_1:3a) profile was also assessed to ensure stutter was modeling appropriately within a mixture. For this profile, all loci except for SE33 and D22S1045 had stutter peak heights that were less than 50% of the peak heights of the minor contributor alleles such that pairing would not be supported using the laboratory's current binary interpretation method. The peak heights for the alleles and stutter peaks are shown in the electropherogram in Figure 3. With the exception of the 17 stutter peak at D22S1045, STRmix™ modeled the known N-1 and N+1 stutter peaks as stutter at each locus. This is demonstrated in the minor contributor interpretation produced by STRmix™ and given in Figure 4. SE33 resulted in a single genotype for the minor component with a weight = 1 (or 100%). D22S1045 had multiple genotypes listed; one of which contained the 17 stutter peak with a weighting = 0.38%. These results make sense given the assumption of two contributors and the number of alleles detected at each locus. SE33 was comprised of two heterozygous donors (16,29.2 and 22,31.2) with no allele sharing, resulting in four discreet alleles. Based on their peak heights, along with the heights of the stutter peaks, it was more intuitive that the two alleles from the minor contributor would pair together rather than with a lower RFU stutter peak. The peak height ratio between the lowest RFU allele (22) and highest RFU stutter peak (22.2) was 50.6%, and therefore any other pairing between an allele and stutter peak would be even less. The peak height ratio between the two lowest RFU alleles (22 and 31.2) was more favorable at 87.6%. D22S1045 was also comprised of two heterozygous donors (15,18 and 15,16), however, the 15 allele was shared resulting in three discreet alleles. The peak height ratio between the lowest RFU allele (16) and highest RFU stutter peak (17) was 56.7%. STRmix™ appropriately assigned a low weighting to the 16,17 genotype given the height of the 17 stutter peak relative to the 16 allele and the lack of a fourth discreet allele available for pairing.

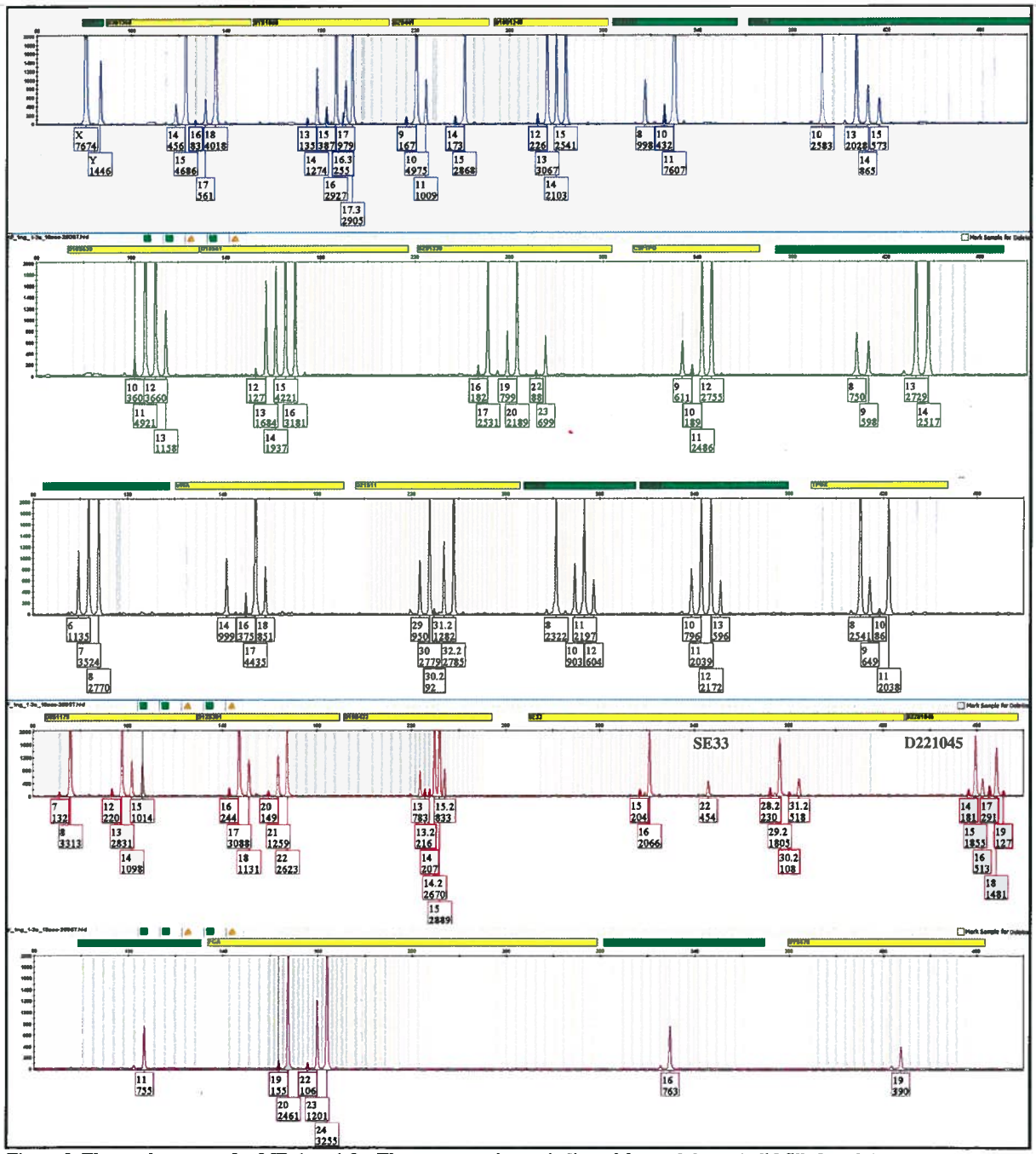


Figure 3. Electropherogram for MF_1ng_1:3a. The stutter peaks are indicated for each locus (solid filled peaks).

Contributor 2 (29.00%)

Locus	Genotype	Weighting	Genotype >=99%
D3S1358	15, 18	100.00 %	15,18
D1S1656	14, 17	100.00 %	14,17
D2S441	10, 11	100.00 %	10,11
D10S1248	14, 14	100.00 %	14,14
D13S317	8, 11	100.00 %	8,11
Penta E	14, 15	100.00 %	14,15
D16S539	11, 13 12, 13	99.999 % 00.001 %	11,13
D18S51	13, 14	100.00 %	13,14
D2S1338	19, 23	100.00 %	19,23
CSF1PO	9, 12 9, 11	99.65 % 00.35 %	9,12
Penta D	8, 9	100.00 %	8,9
TH01	6, 7	100.00 %	6,7
vWA	14, 18	100.00 %	14,18
D21S11	29, 31.2	100.00 %	29,31.2
D7S820	10, 12	100.00 %	10,12
D5S818	10, 13	100.00 %	10,13
TPOX	8, 9 9, 11 9, 9	99.78 % 00.17 % 00.05 %	8,9
D8S1179	14, 15	100.00 %	14,15
D12S391	18, 21	100.00 %	18,21
D19S433	13, 15.2	100.00 %	13,15.2
SE33	22, 31.2	100.00 %	22,31.2
D22S1045	15, 16 16, 18 16, 16 16, 17	94.11 % 03.56 % 01.96 % 00.38 %	16,0
DYS391			0,0
FGA	23, 24	100.00 %	23,24
DYS576			0,0
DYS570			0,0

Figure 4. Minor contributor STRmix™ component interpretation for MF_1ng_1:3a. The red box highlights the genotype at D22S1045 that includes the 17 stutter peak.

These results demonstrate that STRmix™ appropriately models stutter peaks in single source and mixture profiles. STRmix™ does not currently model N-½, N+½ and N-2 repeat stutter. These stutter peaks need to be removed prior to inputting the profile into STRmix™. Failure to do so could result in either a reduced LR or an exclusion at that particular locus.

Section I: Intra-Locus Peak Height *(SWGDAM 4.1.10)*

This section covers the following standard:

4.1.10. Intra-locus peak height variance

STRmix™ models the variability of single peaks. The variance of this model is determined by directly modeling laboratory data within STRmix™ using the Model Maker function described in Part I of the validation. Traditionally, the expected heterozygote balance (Hb) between a pair of peaks rather than single peak height variability is established for the purposes of deconvolution. The peak height variability obtained from STRmix™ from the MCMC process can be used to estimate the level of expected Hb in a dataset. The performance of Model Maker is checked by plotting the log(Hb) versus average peak height (APH) and adding the expected 95% bounds informed by the Model Maker results.

The expected 95% bounds are calculated by $\pm\sqrt{2}\,X\,\sqrt{1.96}\,X\sqrt{\frac{C^2}{APH}}$

where c^2 is the allele variance from the gamma distribution determined from Model Maker. The c^2 mode from the distribution is 2.105. When using the 50th percentile c^2 value from the gamma distribution (c^2 = 2.84), the 95% bounds encapsulate sufficient data (coverage is 96.7% in v. 2.5.11) demonstrating that the values for variance are sufficiently optimized. The reported variance value was monitored for the 107 samples analyzed in Section D. The values did not veer far from the mode and ranged from 1.1 to 4.8. This further supports appropriate peak height modeling. As discussed in Section B, the allele variance increased in saturated profiles and those with extreme peak height imbalance.

Section J: Inter-Locus Peak Heights *(SWGDAM 4.1.11, 4.1.7, 4.1.7.3)*

This section covers the following standard:

4.1.11. Inter-locus peak height variance

4.1.7. Partial profiles, to include the following:

4.1.7.3. Inhibition

Inter-locus peak variance is modeled in STRmix™ using locus specific amplification efficiencies (LSAE). The variance of this model was determined during the Model Maker analysis as described in Part I of this validation. The LSAE values should mimic the average peaks heights (APH) of the locus if degradation is minimal, otherwise a trend across loci of similar molecular weight should be seen[1]. In order to assess if the LSAE values appropriately reflect the observed profile, three single source profiles from laboratory donors were used. One profile (PQ207) was a complete robust profile while the other two profiles were either inhibited (PQ315) or degraded (PQ86). PQ315 was a hematin (0.75mM) inhibited profile created in the laboratory. PQ86 was a severely degraded profile from a 1999 dried bloodstain that showed the typical ski slope effect on peak heights from the smaller to larger loci. The LSAE values were plotted against locus specific APH per dye color. The results for the robust single source profile given in Figure 1 and the inhibited profile given in Figure 2 show that the LSAE values are consistent with the APH observed at each locus throughout the profile. As shown in Figure 3, the relationship between LSAE and APH is not as apparent when plotted by dye color for the degraded sample. However, when the loci of the degraded sample are plotted according to molecular weight (smallest to largest), as given in Figure 4, the LSAE values better mimic APH at each locus.

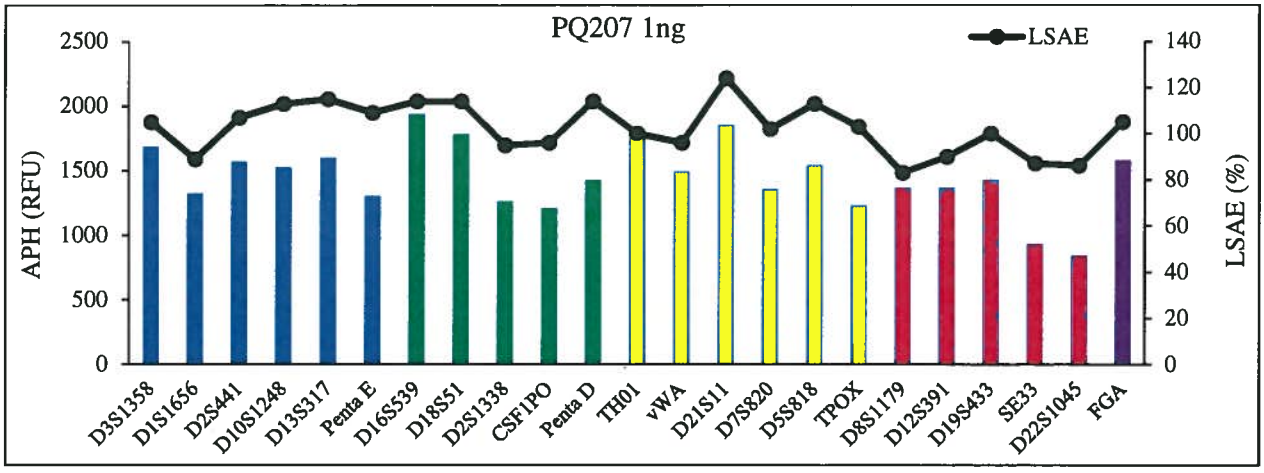


Figure 1. Plot of APH and LSAE for each locus for a robust single source profile

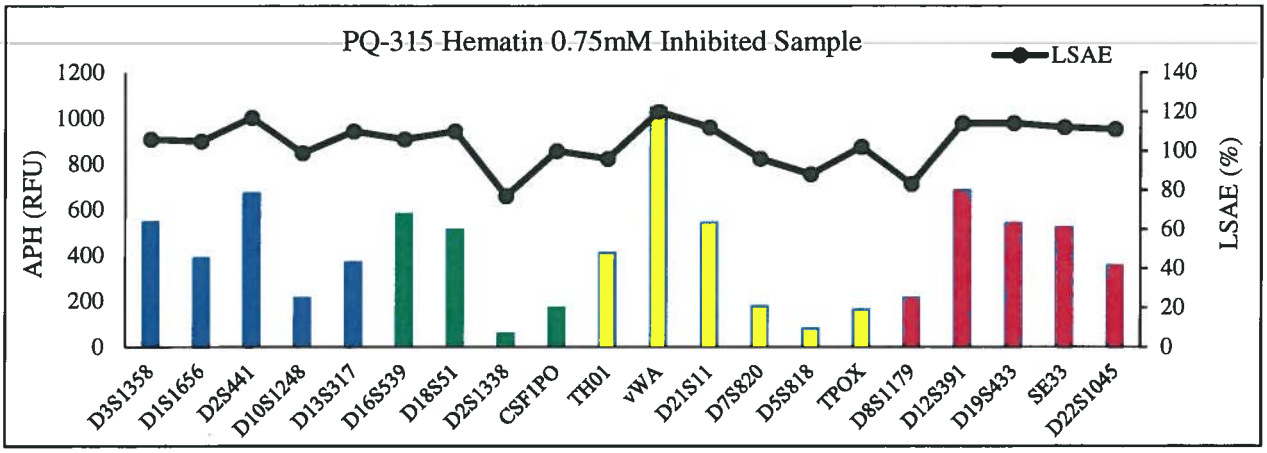


Figure 2. Plot of APH and LSAE for each locus for an inhibited single source profile. Penta E, Penta D and FGA are not shown due to complete inhibition.

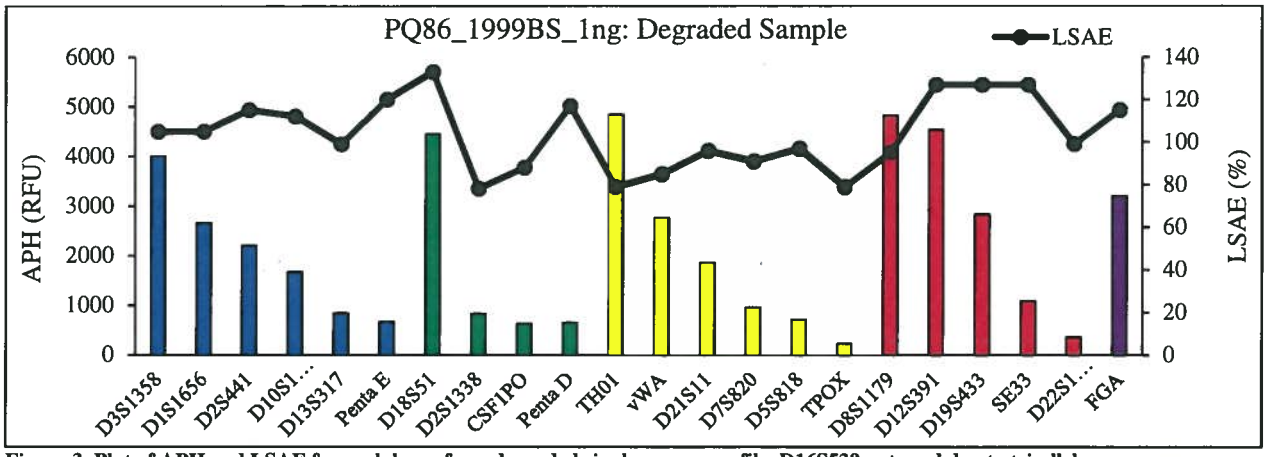


Figure 3. Plot of APH and LSAE for each locus for a degraded single source profile. D16S539 not used due to tri-allele.

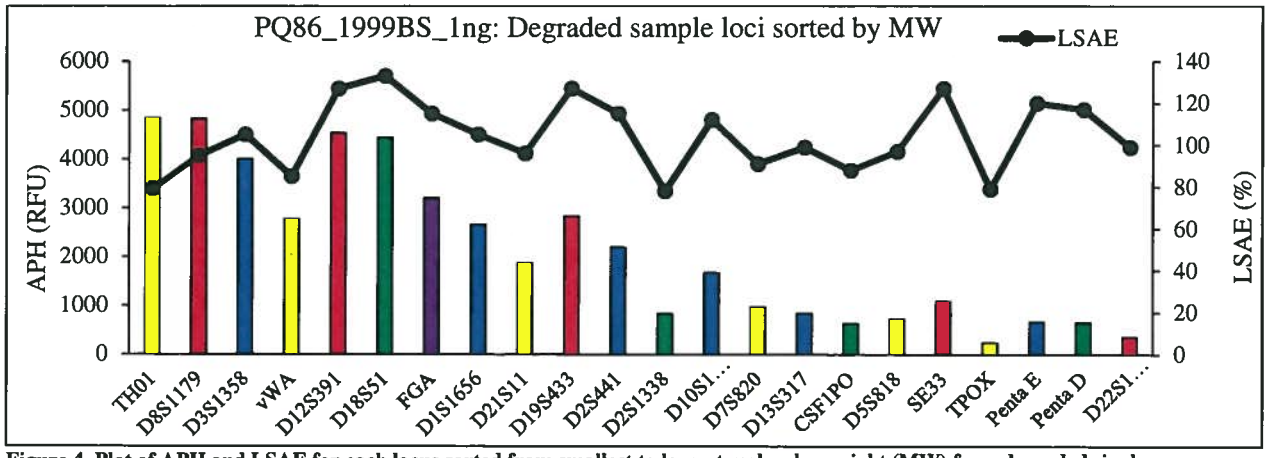


Figure 4. Plot of APH and LSAE for each locus sorted from smallest to largest molecular weight (MW) for a degraded single source profile. D16S539 not used due to tri-allele.

Reference

(1) Bright, J. et al. Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. FSI: Genetics 23 (2016).

Section K: Challenge Testing *(SWGDAM 4.1.14)*

This section covers the following standard:

4.1.14. Additional challenge testing (e.g., the inclusion of non-allelic peaks such as bleed through and spikes in the typing results)

STRmix™ requires that only numeric values are retained within the input file. Any value that is not numeric (such as alleles labeled as off-ladder) will cause STRmix™ to halt the interpretation. Furthermore, the presence of a non-allelic peak called as an allele (such as a labeled pull-up or N - ½ stutter peak) that is retained within the input file or an unresolved allele with only one base pair of separation can cause problems with interpretation. These problems include false exclusionary LR values, sub-optimal run diagnostics, or a failure to interpret. To evaluate the effects of inputting a profile that includes a peak that should have been deleted or edited, eight profiles that contain off-ladder alleles, non-allelic peaks, or an unresolved peak were analyzed with STRmix™. The HPD LR values and run diagnostics were compared to the same profile properly edited before input.

The software completed analysis for five of the profiles while the other three generated error messages after the interpretation was halted. Samples for which analysis could proceed either had a non-allelic peak less than 100 RFU that was labeled as an allele or an unresolved peak (as shown in Figure 1). In all instances the non-allelic peak was modeled as drop-in. A summary of the results for these samples analyzed both edited and unedited is shown in Table 1. The mixture proportions and variances were largely unchanged in either analysis. The HPD LR changed at most by two orders of magnitude for the three of the four samples with a non-allelic peak modeled as drop in. Sample FMMM1:5:5:5_1ng_a experienced a greater change at five orders of magnitude. The sample with the unresolved peak resulted in a false exclusion. If non-allelic peaks are left labeled as alleles then the affected locus may return an LR = 0. Profiles with aberrant results at a single locus or poor diagnostics overall should be reviewed for unresolved peaks, possible artifacts, or un-modeled stutter peaks.

Sample ID	c^2 no Artifact	c^2 w/ Artifact	k^2 no Artifact	k^2 w/ Artifact	HPD LR no Artifact for PQ183	HPD LR w/ Artifact for PQ183	Notes
FM_1ng_1:20a	1.9	2.568	8.5	10.703	1.69E+29	9.39E+28	(9) at D1S1656 is79 RFU: Raised Baseline
MF_1ng_1:5a	2.9	3.033	6.4	6.301	2.01E+28	1.84E+28	(18.3) at D12S391 is 79 RFU: Pull-up
MF_1ng_1:20a	2.6	3.559	6.4	6.432	1.35E+24	5.17E+22	(18.3) at D12S391 is 99 RFU: Pull-up
Sample ID	**c^2**	**c^2 w/ Artifact**	**k^2**	**k^2 w/ Artifact**	**HPD LR no Artifact for OPQ538**	**HPD LR w/ Artifact for OPQ538**	**Notes**
FMMM1:5:5:5_1ng a	2.50	2.8	5.10	4.3	2.70E+18	2.36E+13	(17.2) at SE33 is 78 RFU: n – ½ stutter
Sample ID	**c^2**	**c^2 w/ Artifact**	**k^2**	**k^2 w/ Artifact**	**HPD LR no Artifact for Son1**	**HPD LR w/ Artifact for Son1**	**Notes**
Dad-Son1-Son2 1ng b (IMPP)	4.067	3.957	5.677	5.473	1.05E+19	0.00E+00	9.3/10 not resolved, 9.3 not called at D7S820; ZI is 9.3, 10

Table 1. Challenge samples analyzed by STRmix™. IMPP = informed mixture proportion priors. c^2 mode = 2.106 and k^2 mode = 4.629

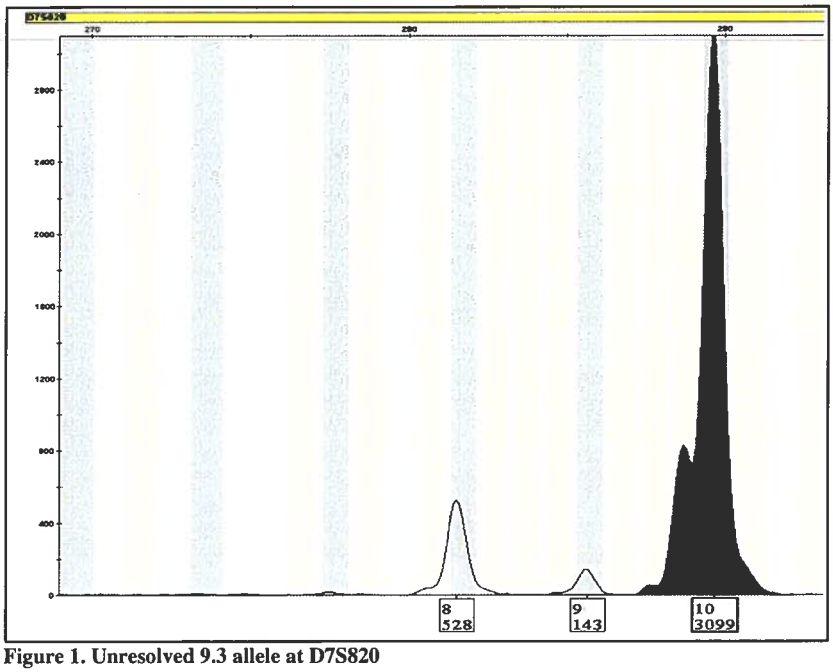


Figure 1. Unresolved 9.3 allele at D7S820

If a peak labeled as off-ladder ("OL") was in the input file, then the error message shown in Figure 2 was returned by the software. Two of the samples had a pull-up peak that was labeled as off-ladder.

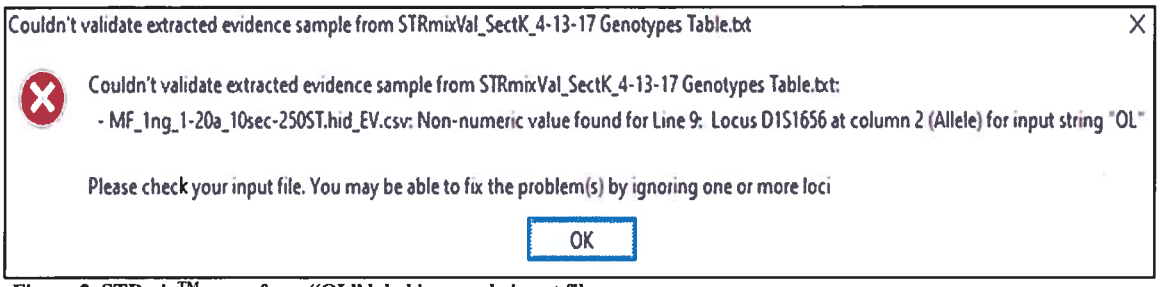


Figure 2. STRmix™ error from "OL" label in sample input file

The other sample that produced an error message was a laboratory donor profile with a tri-allelic pattern at D16S539 (see Figure 3). When the sample is analyzed with number of contributors equal to one, the profile cannot be modeled or explained at that locus. A solution for profiles with known tri-allelic patterns is to ignore the locus.

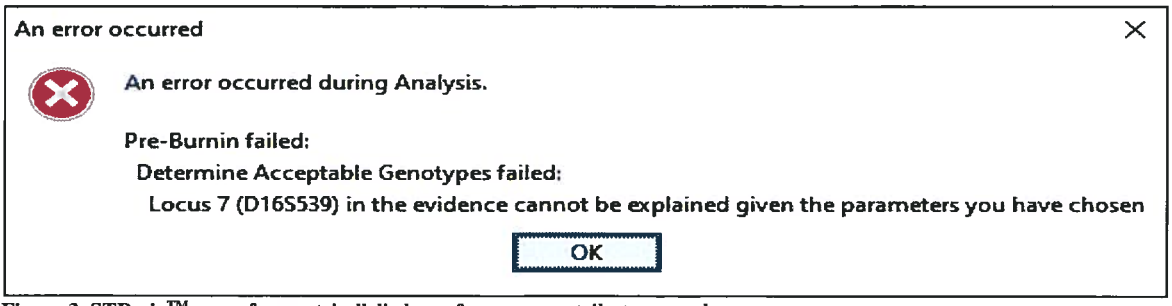


Figure 3. STRmix™ error from a tri-allelic locus for a one contributor sample

Section L: Known Mock or Non-Probative Casework Samples *(QAS 8.3.1 & SWGDAM 4.1.1, 4.2, 4.2.1, 4.2.1.1)*

This section covers the following standards:

4.1.1 Specimens with known contributors, as well as case-type specimens that many include unknown contributors.

4.2. Laboratories with existing interpretation procedures should compare the results of probabilistic genotyping and of manual interpretation of the same data, notwithstanding the fact that probabilistic genotyping is inherently different from and not directly comparable to binary interpretation. The weights of evidence that are generated by these two approaches are based on different assumptions, thresholds and formulae. However, such a comparison should be conducted and evaluated for general consistency.

> *4.2.1. The laboratory should determine whether the results produced by the probabilistic genotyping software are intuitive and consistent with expectations based on non-probabilistic mixture analysis methods.*
>
> > *4.2.1.1. Generally, known specimens that are included based on non-probabilistic analyses would be expected to also be included based on probabilistic genotyping.*

A comparison between the results obtained with STRmix™ and the laboratory's binary interpretation method was performed. One CTS proficiency test (17-5705) and three mock cases were analyzed using the binary interpretation protocol. The four cases consisted of a total of 11 mock evidence samples comprised of single source profiles and various two-, three- and four- person mixture profiles with optimal and low template amounts. Each case had two person of interest (POI) reference samples for comparison, which covered a range of scenarios where the POI was excluded, included and/or deemed inconclusive. Qualitative conclusions were made regarding each POI and the random match probability (RMP) was calculated for inclusions. These same cases were interpreted using STRmix™ under the following propositions:

- H_p: The DNA originated from the POI and N-1 unknown individuals
- H_d: The DNA originated from N unknown individuals

The HPD LR results produced by STRmix™ were compared to the binary derived conclusions and RMP values to determine whether the STRmix™ results were intuitive and if the results were consistent between the two methods. The results are summarized in Table 1.

In each case, when a contributor was included with the binary method an inclusionary HPD LR value was obtained with STRmix™. The RMP and HPD LR values varied in these samples as expected and were within approximately four orders of magnitude in all samples except for one. The RMP for the firearm sample in Case 3 was 7.40×10^{14}, while the STRmix™ HPD LRs were substantially higher for Suspect 1 and Suspect 2 with values equal to 1.56×10^{28} and 5.58×10^{28}, respectively. The binary interpretation of this sample was a distinct group of two contributors that could not be further deconvolved from each other. STRmix™ was able to correctly resolve Suspect 1 and 2 into two separate contributors at proportions of 26% and 63%, respectively, as shown in Figure 1. All but four loci in each contributor resulted in a genotype weight of 99% or greater.

Case 1 (CTS-Sexual Assault)		STRmix™ HPD LR		Binary Conclusions		Binary RMP	
Sample Name	Expected Donors	Suspect Ref (Item 2)	Victim Ref (Item 1)	Suspect Ref (Item 2)	Victim Ref (Item 1)	Suspect Ref (Item 2)	Victim Ref (Item 1)
Semen /BS from Victim's skirt EP	Item 2 & Unknown male	1.89E+30	0.00	Included to major	Excluded	5.90E+32	N/A
Semen /BS from Victim's skirt SP	Unknown male	0.00	0.00	Excluded	Excluded	N/A	N/A
BS from Suspect's shirt	Item 1	0.00	1.45E+35	Excluded	Included	N/A	3.30E+37
Case 2 (Mock Homicide)		**STRmix™ HPD LR**		**Binary Conclusions**		**Binary RMP**	
Sample Name	**Expected Donors**	**Suspect Ref (PQ183)**	**Victim Ref (PQ212)**	**Suspect Ref (PQ183)**	**Victim Reference (PQ212)**	**Suspect Reference (PQ183)**	**Victim Reference (PQ212)**
Blade	PQ212 PQ183	1.06E+12	2.68E+29	Included to minor	Included to major	6.20E+08	3.10E+31
Handle	PQ212 PQ183	2.29E+29	6.70E+28	Included to major	Included to minor	5.20E+30	9.80E+26
Case 3 (Mock Robbery)		**STRmix™ HPD LR**		**Binary Conclusions**		**Binary RMP**	
Sample Name	**Expected Donors**	**Suspect 1 Ref (PQ243)**	**Suspect 2 Ref (PQ94)**	**Suspect 1 Ref (PQ243)**	**Suspect 2 Ref (PQ94)**	**Suspect 1 Ref (PQ243)**	**Suspect 2 Ref (PQ94)**
Phone	PQ212 PQ183 PQ243	6.42E+11	9.67E-07	Both excluded from DG =1, but minor uninterpretable due to limited data		N/A	N/A
Glove	PQ212 PQ183 PQ243	7.74E+02	2.96E+00	Both excluded from major, but minor uninterpretable due to limited data		N/A	N/A
Firearm	PQ212 PQ183 PQ243 PQ094	1.56E+28	5.58E+28	Included to DG=2	Included to DG=2	7.40E+14	
Hat	PQ212 PQ183 PQ243 PQ094	1.99E+09	2.96E+29	Minor uninterp. too complex	Included to major	N/A	2.50E+32
Case 4 (Mock Homicide		**STRmix™ HPD LR**		**Binary Conclusions**		**Binary RMP**	
Sample Name	**Expected Donors**	**Suspect Ref (PQ212)**	**Victim Ref (PQ183)**	**Suspect Ref (PQ212)**	**Victim Ref (PQ183)**	**Suspect Ref (PQ212)**	**Victim Ref (PQ183)**
Door Knob	PQ212 PQ183 PQ243 PQ094	1.48E+08	8.92E+08	Both excluded from major, but minor uninterpretable due to complexity		N/A	N/A
Pipe	PQ212 PQ183	2.99E+29	2.43E+29	Included to minor	Included to major	3.50E+26	4.20E+29

Table 1: Comparison of STRmix™ HPD LR results and binary interpretation results. Blue font = inclusions with both methods. Orange font = exclusions with both methods.

SUMMARY >=99%

Locus	Contributor 1 (63.00%)	Contributor 2 (26.00%)	Contributor 3 (7.00%)	Contributor 4 (5.00%)
D3S1358	15.18	15.16	0.0	0.0
D1S1656	12.17.3	16.17.3	0.0	0.0
D2S441	9.1.11.3	10.14	0.0	0.0
D10S1248	12.13	14.15	0.0	0.0
D13S317	8.12	8.12	11.0	11.0
Penta E	18.18	18.19	0.0	0.0
D16S539	11.0	0.0	0.0	0.0
D18S51	15.16	12.14	0.0	0.0
D2S1338	19.23	19.25	0.0	0.0
CSF1PO	9.11	10.11	0.0	0.0
Penta D	9.13	9.13	0.0	0.0
TH01	9.0	0.0	0.0	0.0
vWA	17.18	14.18	0.0	0.0
D21S11	0.0	0.0	0.0	0.0
D7S820	8.11	9.11	0.0	0.0
D5S818	0.0	0.0	0.0	0.0
TPOX	8.11	8.11	0.0	0.0
D8S1179	11.12	10.11	0.0	0.0
D12S391	21.23	17.23	0.0	0.0
D19S433	14.15	14.15	0.0	0.0
SE33	27.2.28.2	18.25.2	0.0	0.0
D22S1045	16.0	11.0	0.0	0.0
DYS391	0.0	0.0	0.0	0.0
FGA	22.25	23.25	0.0	0.0
DYS576	0.0	0.0	0.0	0.0
DYS570	0.0	0.0	0.0	0.0

Figure 1: STRmix™ results for the firearm sample in Case 3. Contributors 1 and 2 are resolved and correctly correspond to the known contributors (Suspect 2 and Suspect 1, respectively).

For the single source and distinguishable mixtures with interpretable major and minor components, as demonstrated in Case 1, the binary method resulted in exclusions for non-contributors. STRmix™ yielded concordant results for these samples with an HPD LR = 0 for the non-contributor comparisons. The three and four person mixture samples had varying results between the two methods. The major component or distinct group of contributors for the phone, glove and hat samples from Case 3 and the door knob sample from Case 4 were interpretable with the binary method; however, the minor component for each was deemed uninterpretable due to limited data or complexity. The references were excluded from the major or distinct group portion of these mixtures but no conclusions could be made regarding the minor component.

Unlike the binary method, STRmix™ is able to use all of the data detected above the analytical threshold in an electropherogram for interpretation. For the phone, hat and door knob samples, STRmix™ returned an inclusionary HPD LR value for all true contributors, even when the true contributor was associated with the minor component. For example, Suspect 1 (true minor contributor) in Case 3 had an HPD LR of 6.42×10^{11} and 1.99×10^{9} for the phone and hat samples, respectively, while no conclusions could be made about the minor component with the binary interpretation method. STRmix™ was able to exclude Suspect 2 (HPD LR = 6.42×10^{-07}), a non-contributor, from the entire mixture of the phone sample in Case 3, whereas the binary method was only able to exclude Suspect 2 as a major contributor since no conclusions were made about the minor portion of the mixture.

The glove sample for Case 3 resulted in an uninformative HPD LR value for both Suspect 1 (true minor contributor) and Suspect 2 (non-contributor). These results were intuitive given the qualitative assessment of the profile. The profile was a low level, three-person, distinguishable mixture with two minor contributors. Each minor contributor was comprised of approximately 25pg of DNA and many of the minor alleles resided in the stutter position. The results were consistent with the sensitivity studies discussed in Section D, which showed that as the input amount of DNA decreased and number of contributors increased the LR values trended toward the uninformative range.

In general, STRmix™ produced results that were consistent with expectations based on the laboratory's binary interpretation method. True contributors that were included using the binary method, resulted in inclusionary HPD LR values with STRmix™. Likewise, non-contributors that were excluded using the binary method, yielded exclusionary HPD LR values using STRmix™. Differences did occur between the two methods in the minor component interpretation for a few of the mixture samples. These differences can be attributed to the limitations of a binary interpretation method, which restricts what data is deemed suitable for interpretation given the assumed number of contributors. STRmix™ uses all of the available data in the profile for interpretation. It is not constrained by peak height ratios, stochastic thresholds, loci suitability, and stutter peaks, and therefore minimizes the wasting of evidentiary data within the profile.

Section M: Precision

This section covers the following standards:

4.1.13 Precision as described for Developmental Validation (3.2.3)

3.2.3 Precision- Studies should evaluate the variation in Likelihood Ratios calculated from repeated software analysis of the same input. This should be evaluated using various sample types (e.g. different number of contributors, mixture proportions and template quantities).

3.2.3.1 Some probabilistic genotyping approaches may not produce the same LR from repeat analyses. Where applicable, these studies should therefore demonstrate the range of LR values that can be expected from multiple analyses of the same data and are the basis for establishing an acceptable amount of variation in LRs.

3.2.3.2 Any parameter setting (e.g. increasing the iterations of MCMC) that can reduce variability should be evaluated.

The MCMC process is used to generate the weights within STRmix™ for different genotype combinations. This is a sampling procedure and therefore the weights will vary slightly between each run, resulting in variations in reported likelihood ratios. In order to evaluate the variation in the calculated likelihood ratios (point estimate LR and HPD LR values) between replicate interpretations, a subset of samples from Section D of the validation was analyzed a total of six times. The samples selected included a single source profile at optimal and low concentrations (PQ183 at 60pg and 1.0ng) as well as various two, three and four-contributor mixtures (MF1:40_1ng_a, MF1:1_0.2ng_a, MMF1:1:1_0.3ng_a, MMF1:1:20_1ng_b, FMMM1:1:1:1_0.3ng_a and FMMM1:2:3:4_0.3ng_a). A minor contributor associated with each mixture was selected for all comparisons. The LR was calculated using the following propositions:

- H_p: The DNA originated from the person of interest and *N-1* unknown individuals
- H_d: The DNA originated from *N* unknown individuals

A parameter within STRmix™ that affects run variability is the number of iterations used during the MCMC process. The default number of iterations, which is set to 100,000 burn-ins and 400,000 post burn-in (accepts), is suitable for many different types of profiles. However, for some complex mixtures, the default settings may not be enough to achieve convergence. Therefore, to give analysts the option to increase the iterations for complex mixtures, a three-contributor (MMF1:1:20_1ng_b) and a four-contributor (FMMM1:2:3:4_0.3ng_a) mixture from Section D were interpreted using 1,000,000 burn-in and 4,000,000 post burn-in accepts. Five replicate runs were performed using these increased settings.

The likelihood variability, mixture proportion variability as well as the diagnostics for each sample were evaluated and are summarized below.

Single Contributor Samples

Two single contributor samples at 60pg and 1.0ng targets were interpreted six times. Table 1 and Figure 1 below summarize the results for the six replicates for each sample.

PQ183	Avg. Log likelihood	GR	c^2 *(Mode 2.106)*	k^2 *(Mode 4.629)*	HPD LR *(PQ183)*	Point Estimate LR *(PQ183)*
0.060ng	17.65	1.01	1.900	8.900	2.25E+17	1.40E+18
	18.24	1.02	1.919	7.874	2.49E+17	1.67E+18
	18.29	1.03	2.071	9.759	2.74E+17	1.90E+18
	18.3	1.12	1.929	10.493	2.45E+17	1.70E+18
	17.94	1.08	2.137	10.923	3.01E+17	2.23E+18
	18.35	1.01	1.991	6.345	2.57E+17	1.61E+18
1.0ng	66.24	1.02	1.800	6.500	3.92E+29	1.05E+30
	66.42	1.01	1.851	6.044	1.72E+29	1.05E+30
	66.37	1.00	1.849	6.298	1.40E+29	1.05E+30
	66.35	1.03	1.840	6.699	1.75E+29	1.05E+30
	66.26	1.02	1.919	6.173	1.55E+29	1.05E+30
	66.27	1.02	2.007	6.213	1.50E+29	1.05E+30

Table 1. Gelman-Rubin (GR), variances and LR values from replicate runs of single source samples. Replicates correspond to a – f in Figure 1.

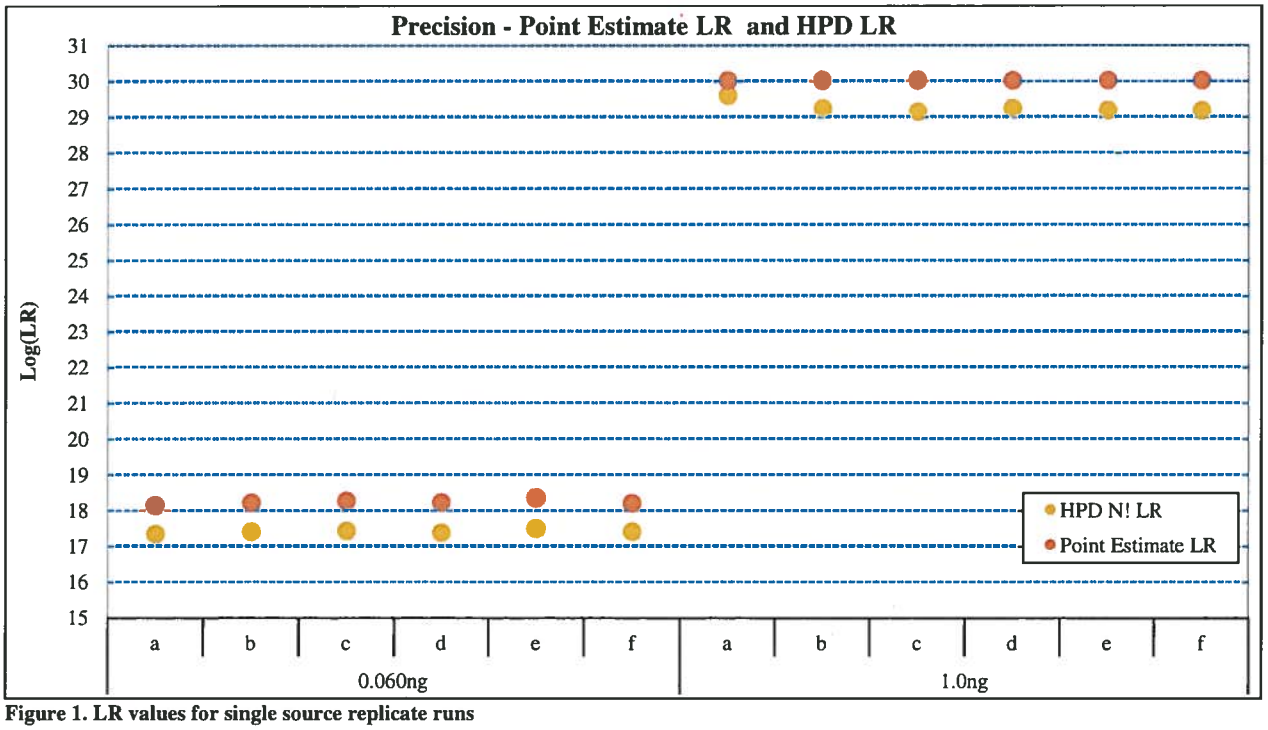


Figure 1. LR values for single source replicate runs

Two-Contributor Samples

Two two-contributor samples at 1.0ng and 200pg were interpreted six times. Table 2 and Figure 2 below summarize the results for the six replicates for each sample.

	STRmix™ Mixture Proportions							
Sample	**PQ183** ♂	**PQ212** ♀	**Avg. Log Likelihood**	**GR**	**c^2** *(Mode 2.106)*	**k^2** *(Mode 4.629)*	**HPD LR** *(PQ183)*	**Point Estimate LR** *(PQ183)*
MF_1:40_1ng_a	2%	98%	89.68	1.03	2.1	7.800	1.06E+12	8.15E+12
	2%	98%	89.78	1.01	2.077	7.813	1.73E+12	1.20E+13
	2%	98%	89.24	1.03	2.271	8.351	5.68E+11	3.73E+12
	2%	98%	90.36	1.04	2.279	7.377	1.01E+12	7.31E+12
	2%	98%	89.21	1.03	2.235	7.704	1.29E+12	7.96E+12
	2%	98%	89.64	1.07	2.136	8.003	1.14E+12	7.81E+12
MF1:1_0.2ng a	55%	45%	59.24	1.01	1.5	8.000	6.81E+15	6.95E+16
	55%	45%	59.46	1.01	1.714	6.618	3.30E+15	5.34E+16
	55%	45%	58.74	1.02	1.679	8.366	5.73E+15	6.74E+16
	55%	45%	60.24	1.02	1.645	6.309	3.47E+15	5.20E+16
	55%	45%	59.83	1.01	1.702	5.957	4.80E+15	6.24E+16
	55%	45%	58.71	1.01	1.69	5.826	3.83E+15	5.77E+16

Table 2. Gelman-Rubin (GR), variances and LR values from replicate runs of two-contributor samples. Replicates correspond to a – f in Figure 2.

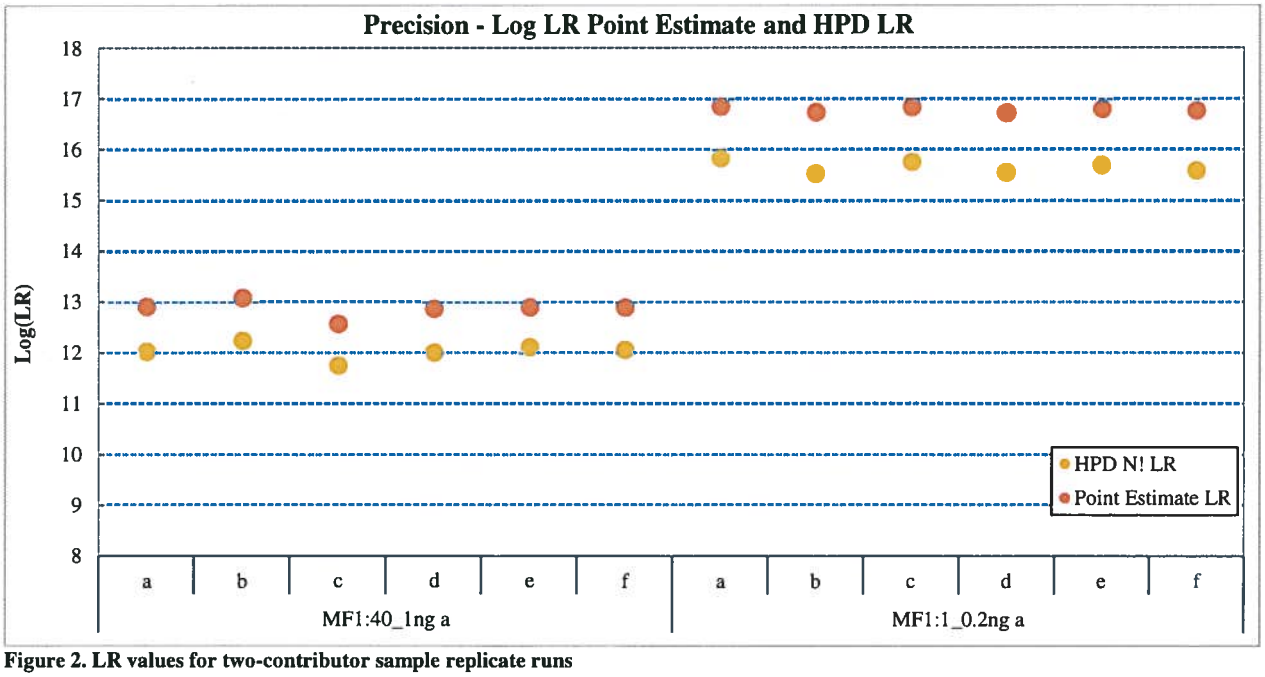


Figure 2. LR values for two-contributor sample replicate runs

Three-Contributor Samples

Two three-contributor samples at 300pg and 1.0ng were interpreted six times. Table 3 and Figure 3 below summarize the results for the six replicates for each sample.

	STRmix™ Mixture Proportions								
Sample	**PQ243** ♂	**PQ183** ♂	**PQ212** ♀	**Ave Log likelihood**	**GR**	**c^2** *(Mode 2.106)*	**k^2** *(Mode 4.629)*	**HPD LR** *(PQ183)*	**Point Estimate LR** *(PQ183)*
MMF1:1:1_0.3ng a	26%	40%	34%	68.50	1.01	1.700	6.700	1.46E+09	1.10E+10
	26%	41%	33%	70.19	1.01	1.730	6.287	1.07E+09	1.36E+10
	26%	40%	34%	69.53	1.01	1.753	5.821	9.71E+08	1.54E+10
	26%	40%	34%	68.95	1.02	1.792	5.117	8.44E+08	9.66E+09
	27%	40%	34%	69.03	1.03	1.798	5.674	1.03E+09	1.18E+10
	26%	40%	34%	69.05	1.01	1.778	5.562	5.89E+08	9.01E+09
MMF1:1:20_1ng b	3%	4%	93%	83.96	1.06	2.100	7.400	1.83E+05	1.53E+06
	3%	4%	93%	82.03	1.01	2.229	7.525	6.00E+05	6.38E+06
	3%	4%	93%	81.78	1.02	2.274	7.325	2.52E+05	3.02E+06
	3%	4%	93%	81.84	1.10	2.292	7.142	4.74E+05	5.02E+06
	3%	4%	93%	84.86	1.05	2.397	7.314	3.34E+05	2.92E+06
	3%	4%	93%	81.97	1.01	2.287	7.280	3.10E+05	3.26E+06
MMF1:1:20_1ng b (4M)*	3%	4%	93%	83.03	1.00	2.241	7.478	3.27E+05	3.67E+06
	3%	4%	93%	82.94	1.00	2.194	7.080	2.66E+05	2.81E+06
	3%	4%	93%	82.91	1.00	2.261	7.033	2.22E+05	2.85E+06
	3%	4%	93%	82.75	1.00	2.264	7.265	3.85E+05	4.17E+06
	3%	4%	93%	82.94	1.00	2.228	6.944	3.78E+05	3.61E+06

Table 3. Gelman-Rubin (GR), variances and LR values from replicate runs of three-contributor samples. *(4M) = analysis using 1,000,000 burn-in and 4,000,000 post burn-in accepts. Replicates correspond to a – f, or a – e for (4M), in Figure 3.

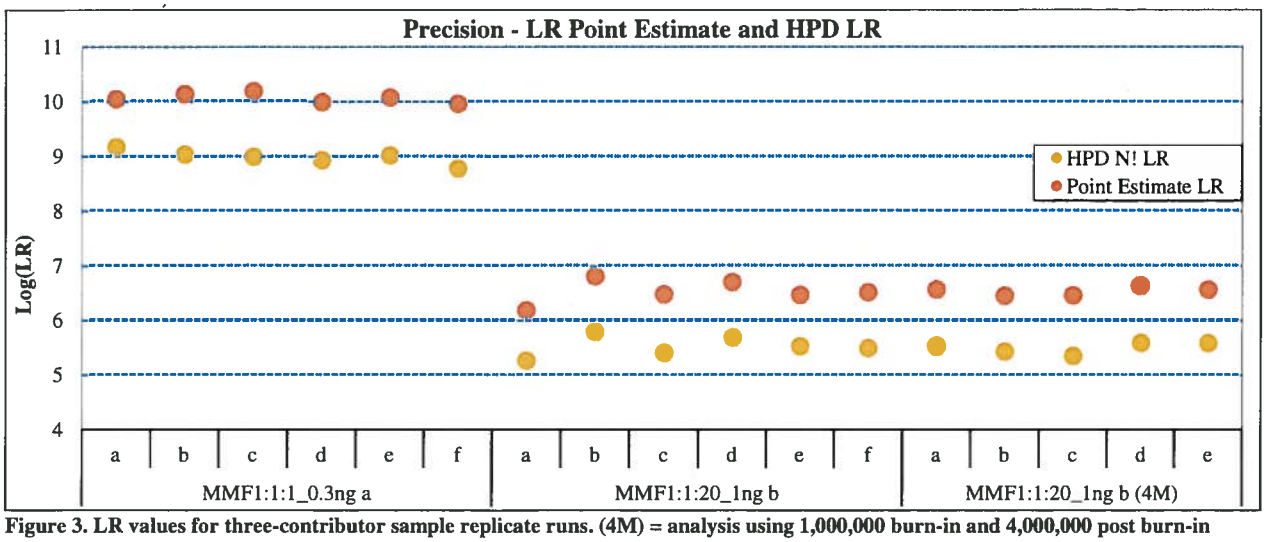


Figure 3. LR values for three-contributor sample replicate runs. (4M) = analysis using 1,000,000 burn-in and 4,000,000 post burn-in accepts.

Four-Contributor Samples

Two four-contributor samples each at 300pg were interpreted six times. Table 4 and Figure 4 below summarize the results for the six replicates for each sample.

	STRmix™ Mixture Proportions									
Sample	**PQ212 ♀**	**PQ183 ♂**	**PQ243 ♂**	**PQ94 ♂**	**Ave Log likelihood**	**GR**	**c^2** *(Mode 2.106)*	**k^2** *(Mode 4.629)*	**HPD LR** *(PQ183)*	**Point Estimate LR** *(PQ183)*
FMMM1:1:1:1_0.3ng a	26.0%	29.0%	24.0%	21.0%	81.02	1.00	1.300	5.400	7.76E+05	6.62E+06
	26.0%	29.0%	24.0%	21.0%	82.80	1.01	1.412	4.953	6.86E+05	1.12E+07
	26.0%	29.0%	24.0%	21.0%	79.60	1.01	1.411	5.114	4.12E+05	6.90E+06
	27.0%	29.0%	24.0%	21.0%	80.65	1.01	1.382	4.891	7.15E+05	9.56E+06
	26.0%	29.0%	24.0%	21.0%	81.49	1.05	1.403	5.491	8.51E+05	9.84E+06
	26.0%	29.0%	24.0%	21.0%	80.17	1.01	1.412	5.619	5.75E+05	1.09E+07
FMMM1:2:3:4_0.3ng a	15.0%	21.0%	28.0%	36.0%	53.07	1.08	2.070	10.700	1.13E-01	1.45E+00
	16%	21%	27%	37%	53.40	1.10	2.998	10.811	1.07E+00	7.73E+00
	15%	21%	27%	36%	49.06	1.21	3.613	12.088	1.30E-01	1.16E+00
	16%	21%	27%	36%	49.82	1.20	3.701	9.864	1.85E-01	1.49E+00
	16%	21%	27%	37%	53.33	1.24	3.607	9.892	2.13E-01	1.96E+00
	15%	22%	27%	36%	53.70	1.12	3.452	9.729	2.19E-01	1.89E+00
FMMM1:2:3:4_0.3ng a (4M)*	16%	21%	27%	37%	53.71	1.01	3.224	9.500	2.90E-01	2.09E+00
	16%	21%	27%	37%	54.58	1.00	3.032	8.793	2.53E-01	1.90E+00
	16%	21%	26%	37%	54.49	1.01	3.289	9.172	3.04E-01	2.28E+00
	16%	21%	26%	37%	54.26	1.02	3.169	8.923	3.03E-01	2.51E+00
	16%	21%	26%	37%	54.62	1.03	3.097	9.413	2.63E-01	1.99E+00

Table 4. Gelman-Rubin (GR), variances and LR values from replicate runs of four-contributor samples. *(4M) = analysis using 1,000,000 burn-in and 4,000,000 post burn-in accepts. Replicates correspond to a – f, or a – e for (4M), in Figure 4.

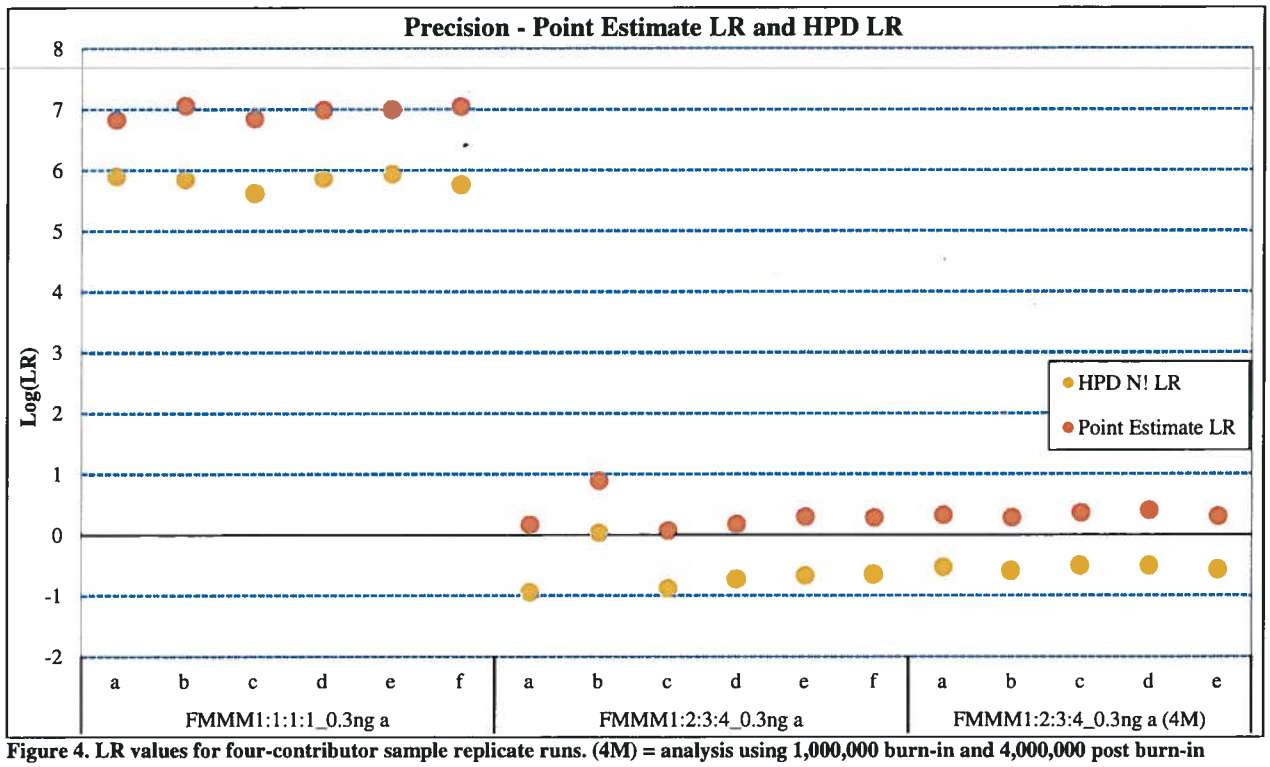


Figure 4. LR values for four-contributor sample replicate runs. (4M) = analysis using 1,000,000 burn-in and 4,000,000 post burn-in accepts

These experiments show that the variability due to the MCMC process is random. Variation between the replicates remains consistent and within approximately one order of magnitude between the six replicates analyzed at the default setting and the five replicates analyzed using increased iterations. The HPD LR was always lower than the point estimate LR demonstrating its ability to account for MCMC and allele frequency uncertainty. The Gelman-Rubin diagnostic, mixture proportions, and variance values produced from the replicate runs are similar for each sample. The default number of iterations is suitable for many different types of profiles. Increasing the number of iterations may be needed to help resolve complex mixture profiles.

Section N: NIST and NIST Traceable Samples (QAS 9.5.5)

In compliance with the current (2011) FBI QAS standard 9.5.5 two NIST certified profiles (NIST 2391c components B and C) and two LASD generated NIST traceable profiles (LASD NIST 09-1 BS5 and LASD NIST 12-1 SS3) were analyzed to verify concordance between STRmix™ deconvoluted profiles and the expected profiles. For all four profiles, STRmix™ returned a weight = 1 for the correct genotype at each locus. The results for NIST 2391c components B and C are given below. Table 1 lists the expected genotypes as provided by the Certificate of Analysis. Table 2 shows the STRmix™ v2.5.11 genotype probability distribution outputs. The results for the LASD NIST traceable samples are maintained electronically.

Locus	Component					
	A	B	C	D	E	F
D1S1656	17.3, 17.3	11, 14	11, 15	11, 15, 17.3	11, 16.3	17.3, 17.3
D2S1338	18, 23	17, 17	19, 19	18, 19, 23	19, 20	17, 17
D2S441	10, 10	10, 14	10, 10	10	10, 10	14, 14
D3S1358	15, 16	15, 19	16, 18	15, 16, 18	14, 15	16, 17
D5S818	11, 12	12, 13	10, 11	10, 11, 12	11, 13	11, 13
D6S1043	11, 18	14, 19	11, 14	11, 14, 18	11, 11	11, 16
D7S820	11, 11	10, 10	10, 12	10, 11, 12	8, 10	8, 12
D8S1179	13, 14	10, 13	10, 17	10, 13, 14, 17	11, 13	10, 13
D8S1115	15, 16	15, 17	9, 9	9, 15, 16	9, 16	9, 17
D10S1248	15, 16	13, 13	12, 16	12, 15, 16	14, 14	14, 15
D12S391	18.3, 22	19, 24	19, 23	18.3, 19, 22, 23	17, 22	18, 19
D13S317	8, 8	9, 12	11, 11	8, 11	8, 12	8, 11
D16S539	10, 11	10, 13	10, 10	10, 11	11, 12	9, 11
D18S51	12, 15	13, 16	16, 19	12, 15, 16, 19	14, 17	17, 22
D19S433	13, 14	16, 16.2	13.2, 15.2	13, 13.2, 14, 15.2	14, 14	13, 14
D21S11	28, 32.2	32, 32.2	29, 30	28, 29, 30, 32.2	29, 30	29, 32.2
D22S1045	15, 15	15, 17	16, 16	15, 16	16, 17	11, 15
CSF1PO	10, 10	10, 11	10, 12	10, 12	10, 11	10, 11
FGA	21, 23	20, 23	24, 26	21, 23, 24, 26	20, 23	21, 25
Penta D	9, 13	8, 12	10, 11	9, 10, 11, 13	14, 14	9, 10
Penta E	5, 10	7, 15	12, 13	5, 10, 12, 13	13, 19	11, 15
SE33	16, 18	17, 18	28.2, 31.2	16, 18, 28.2, 31.2	22, 30.2	12, 21
TH01	8, 9.3	6, 9.3	6, 8	6, 8, 9.3	6, 9.3	7, 9.3
TPOX	8, 8	8, 11	11, 11	8, 11	8, 11	8, 8
vWA	18, 19	17, 18	16, 18	16, 18, 19	17, 18	16, 18
Amelogenin	X, X	X, Y	X, Y	X, Y	X, X	X, Y

Table 1. NIST 2391c genotypes table from the Certificate of Analysis (expiration 02/03/2020)

GENOTYPE PROBABILITY DISTRIBUTION Component B

LOCUS	CONTRIBUTORS	WEIGHT
	1 (100%)	(HIGHLIGHT ≥ 0.99)
D3S1358	15, 19	1
D1S1656	11, 14	1
D2S441	10, 14	1
D10S1248	13, 13	1
D13S317	9, 12	1
Penta E	7, 15	1
D16S539	10, 13	1
D18S51	13, 16	1
D2S1338	17, 17	1
CSF1PO	10, 11	1
Penta D	8, 12	1
TH01	6, 9.3	1
vWA	17, 18	1
D21S11	32, 32.2	1
D7S820	10, 10	1
D5S818	12, 13	1
TPOX	8, 11	1
D8S1179	10, 13	1
D12S391	19, 24	1
D19S433	16, 16.2	1
SE33	17, 18	1
D22S1045	15, 17	1
FGA	20, 23	1

GENOTYPE PROBABILITY DISTRIBUTION Component C

LOCUS	CONTRIBUTORS	WEIGHT
	1 (100%)	(HIGHLIGHT ≥ 0.99)
D3S1358	16, 18	1
D1S1656	11, 15	1
D2S441	10, 10	1
D10S1248	12, 16	1
D13S317	11, 11	1
Penta E	12, 13	1
D16S539	10, 10	1
D18S51	16, 19	1
D2S1338	19, 19	1
CSF1PO	10, 12	1
Penta D	10, 11	1
TH01	6, 8	1
vWA	16, 18	1
D21S11	29, 30	1
D7S820	10, 12	1
D5S818	10, 11	1
TPOX	11, 11	1
D8S1179	10, 17	1
D12S391	19, 23	1
D19S433	13.2, 15.2	1
SE33	28.2, 31.2	1
D22S1045	16, 16	1
FGA	24, 26	1

Table 2. STRmix™ genotype probability distribution outputs for NIST 2391c components B & C