OREGON STATE POLICE
FORENSIC SERVICES DIVISION

PORTLAND METRO LABORATORY

VALIDATION STUDY FOR STR ANALYSIS

Volume 67—2016
Validation – STR Casework Analysis Using GlobalFiler, the 3500xl, and STRmix

MEMORANDUM
OREGON STATE POLICE

DATE: November 2, 2016

TO: DNA Casework Analysts
Portland-Metro Forensic Laboratory

FROM: Chrystal Bell, Quality Assurance Manager
Portland-Metro Forensic Laboratory

SUBJECT: **Memo of Authorization: STR Casework Analysis Using GlobalFiler, the 3500xL, and STRmix**

I have reviewed the validation summary report for the implementation of GlobalFiler using the 3500xL and STRmix. Sufficient validation has been conducted, and implementation can move forward pending the appropriate procedural updates and analyst training.

CB:mfk

cc: Marla Kaplan—DNA Technical Leader
Stephenie Winter Sermeno—Casework DNA Supervisor

Autosomal Casework Analysis Transition to New CODIS Core Loci

Validation Summary
MFK 2016

Introduction

This document summarizes the experiments performed to transition autosomal casework analysis to amplification via GlobalFiler, genotyping on a 3500xL using GeneMapper ID-X v1.4, and profile interpretation assisted by STRmix with a likelihood ratio output. Note that this validation draws upon the work performed to transition convicted offender sample analysis to GlobalFiler and the 3500xL, and portions of the validation relevant to casework sample analysis may be contained in the validation summary 2015 DNA Validation – Database Sample Analysis Transition to New CODIS Core Loci.

As noted in the database sample analysis validation summary: "*In the middle of this portion of the validation, we were notified by Life Technologies of a reformulation of the GlobalFiler kit. This reformulation was performed in order to improve the long-term stability of the GlobalFiler master mix and to address some peak height and pattern issues in the NED (yellow) and SID (purple) dye channels. The reconfigured kit contains a raw material BSA (Bovine Serum Albumin) prepared with a Tris buffer as opposed to a phosphate buffer. Studies at LifeTech revealed that the old BSA formulation led to the development of a magnesium phosphate complex, thereby reducing the Mg2+ accessible during amplification."* As with the databasing validation, the results obtained from validation experiments already completed were assessed to determine whether additional work was necessary using the reformulated kit. Summaries of any additional work performed are included alongside each relevant experiment.

Per the plan, extensive work was done during this validation to explore the Normalization feature on the 3500xL. Studies showed, however, that the expected signal strength of the 11 LIZ peaks required to Normalize a sample was not stable enough from run to run to accurately set a Normalization target. The data is available for review and is summarized via charts and graphs in the electronic validation documentation, but no formal summary of the assessment of sample Normalization is contained herein.

All amplifications from DNA extracts were achieved via the following thermalcycling parameters (max ramp speed):

Initial Incubation Step	29 cycles		Extension	Final Soak
	Melt	Anneal		
95°C 1 min.	94°C 10 sec.	59°C 1 min. 30 sec.	60°C 10 min.	4°C ∞

Unless otherwise noted, GeneMapper analysis for experiments performed with the original GlobalFiler formulation was achieved using an analytical threshold of 120RFU. Experiments performed using the reformulated kit were genotyped using an analytical threshold of 100RFU. See the Experiment 22 summary for further details.

Experiment 6: Workflow integration study: Amplification of samples concentrated via Vacufuge

See 2015 DNA Validation – Database Sample Analysis Transition to New CODIS Core Loci.

Experiment 9: Direct amplification of oral swab reference samples

This experiment was performed using the original GlobalFiler formulation. Based upon the results obtained, we determined it was unnecessary to repeat this experiment using the reformulated kit.

For this experiment, twenty oral swab reference standards, each from a different individual, were collected for analysis. To determine the efficacy of a direct amplification method for casework reference standards, approximately ½ of each swab was digested for 20 minutes at 37˚C on a thermal mixer using 400µL of Prep-n-Go buffer (Applied Biosystems). Following incubation, amplification was achieved by using 3µl of sample lysate, 12µl of TE, and 10µl of GlobalFiler master mix cocktail.

The samples were run on a 3500xL at 24, 12, and 6 second injection times. Briefly, results suggest that direct amplification using the Prep-n-Go buffer is appropriate for implementation as complete profiles for all 20 reference standards were obtained.

	DIRECT AMP 24 SEC	DIRECT AMP 12 SEC	DIRECT AMP 6 SEC
	Profile complete	Profile complete	Profile complete
Ex.1	yes	yes	yes
Ex.2	yes	yes	yes
Ex.3	yes	yes	yes
Ex.4	yes	yes	yes
Ex.5	yes	yes	yes
Ex.6	yes	yes	yes
Ex.7	yes	yes	yes
Ex.8	yes	yes	yes
Ex.9	yes	yes	yes
Ex.10	yes	yes	yes
Ex.11	yes	yes	yes
Ex.12	yes	yes	yes
Ex.13	yes	yes	yes
Ex.14	yes	yes	yes
Ex.15	yes	yes	yes
Ex.16	yes	no	no
Ex.17	yes	no	no
Ex.18	yes	yes	yes
Ex.19	yes	yes	yes
Ex.20	yes	yes	yes

At the default injection time of 24 seconds, 75% of the samples exhibited off-scale data. The 12 second injection was considerably better with only 25% of samples exhibiting off-scale data. Finally, the 6 second injection showed no off-scale data in any of the samples. In the 12 and 6 second injections, however, there were two profiles with uncalled alleles.

As expected, the off-scale data produced artifacts such as pull-up, raised baseline, and elevated stutter. These artifacts were much more severe and problematic with the default injection due to all the off-scale data. The 12 and 6 second injections also exhibited some artifacts, but these were generally very minimal.

All of the sample digests were taken through extraction on an EZ1 (Large Volume/50) and quantitated per protocol. Quantitation results were assessed, and four samples were selected and taken to GlobalFiler amplification targeting 0.5ng. No genotyping issues were observed for these samples, and they all had complete profiles.

The quantitation results (data not shown), were low for every sample, relative to what is expected for a reference sample. Out of the 20 samples, ten would require concentration if an analyst had to go back to the Prep-n-Go digest due to a failed direct amp (assuming the analyst is only targeting 0.5ng). Three would likely have to be re-sampled and extracted due to extremely low quantitation results.

After the above results were assessed, additional experimentation was prescribed to (1) assess sample preparation in an elution tube as opposed to a flip-top tube and (2) determine if concentration of a Prep-n-Go sample subsequently extracted on an EZ1 via Vacufuge is appropriate. Specifically, the concern was that the Prep-n-Go solution—even after extraction on an EZ1—might have an inhibitory effect on the amplification.

The expectation regarding the elution tubes was that the tube type would not impact the final direct -amp results and, whenever possible, we should perform this step directly in the elution tube so as not to have a separate transfer step into the elution tube from a flip-top tube. This would reduce the amount of sample handling and time prior to placing extracts on the JANUS robot. As an adjunct to this experiment the question of using the 56˚C soak instead of the 37˚C soak was raised. Six samples were chosen and roughly equal halves of the oral swabs were placed into two separate elution tubes—one set soaked at 37˚C and one set at 56˚C for 20 minutes, after which direct amplification as described above was performed.

The direct amplification of the twelve samples digested directly in the elution tubes produced equivalent results to the previous experiment utilizing the flip-top tubes. Also, there was no observable difference in RFUs or number of off-scale samples in either the 24sec or 12sec injections between the different temperature soaks. All concentrated samples amplified as expected.

Upon verification of this method as described above, a series of samples were processed via this method to populate the elimination database. This sample set included a wide range of samples of varying strengths and ages. The samples were run alongside a series of TE blanks from Experiment 19. Review of the TE blanks indicated some low-level peaks, some above and some below the analytical threshold. Further investigation determined that these peaks were likely carryover from the elimination data sample set. For this reason, oral swab processing via Prep-n-go buffer will not be implemented at this time. At such time as the laboratory adjusts workflows to include batch processing of large numbers of references, we will revisit this decision and ensure that the appropriate carryover-control measures are put in place alongside.

Instead, the procedures for analyzing a reference sample without prior quantitation will stay as they are (0.5µL and 2µL for Normalization/50µL, 1µL and 4µL for Normalization/200µL). These values were derived from Quantifiler Duo data with a 1ng target in mind. Since the target for GlobalFiler is also 1ng, no further investigation regarding input volumes is necessary.

Experiment 10: *Precision and accuracy*

See 2015 DNA Validation – Database Sample Analysis Transition to New CODIS Core Loci.

Experiment 11: *Workflow integration study: ND results with Quantifiler Duo*

This experiment was originally performed using the original GlobalFiler formulation (data not described). The summary below describes the results of this experiment using the reformulated kit.

DNA extracts from two male individuals were retrieved for use in this experiment. Each sample was diluted to 0.1ng, based upon previously obtained quantitation results, and then further diluted in serial (1:1 each time) for a total of 12 dilutions. Each dilution was quantitated, and any sample in which no DNA was detected OR any sample in which only a human DNA quantitation result (no male) was obtained was selected for amplification.

The samples which yielded no human and male DNA or those in which no male DNA was detected were amplified twice – once using 10µl, mimicking current processing protocols, and once using 15µl, the maximum input allowed by the GlobalFiler kit. The total number of autosomal STR loci yielding genotyping data was assessed. Results were as follows:

	10µl input		**15µl input**	
Dilution Number	Sample 43	Sample 101	Sample 43	Sample 101
3		21 [human DNA] = 0.008 ng/µl		21 [human DNA] = 0.008 ng/µl
4				
5	13 [human DNA] = 0.007 ng/µl	5	21 [human DNA] = 0.007 ng/µl	13
6	4	1	14	9
7	4	1	0	0
8	1	0	0	0
9		0		0
10	0	0	0	0
11	0	0	0	0
12	0	0	0	0

■ = only male or both human and male quant values obtained, no amp of sample

■ = human quant only obtained (male value negative)

□ = no human quant nor male quant value obtained

Based upon the results obtained, negative quantitation results obtained from a sample that can be amplified using 15ul of input DNA does not correlate to a lack of interpretable genotyping data. When examining those samples that mirror our current workflow (concentration to 11ul prior to quantitation), only those samples showing a human DNA quantitation value yielded results that could be deemed interpretable (in a single source sample) while samples that were completely negative did not yield interpretable genotyping results. Therefore, upon GlobalFiler implementation, processing will be stopped for those samples concentrated to 11ul prior to quantitation for which no DNA is detected. Following implementation, Y-screen samples yielding negative male quantitation values but human

quantitation values of ≤0.01ng/ul will not automatically be stopped, and the amplification results will be tracked for a period of time so that more informed decisions about discontinuing analysis based upon quantitation results can be made.

Experiment 12: *Workflow integration study: Human to male ratios*

This experiment was originally performed using the original GlobalFiler formulation (data not described). The summary below describes the results of this experiment using the reformulated kit.

Previously quantitated DNA extracts from a known female and known male contributor were mixed in ratios of 1:20, 1:30, 1:40, 1:50, 1:60, and 1:70 in which the female is the major contributor.

Each mixture was amplified in duplicate using the GlobalFiler amplification kit at 1ng, 0.1ng, and 0.01ng and run on a 3500xL at a 24 second injection time. Profiles were assessed for total number of minor contributor alleles listed in the 99% weights from STRMix (see below for further details on STRmix validation) and whether an association via an LR could be made to the minor contributor.

Sample	Amplification Target (ng)	Mixture Observed?	Number of Minor Loci with 99% Weighting	CODIS Entry for Minor? (Number of Old Core with 99% Weighting)	Association to Known Minor
1:20 – amp 1	1	Yes	11	No (6)	Yes
1:20 – amp 2	1	Yes	11	No (7)	Yes
1:30 – amp 1	1	Yes	10	No (5)	Yes
1:30 – amp 2	1	Yes	11	No (6)	Yes
1:40 – amp 1	1	Yes	8	No (6)	Yes
1:40 – amp 2	1	Yes	7	No (5)	Yes
1:50 – amp 1	1	Yes	10	No (5)	Yes
1:50 – amp 2	1	Yes	6	No (3)	Yes
1:60 – amp 1	1	Yes	7	No (3)	Yes
1:60 – amp 2	1	Yes	9	No (6)	Yes
1:70 – amp 1	1	Yes	6	No (4)	Yes
1:70 – amp 2	1	Yes	4	No (3)	Inconclusive
1:20 – amp 1	0.1	Yes	1	No (0)	Inconclusive
1:20 – amp 2	0.1	Yes	3	No (2)	Inconclusive
1:30 – amp 1	0.1	Yes	1	No (0)	Inconclusive
1:30 – amp 2	0.1	Yes	1	No (0)	Inconclusive
1:40 – amp 1	0.1	Yes	3	No (2)	Inconclusive
1:40 – amp 2	0.1	Yes	2	No (1)	Inconclusive
1:50 – amp 1	0.1	No	n/a	n/a	n/a
1:50 – amp 2	0.1	No	n/a	n/a	n/a
1:60 – amp 1	0.1	No	n/a	n/a	n/a
1:60 – amp 2	0.1	Yes	2	No (2)	Inconclusive
1:70 – amp 1	0.1	No	n/a	n/a	n/a
1:70 – amp 2	0.1	No	n/a	n/a	n/a

All samples amplified at 1ng and the majority of the samples amplified at 0.1ng yielded mixture components, however only those amplified at 1ng yielded results for which an association could be made to the minor contributor. No CODIS entries could be achieved for the minor contributor from these samples, even those amplified at 1ng; however, this is, in part, impacted by the amount of allele sharing between the two contributors. It would not be appropriate to draw a direct correlation between these results from these amplifications and the likelihood of obtaining a CODIS-eligible profile for the minor.

For the 0.01ng amplifications, no minor contributor alleles were observed with one exception: Inverted dropout was observed at D10S1248 in one of the duplicate 1:40 amplifications. At D10S1248, the known major contributor to this sample is a [redacted]. The known minor is a [redacted]. The observed profile is [redacted] (292RFU), [redacted] (139RFU), and [redacted] (below threshold at ~80RFU).

Putting these results into the context of Y-screening, samples with ratios of 1:>70 are unlikely to yield an interpretable minor profile and, therefore, should not be forwarded for autosomal amplification. If a sample yields a ratio between 1:20 and 1:70 and at least 0.1ng cannot be input into the amplification reaction, that sample should be held in favor of Y-STR analysis. Using GlobalFiler, samples will be forwarded for autosomal analysis if the ratio is 1:<20 or if the ratio is between 1:20 and 1:70 and amplification can be achieved using greater than 0.1ng of input DNA. As with the ND results, data from GlobalFiler amplifications of casework samples will be tracked for a period of time in order to determine whether adjustments to these parameters are appropriate.

Experiment 13: Validation of the expert system/expert assistant

See 2015 DNA Validation – Database Sample Analysis Transition to New CODIS Core Loci.

Experiment 14: *Known and non-probative samples*

This experiment was performed using the original GlobalFiler formulation. Based upon the results obtained, we determined it was unnecessary to repeat this experiment using the reformulated kit.

This experiment was designed to assess the outcomes of amplifying the same DNA extracts with both Identifiler Plus and GlobalFiler. Six samples were compared, one of which was extracted organically. Various analysts amplified the samples using Identifiler Plus, and due to the variability in the initial quant dates, the samples were requantitated prior to GlobalFiler amplification; however, minimal differences were noted between the input volumes needed for the two amplifications (data not shown). Additionally, two samples amplified with Identifiler Plus exhibited off scale data; however, the resulting profiles would still have been used for interpretation and comparisons in casework. Therefore, they were retained for comparison in this experiment.

A comparison of the six samples amplified using both Identifiler Plus and GlobalFiler amplification kits yielded profiles that were largely concordant at the overlapping loci. Overall, GlobalFiler did appear to be slightly more sensitive than Identifiler Plus; however, the difference appears to be minimal when taking possible stutter effects and amplification input volumes into consideration. The table below details a comparison of the results from the two kits for each sample with reference to allele concordance.

Sample	Minimum Number of Contributors	TOTAL Alleles MISSING with GlobalFiler	Alleles Missing with GlobalFiler in a -4 Position	Alleles Missing with GlobalFiler in Both a -4 and +4 Position	TOTAL ADDITIONAL Alleles NOTED with GlobalFiler	Additional alleles noted with GlobalFiler in a +4 position only	Additional alleles noted with GlobalFiler in a -4 position only	Additional alleles noted with GlobalFiler in Both a -4 and +4 positions
14.1	3+	2*	2	0	10	3	3	1
16.3	4+	1	1	0	3	1	1	1

Sample	Minimum Number of Contributors	TOTAL Alleles MISSING with GlobalFiler	Alleles Missing with GlobalFiler in a -4 Position	Alleles Missing with GlobalFiler in Both a -4 and +4 Position	TOTAL ADDITIONAL Alleles NOTED with GlobalFiler	Additional alleles noted with GlobalFiler in a +4 position only	Additional alleles noted with GlobalFiler in a -4 position only	Additional alleles noted with GlobalFiler in Both a -4 and +4 positions
1sp	2	2	1	1	4**	1	0	2
3epi	2	0	0	0	3**	1	0	1
22L-88	2	0	0	0	1	0	0	1
Organic	1	0	0	0	0	0	0	0

**Both missing alleles were in a -4 position, however one was almost the same height as the +4 allele.*
***One of the listed alleles is a possible artifact*

One interesting difference between the two kits was noted at TH01 in two samples. These samples were used for new employee training, so the contributors were known. For these two samples, the female minor contributor should be the same individual. An artifact was noted and called as such by the trainee in one profile when amplified with Identifiler Plus. The other Identifiler Plus amplification showed an uncalled peak in a similar position. GlobalFiler, however, assigns the peak an allelic designation of [redacted]. The peak appears to have good peak morphology in both kits. The known minor contributor has a heterozygous [redacted] genotype at TH01. As this contributor is unlikely to be a triallelic [redacted] at TH01 given the RFU's noted for all peaks – particularly given the rarity of a [redacted] allele at TH01 in the population – this peak was appropriately called as an artifact using Identifiler Plus and should also be in GlobalFiler. See below.

x1sp-Identifiler Plus,TH01 x1sp-GlobalFiler, TH01

x3epi Identifiler Plus, TH01 x3epi-GlobalFiler, TH01

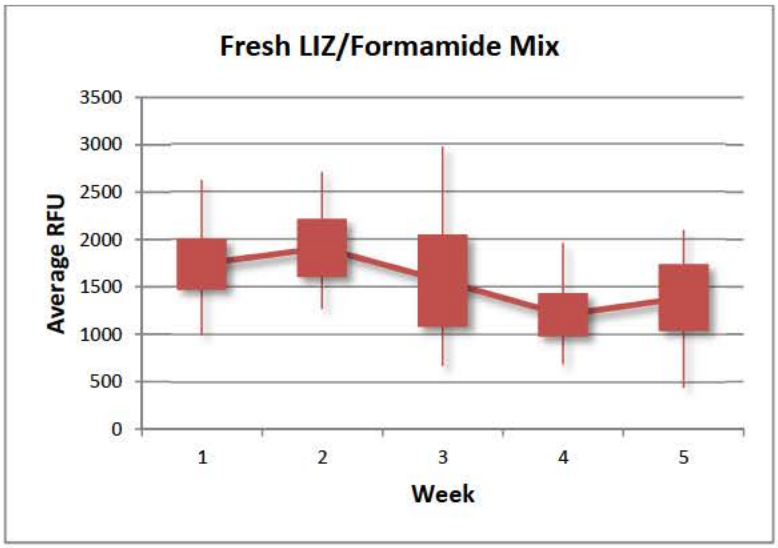
Notwithstanding the degenerate primers added to the GlobalFiler amplification kit, analysts should expect concordance with previously-used amplification chemistries when using the GlobalFiler amplification kit.

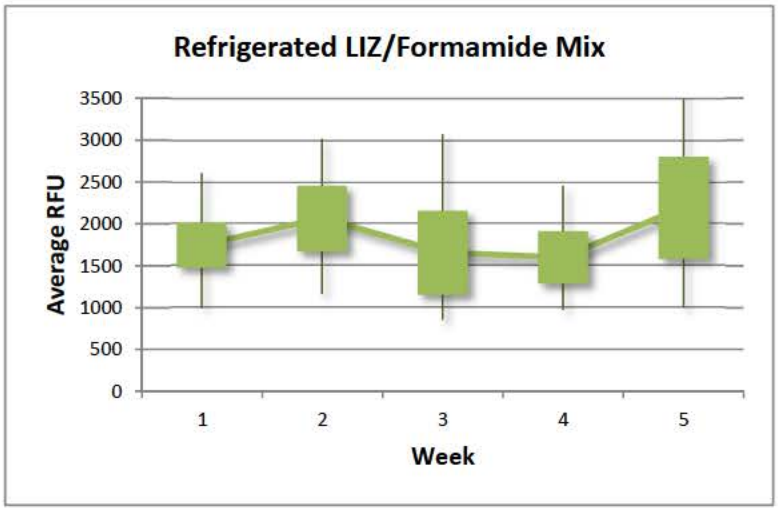
Experiment 15: *Contamination assessment: Capillary cross-talk and sample carryover*
See 2015 DNA Validation – Database Sample Analysis Transition to New CODIS Core Loci.

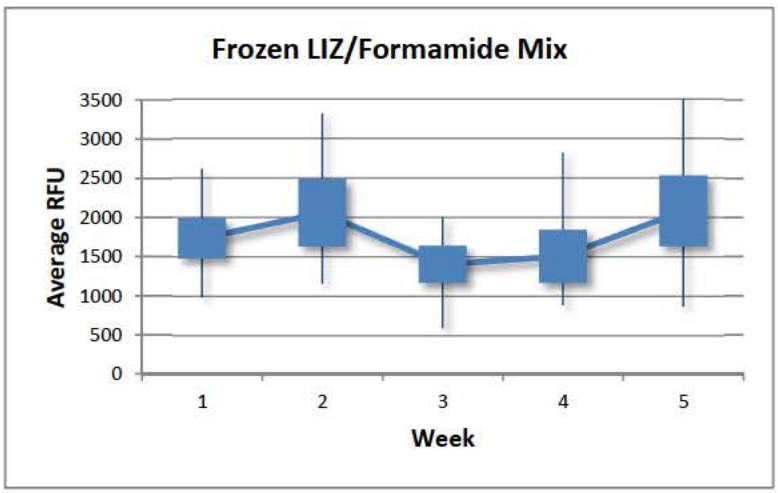
Experiment 17 (from Phase II Validation Plan): *NIST-traceable sample*
See 2015 DNA Validation – Database Sample Analysis Transition to New CODIS Core Loci.

Experiment 19: *LIZ/formamide storage condition assessment*
While this experiment was originally intended to assist in the establishment of a Normalization Target, once the determination was made that sample Normalization via the 3500xL data collection software was not going to be implemented, the data is instead being used to establish the appropriate storage conditions for pre-mixed LIZ/Formamide aliquots.

Twenty-two TE blanks were amplified with the GlobalFiler kit. These TE blanks were run on each 3500xL instrument using freshly-prepared, previously prepared and frozen, and previously prepared and refrigerated LIZ/formamide master mix aliquots in a ratio of 0.4ul LIZ added to 10ul formamide. The blanks were run once using freshly-prepared LIZ/formamide, once using refrigerated pre-mixed LIZ/formamide, and once using frozen and thawed pre-mixed LIZ/formamide for four weeks after the original preparation date (for a total of five data sets). The average peak height of the LIZ 200, 220, 240, 260, 280, 300, 314, 320, 340, 360, and 400bp peaks was calculated for each capillary in the array for each individual run and the average peak heights compared.







The box plots above show the average RFU across all 24 capillaries of each of the 11 LIZ peaks +/- 1 standard deviation. Vertical bars represent the maximum and minimum values observed. The data

shows that regardless of whether the LIZ/Formamide mix is prepared fresh with each run or if it is previously prepared and stored frozen or refrigerated, the height of the LIZ is more than sufficient to achieve sample sizing. Therefore, upon implementation, the LIZ/formamide mixes will be pre-prepared and stored frozen to be thawed upon first use, after which they will be stored refrigerated.

Experiment 20: *Final assessment of sample normalization*

Sample Normalization was determined to not be a viable solution for the "edge effect" issues seen during previous validation experiments and therefore will not be implemented. Strategic placement of positive and negative amplification controls as well as allelic ladders should mitigate this effect.

Experiment 21: *STRmix parameter setting – Dilution series*

This and all subsequent experiments were performed using the reformulated GlobalFiler kit.

The purpose of this experiment was to provide data input for some of the remaining experiments in the validation. Ten references (designated A-J) were quantitated in triplicate using Quantifiler Duo and amplified using the reformulated GlobalFiler kit at the following template targets: 1.5ng, 1ng, 0.9ng, 0.8ng, 0.7ng, 0.6ng, 0.5ng, 0.4ng, 0.3ng, 0.2ng, 0.1ng, 0.09ng, 0.08ng, 0.07ng, 0.06ng, 0.05ng, 0.04ng, 0.03ng, 0.02ng, 0.01ng.

These samples were injected one time each at 24 seconds, 48 seconds, and 12 seconds, with and without Normalization. The normalized data was not used further. The non-normalized injections were used as described below.

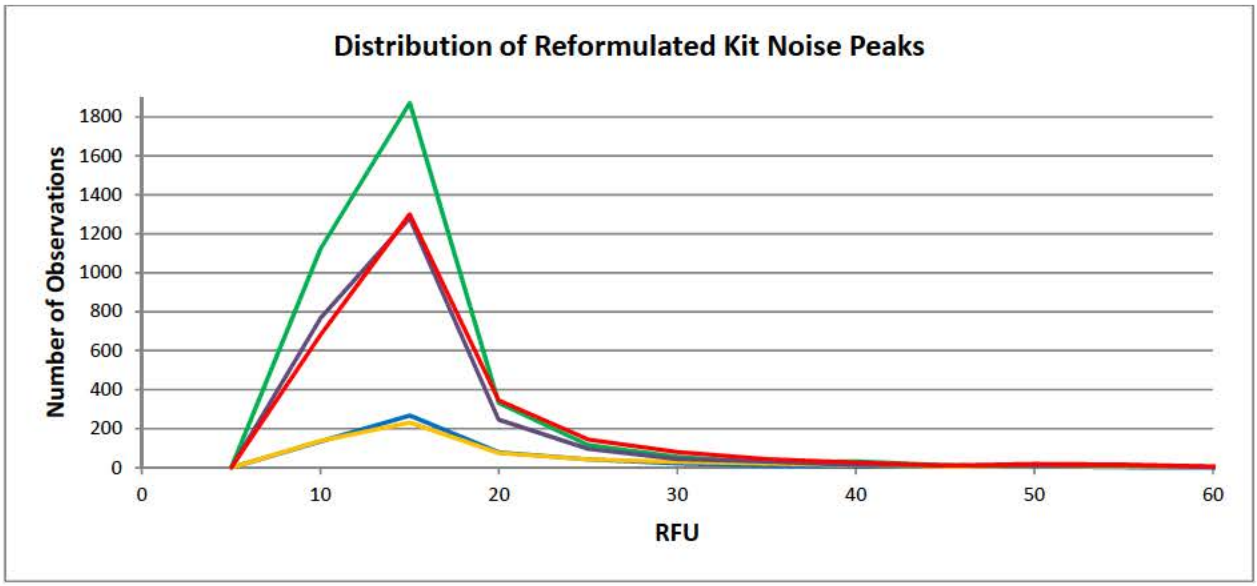
It should be noted that validation of STRmix was undertaken using two separate versions of STRmix. Experimentation began using version 2.3 and was completed using version 2.4. The most notable enhancement in version 2.4 was the introduction of forward stutter modelling capabilities. Where necessary, in the summaries below, we will describe the experimentation and results pertaining to versions 2.3 and 2.4 separately.

Experiment 22: *STRmix parameter setting – Reassessment of analytical threshold*

The laboratory transition to the utilization of STRmix has been coupled with the validation of the new GlobalFiler kit (since re-formulated) and the 3500xL Genetic Analyzer. STRmix analysis parameters are directly dependent on the DNA amplification kit and CE instrumentation being utilized within our lab. For this reason, an analytical threshold for the 3500xL CE instruments using the re-formulated GlobalFiler kit was determined.

Using the dilution series amplifications created in Experiment 21, a detection threshold of 10 rfu was set and the samples analyzed. The data was assessed for true alleles and amplification-related artifacts and these peaks were removed, leaving behind only the called baseline noise peaks. This was achieved one amplification at a time, one dye lane at a time utilizing both the standard EPG window and Raw Data view to confirm artifacts from baseline noise. The data from all three instruments was combined and the noise peaks were tabulated by color and overall.

The results for each dye channel were as follows:



As expected, a large proportion (89%) of the noise peaks had heights fewer than 25 RFU; however, the maximum observed noise peaks were 83 RFU for blue and yellow, 90 RFU for green, 96 RFU for purple, and 109 RFU for red. Based upon these heights, using twice the maximum observed (as has been done in the past) would result in an analytical threshold that would needlessly filter out allelic data. For that reason, twice the RFU that would filter 99% of the noise peaks was employed as a means to determine the analytical threshold as follows:

Dye	99%	99% * 2	Dye	99%	99% * 2
Blue	50	100	Red	45	90
Green	40	80	Purple	55	110
Yellow	60	120	**All Data**	**50**	**100**

Thus, the analytical threshold for the reformulated GlobalFiler kit was set at 100RFU for all dye channels.

Experiment 23: *STRmix parameter setting – Stutter assessment*

In order to create both STRmix stutter input files and to determine the GeneMapper stutter filters to be used in our laboratory, staff member elimination profiles for which permission has been given to use for validation purposes and offender samples having small and large alleles at each of the GlobalFiler loci were located. These samples were extracted and run a single time targeting 0.5ng of input DNA.

These samples and the samples from the dilution series in Experiment 21 were loaded into GeneMapper and analyzed with the stutter filters turned off at an analytical threshold of 30 RFU. In total, the electropherograms from 314 amplifications were assessed for -2 repeat, -1 repeat, +1 repeat, and -1/2 repeat stutter products. Following the determination of the GeneMapper stutter filter ratios using the single-source data described above, the data from the mixtures prepared in Experiment 32 were used to fine-tune the GeneMapper settings to avoid the labeling of stutter products.

The final GeneMapper settings were set using the maximum observed values, with the exception of any clear outliers. -2 repeat filters and +1 repeat filters were implemented at each locus for which these

types of stutter were observed. - ½ repeat stutter was observed at SE33 and D1S1656 as described in the GlobalFiler User's Manual. +1 repeat filters were implemented at all loci.

Locus	-2 repeat filter	-1 repeat filter	- ½ repeat filter	+1 repeat filter
D3S1358	0.0143	0.1720		0.0421
vWA	0.0172	0.1395		0.0198
D16S539	0.0069	0.1130		0.0397
CSF1PO	0.0162	0.0948		0.0502
TPOX		0.0541		
D8S1179	0.0289	0.1176		0.0390
D21S11	0.0158	0.1360		0.0313
D18S51	0.0137	0.1423		0.0501
DYS391		0.0901		
D2S441	0.0143	0.1333		0.0705
D19S433	0.0323	0.1152		0.0278
TH01		0.0455		
FGA	0.0189	0.1468		0.0491
D22S1045	0.0526	0.1391		0.0626
D5S818	0.0270	0.0944		0.0422
D13S317	0.0067	0.1273		0.0444
D7S820	0.0224	0.0942		0.0249
SE33	0.0330	0.2000	0.0435	0.0389
D10S1248	0.0260	0.1328		0.0399
D1S1656	0.0252	0.1614	0.0242	0.0879
D12S391	0.0270	0.1799		0.1449
D2S1338	0.0135	0.2174		0.0385

In addition to using the stutter data to establish the GeneMapper stutter filters, the data was used to create the -1 repeat and +1 repeat stutter files and stutter exceptions files as described in the STRmix v2.3 Implementation and Validation guide and in the STRmix v2.4 Operation Manual. For all loci, the repeat value was used to determine a linear regression of stutter vs allele for -1 repeat stutter data. These values were used to create the stutter input file. For D3S1358, vWA, D8S1179, D21S11, D19S433, TH01, FGA, D1S1656, D12S391, and D2S1338, linear regressions were run for stutter versus the longest-uninterrupted sequence to determine the values for the STRmix stutter exceptions file. As recommended by STRmix support personnel, for D2S441, D22S1045, and SE33, the average observed value for each allele was used to populate the stutter exceptions file. (Exceptions file data not shown)

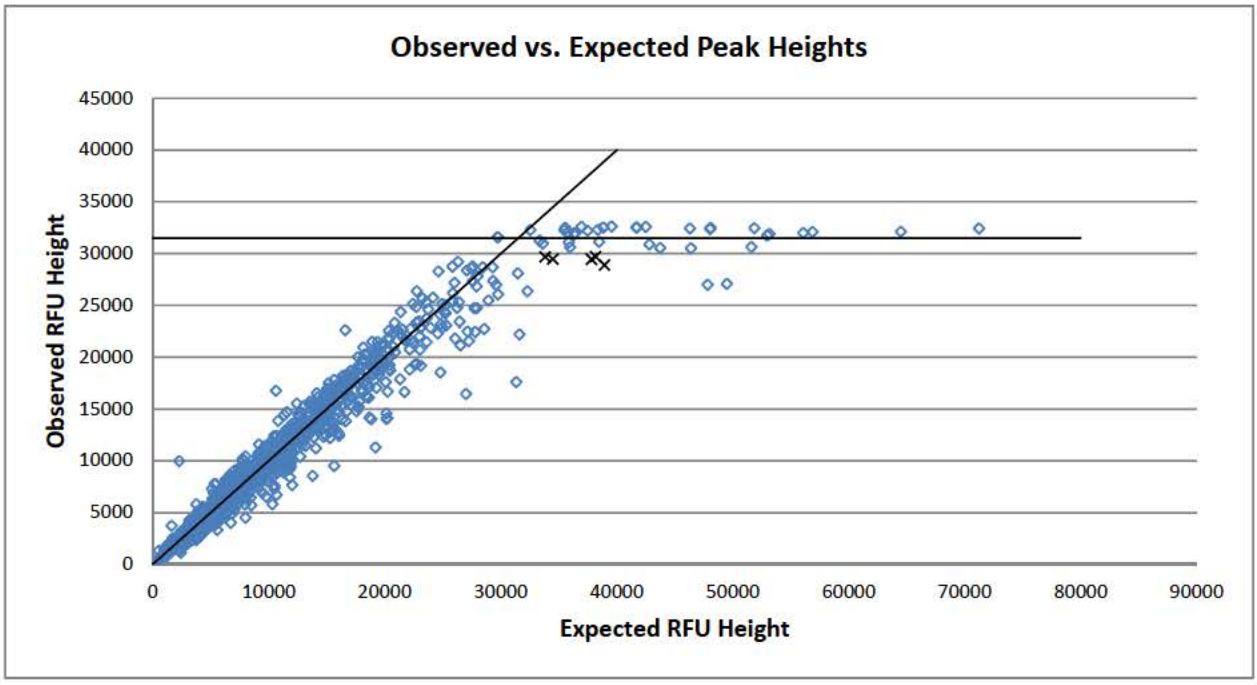
Forward (+1 repeat) stutter input files were determined using the average observed stutter ratio for all loci except D22S1045. Linear regression data for this locus (simple repeat) was used to determine the forward stutter input file for STRmix.

GF_stutter_OSP input file	**GF_N+1_stutter_OSP input file**
Locus,Intercept,Slope	Locus,Intercept,Slope
1,-0.0624391,0.0086147	1,0.0079295,0
2,-0.0930822,0.0094738	2,0.0085977,0
3,-0.0532475,0.0096505	3,0.0078717,0
4,-0.0448656,0.0089540	4,0.0112241,0
5,-0.0263385,0.0053124	5,0,0
6,0,0	6,0,0

GF_stutter_OSP input file	GF_N+1_stutter_OSP input file
7,0.0054354,0.0042960	7,0.0115150,0
8,-0.0553377,0.0040605	8,0.0096596,0
9,-0.0376515,0.0069128	9,0.0097778,0
10,-0.0368823,0.0089043	10,0,0
11,0.0383696,0.0004994	11,0.0136485,0
12,-0.0681280,0.0094497	12,0.0120487,0
13,-0.0133461,0.0047960	13,0,0
14,-0.0717929,0.0063540	14,0.0084480,0
15,-0.1213610,0.0132905	15,-0.050172983,0.00524985
16,-0.0417032,0.0086023	16,0.0103898,0
17,-0.0531090,0.0090462	17,0.0088363,0
18,-0.0467929,0.0089586	18,0.0070356,0
19,0.0337142,0.0023892	19,0.0158989,0
20,-0.0512930,0.0089586	20,0.0103393,0
21,0.0140849,0.0037751	21,0.0119330,0
22,-0.0741925,0.0078784	22,0.0415102,0
23,-0.0088928,0.0040034	23,0.0243705,0

Experiment 24: *STRmix parameter setting – Saturation setting*

Data at CSF1PO, D13S317, D16S539, D18S51, D19S433, D21S11, D3S1358, D5S818, D7S820, vWA, TH01 (excluding the .3 variant alleles) and D10S1248 used to generate the following scatter plot, depicting the expected versus observed peak heights as determined by the observed stutter ratio and the stutter regression formulae determined in the above experiment.



Per the Implementation and Validation Guide, a common saturation value for a 3500 series instrument is 30,000 RFU. The data obtained in this experiment generally supports that value. Some data points, denoted with an X in the graph above, appeared to depart from the general trend but were still under the 30,000 RFU saturation value. Upon further analysis, four of these five peaks were off scale.

Therefore, the 30,000 RFU saturation value will catch most, but not all off scale peaks imported into STRmix. This was expected due to the rare observation of a peak denoted as off scale in GMIDX that is resultant of saturation in a different dye channel than the peak resides.

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 by STRmix are consistent with qualitative expectations.

Experiment 25: *STRmix parameter setting – Probability of drop in*

The reformulated kit genotyping data obtained from Experiments 11, 12, 21, 29, and 32 was analyzed in GeneMapper at an analytical threshold of 30RFU, and the data was assessed for drop-in events. -2 repeat stutter alleles observed over the GeneMapper filter were included in this assessment. In total, 104 drop in events were observed in the 10633 loci assessed.

The maximum observed drop-in height was 97RFU; thus, the maximum allowed drop-in height was set to 100RFU. The drop-in frequency and the α and β drop-in parameters were determined by inputting the observed data into a probability of drop-in spreadsheet provided by the STRmix developers. The Solver function within Microsoft Excel was utilized to establish a frequency of 0.1080 (to include drop-in below the analytical threshold), and α and β parameters of 0.65 and 17.20.

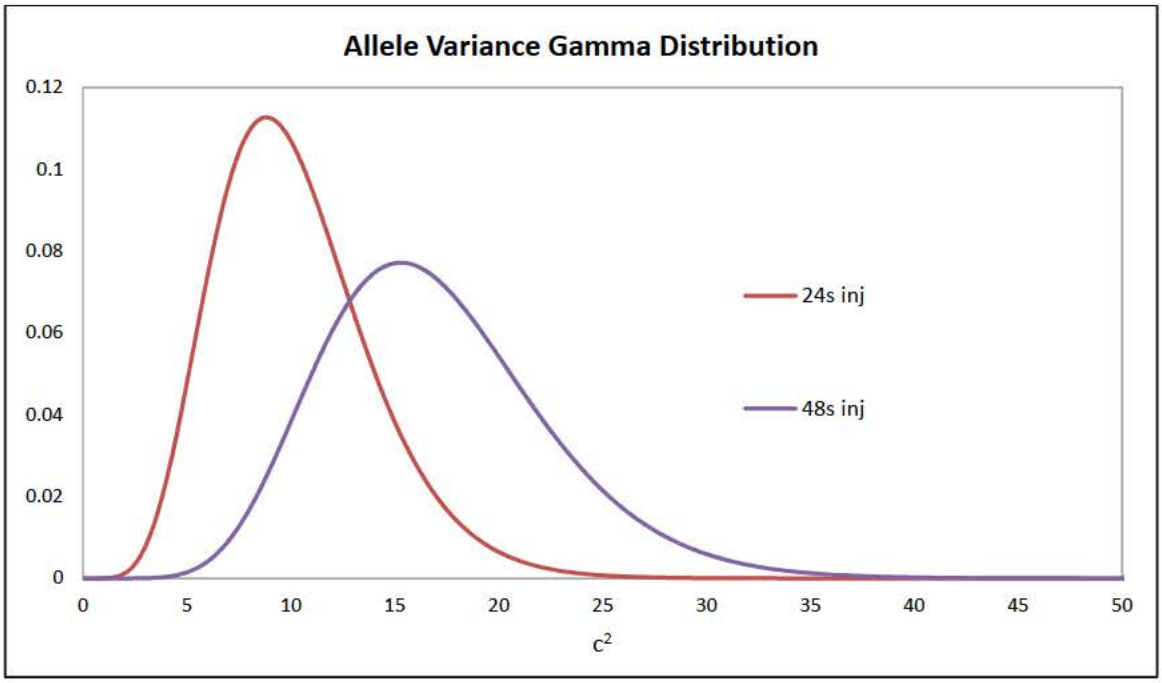
It is important to note that this assessment serves as a proxy for a more ideal assessment using a large volume of negative control data. For that reason, we will require that analysts raise the drop-in cap for any STRmix deconvolutions run where a drop-in event is observed in a corresponding negative control. In addition, following implementation, we will reevaluate the probability of drop-in parameter once sufficient reagent blank and negative control data has been gathered.

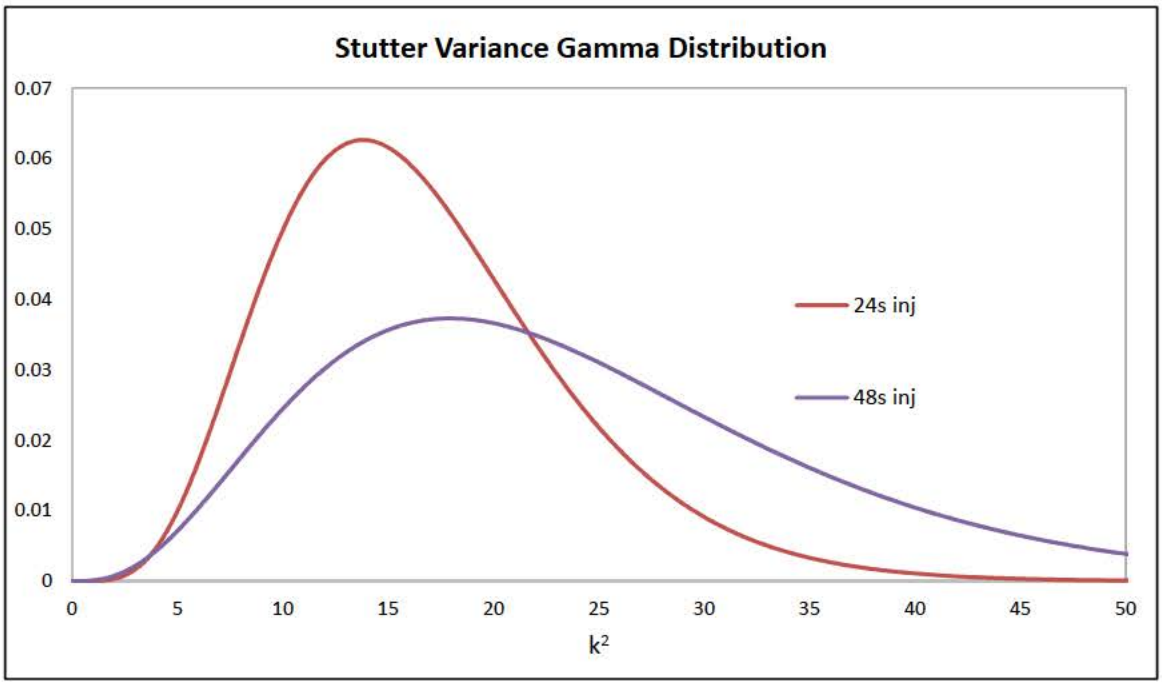
Experiment 26: *STRmix parameter setting – Model Maker*

A complete summary of the Model Maker experiment is found in the supplemental document STRmix parameters V2 3 Oregon, provided by the STRmix support team.

The variance results for the Model Maker data were assessed, and it was determined that two separate sets of STRmix parameters were necessary to fully account for the range of casework profiles anticipated. One set of parameters, derived from the 24 second injection data, will be applied to all 24 and 12 second injections. A second set of parameters, derived from the 48 second injection data, will be applied to all 48 second injections of single-source samples. No 48 second injections of mixture data will be allowed per protocol.

GlobalFiler™ 29 cycle 3500	Number profiles analysed	Allele variance parameters (Mode)	Stutter variance parameters (Mode)	Mean LSAE variance
24s inj	567	7.357,1.386 (8.811)	5.866,2.839 (13.811)	0.0144
48s inj	511	9.899,1.717 (15.282)	3.974,6.032 (17.936)	0.0150





Experiment 27: *STRmix parameter setting – Populations and allele frequency files*

The allele frequencies from the FBI Dataset for the Caucasian, African American, and Hispanic populations were used for this experiment instead of the NIST 1036 US Population Dataset. This allele frequency data is currently available via "Tamyra R. Moretti, Lilliana I. Moreno, Jill B. Smerick, Michelle L. Pignone, Rosana Hizon, John S. Buckleton, Jo-Anne Bright, Anthony J. Onorato, Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States, Forensic Science International: Genetics http://dx.doi.org/10.1016/j.fsigen.2016.07.022."

The allele frequency tables were generated as described in the STRmix User's manual and loaded into STRmix. Initially the N value included in the table was set to the individual population number rather than the number of alleles but this was corrected and the tables were reloaded. The table reload was timed to ensure no other experiments were adversely impacted by the change.

The population groups were edited using information from the U.S. Census data website http://quickfacts.census.gov/qfd/states/00000.html. The number of children per family was set to four and the point estimate for theta was set to 0.01.

Experiment 28: *STRmix performance verification – Check of the likelihood ratio*

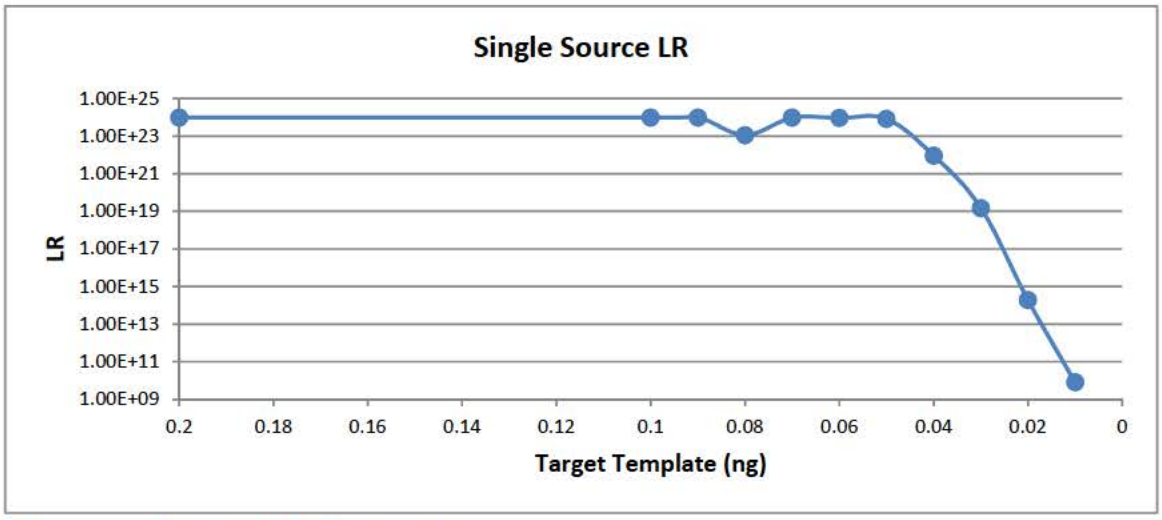
The likelihood ratios (LRs) for a single source profile without dropout were calculated in Excel and match the STRmix (v2.3) outputs. The LRs were also calculated for a single source profile with dropout. While the Excel values don't match the STRmix outputs exactly, the differences can likely be attributed to rounding per the STRmix Implementation and Validation guide. More calculations are involved in generating the LRs for profiles with dropout so the rounding differences will be compounded for these profiles.

Hand-calculations were achieved using spreadsheets provided by the STRmix training team. The single source calculation with dropout was repeated using STRmix (v2.4), to ensure the continuation of expected functionality.

	African American		Caucasian		Hispanic	
Sample	STRmix	Excel	STRmix	Excel	STRmix	Excel
No dropout, v2.3 and 2.4	7.44E+28	7.44E+28	3.66E+26	3.66E+26	3.19E+28	3.19E+28
With dropout v2.3	5.32E+21	6.00E+21	2.96E+19	3.23E+19	8.27E+20	9.80E+20
With dropout v2.4	7.73E+21	8.70E+21	4.45E+19	4.90E+19	1.15E+21	1.38E+21

Experiment 29: *STRmix performance verification – Check of the weights*

A single source profile was amplified across a range of template targets (0.2ng to 0.01ng) and analyzed with STRmix to generate an LR using the Caucasian database. As template decreases, the LR is expected to also decrease.

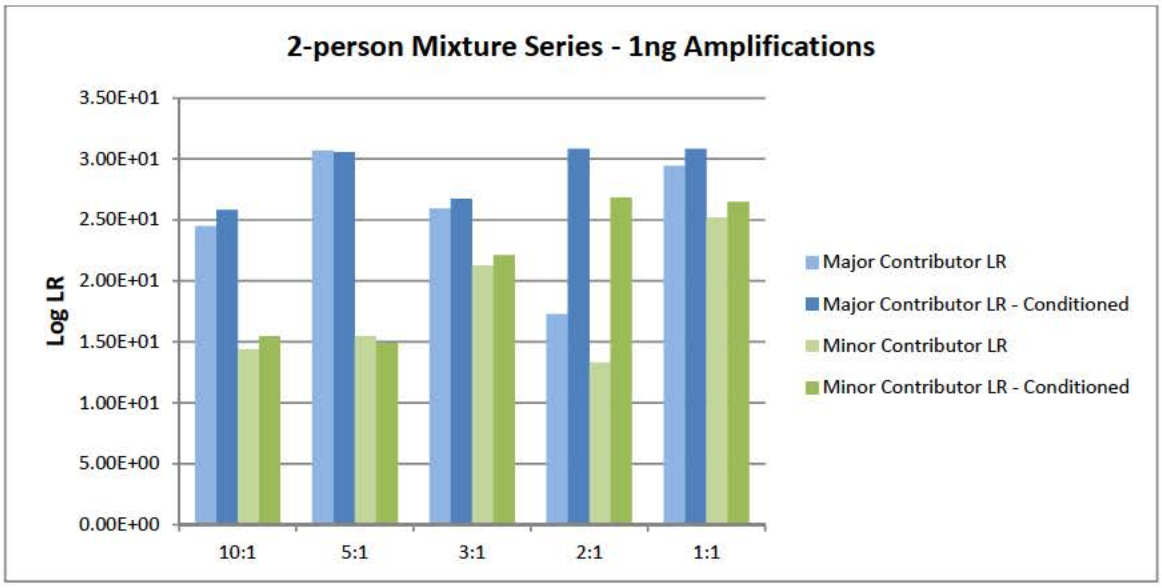


*Calculations achieved with STRmix v2.3

In addition, five 2-person mixtures were created in ratios of 10:1, 5:1, 3:1, 2:1, and 1:1. The mixtures were amplified at template targets of 1ng and 0.5ng. The amplifications were run in STRmix and the likelihood ratios of the known contributors were calculated using the Caucasian database (STRmix v2.4).

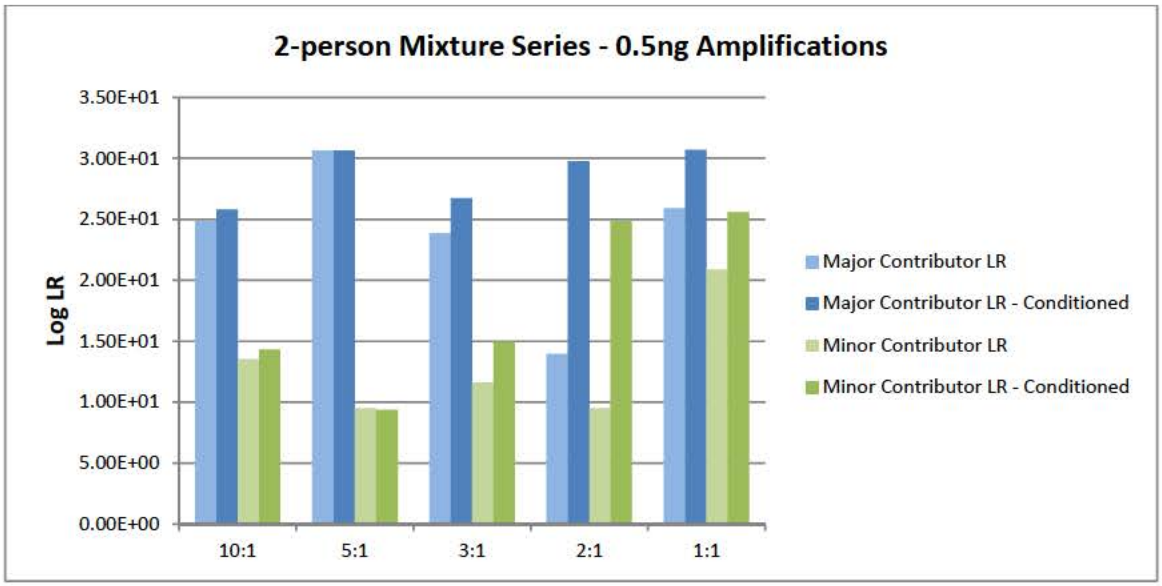
While the experiment was designed to show that as the ratios become closer, the major contributor LRs should trend downwards, the use of different individuals as the major contributor made this assessment difficult to perform across all five ratios; therefore, that data is not shown. Instead, the focus of the assessment of these results is to demonstrate the effect that conditioning the deconvolution on a given contributor can have on the resultant LR.

For the 1ng amplifications, the log of the resulting LRs are as follows:



Overall, conditioning had a positive effect on the LR, resulting in a higher value. In addition, the major contributor LRs were consistently higher than those of the minor contributors. As previously mentioned, caution should be taken in attempting to determine any trends in the LR related to contributor ratios, as each of these mixtures is comprised of different individuals with different allele frequencies.

The 0.5ng amplifications demonstrated the same trend as the 1ng amplifications.



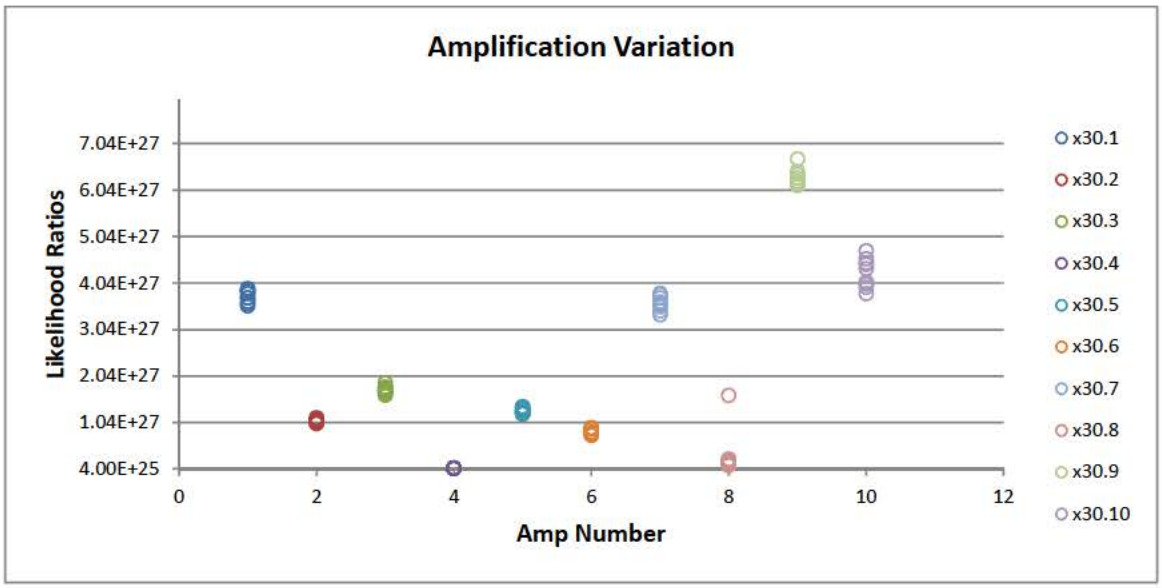
Upon implementation, analysts will be directed to condition deconvolutions on known contributors who can reasonably be assumed to be present on evidentiary items (e.g. conditioning intimate samples or items such as steering wheel swabs), to ensure that the resulting LR is appropriately representative of the competing hypotheses in the case.

Experiment 30: *STRmix performance verification – Reproducibility*

An analysis of *amplification variation* was performed using ten amplifications of a two person major/minor (1:5) mixture analyzed on a 3500xL genetic analyzer. The ten amplifications were then interpreted with STRmix (v2.3), each ten times, comparing the DNA reference standard of the minor contributor to each mixture. The resultant interpretations and likelihood ratios (Caucasian database only) were checked for correctness and statistical consistency. The likelihood ratios were as expected (see table and plot below). There was minimal variation for each interpretation by STRmix, but the order of magnitude never changed—the exception being amplification no. 8 (x30.8), where one interpretation gave a likelihood ratio of E+27 whereas the other nine interpretations were E+26. At values this high, however, this difference would not likely change the reported weight of the match/inclusion. There was also some minimal variation between the amplifications, but again, not enough variation to change a reported match/inclusion.

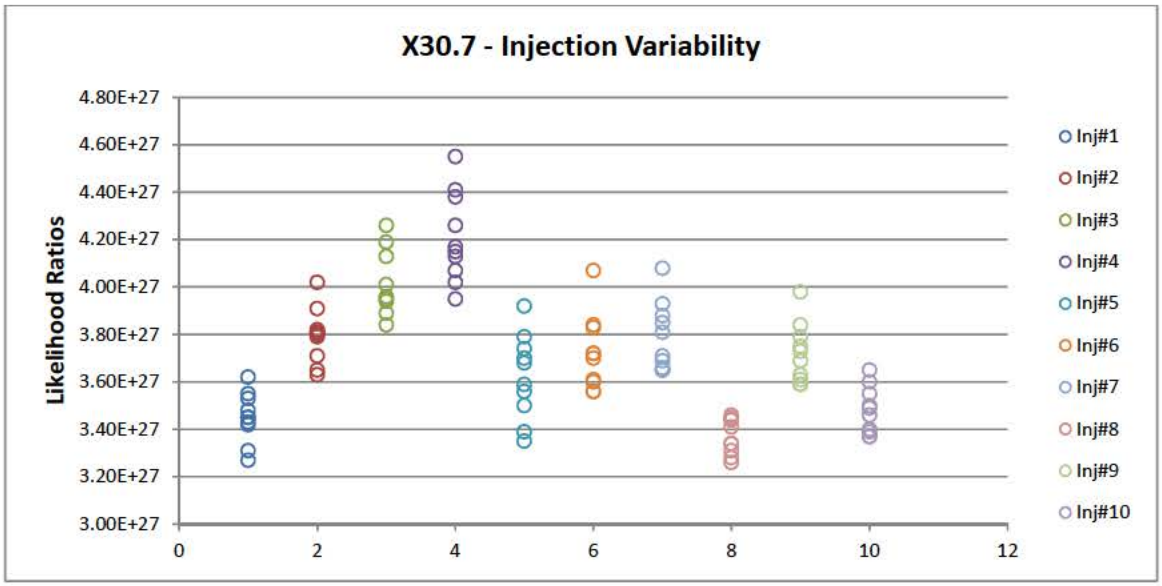
Amp	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
x30.1	3.9E+27	3.9E+27	3.6E+27	3.7E+27	3.6E+27	3.6E+27	3.9E+27	3.9E+27	3.8E+27	3.7E+27
x30.2	1.0E+27	1.0E+27	1.0E+27	1.1E+27	1.1E+27	1.1E+27	1.1E+27	1.1E+27	1.1E+27	1.1E+27
x30.3	1.9E+27	1.8E+27	1.8E+27	1.7E+27	1.7E+27	1.7E+27	1.8E+27	1.7E+27	1.6E+27	1.7E+27
x30.4	6.2E+25	6.2E+25	5.3E+25	6.0E+25	4.5E+25	6.0E+25	5.3E+25	5.7E+25	5.2E+25	5.1E+25
x30.5	1.2E+27	1.3E+27	1.3E+27	1.4E+27	1.3E+27	1.4E+27	1.3E+27	1.3E+27	1.2E+27	1.3E+27
x30.6	8.2E+26	8.2E+26	7.6E+26	8.1E+26	8.3E+26	9.0E+26	9.2E+26	7.9E+26	9.3E+26	9.0E+26
x30.7	3.6E+27	3.6E+27	3.7E+27	3.5E+27	3.8E+27	3.4E+27	3.8E+27	3.6E+27	3.4E+27	3.6E+27
x30.8	1.2E+26	2.5E+26	2.0E+26	2.3E+26	1.5E+26	1.6E+26	2.0E+26	1.6E+27	1.6E+26	2.3E+26
x30.9	6.2E+27	6.1E+27	6.3E+27	6.4E+27	6.2E+27	6.2E+27	6.2E+27	6.3E+27	6.4E+27	6.7E+27
x30.10	3.8E+27	4.4E+27	4.7E+27	4.6E+27	4.0E+27	4.1E+27	4.5E+27	4.4E+27	4.0E+27	4.0E+27

Represented graphically, the likelihood ratios obtained from the replicate analyses of each of the amplifications is as follows:



In addition, an analysis of CE *injection variation* was performed using one amplification from the previously mentioned ten. This single amplicon was injected on a 3500xl genetic analyzer over the course of ten separate injections. Each injection was then interpreted using STRmix ten times each. The variability, for the CE injections, was less than it was between the amplifications (see table and graph below).

Injection	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
1	3.4E+27	3.5E+27	3.6E+27	3.5E+27	3.4E+27	3.5E+27	3.6E+27	3.3E+27	3.3E+27	3.5E+27
2	3.8E+27	3.8E+27	3.7E+27	3.7E+27	3.8E+27	3.9E+27	4.0E+27	3.6E+27	3.8E+27	3.9E+27
3	4.0E+27	4.3E+27	4.0E+27	3.9E+27	3.9E+27	4.1E+27	3.8E+27	4.2E+27	4.0E+27	4.0E+27
4	4.0E+27	4.3E+27	4.1E+27	4.1E+27	4.4E+27	4.4E+27	4.6E+27	4.0E+27	4.2E+27	4.2E+27
5	3.4E+27	3.6E+27	3.5E+27	3.8E+27	3.4E+27	3.9E+27	3.7E+27	3.7E+27	3.7E+27	3.6E+27
6	3.7E+27	3.8E+27	3.8E+27	3.6E+27	3.6E+27	3.6E+27	3.6E+27	3.7E+27	4.1E+27	3.7E+27
7	3.7E+27	3.8E+27	4.0E+27	3.7E+27	3.7E+27	3.7E+27	3.9E+27	3.7E+27	4.1E+27	3.9E+27
8	3.4E+27	3.3E+27	3.4E+27	3.4E+27	3.5E+27	3.5E+27	3.3E+27	3.3E+27	3.3E+27	3.3E+27
9	3.6E+27	3.8E+27	3.7E+27	3.6E+27	3.6E+27	3.6E+27	3.8E+27	4.0E+27	3.8E+27	3.7E+27
10	3.7E+27	3.5E+27	3.4E+27	3.5E+27	3.6E+27	3.4E+27	3.5E+27	3.6E+27	3.4E+27	3.4E+27



The variability observed using STRmix is as expected and as reported by the developers, and is acceptable for the purposes of interpreting DNA profiles.

Experiment 31: *STRmix performance verification – Repeatability*

One of the amplifications (x30.3) was interpreted with STRmix ten times, with the seed manually set before the analysis began. Likelihood ratios for the Caucasian, African American, and Hispanic populations were calculated. The weights and likelihood ratios were identical across all ten calculations.

Experiment 32: *STRmix performance verification – Mixture interpretation and case-type samples*

A multi-pronged approach was taken with respect to the mixture samples. Output diagnostics and genotype weights were assessed and a determination of how the known contributor's genotypes fall within the 99% output was made. The final LRs were calculated, and a comparison of the LR results obtained with the known number of contributors and with the number of contributors observed using maximum allele count was made for a subset of the samples. Mixtures of related individuals (parents and offspring) were assessed with and without conditioning.

In the middle of this assessment, a new version of STRmix was released (v2.4). This new version includes modelling of forward (+1 repeat) stutter, which is critical to the proper evaluation of GlobalFiler samples. Some of the experimentation described below was performed in version 2.3 (the previous version), and the results of those experiments were used to inform the additional experiments to be performed with v2.4. Unless specifically noted, readers can assume that the results described were obtained using v2.4.

In preparation for these assessments, the following amplifications were achieved. Three person mixtures of unrelated individuals were prepared in ratios of 70 / 20 / 10, 60 / 20 / 20, 50 / 25 / 25, 45 / 45 / 10, and 80 / 10 / 10. Four person mixtures were prepared in ratios of 40 / 40 / 10 / 10, 50 / 25 / 12.5 / 12.5, 60 / 20 / 10 / 10, and 70 / 20 / 5 / 5. These mixtures were each amplified at 1ng, 0.5ng, and 0.1ng, and will be referred to hereinafter as the UNR series (UNRelated). Following assessment of the 0.1ng data, we determined that additional data points in the lower amplification ranges were necessary. Therefore, new mixtures at these same ratios (using different contributors) were created and amplified at 0.1ng, 0.05ng, and 0.01ng. These samples will hereinafter referred to as the DS (Dropout Series).

Mixtures of related individuals included the following. Two-person sibling mixtures were created in ratios of 3 / 1, 2 / 1, and 1 / 1. Three-person Parent / Offspring1 / Offspring2 mixtures were created in ratios of 70 / 20 / 10, 20 / 60 / 20, 45 / 45 / 10, and 10/ 80 / 10. Parent1 / Parent2 / Offspring mixtures were created in ratios of 70 / 20 / 10, 20 / 20 / 60, 25 / 50 / 25, 10 / 45 / 45. Finally, four-person Parent/Parent/Offspring1/Offspring2 mixtures were created in ratios of 40 / 40 / 10 / 10, 12.5 / 12.5 / 50 / 25, 10/ 60 / 20 / 10, and 70 / 5 / 20 / 5. Each of these mixtures was also amplified at 1ng, 0.5ng, and 0.1ng.

In addition to the above samples, the mixture samples from Experiments 12 and 29 were included in many of the assessments STRmix interpretation and LR assessments.

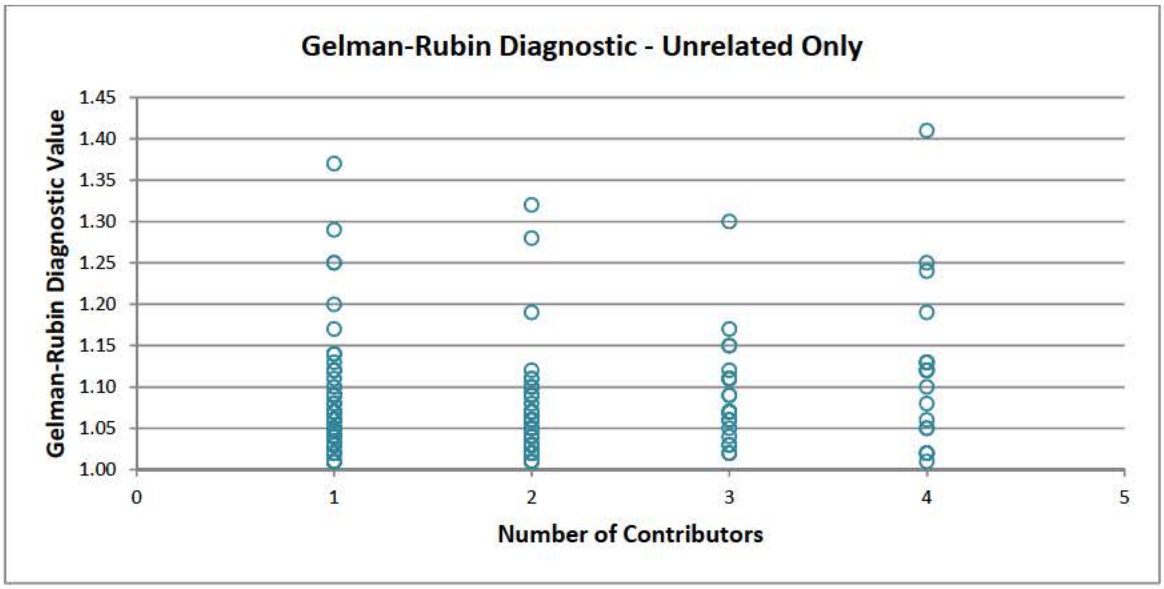
DIAGNOSTICS

The STRmix training team discussed the different outputs that analysts can use to assist in reviewing STRmix outputs. Those that can be most demonstrative of a problem with a specific deconvolution are the average log likelihood, the allele and stutter variance values (and how far they depart from the mode of our Model Maker data), the Gelman-Rubin diagnostic, and the Effective Sample Size Output. Detailed definitions of each of these diagnostics can be found in the STRmix User's Manual.

Starting with the Effective Sample Size, several deconvolutions returned a "NaN" error for the effective sample size. This value is used by STRmix when calculating the HPD value (see Experiment 33 for further detail on HPD). This error is generally the result of a bad STRmix input file; one that is missing a stutter peak. A review of each of the STRmix input files that returned this error showed that, in fact, a stutter peak was missing. Adding this stutter peak to the input file corrected this "NaN" error.

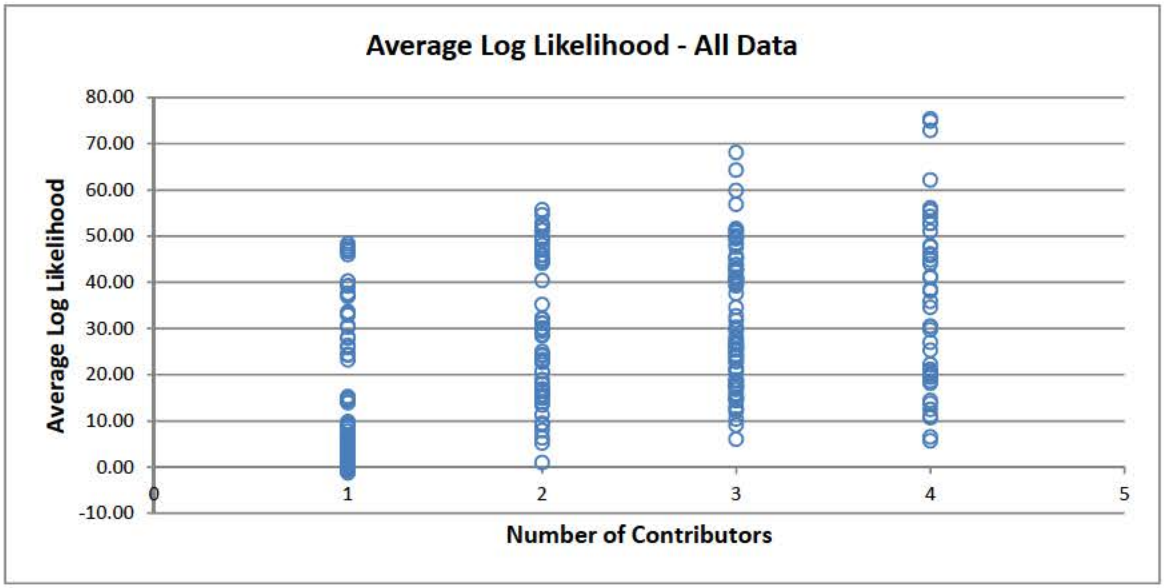
The Gelman-Rubin diagnostic is used to assess how well each of the chains aligned at the end of the prescribed number of MCMC accepts. If the chains are far apart, indicated by Gelman-Rubin value of >1.2, this can be an indication of a problem with the MCMC. It can also be a reflection of the complexity of the input file and an indication that re-running the profile with a greater number of MCMC accepts might be appropriate. Experimentation with v2.3 showed this to be the case (data not described).

For the unrelated data, no Gelman-Rubin value of greater than 1.41 was observed. Six Gelman-Rubin values were observed that were greater than two (2.12, 2.41, 2.43, 3.37, 3.53, 4.44). All of these were from some iteration of the Parent/Offspring mixtures. The fact that the Gelman-Rubin value was high for these mixtures is not surprising, given the complexity of deconvoluting this type of mixture.

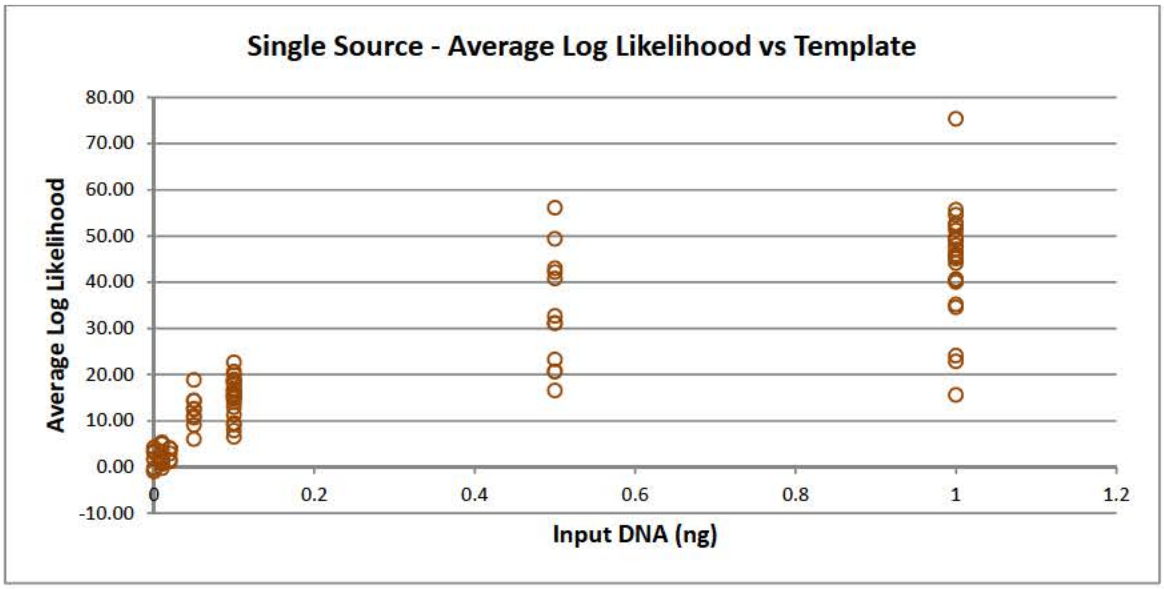


The average log likelihood was assessed for each number of contributors, and also by template for the single-source samples. Generally, a large average log likelihood signals that STRmix was able to model the observed data; however, that alone is not necessarily an indication of a successful deconvolution. Conversely, low average log likelihood values can simply be the result of low-level data, and are not necessarily indicative of a problem with the deconvolution.

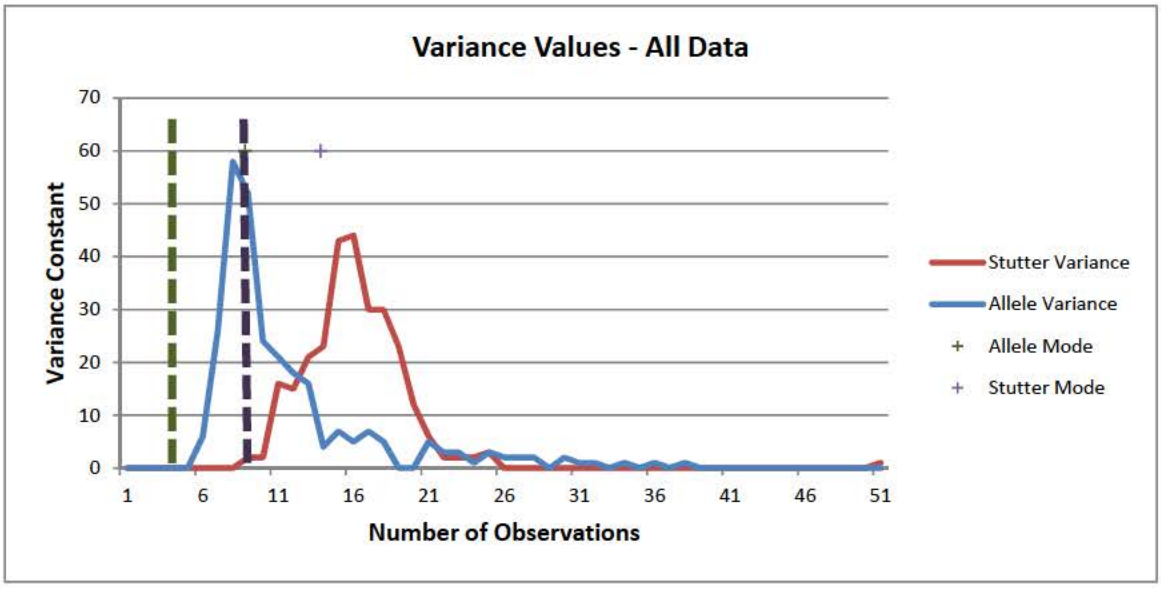
For the single-source and two-, three-, and four-person data, the average log likelihood values obtained are shown below.



Perhaps more demonstrative of the meaning of the average log likelihood, when plotted against the input template for single-source samples, we can see that the value will rise alongside DNA input.



Finally, the stutter and allele variance was assessed. The allele and stutter variance constants were set during Model Maker. The STRmix output includes allele and stutter variance as observed in the entire post burn-in MCMC analysis. If this constant is high relative to the mode of the distribution observed during Model Maker, it can be indicative of a problem with the deconvolution.



While the data largely stayed on-trend with the modes set during Model Maker, specific to the allele variance, large deviations from the mode were observed with the various mixtures of parents and offspring.

It is worth noting that taken individually, these diagnostics should not be used to determine an unsuccessful deconvolution; however, in combination with a review of the genotype weights and final LR, they may assist an analyst in determining that an input file should be scrutinized and/or a sample be rerun.

CHECK OF THE WEIGHTS

There were two separate assessments of the weights. The first was a determination of the exact weight a known contributor's genotype was given within the deconvolution, and the second was whether the known contributor's genotype was represented within the genotypes listed has having >99% weighting in the contributor summary of the STRmix output.

The exact weight for each known contributor was assessed for 177 different mixtures covering 447 separate contributors as follows. The STRmix deconvolution was run, followed by a likelihood ratio for the contributor. Because there are times when two contributors can be linked back to the same STRmix contributor order (e.g., in a mixture of A, B, and C, both B and C give the highest LR as contributor 2, even though in the known mixture makeup, C should be contributor 3), the contributors were run together and the weights assessed based upon the contributor order returned for all pertinent individuals in Hp. Once the contributor order was established, the weight given to the known contributor's genotypes within that contributor order was determined.

Of the 9294 total loci assessed, there were 143 loci (1.5%) covering 82 different contributors to samples where the known contributor's genotype was weighted less than 1%. 56 of the 82 (68%) contributors were iterations of the three- and four-person parent/offspring mixtures. Of the remaining 26 contributors where one or more loci yielded a known contributor genotype of less than 1%, all contributors were estimated to have contributed less than 0.3ng of total input DNA to the sample, and a maximum of three affected loci per sample was observed. Overall, with the exception of the parent/offspring mixtures, the loci having less than 1% weighting had little bearing on the final contributor likelihood ratio.

Regarding the contributor summary, the 99% outputs were assessed against the known contributors. In the 99% weighting, there were 64 of 2963 loci (2.2%) that returned 99% values that were inconsistent with the known contributor genotypes. Focusing on the profiles that would be CODIS eligible (having the 8 "old" CODIS Core loci), all of the mixtures comprised of unrelated individuals yielded 99% contributor summaries that were consistent with the known contributor(s). Four mixtures (two Parent1/Parent2/Offspring and two Parent/Offspring1/Offspring2) yielded 99% contributor weightings of 9 and 10 loci that did not include the genotypes of the known contributor and would lead to an incorrect CODIS entry. This is not surprising, however, as these mixtures are not well-deconvoluted by STRmix.

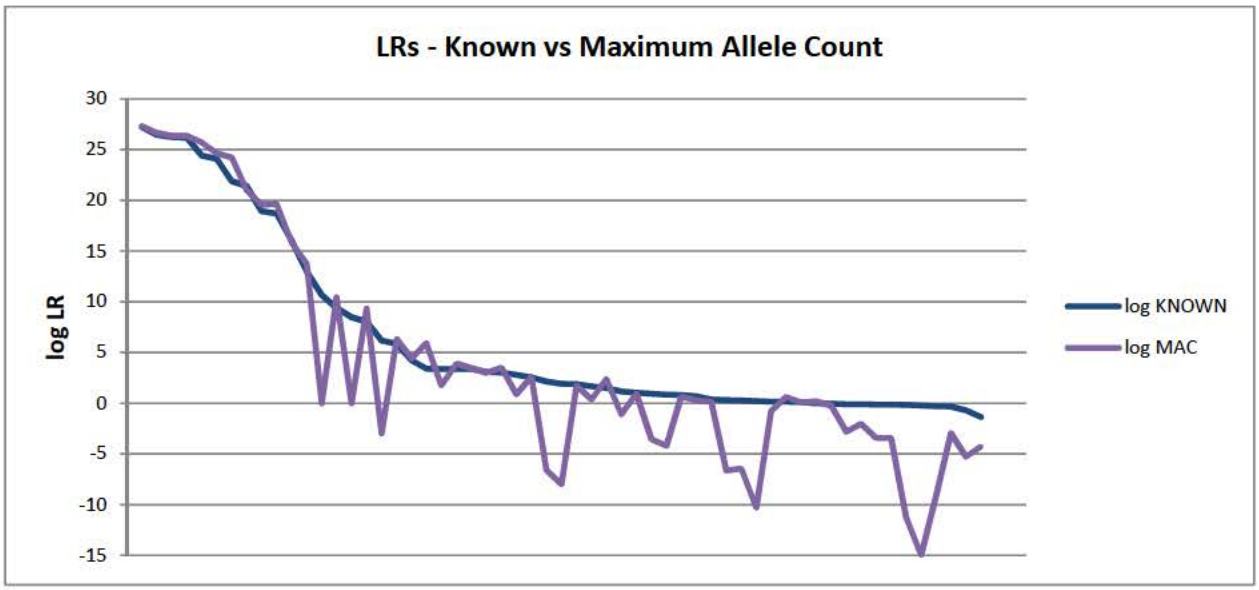
IMPACT OF INCORRECT NUMBER ON CONTRIBUTORS ON LIKELIHOOD RATIO

As is noted in the literature, inputting the incorrect number of contributors into the STRmix deconvolution can impact the resulting likelihood ratio (LR). Too few contributors can lead to false exclusions and an artificial increase in the LR for a true contributor. Too many contributors can decrease the LR for the known contributor and increase the LR for a non-contributor.

In-house, using v2.3, 19 3-person mixtures were analyzed using the known number of contributors and the number of contributors determined by maximum allele count (MAC)[1]. Likelihood ratios were

[1] Note that for a number of these mixtures peak-height ratios would have led to the determination that there were, in fact, three contributors; however, for the sake of the experiment, an assumption of two contributors was made.

calculated for each known contributor based upon the STRmix deconvolution. The resulting LRs are shown below.



As described in the literature, in general, inclusionary LRs were largely unchanged if not slightly elevated by the incorrect assumption of the number of contributors. Lower LRs, those in the inconclusive range, showed a more significant change. Notably, two LRs went from an inclusionary value (4.56E+10 and 2.88E+08) to an LR of 0. These were two contributors to the same mixture (80 / 10 / 10, amplified at 0.5ng).

MIXTURES OF RELATED INDIVIDUALS

Overall, the performance of STRmix on mixtures of related individuals was mixed. One of the challenges faced in deconvolution of mixtures consisting entirely of parents and offspring is the assignment of stutter peaks as true contributor peaks, given that STRmix is more likely to weight a stutter peak as a contributor peak as opposed to borrowing allelic product from a slight imbalance in the known contributor peaks. As described above, a large proportion of the incorrect weighting observed was within the data set comprised by these related mixtures.

Surprisingly, the parent/offspring mixtures amplified at the lower template targets (0.5ng and 0.1ng) performed better than those amplified at 1ng. Theoretically, the RFU of the higher inputs would be such that variability in total allelic product would be less likely; therefore, the shared genotypes (e.g. contributor 1 and 2 both as [redacted]) would be given little to no weight and the stutter peaks (e.g. contributor 1 as [redacted] and contributor 2 as [redacted]) would be given greater weights. This was specifically observed at D2S441 where parent 1's genotype is [redacted], parent 2's genotype is [redacted], and the genotype of both offspring is [redacted]. The pairing of the [redacted] as a minor contributor to the mixture was not given weight by STRmix, especially in those that included conditioning on one or both parents. Therefore, even though LRs supporting inclusion were observed at all other loci for offspring 2, a final LR result of zero was returned. Further discussion on the likelihood ratios observed with mixtures of related individuals can be found in the summary for Experiment 34 (Defining the inconclusive range).

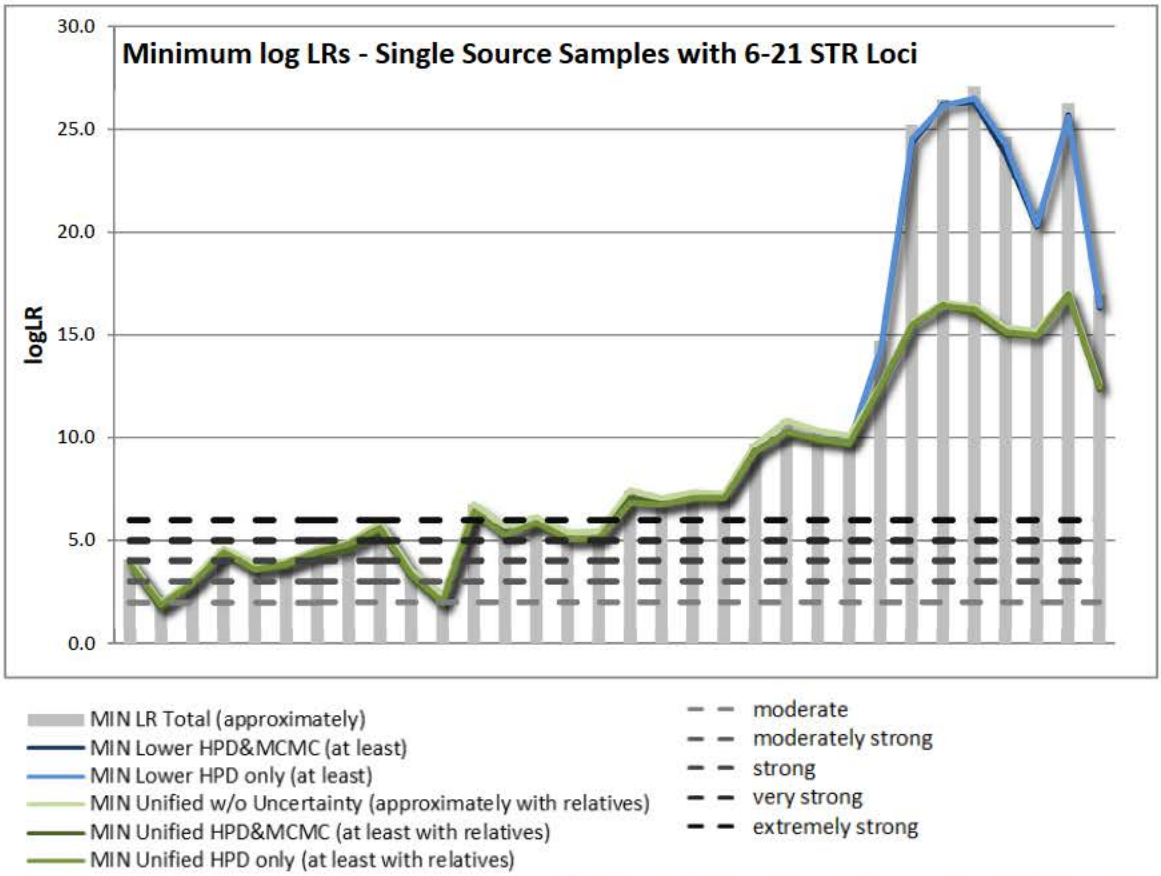
In all, we observed that STRmix can provide appropriate deconvolution of mixtures of parents and offspring on a case-by-case basis. As we develop our corresponding STRmix protocols, we will ensure that the limitations of these types of mixtures are adequately addressed.

Experiment 33: *STRmix performance verification - Assessment of likelihood ratios*

This experiment was performed using STRmix version 2.3. Forty-nine single source samples were assessed with STRmix. Each available type of LR was assessed: The point-estimate likelihood ratio (LR total), the unified likelihood ratio, which accounts for relatives in the population, the stratified likelihood ratio, which corrects for the distribution of each racial group in the population, the 99% 1-sided lower highest posterior density (HPD) likelihood ratio, which accounts for sampling uncertainty within the population with respect to allele frequencies, and likelihood ratios that include an adjustment for the uncertainty in the MCMC process.

The first assessment was of the stratified LR option. The stratified LR was consistently higher than the minimum LR across the three population groups assessed; sometimes by several orders of magnitude. Reporting the stratified could lead to an LR being reported for a defendant of a given racial background that is higher than that of his or her own population group; therefore, we determined that the stratified LR will not be implemented and it was not assessed further.

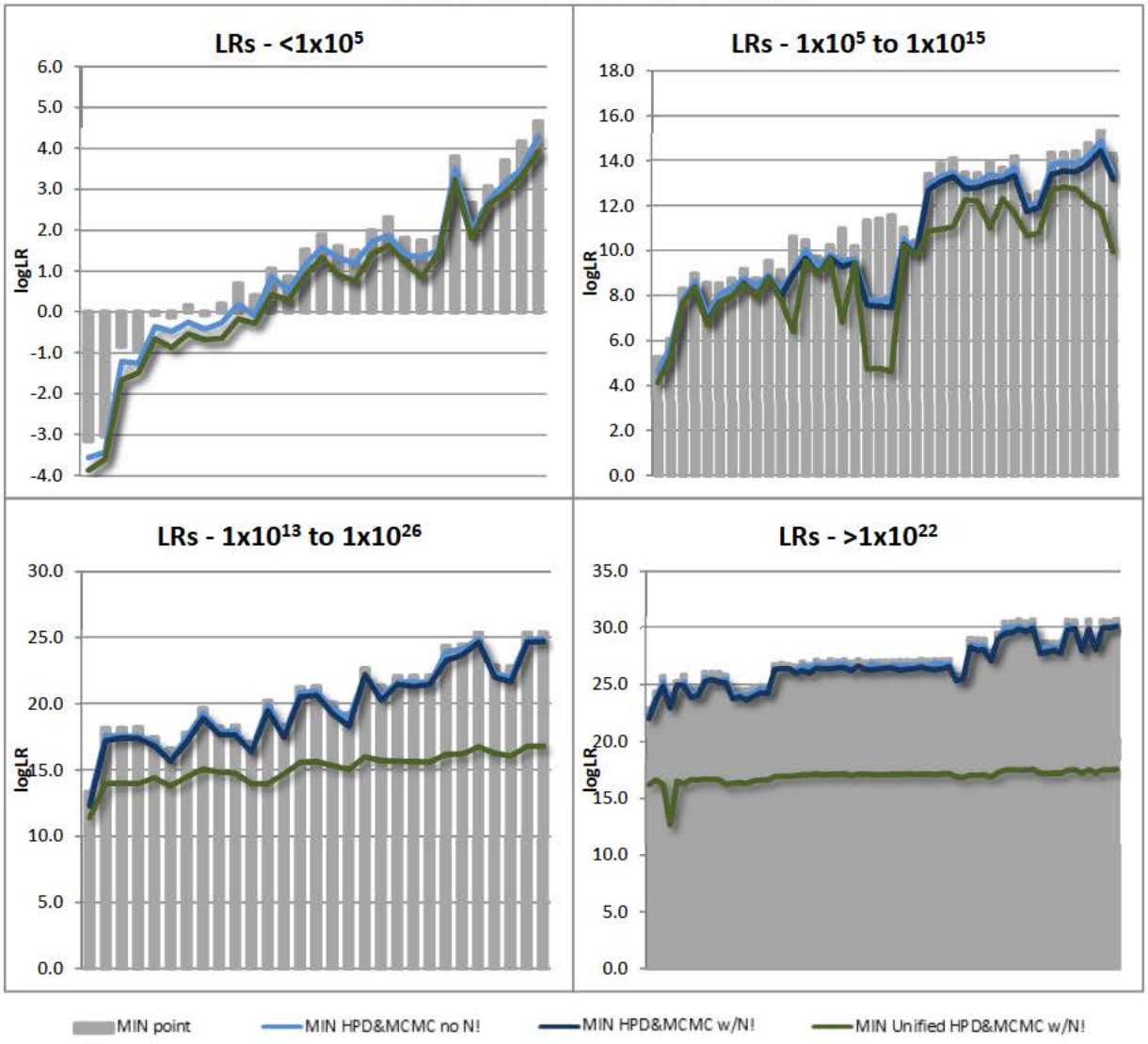
The remaining LR values were assessed. As expected, as the number of loci increased, so did the LR. In addition, there was very little difference in the LR total, and the LRs with HPD and with and without MCMC. On average, the LR including the HPD and MCMC uncertainty was approximately 6% less than the point-estimate.



The biggest impact on the LR was determined to be the incorporation of the unified option although, unexpectedly, it didn't diverge from the point-estimate until the LR became in the range of $1x10^5$ to $1x10^{10}$. Upon consultation with the STRmix training team, we learned the reason why. Depending on the population size and the number of siblings per family input into STRmix, the unified LR assumes that greater than 99.99% of the population is unrelated, and that the remaining portion of the population is related to the person of interest. That 0.01% has a small impact on a small LR, but can have a bigger impact on a bigger LR.

Based upon the assessment with and without the MCMC uncertainty incorporated into the HPD LR, the comparison of LR total to MCMC&HPD uncertainty, to include the unified values, was undertaken for 2-person mixtures. This assessment of 163 LRs included examining LR values with and without the factor of N!.

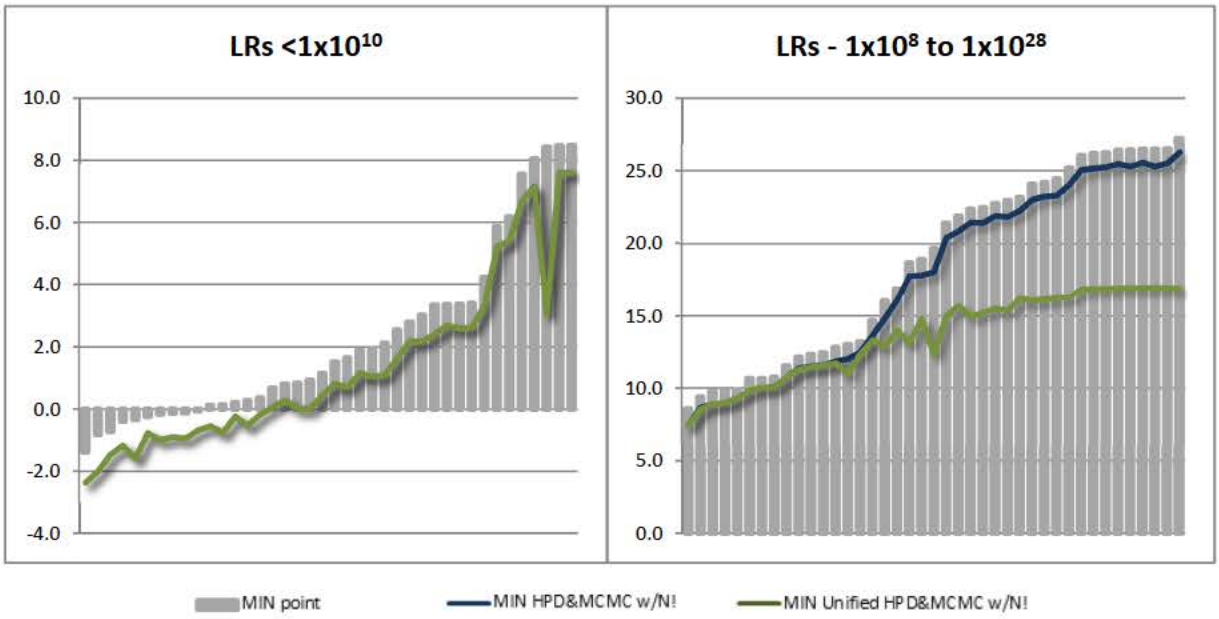
MINIMUM 2-PERSON MIXTURE LRs



As with the single-source samples, the HPD&MCMC uncertainty did not significantly lower the LR relative to the point estimate, although at an average of 8% difference, it was slightly more pronounced. The introduction of the N! also did not dramatically impact the LR.

To complete the assessment, the 3-person mixture samples were examined. Because of some of the challenges with deconvolution of the 3-person mixtures of parents and offspring, only the LRs resulting from analysis of the 3-pesron unrelated contributor mixtures are described. To that end, 81 LRs were calculated for the 3-person mixtures.

MINIMUM 3-PERSON MIXTURE LRs



For the 3-person unrelated mixtures, the average % difference in the point-estimate and the LR with HPD and MCMC (and N!) was 18%. This increase over what was observed in the single-source and 2-person samples was not unexpected, given that the N! now considers 6 combinations (3!=3x2x1=6) instead of 2 combinations (2!=2x1=2), and that we expect there to be greater uncertainty in the MCMC as the mixture becomes more complex.

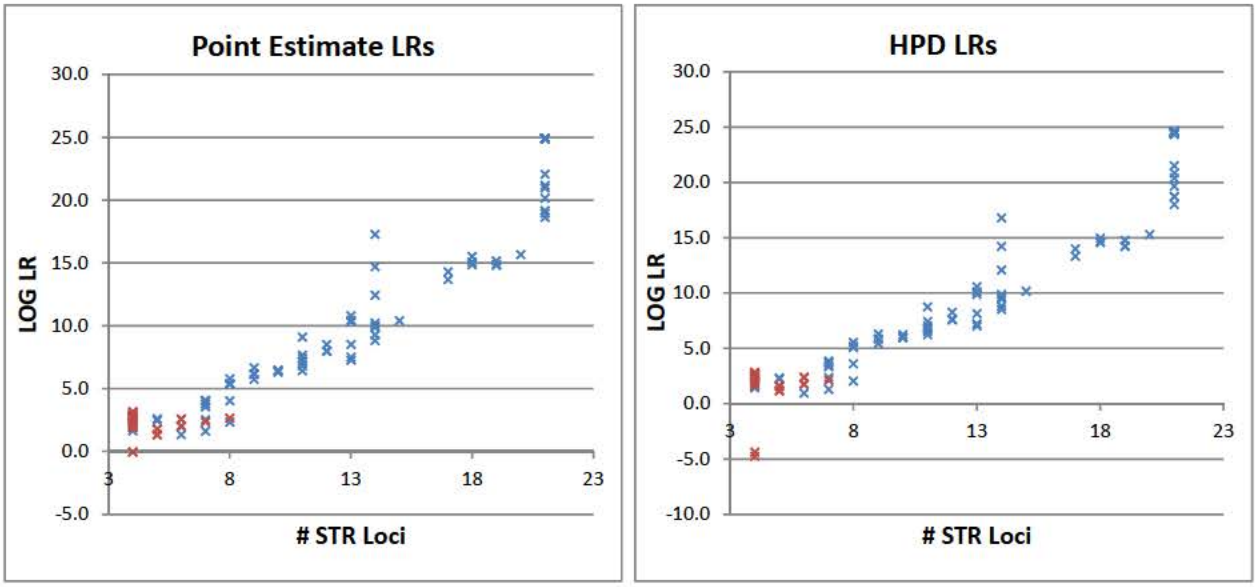
The unified LR, while appealing in that it better-represents the real-world population that includes relatives of the person of interest, relies heavily on user-input values for the number of children per family and the size of the population. Based upon these assessments, we will report the minimum HPD/MCMC/N! LR for probative associations. The HPD and MCMC allow for reporting language of "at least" versus "approximately," and we will incorporate some language into the report appendix describing the LRs from individual ethnic groups are available in the case notes and are available upon request. The N! adjustment addresses the individual's contribution at the sub-source level (the DNA in the sample came from the person of interest) as opposed to the sub-sub-source level (the major profile in the sample came from the person of interest), although we may continue to report associations to a major contributor as dictated per protocol.

Experiment 34: *STRmix performance verification –Defining the inconclusive range*

Each STRmix v2.4 deconvolution from Experiment 32 was assessed against an elimination database containing 507 profiles. Contained within the database were the known contributors to the mixtures and also relatives of the individuals comprising the various parent/offspring mixtures. Point estimate and HPD/MCMC/N! LRs (hereinafter referred to as "HPD LRs") were calculated for all known contributors and for up to 10 non-contributors, the non-contributors identified via a search of the

elimination database. "False negatives" were defined as LR values less than 1 for known contributors. "False positives" were defined as LR values greater than one for non-contributors. The data was ordered in a manner to assess the lower-bound of the values obtained for known contributors and the upper-bound of the values obtained for the non-contributors.

Single-source samples having four or more loci were analyzed for this experiment. No false negative LRs were returned for this data set. False positive LRs were observed in profiles containing fewer than 9 loci.



For the single source data set, false positive LRs were obtained for samples with eight or fewer loci. The maximum false positive point-estimates and the maximum false-positive HPD LRs are as follows:

	Maximum False Positive Point Estimate	Maximum False Positive HPD LR
4 loci	1490	688
5 loci	60.5	43.7
6 loci	384	247
7 loci	231	126
8 loci	468	264

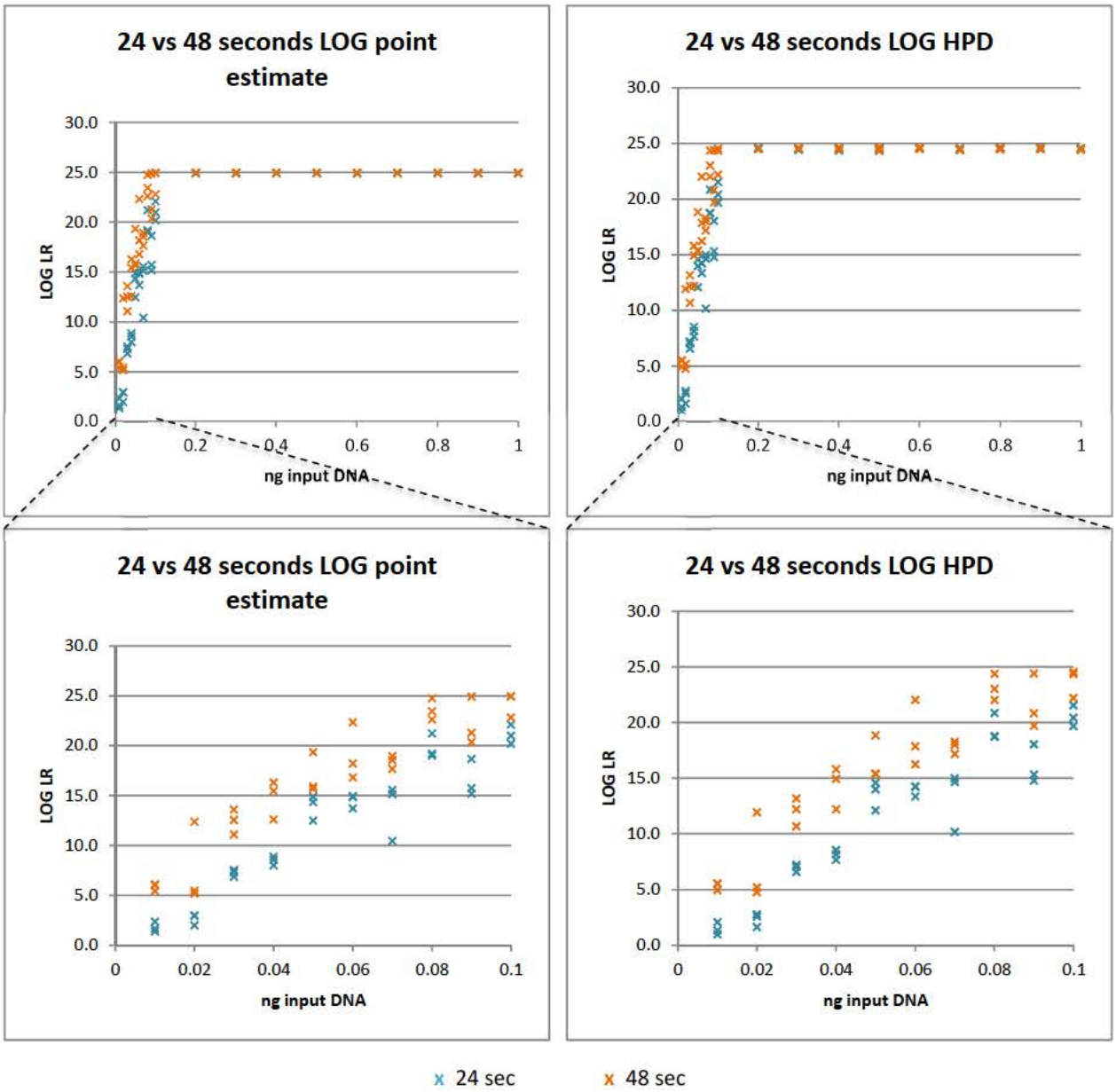
Because the HPD LR is contingent upon uncertainty calculations and is reduced by a factor dependent upon the number of contributors, we are choosing to base the inconclusive range on the point estimate LR as opposed to the HPD. It is worth noting that the minimum point estimate LR can sometimes be in a different population group than the minimum HPD LR.

Based upon the data obtained, we will define LR values of <500 as inconclusive for single-source samples. We will also put some constraints upon the number of loci that can be interpreted. Samples with 5 or fewer STR loci may not be interpreted. Single-source samples having 6 or 7 STR loci may be interpreted at the analyst's discretion, and samples with 8 or more STR loci, be they single-source or

mixed samples, should be further interpreted. In addition, samples must have a minimum of 6 STR loci to qualify for a 48-second injection.

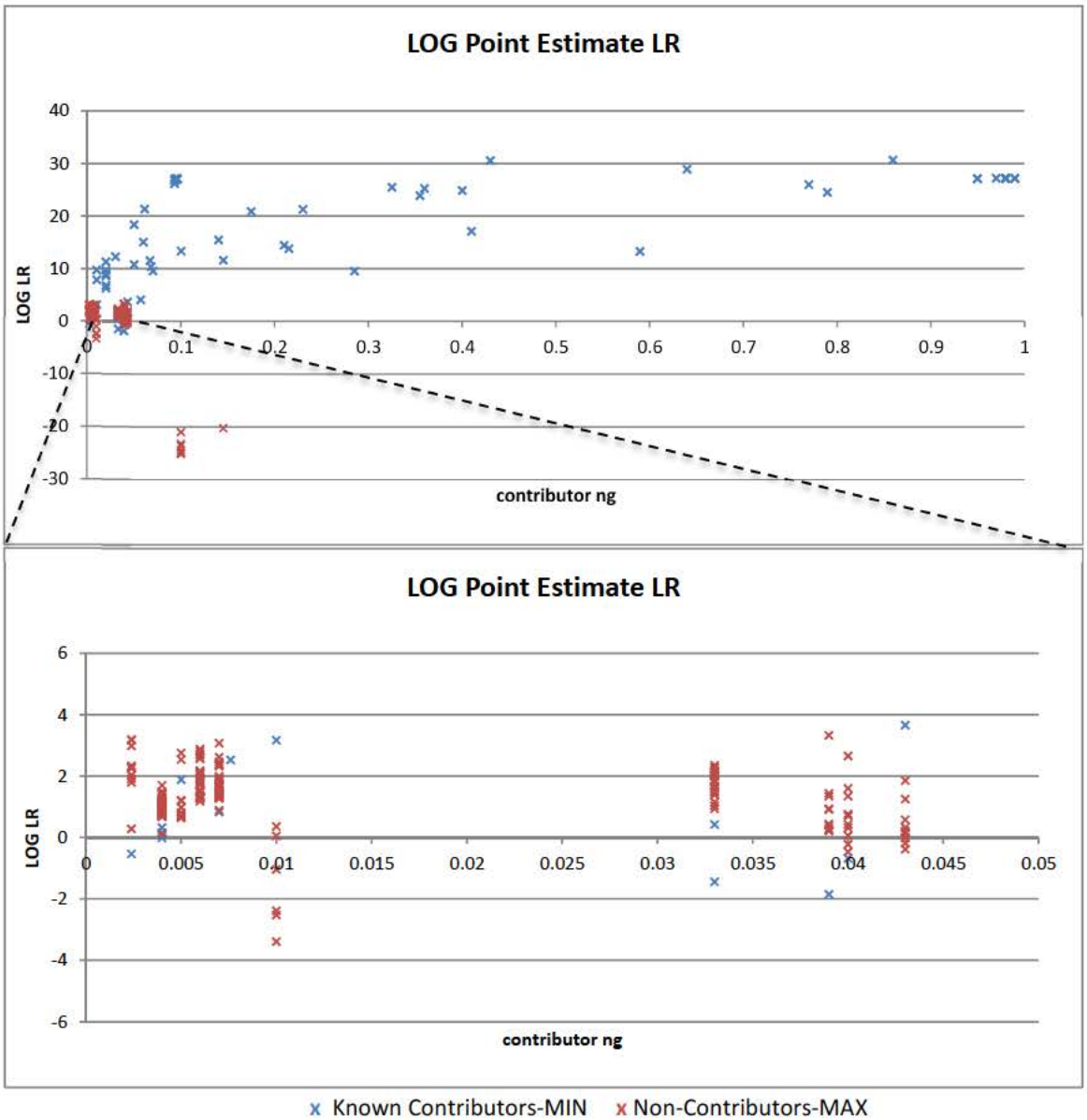
Because no false-negative LRs were returned, there will be no limit on reporting an exclusion to a single-source sample. Any LR of less than 1 supports an exclusion on a single-source sample, and we will allow exclusions from single-source samples to be performed in casework without requiring a STRmix run.

It was during this experiment that the 48-second option for single-source profiles was also further assessed. Both the 24- and 48-second data from the entire dilution series created for one of the contributors in Experiment 21 (1.5ng to 0.01ng), including the data from each 3500xl, was run in STRmix and the resulting minimum point estimate and HPD LRs were collated.



As expected, as the input DNA went up so did the LRs. The 48-second LRs were consistently higher than the 24-second LRs, as they contained more genotyping data. Both the 24- and 48-second converged once complete contributor genotypes were obtained of sufficient strength to eliminate any uncertainty regarding homozygosity. In keeping with the rules for 24-second injection data, we will require a minimum of 6 loci for interpretation and an LR of <500 for any genotyping data resulting from a 48-second injection.

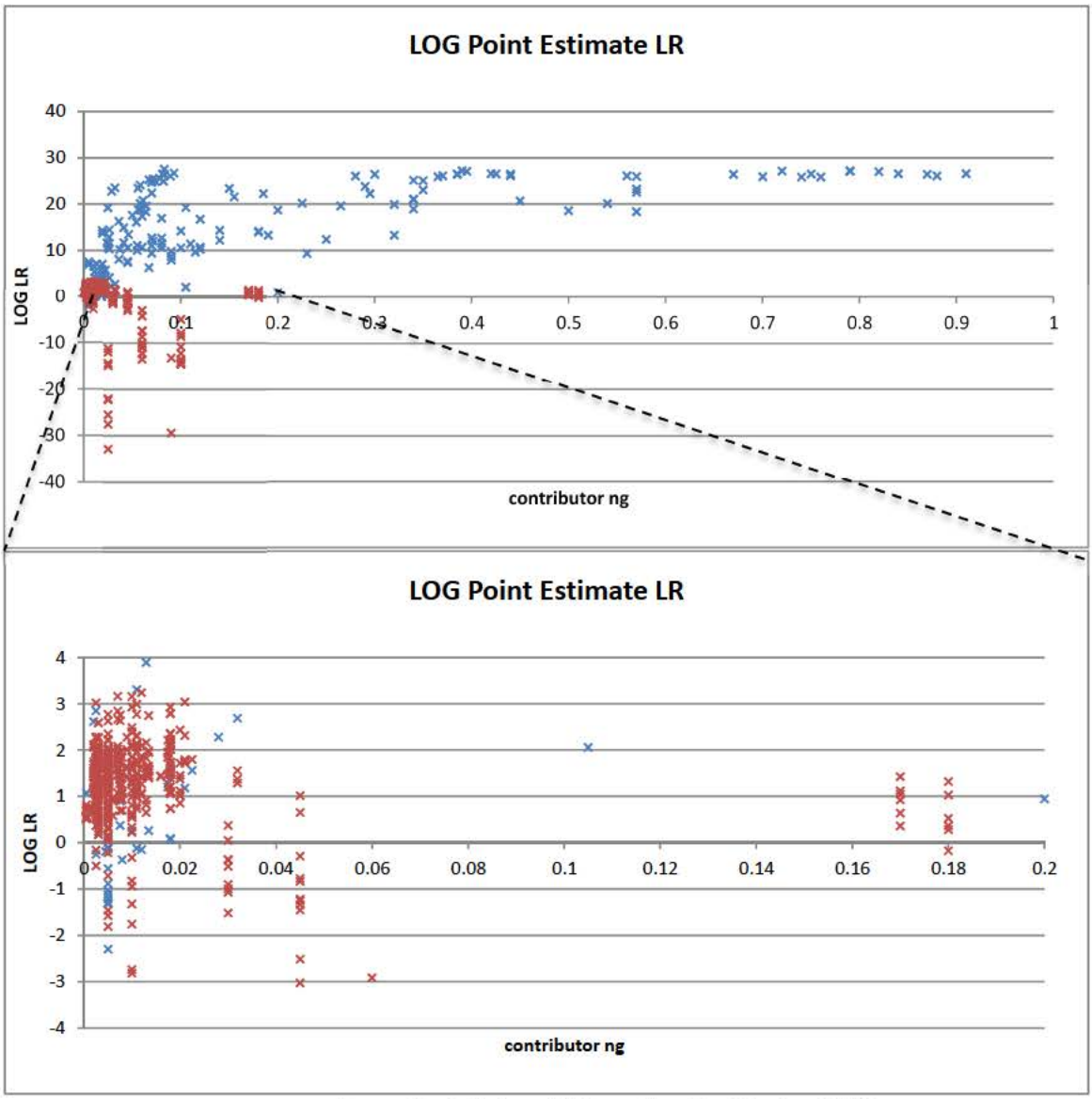
The assessment of the 2-person mixtures was undertaken using estimated ng of total contributor DNA, which was calculated using the STRmix contributor % and the known input of total DNA.



For the two-person mixtures, the maximum false positive point estimate LR obtained was 2130 (2.13E+03) and the minimum false negative was 0.0141 (1.41E-02). These values were both obtained at estimated 0.039ng of input DNA for each contributor. The maximum estimated input DNA yielding a

false-positive LR was 0.043ng. The maximum estimated input DNA yielding a false-negative LR for a known contributor was 0.040ng. This is well within the stochastic range.

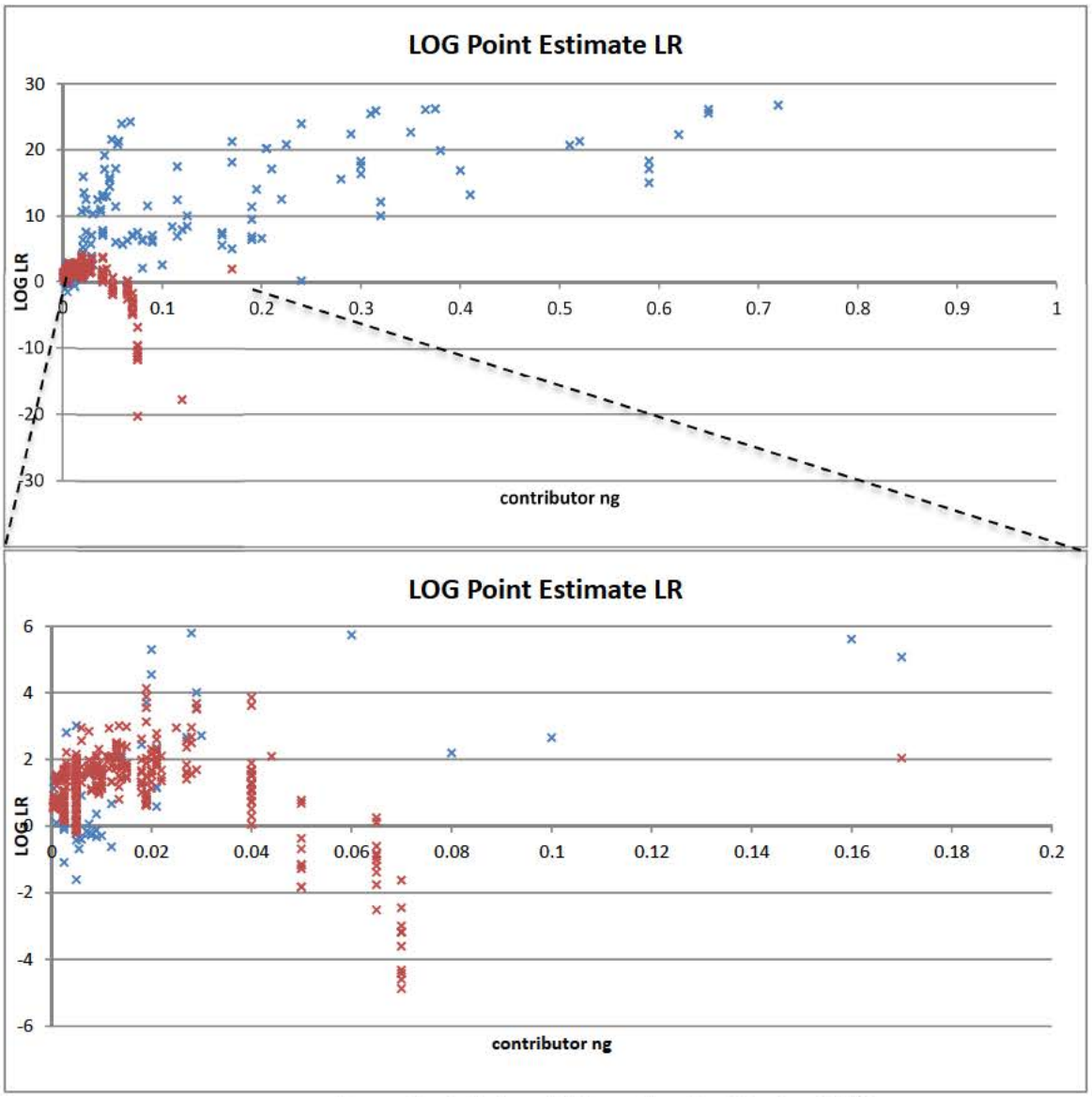
The three-person mixtures were assessed in the same manner as the two-person mixtures. Initially, only the LRs from contributors with a first-order (e.g. parent/offspring) relationship to the known contributors were removed from assessment of the maximum false-positive. Once it was determined that siblings could also return high false-positive LRs (maximum observed value of 8,760,000), all related individuals were removed from the determination of the inconclusive range.



For the three-person mixtures, the maximum false positive point estimate LR obtained was 1710 (1.71E+03) and the minimum false negative was 0.00498 (4.98E-3). These values were both obtained at estimated 0.025ng and 0.005ng, respectively.

It is worth noting here that there were several instances of one of the offspring in the Parent/Parent/Offspring and Parent/Offspring/Offspring mixtures returning an LR of 0. These deconvolutions were all undertaken without any user-informed mixture ratio priors, and the offspring question, by design, was contributing 25% or less of the total DNA input. We will ensure that analyst training includes education on the risk of a false-exclusion in mixtures that can reasonably be assumed to consist of first-order relatives.

The four-person mixtures were handled in the same manner as the three-person mixtures.



For the four-person mixtures, the maximum false positive point estimate LR obtained was 13,500 (1.35E+04) and the minimum false negative was 0.02470 (2.47E-02). These values were both obtained at estimated 0.027ng and 0.005ng, respectively.

To summarize the false positive and false negative data:

Number of contributors	Maximum False Positive Observed	Minimum False Negative Observed
1	468	Not observed
2	2130	0.0141
3	1710	0.00498
4	13,500	0.02470

Follow-up with other laboratories online with STRmix has shown that their inconclusive ranges are symmetrical about the true inconclusive value of 1. Therefore, the following inconclusive ranges will be implemented:

Number of contributors	Inconclusive between		
1	500 (5.00E+02)	and	0.00200 (2.00E-03)
2 and 3	2500 (2.50E+03)	and	0.000400 (4.00E-04)
4	13,500 (1.35E+04)	and	0.0000741 (7.41E-05)

Experiment 35: *STRmix performance verification – Comparison to existing interpretation methods*

While we acknowledge that any direct comparisons between GlobalFiler interpretations and Identifiler Plus interpretations are somewhat flawed given that LR and CPI are vastly different, it is important to understand the GlobalFiler/STRmix combination in the context of the knowledge we already possess about mixture interpretation. Using known three and four person mixtures at differing contributor ratios and template input, samples were amplified using both the GlobalFiler and Identifiler Plus amplification kits and genotyped on their respective instruments. Identifiler Plus profiles were interpreted using current protocols and compared to 10 known non-contributor profiles. GlobalFiler Plus profiles were initially analyzed using the STRmix software, then again against the 10 known non-contributors.

Under current Identifiler Plus interpretation methods, no comparable mixture profile had a non-contributor profile that was not excluded or inconclusive. Comparing known contributors, only one three-person profile included all known contributors in a full mixture profile. The remaining comparable three person mixtures had at least one known contributor excluded because of inconclusive minors. Two four person profiles were full mixes and all known contributors were not excluded, and again, the remaining comparable profiles had at least one known contributor excluded because of inconclusive minors.

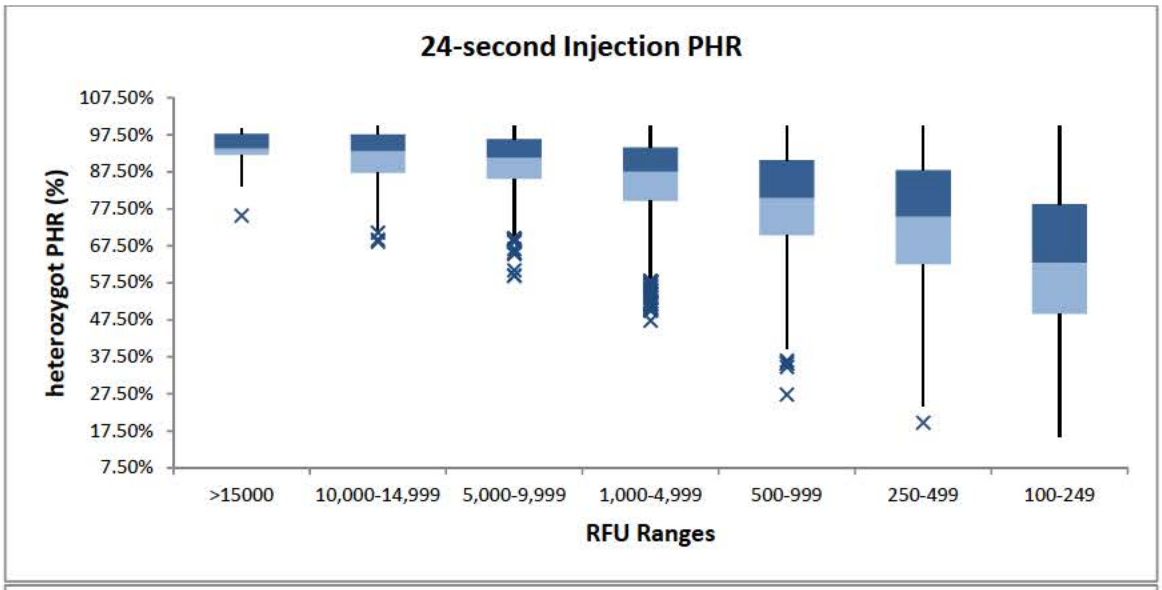
Using GlobalFiler and STRmix, no non-contributor profiles yielded LRs that would indicate inclusion to the mixture profiles. There were three positive LR values at low level template amounts and the LRs were well below the inconclusive range. One known contributor LR was in the inconclusive range in the four person mixtures (0.1ng target profiles were not run), all other known contributors had strong positive LRs. Four out of five low template (0.1ng) three person profiles gave inconclusive LR for minor known contributors. All other three person profiles gave strong positive LRs for the known contributors

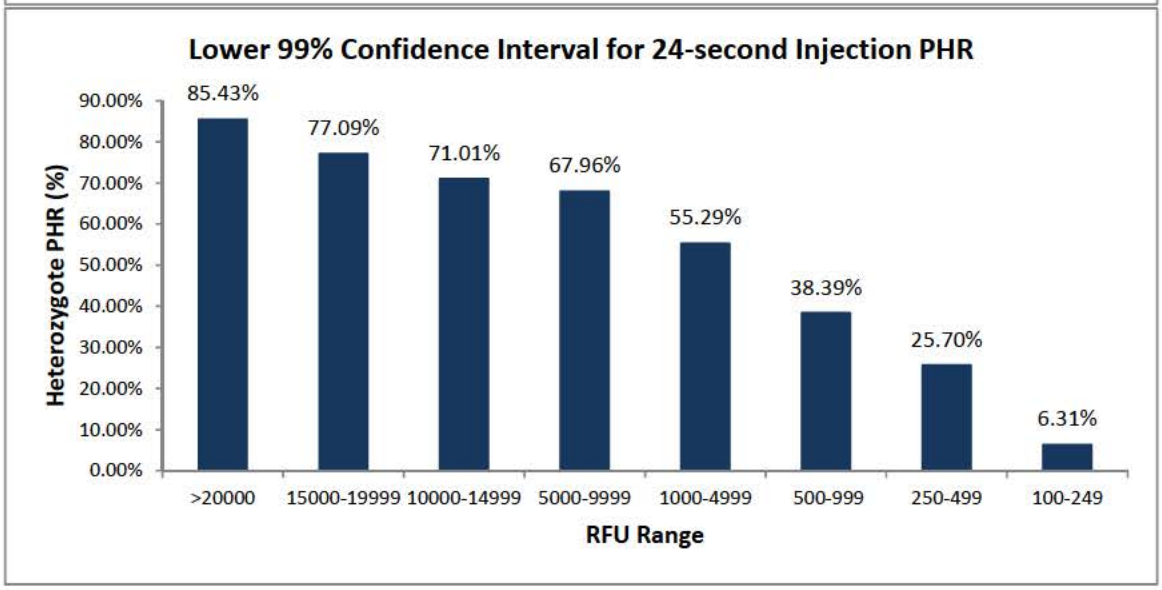
Experiment 36: *Workflow integration – Determination of stochastic threshold and peak height ratio expectations*

PEAK-HEIGHT RATIOS

Peak height ratio (PHR) determinations are critically important when assessing the minimum number of contributors to a DNA profile. The laboratory transition to the utilization of STRmix will greatly enhance the ability to deconvolute mixtures but still requires the analyst to input the number of contributors to a mixture prior to deconvolution.

For this experiment, all of the heterozygote pairs in the samples analyzed in Experiment 21 were grouped into RFU ranges. The height of the smaller peak of the heterozygote pair was used to calculate the PHR. The average, standard deviation, and minimum PHR were also calculated for each RFU range.

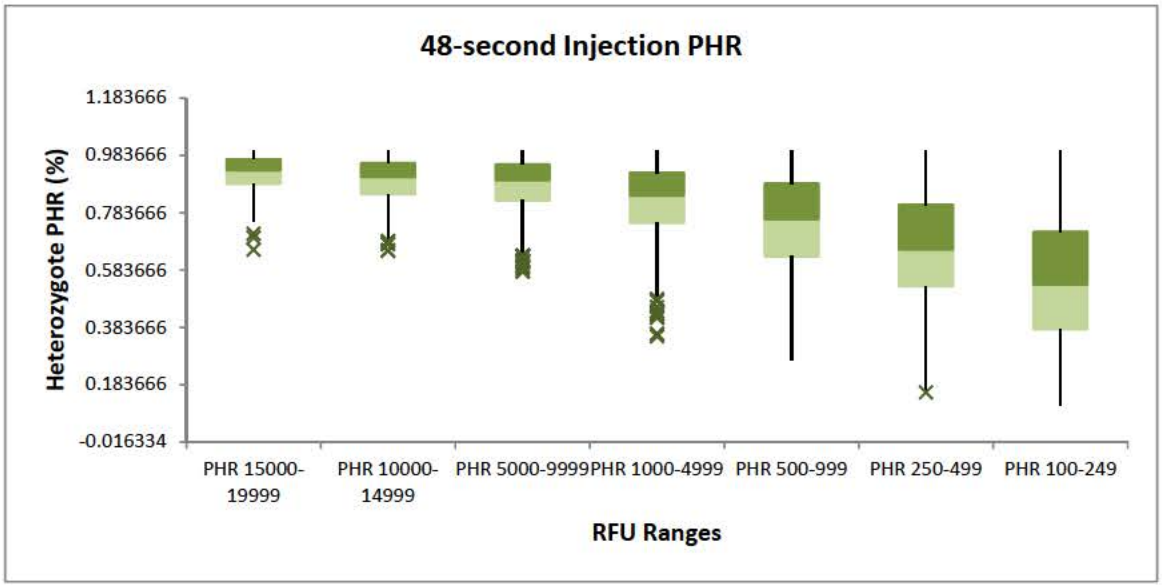


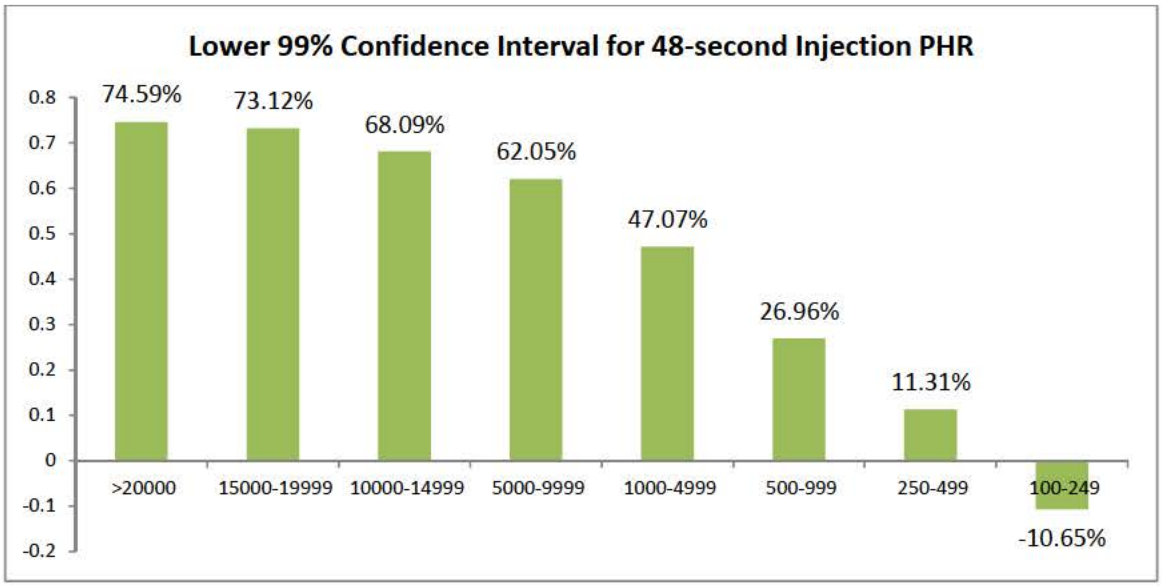


Upon review of the data, we determined that the following ranges will be described in the protocol as general expectations for minimum PHR within each of the ranges. Since the PHR will not be explicitly used in mixture deconvolution but instead will be used to assist in the assessment of the number of contributors, it is important to note that PHR outside of these ranges may be observed in casework samples.

RFU Range	Default Injection PHR Threshold
>15,000	75%
5,000-14,999	65%
1,000-4,999	55%
500-999	40%
250-499	25%
100-249	5%

The same assessment was undertaken for genotyping data obtained from 48-second injections.





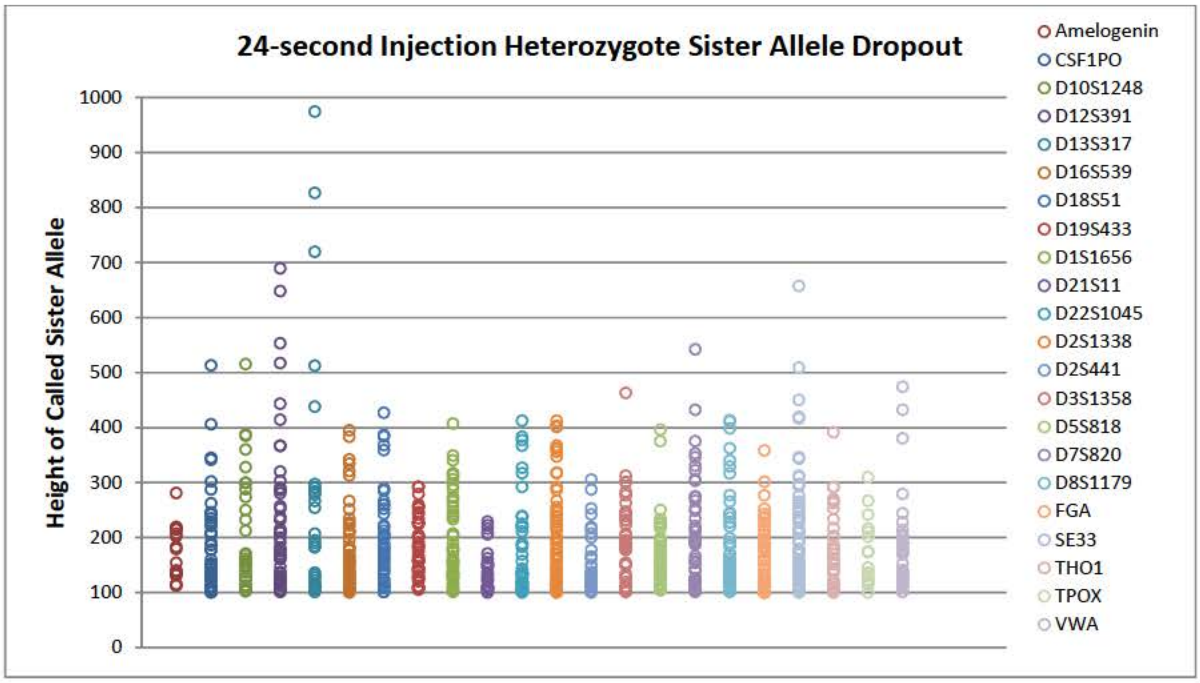
For the 48-second data, the PHR ranges will be described as follows.

RFU Range	Extended Injection PHR Threshold
>15,000	70%
5,000-14,999	60%
1,000-4,999	50%
500-999	25%
250-499	10%
100-249	No minimum PHR expectations

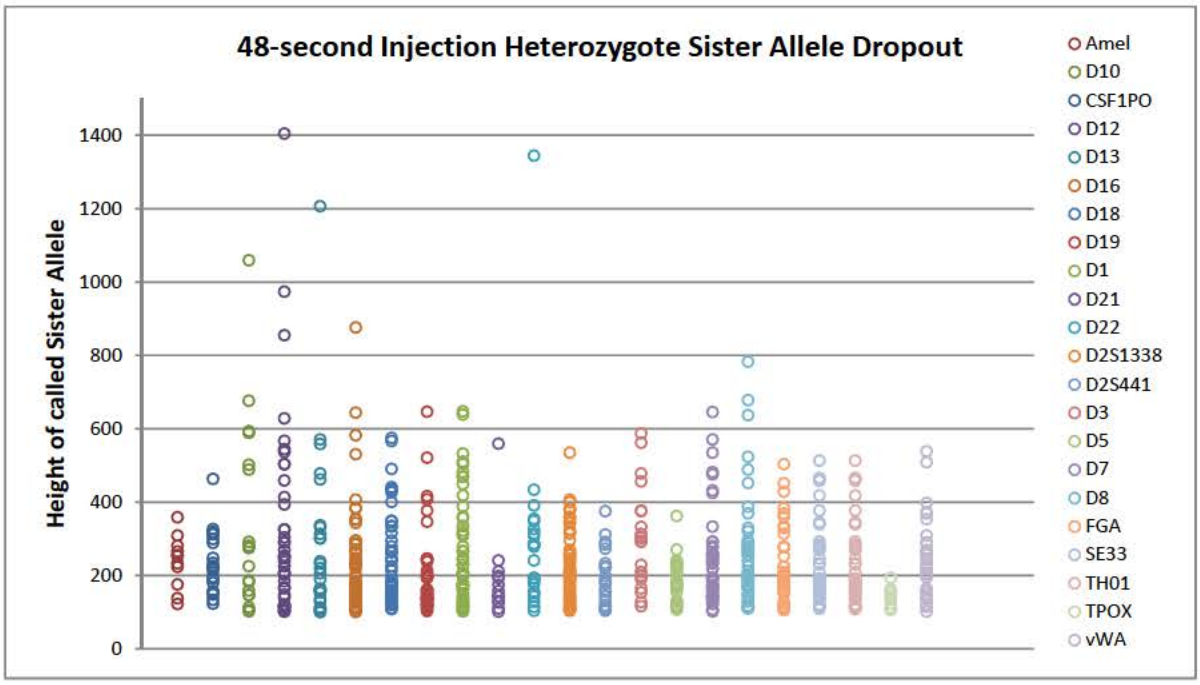
STOCHASTIC THRESHOLD

Stochastic thresholds for the 24- and 48-second injections were determined by using the upper 99% confidence interval for the average height of a peak with an uncalled sister allele. This stochastic threshold will be used in the STRmix workflow in terms of assessing both single-source and mixed samples.

For the 24-second injection data, the average RFU of a peak missing its heterozygote sister was 189RFU. The upper 99% confidence interval (average +3standard deviations) is 470RFU. 500RFU will be implemented as the 24-second injection stochastic threshold. Note that even at 500RFU, there were 13 observations of false homozygotes above the stochastic threshold. For each of these observations, however, the sister peak can be seen below the 100RFU analytical threshold.



For the 48-second injection data, the average RFU of a peak missing its heterozygote sister was 248RFU. The upper 99% confidence interval (average +3standard deviations) is 709RFU.



700RFU will be implemented as the 48-second injection stochastic threshold. Note that even at 700RFU, there were 10 observations of false homozygotes above the stochastic threshold. For each of these observations, however, the sister peak can be seen below the 100RFU analytical threshold.

Conclusions

The transition from binary interpretation using Identifiler Plus and a 3130*xl* to probabilistic genotyping using GlobalFiler on a 3500xL is a significant one. The implementation of both GlobalFiler and STRmix is an improvement in many aspects over our current methodologies and practices. Protocol development based upon this validation will provide further opportunity to standardize DNA analysis in the Oregon State Police system where appropriate, thereby improving the quality of reported results to the criminal justice community. Additionally, multiple laboratories in the U.S. forensic community and some laboratories globally are moving to probabilistic genotyping software for assistance in DNA result interpretation. The incorporation of both GlobalFiler and STRmix will aid Oregon in keeping step with the forensic community.

References

Institute of Environmental Science and Research, *STRmix™ V2.3 Implementation and Validation Guide*, 2014.

Institute of Environmental Science and Research, *STRmix™ V2.3 User's manual,* 2015.

Institute of Environmental Science and Research, *STRmix™ V2.4 Operation manual*, 2016.

Institute of Environmental Science and Research, *STRmixTM V2.4 User's manual,* 2016.

Applied Biosystems, *3500/3500xL Genetic Analyzer User Guide*, Life Technologies Corporation, 2009.

Applied Biosystems, *GlobalFiler™ PCR Amplification Kit User Guide*, Thermo Fisher Scientific Inc., 2015.

Kelly, et al., *A comparison of statistical models for the analysis of complex forensic DNA profiles,* Science and Justice 54 (2014) 66-70.

Taylor et al., *The interpretation of single source and mixed DNA profiles*, Forensic Science International: Genetics 7 (2013) 516-528.

Taylor et al., *Using continuous DNA interpretation methods to revisit likelihood ratio behaviour*, Forensic Science International: Genetics 11 (2014) 144-153.

Bright et al., *Developing allelic and stutter peak height models for a continuous method of DNA interpretation,* Forensic Science International: Genetics 7 (2013) 296-304.

Bright et al., *Degradation of forensic DNA profiles*, Australian Journal of Forensic Sciences 45:4 (2013) 445-449.

Taylor et al., *An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations,* Forensic Science International: Genetics 11 (2014) 56-63.

Brookes et al., *Characterising stutter in forensic STR multiplexes,* Forensic Science International: Genetics 6 (2012) 58-63.

Bright et al., *The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation*, Forensic Science International: Genetics 12 (2014) 208-214.