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

Validation of the GlobalFiler PCR Amplification Kit from Applied Biosystems

Validation Summary

Introduction

The GlobalFiler PCR Amplification Kit from Life Technologies is a multiplex DNA typing kit that amplifies 24 STR loci. In January 2017, the CODIS Core Loci will expand from 13 to 20 (1). The use of this kit will satisfy these requirements (2). The potential for profile discrimination is increased with this many markers. In addition to the new expanded CODIS Core (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391, D2S1338), GlobalFiler contains a sex determining locus (Amelogenin), two Y chromosome loci (DYS391, Yindel) and a CODIS accepted but not required autosomal locus (SE33). D22S1045 has a tri-nucleotide repeat structure, and some of the new STR markers have complex/compound repeat structures and a high level of heterogeneity.

To accommodate the increase of genetic STR markers, this kit utilizes a 6^{th} dye, which minimized primer redesign (important for concordance), kept all required amplicons below 400 nucleotides, and increased the number of mini-STRs (both of which improve performance on degraded samples). Although GlobalFiler utilizes mini-STRs below 220 base pairs, it does not create amplicons smaller than 75 base pairs.

Development of the GlobalFiler kit included: optimization of the master mix formulation (3) and PCR cycling to overcome inhibition, increase sensitivity, and increase efficiency. TPOX and DYS391 are the only primer sequences that were re-designed, and additional SNP-specific primers were added to 8 STR markers to reduce false homozygosity (4).

In addition to the validation of GlobalFiler, this set of studies also includes validation of 3500 Genetic Analyzers, which have the capability to collect data from an additional dye channel. Data from 3500 instruments has an increased dynamic range compared to 3130 instruments, so peak heights and thresholds are all higher than what is utilized with AmpFℓSTR® Identifiler® Plus on a 3130 genetic analyzer. All of the data over the course of this validation was collected using one of two 3500 instruments in the lab, and analyzed using GeneMapper® ID-X Software v1.4.

The following validation experiments were designed to assess the performance of GlobalFiler: Baseline, Sensitivity, Stochastic, Peak Height, Stutter, Challenged Samples, Injection Time, Contamination Assessment, Normalization, Concordance, Reproducibility, and Precision.

Mixtures were also examined after amplification with GlobalFiler. Refer to the STRmix Validation write-up for a detailed description of these results.

Baseline Study
The baseline study established the analytical threshold. Negative controls and samples containing DNA were amplified for 29 cycles and capillary electrophoresis was performed on both 3500 instruments. The noise was evaluated for each dye channel and across the entire profile. Three methods were considered for determining analytical threshold. In addition to calculating based the mean and standard deviation of the noise peaks, some weight was given to the highest noise peaks over the course of the validation. With the goal of maximizing detection of alleles while minimizing artifacts, a threshold of 100 RFU was chosen and assessed throughout the validation. Spectral artifacts were occasionally observed above this threshold, but the magnitude and frequency of these did not warrant an increase in the analytical threshold.

Therefore, 100 RFU was determined to be an appropriate value to use as the analytical threshold for all channels.

Sensitivity Study
The sensitivity study determined the lowest level of obtaining data with the kit and the levels above which we start to see artifacts routinely detection limit, saturation point, and optimal DNA input range. Full DNA profiles were obtained down to ~50 pg in several samples in the dilution series. Partial DNA profiles were observed down to the minimal input tested (~12 pg). Based on ideal peak heights, artifacts and heterozygote peak height balance, the optimal DNA input range is between 0.4 and 0.8 ng.

Stochastic Study
Known samples were amplified multiple times at different target input amounts down to a very low level of DNA in an effort to generate DNA profiles that straddle the 100 RFU analytical threshold. Peak heights as tall as 596 RFU were observed at known heterozygous loci where one peak was detected and the sister allele dropped below the 100 RFU analytical threshold. It is also possible that peaks in this range will have a sister allele that cannot even be distinguished from the baseline noise, therefore, sister peaks to detected peaks are not always expected to be visibly detectable. Based on these observations, below ~600 RFU, drop-out is a possibility. The lower an allele peak height, the more probable that dropout has occurred. The probability of dropout was calculated as part of this study with probabilistic genotyping in mind.

Peak Height Study
Samples amplified within the optimal DNA input range with GlobalFiler demonstrated good inter and intra-locus balance. Robust samples can be expected to routinely generate peak height ratios of greater than 80% with imbalance occasionally reaching as low as 50%. As expected, samples amplified with DNA input amounts below optimal range will begin to show more

imbalance. This relationship is both DNA input amount and peak height dependent. As input amounts decrease and peak heights enter the range in which stochastic effects are expected to be pronounced, peak heights start to drop, and peak height ratios of less than 30% may be observed.

Stutter Study

The levels of N-1 repeat unit (N-4 or N-3 base pairs) stutter were evaluated for known samples amplified within the optimal DNA input range as well as sub-optimal range, since stochastic effects can affect stutter. Allele specific stutter was established at all markers demonstrating stutter. Additional types of stutter were also observed following amplification with the GlobalFiler kit. N+1 repeat unit stutter was seen at several loci, and N-0.5 repeat unit (N-2 base pairs) was seen at three of the loci. These two types of stutter do not seem to be repeat unit specific, and a flat filter is more appropriate for filtering these.

Minus stutter (1 repeat unit) in contrast, is very dependent upon the number of repeats at a locus. GMID-X (v1.4) does not have the capability to filter stutter in an allele specific manner, so analysts should assess stutter in every instance. GMID-X filters were set based on internal validation values such that the GMID-X filters are sufficient for filtering out the vast majority of stutter without being too high, but may not sufficiently filter stutter at larger alleles.

Challenged Samples Study

Challenging adjudicated casework and mock casework samples previously analyzed with the Identifiler, Identifiler Plus, and Minifiler kits were re-analyzed with GlobalFiler. The GlobalFiler results provided the most genetic information of all the kits. At least a partial profile was obtained from all samples analyzed with GlobalFiler.

Mixture Study

Two, three, four, and five person DNA mixtures were created in a variety of ratios. In order to examine the affects of allele stacking, two different combinations of people were used for every different mixture ratio combination. Because electropherograms of single source reference samples were examined closely before mixture design, average peak height of the alleles from the single source samples was taken into account when combining them to make mixtures. The mixture ratio results and peak heights reflected the target to a large degree. Stutter was not taken into account when determining these ratios manually, so there is some variation from locus to locus. The mixtures range from being very robust (peak heights above the target range creating spectral artifacts) to very low (considerable dropout occurring in one or more of the contributors). The results of this mixture set provide a sufficiently wide range of contributor ratio and input amount to extensively test the deconvolution ability, accuracy and limits of STRmix.

Injection Time Study

The default injection time on the 3500 instrument is 15 seconds. Two decreased injection times (5 and 10 seconds), and one increased injection time (24 second) were evaluated in this study.

Samples with higher input amounts were evaluated at injection times of 5 and 10 seconds compared to 15 seconds. The peak heights in these samples decreased in a linear way, such that peak heights from 15 second injections are, on average, 1.59 times higher than peak heights from 10 second injections. Peak heights from 10 second injections are, on average, 2.06 times higher than peak heights from 5 second injections.

Samples with peak heights in the stochastic range were evaluated at injection times of 24 seconds compared to 15 seconds. Peak heights from 24 second injections are, on average, 1.7 times higher than peak heights from 15 second injections. Increasing injection time also increase the height of peaks whose sister alleles have dropped out. This indicates that the stochastic range is increased, and therefore, different STRmix model parameters (allele, stutter, and LSAE variance) should be developed to analyze any sample that has been injected for 24 seconds.

Contamination Assessment Study

All extraneous peaks were investigated and documented over the course of this validation. There were many instances of extraneous peaks, and this study extensively documents the types of artifacts that are seen with GlobalFiler. These events appear to be sample or GlobalFiler kit lot number specific and not a systematic problem.

Contamination is not an inherent problem associated with the GlobalFiler kit. Each new lot of GlobalFiler kits received into the laboratory will have a quality control ("QC") run performed prior to use in casework. This QC and the TE amplification blank included in each amplification will serve as ongoing verification that the components of the GlobalFiler kits do not exhibit inherent contamination.

Normalization Study

Normalization is a new feature included in the 3500 data collection software. It is designed to attenuate signal variations associated primarily between instruments. In this study, the correlation between the two instruments is very good ($R^2 = 0.9846$), suggesting low variability between injections on 3500A and 3500B. Normalization is a feature that is not recommended for analysis in the SDPD lab at this time. The sensitivity difference and variability between the two instruments is so minimal when compared with the variability associated with the other analytical procedures that allowing for normalization would potentially create unwanted variability when it comes to interpretation. Using the normalization feature is not recommended until a time where the sensitivity of the two instruments differs more than sample to sample variability.

Concordance Study

Known samples previously typed with the Identifiler Plus kit were re-analyzed with GlobalFiler. The genotypes obtained following GlobalFiler® amplification of the 70 reference samples were 100% concordant at the loci previously analyzed using Identifiler Plus. A total of 1991

concordant alleles were observed. Because GlobalFiler contains additional loci to Identifiler Plus, these loci could not be tested against concordance with the previous kit. Genotypes for the NIST standard reference material (NIST SRM2391c Component A, B, and C) were 100% concordant at all loci after being amplified with the GlobalFiler kit.

Reproducibility Study

The 007 positive control and many known samples were amplified by multiple analysts over the course of this validation. The results for these samples were always consistent between amplifications. Therefore, DNA typing results obtained using GlobalFiler have proven to be reproducible.

Precision Study

Precision was assessed by examining the variability of every ladder allele across 25 injections of ladder. Sizing precision and accuracy are dependent on amplicon size, and the smaller loci had a much smaller standard deviation than larger amplicons across these 25 injections.

This study reveals that reliable results will be obtained by performing capillary electrophoresis on both genetic analyzers. Precision is more closely associated with the capillary electrophoresis instrument than the DNA typing kit. This data confirms that GlobalFiler demonstrates acceptable levels of precision on a 3500 Genetic Analyzer.

Conclusions

The GlobalFiler PCR Amplification Kit from Life Technologies on the 3500 Genetic Analyzer has undergone extensive internal validation studies in accordance with the SWGDAM guidelines and DAB requirements. The kit has been validated for 29 amplification cycles using manual set up as well as automated set up on the Hamilton Nimbus. The kit is therefore approved for use in DNA casework in the SDPD Forensic Biology Unit.

References

1. CODIS Bulletin #: BT031915. Implementation of the New Expanded CODIS Core Loci
2. CODIS Bulletin #: BT061814. Approval of a New PCR STR Kit (GlobalFiler)
3. CODIS Bulletin #: BT050615. Modified Buffer in GlobalFiler Casework Kit
4. GlobalFiler PCR Amplification Kit Users Manual
5. 3500 Genetic Analyzer User Guide
6. 3500 User Software Manual
7. GeneMapper IDX v1.4 User Bulletin

Validation Analysts

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

For a signed copy please contact the Quality Assurance Manager

Supervisor Review Date	**Supervisor Review Date**

Quality Assurance Review Date

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Life Technologies

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

Purpose

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

Materials and Methods

Negative controls and approximately 30 DNA-containing samples were amplified using the amount of input DNA and cycling parameters recommended by the manufacturer. The recommended parameters were: 15uL of TE for the blanks, 0.8ng of DNA for amplification and 29 cycles with the GlobalFiler kit. The amplification product from 12 TE blanks and the 30 DNA-containing samples were injected for capillary electrophoresis on both 3500 instruments according to the manufacturer's recommended parameters. A detection threshold of 1 RFU was applied to all samples and blanks during GeneMapper ID-X (v1.4) analysis. All known DNA peaks, PCR artifacts (i.e. stutter peaks and minus-A peaks, and spectral artifacts were removed prior to the assessment. The data was evaluated to determine the average baseline peak heights as well as the standard deviation of baseline peaks for each dye channel. Additionally, the data was evaluated to determine the RFU of the highest peaks in all dye channels that could not be identified as a DNA related peak or spectral artifact.

After the start of the internal validation, and after the baseline was established, Applied Biosystems reformulated the GlobalFiler master mix. Developmental validation and factory performance check of this product indicated no difference in detection threshold. However, a comparison between the original and reformulated master mix was made as part of the SDPD internal validation to verify that there were no differences. TE amplification blanks were examined after samples were run with the reformulated GlobalFiler master mix provided by

Applied Biosystems after the initial analytical threshold study. Five TE amplification blanks were chosen from 5 separate injections and examined with the same parameters as above. A statistical comparison (t-test) was made between the noise peaks of the original and reformulated products.

Spectral artifacts (pull-up and pull-down) were assessed in single source and mixture samples amplified with the reformulated GlobalFiler product. Ninety-six samples, with input DNA within the target range (≥400pg), were evaluated for the presence of spectral artifacts. Twenty-three of these samples had either observed pull up, pull-down, or both. The instances of spectral artifacts, their peak heights, and their percentage of pull-up or pull-down off the parent peak were calculated to determine if the baseline threshold required adjustments to account for spectral artifacts.

Results and Discussion

The baseline study evaluated the signal-to-noise ratio for the Applied Biosystems GlobalFiler kit on the 3500 Genetic Analyzer. The instrumental noise (baseline) was examined from amplified TE blanks as well as from amplified samples containing both single source and mixture DNA (allele peaks and artifacts removed). A summary of the average, standard deviation, and maximum peak heights can be seen in Figure 1.

Amplified samples (DNA peaks removed)			
Channel	Average	StDev	Max PH
B	5.39	2.77	49
G	9.61	4.29	63
Y	4.79	2.75	33
R	8.40	3.82	47
P	8.80	4.01	45

TE Blanks			
Channel	Average	StDev	Max PH
B	5.08	2.06	37
G	9.23	3.09	58
Y	4.30	1.90	34
R	4.30	1.90	34
P	7.89	2.60	23

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

Average noise peak height and max peak height were similar from both types of samples (Amplified samples with peaks removed and TE blanks). The average peak height was below 10 RFU, and the standard deviation very low. Occasionally, a peak above 50 RFU was seen, especially in the green channel. The distribution of all of these peaks combined can be seen in Figure 2. Most of the peaks fall between 0 and 10 RFU.

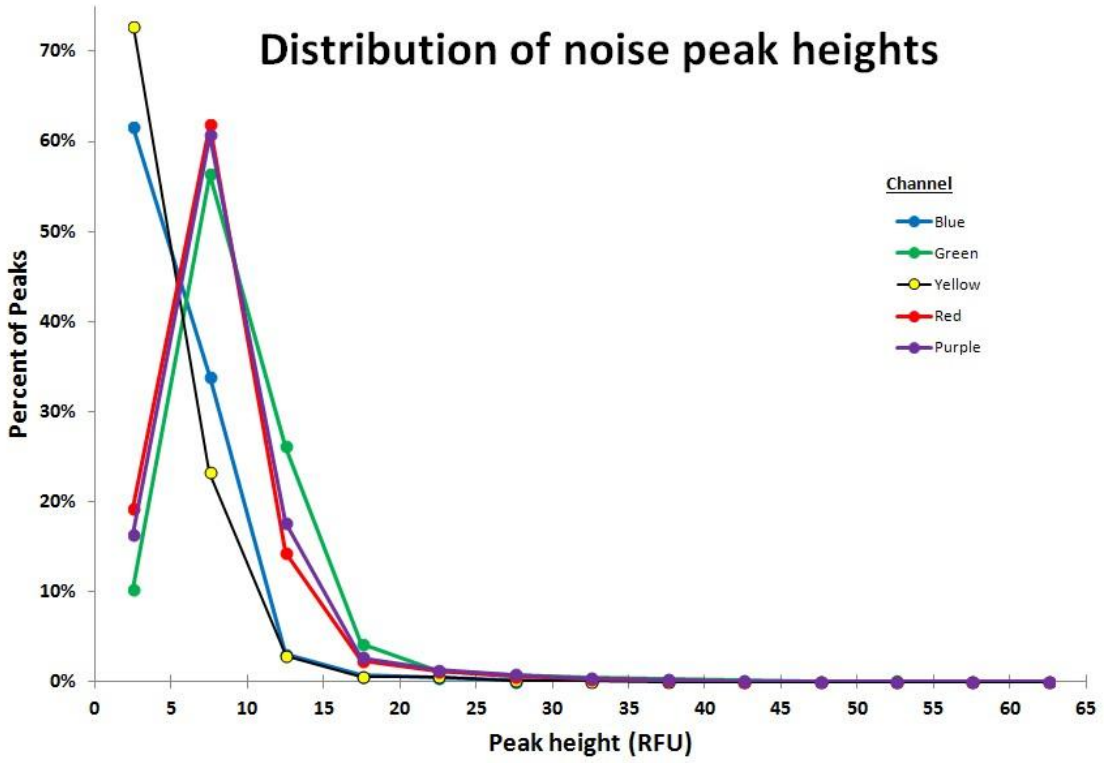


Figure 2 – A distribution graph demonstrating that the majority of noise peaks are very small

There are multiple ways of calculating an empirical analytical threshold from baseline data. Analytical thresholds are generally determined by examining the baseline data and calculating the mean and standard deviation. The limit of detection is defined as the mean plus three standard deviations. The limit of quantification is defined as the mean plus 10 standard deviation, which would be the minimum of what we would consider as an analytical threshold. Michael Coble from the applied testing group at NIST has previously given an example for determining the analytical threshold using two times the intensity difference between the highest peak in the baseline and the lowest trough (1). As an additional consideration, GeneMapper ID-X data generally omits negative baseline data (i.e. data below zero) in its collection of data. In order to compensate for this, one option in determining the analytical threshold would be to use a calculation of three times the highest peak in the baseline. Analytical thresholds can also be calculated globally or per dye color. Figure 3 shows three possible methods to help guide the analytical threshold value.

Method	Channel	Analytical Threshold (RFU)
Mean + 10 standard deviations (LOQ)	Blue	35
	Green	59
	Yellow	34
	Red	50
	Purple	53
	Global	**52**
2 x the highest baseline peak	Blue	98
	Green	126
	Yellow	66
	Red	94
	Purple	90
	Global	**126**
3 x the highest baseline peak	Blue	147
	Green	189
	Yellow	99
	Red	141
	Purple	135
	Global	**189**

Figure 3 – Multiple methods to determine the analytical threshold

Additional studies done as part of this validation project used an analytical threshold of 100 RFU. No extraneous peaks were detected when amplifying single source reference samples. Over the course of the mixture study where 180 samples were amplified, less than 5 peaks were identified that were not attributable to DNA or DNA artifact peaks. It is possible that they were an unknown spectral artifact. The peak heights of these peaks ranged from 123 to 185 RFU.

Reformulated Master Mix

Noise peaks were examined in TE blanks amplified with the reformulated Master Mix. Although there was some variation between average and standard deviation between the reformulated master mix and the original formulation, the average noise peaks and the maximum noise peak were still well below the analytical threshold.

Noise peaks collected from TE blanks (n=12) amplified with the *original* GlobalFiler master mix formulation were not statistically different from noise peaks collected from TE blanks (n=5)

amplified with the *reformulated* GlobalFiler master mix (p=0.707, t-test, two tailed equal variance).

TE Blanks			
Channel	Average	StDev	Max PH
B (6-FAM)	4.72	1.79	18
G (VIC)	9.74	3.04	31
Y (NED)	5.12	2.20	39
R (TAZ)	8.27	2.55	22
P (SID)	9.32	2.95	33

Figure 4 – Peak Heights from TE blanks amplified with the reformulated GlobalFiler kit and injected on 3500 Genetic analyzers

Spectral consideration

A final consideration of the analytical threshold was done by examining the spectral pull up using the reformulated GlobalFiler master mix. This factor merits consideration because of the dynamic range of the 3500 Genetic Analyzer. Data originating from the 3500 Genetic Analyzer is uncompressed data as compared to data originating from the 3130 Genetic Analyzers. When samples amplified with the recommended input amount of DNA is analyzed, the 3500 instrument produces data with peak heights in the range of 3,000 to 10,000 RFU. Considerations should be given to expected artifacts that occur during the routine course of analysis, including spectral overlap.

The presence of artifact peaks should not drive the determination of the analytical threshold. However, consideration of the practical implications of interpreting numerous artifact peaks should be considered prior to the implementation of the analytical threshold. The manufacturer's engineering specifications state that pull-up between 1-5% is within tolerable limits. Using the 3% average pull-up, the large dynamic range will result in potential pull-up peaks between 90 RFU and 300 RFU for samples amplified with the recommended amount of input DNA. If the average pull-up observed is in excess of 3%, a new spectral calibration should be done for the effect it can have on baseline "noise" peaks.

After evaluating the observed spectral artifacts in the twenty-three samples, it was noted that there were 57 instances of pull-up and pull-down peaks with an additional 14 peaks that were not linked to PCR artifacts. Each peak height was assessed against the percentage of pull-up or pull-down to its parent peak. All observed spectral artifacts were below 3%, with the majority (90%) falling below 2%. Figure 4 shows spectral artifacts observed as pull-up, pull-down, or additional

artifacts which were unrelated to pull-up or pull-down. These additional artifacts, therefore, had no calculation for a peak height percentage off the parent peak. The pull-down peak percentages were calculated using the average of the two parent peaks and the average of all detected peak heights associated with one pull-down event.

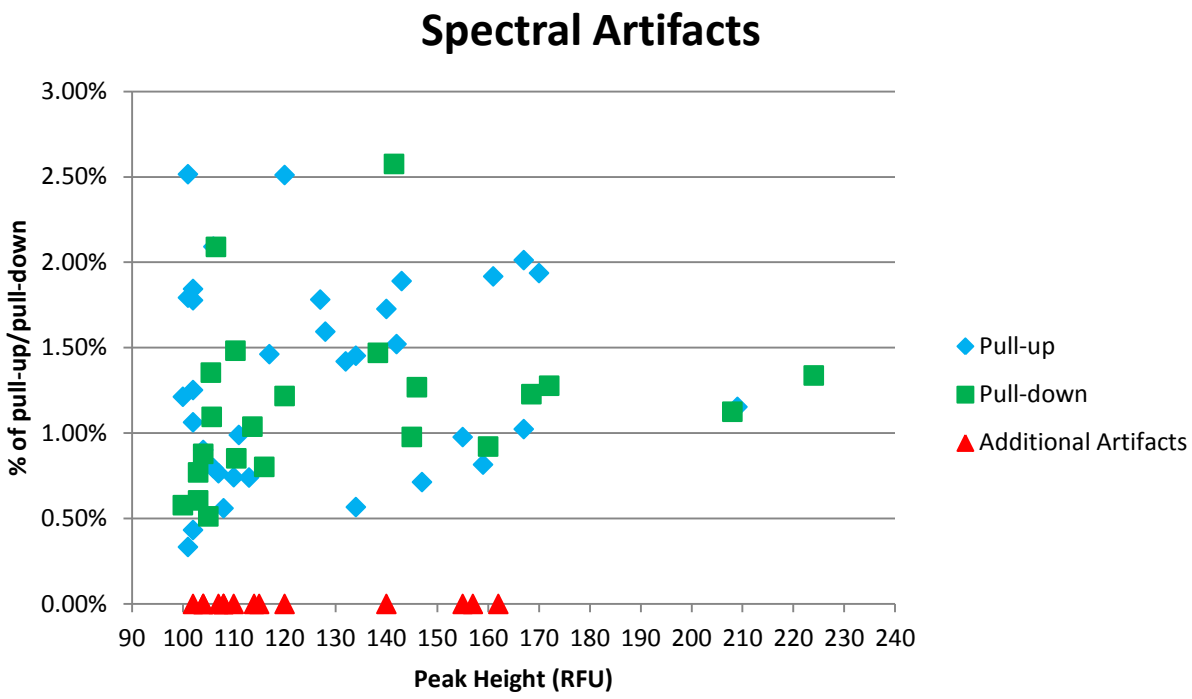


Figure 5 – Spectral artifacts observed in 23 samples, n= 71. Ninety percent of the pull-up and pull-down peaks fell below 2%, n= 57.

Conclusions

The data from this study suggests that a peak amplitude threshold of 100 RFU is an appropriate analytical threshold for DNA casework. The evaluation of spectral artifacts did not justify an adjustment of this threshold. Raising the threshold to eliminate spectral artifacts in less than a quarter of the samples reviewed did not out weight the loss of true peaks at low levels.

References

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

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Applied Biosystems

Sensitivity and Target DNA Input Amount Study

Purpose

The goal of this study is to determine the lowest level of obtaining data with the kit and the levels above which we start to see artifacts routinely detection limit, saturation point, and optimal DNA input range of the GlobalFiler PCR Amplification Kit.

Materials and Methods

Ten reference samples were amplified at variable input volumes for 29 cycles with the GlobalFiler kit, based on quantification results from the Quantifiler Human kit. Three of the samples were amplified with template amounts ranging from ~7ng down to ~30pg, and the remaining seven samples were amplified with template amounts ranging from ~2ng down to ~20pg in order to determine ideal input range and determine at what template amount dropout starts to occur. This was done with the original GlobalFiler formulation first. When the reformulated master mix was released, it was done again using 13 single source reference samples amplified with template amounts ranging from ~800pg down to ~12.5pg. Results presented in this study will focus on data collected using the reformulated GlobalFiler amplification kit, because that is the only product commercially available, however all data is accessible in the lab validation files on the FB network.

The amplification product from these samples was injected for capillary electrophoresis on both 3500 instruments according to the manufacturer's recommended parameters. A detection threshold of 100 RFU was applied during GeneMapper ID-X (v1.4) analysis. Peak heights were observed across the dilution series to determine the DNA target amounts that resulted in peak heights between 3,000 and 10,000 RFU (the ideal peak height range for data processed with the 3500 instrument). Artifacts (spectral and amplification related) were also examined to help to narrow down the ideal target amount. For each set of dilutions, the ideal target amount was highlighted.

Results and Discussion

Electropherograms were examined for each of the samples to determine peak heights, drop out and artifacts across all 24 loci. There was some variability between the 13 samples with regard to the relationship between target amount and peak height. The ideal target amount from each single source sample was chosen based on peak height, peak height balance, and the absence of artifacts. The optimal target is about ~600pg, and always within the range of 400 to 800 pg of DNA. Data for this study can be seen in Figure 1. Peak height ratios remained high until the target input amount dropped below 400 pg.

Sample	Template Amount (QH) (pg/µL)	Alleles Expected	Alleles Observed	Peak Height description (RFU)
1	800	38	38	5,000-25,000, artifacts
	400		38	3,000-15,000, artifacts
	200		38	200-2,500, PH imbal.
	150		38	300-2,500, PH imbal.
	100		36	300-1,500, PH imbal.
	75		36	highest peak 647
	50		32	highest peak 1019
	25		17	highest peak 383
	12.5		8	highest peak 183
2	800	39	39	6,000 to sat., artifacts
	400		39	3,000-15,000
	200		39	1,500-6,500
	150		39	1,000-5,000
	100		39	600-3,000, PH imbal.
	75		39	350-3,000, PH imbal.
	50		39	200-1300, PH imbal.
	25		20	highest peak 405
	12.5		12	highest peak 266
3	800	42	42	4,000-23,000, artifacts
	400		42	2,500-15,000
	200		42	650-5,500
	150		42	500-5,000, PH imbal.
	100		42	200-1,500, PH imbal.
	75		42	150-1,750, PH imbal.
	50		36	highest peak 710
	25		25	highest peak 511
	12.5		9	highest peak 235
4	800	39	39	3,500-14,000, artifacts
	400		36	100-1,600, PH imbal.
	200		34	highest peak 1137
	150		28	highest peak 872
	100		18	highest peak 722
	75		19	highest peak 654
	50		10	highest peak 253
	25		8	highest peak 368
	12.5		4	highest peak 168
5	800	42	42	2,500-10,000
	400		42	1,500-5,000
	200		42	500-1,600, PH imbal.
	100		41	highest peak 1443
	50		26	highest peak 651
	25		19	highest peak 321
6	800	44	44	1,200-6,000
	400		44	800-4,500
	200		44	400-2,000, PH imbal.
	100		44	150-600, PH imbal.
	50		39	highest peak 910
	25		32	highest peak 440

Sample	Template Amount (QH) (pg/µL)	Alleles Expected	Alleles Observed	Peak Height description (RFU)
7	800	41	41	5,000-20,000, artifacts
	400		41	3,000-10,000
	200		41	500-4,000, PH imbal.
	150		41	500-3,000, PH imbal.
	100		41	250-2,000, PH imbal.
	75		41	PH imbalance, 1027
	50		39	highest peak 750
	25		26	highest peak 627
	12.5		12	highest peak 405
8	800	38	38	5,000-20,000, artifacts
	400		38	1,700-6,000
	200		38	600-2,500, PH imbal.
	150		38	500-3,200, PH imbal.
	100		38	highest peak 1208
	75		37	highest peak 1250
	50		34	highest peak 949
	25		34	highest peak 588
	12.5		20	highest peak 381
9	800	40*	40	3,000-13,000, artifacts
	400		40	1,600-7,000
	200		40	400-2,000, PH imbal.
	150		40	400-2,000, PH imbal.
	100		40	200-1,400, PH imbal.
	75		40	150-1,500, PH imbal.
	50		37	highest peak 806
	25		13	highest peak 339
	12.5		9	highest peak 220
10	800	39	39	3,000-1,500
	400		39	1,400-7,300
	200		39	1,100-5,200
	100		39	200-1,700, PH imbal.
	50		39	highest peak 594
	25		31	highest peak 581
	12.5		25	highest peak 466
11	800	40	40	3,000-16,000, artifacts
	400		40	1,600-10,000, artifacts
	200		40	350-2,000, PH imbal.
	100		40	200-1,500, PH imbal.
	50		40	170-1,300, PH imbal.
	25		21	highest peak 251
12	800	39	39	2,200-17,000, artifacts
	400		39	1,500-7,500
	200		39	400-3,000, PH imbal.
	100		37	highest peak 879
	50		14	highest peak 426
	25		14	highest peak 395
13	800	38	38	2,000-10,000
	400		38	900-5,500
	200		38	500-4,000
	100		38	150-1,100, PH imbal.
	50		31	highest peak 732
	25		20	highest peak 335

Ideal target amount

	>1/2 alleles present
	1/4 and 1/2 alleles present
	< 1/4 of alleles present

*excluding partial null allele

Figure 1 – A dilution series with associated peak heights and ideal input amount

The target input amount determined using the original formulation was very similar. It had an idea target amount of ~750 pg of DNA for single source samples. The optimal target was always within the range of 0.5 to 1.0 ng of DNA. The difference between the two different formulations of the kits became apparent when it came to dropout. Samples amplified with the original formulation started have dropout around 200 pg, with partial profiles obtained down to 25 pg, while samples amplified with the reformulated master mix started to have alleles dropping out around 50-100 pg, and partial profiles were obtained down to ~12 pg.

Saturation was determined to be ~32,000 RFU for this instrument. See Figure 1. Amplifications with the original formulation contained samples amplified with enough DNA to test this, and no peaks were seen above the low 30,000's. Stutter peaks associated with these alleles did not saturate, however, and the observed stutter peak was much higher than the expected stutter peak height at some of these alleles, giving another indication of saturation point. Finally, ~30,000 RFU is consistent with the developmental validation and manufacturers specifications of this instrument.

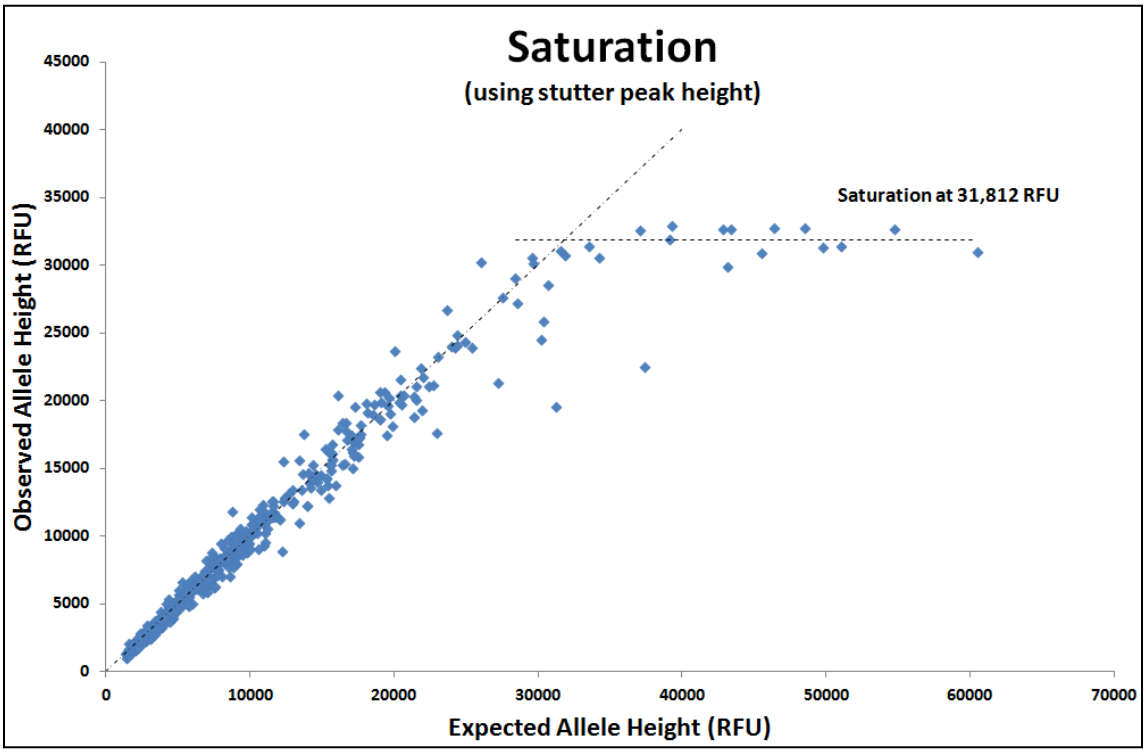


Figure 1 – Saturation for 3500 Genetic Analyzers

Conclusions

The detection limit for full DNA profiles is ~50-100 pg. Partial DNA profiles were observed down to the minimal input tested (~12 pg). The optimal DNA input for single source samples is 600 pg of DNA. Profiles in the range of 400-800 pg demonstrated the most desirable peak

heights and peak height ratios, although spectral artifacts became apparent when peak heights exceeded 10,000 RFU (usually beginning around 800 pg).

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Applied Biosystems

Stochastic Study

Purpose

The goal of this study is to determine at what levels drop-out of genetic information begins to occur after single source samples are amplified using the GlobalFiler PCR Amplification Kit.

Materials and Methods

This study was conducted in two parts. The first part was done using the original GlobalFiler master mix formulation. During the validation, Applied Biosystems released a new master mix formulation. The second part of this study was done using the reformulated GlobalFiler master mix. For the first part, a dilution series of 10 samples were amplified 10 different times at a range of target amounts from >2ng to ~20pg. For the second part, 13 samples were amplified 6 or 7 different times within a target range of ~800 to ~12pg. Samples were amplified using the standard cycling parameters and 29 cycles.

Capillary electrophoresis was performed on the amplified product using both 3500 genetic analyzers and the resultant data was analyzed in GMID-X (v1.4) using a 100 RFU detection threshold (normalization off). Results presented in this study will focus on data collected using only the reformulated GlobalFiler amplification kit, because the original formulation has been discontinued and the reformulation is the only product commercially available. Details from both studies are available in the validation files on the FB network.

Results and Discussion

Having two large dilution series available provided a good indication of the optimal target input amount as well as the sensitivity and detection limit of the GlobalFiler kit using a 3500 genetic analyzer. Details about the optimal input amount can be referenced in the Sensitivity and Target DNA Input Amount Study. This study focuses on results obtained from samples amplified at sub-optimal input amounts.

Overall, detection of alleles in single source samples was very robust down to about 100pg. Some sample to sample variability occurred, likely due to known variability in quantification results and/or amplification preparation. The two GlobalFiler kit formulations performed very similarly on most of the measures tested in the validation. A minor difference between the two different formulations of the kits became apparent when it came to dropout. Samples amplified with the original formulation started have dropout around 200 pg, with partial profiles obtained down to 25 pg, while samples amplified with the reformulated master mix started to have alleles dropping out around 50-100 pg, and partial profiles were obtained down to ~12 pg. Table 1,

below demonstrates the dropout observed with each of the samples on instrument 3500A. Results were similar using 3500B.

In Table 1, each horizontal row represents one single source sample amplified at the indicated target amount. Green indicates that a complete genotype was obtained at that locus, Yellow indicates that one of the two sister alleles is dropping out, and Red indicates full dropout at that locus. An H inside the box indicates a homozygous allele for that individual for that locus, and a number designates the peak height of the remaining sister allele. Highest peaks with sister alleles dropped out is summarized in Table 2 below.

Table 1 – A matrix displaying the prevalence of dropout in a dilution series amplified with GlobalFiler

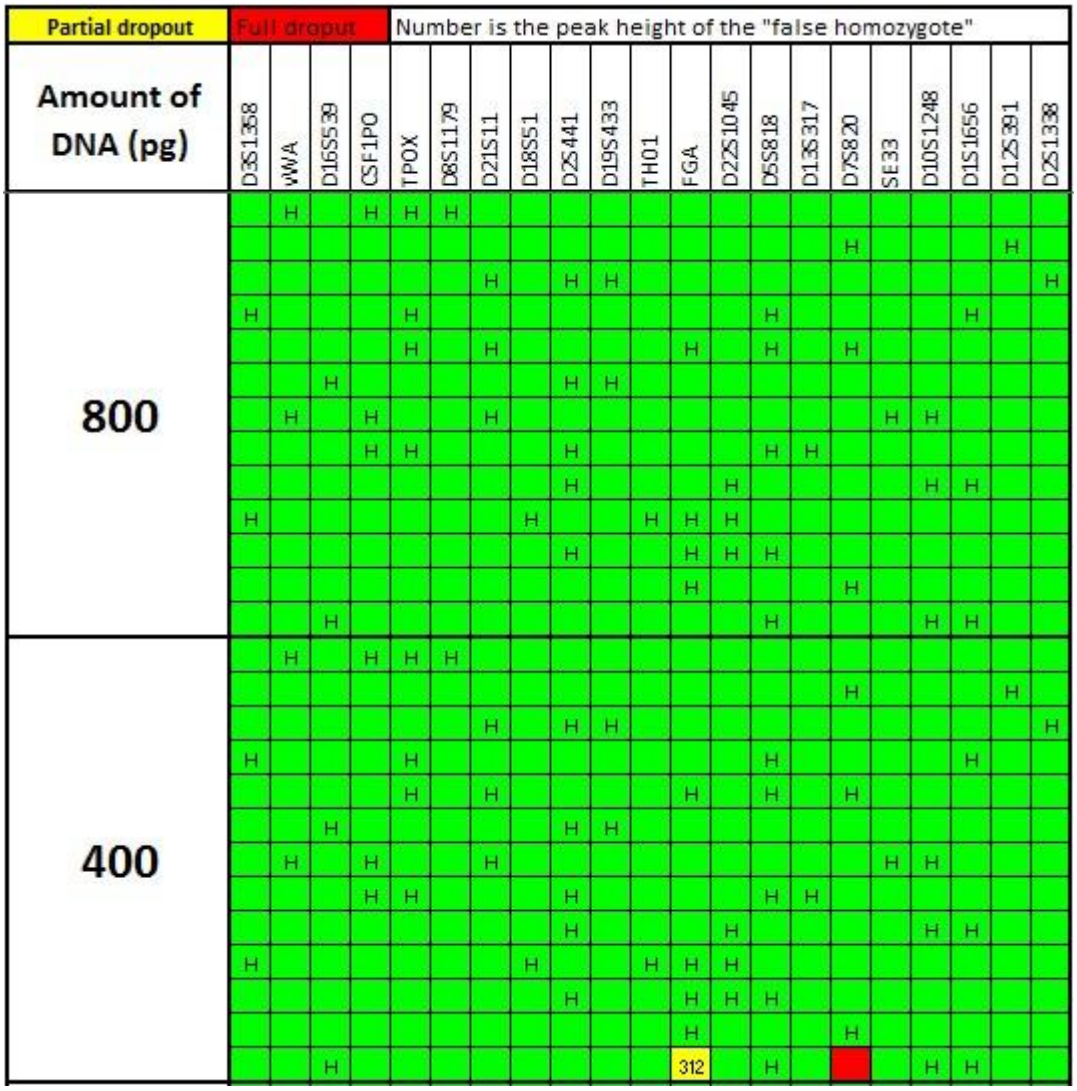
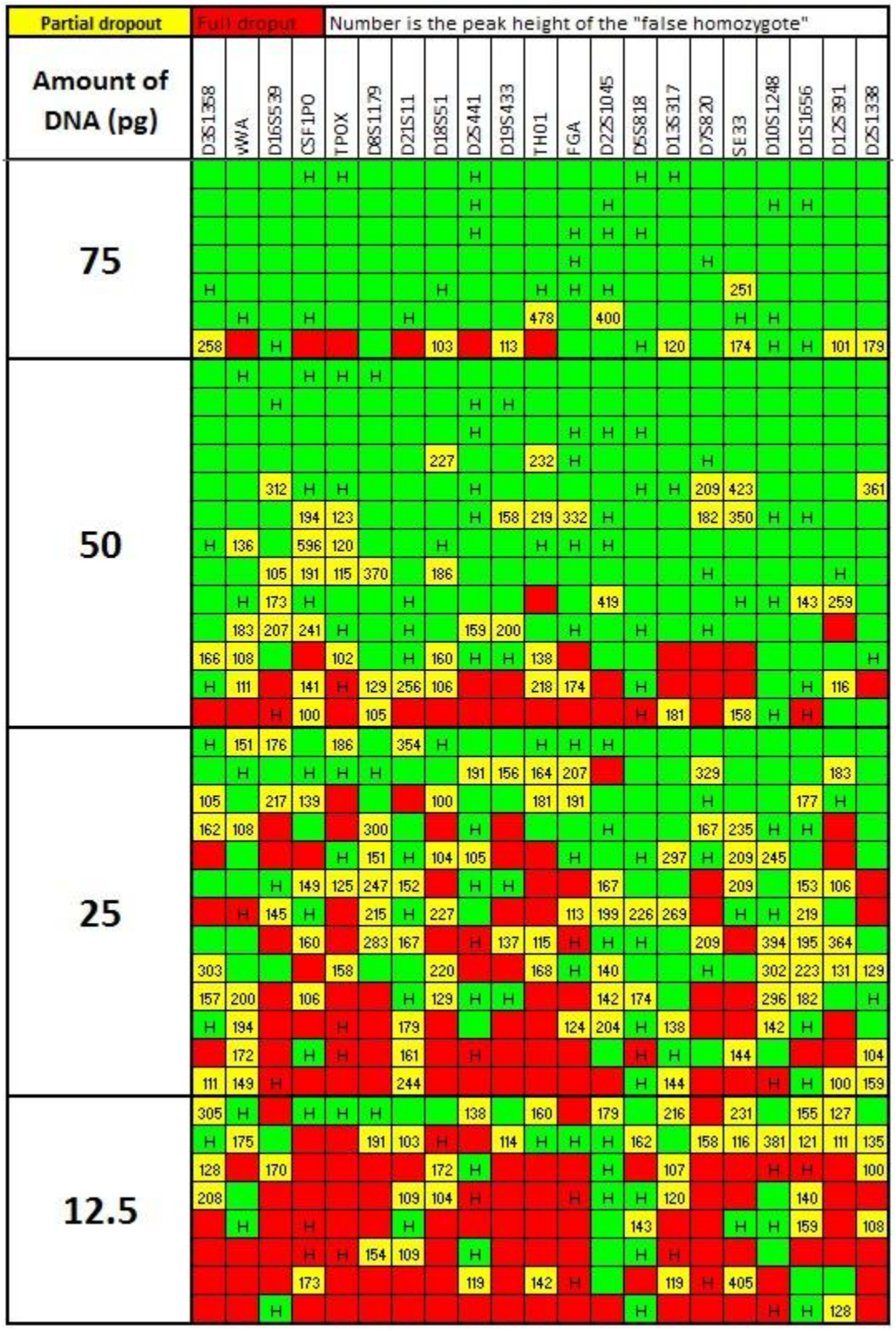


Table 1 – continued

Partial dropout	Full dropout	Number is the peak height of the "false homozygote"																							
Amount of DNA (pg)	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338
200	H	H		H	H	H																			
																H				H	
							H		H	H											H
	H				H									H					H		
					H		H					H		H		H					
			H						H	H											
		H		H			H										H	H			
				H	H				H					H	H						
									H				H					H	H		
	H							H			H	H	H								
									H			H	H	H							
												H				H					
			H	H						130	124	179		H		344		H	H		
150		H		H	H	H															
																H				H	
							H		H	H											H
	H				H									H					H		
					H		H					H		H		H					
			H						H	H											
		H		H			H										H	H			
				H	H				H					H	H						
									H				H					H	H		
	H							H			H	H	H								
									H			H	H	H							
												H				H					
		235	H					149			175			H				H	H	251	230
100		H		H	H	H															
																H				H	
					H		H					H		H		H					
			H						H	H											
				H	H				H					H	H						
									H				H					H	H		
	H							H			H	H	H								
									H			H	H	H							
												H				H					
		H		H	388		H				256						H	H			
							H		H	H						198					H
	H				H						309			H					H	162	
		164	H	149		270	119	274	224			109	100	H	318	109	105	H	H	127	117

Table 1 – continued



These numbers in yellow in this dropout matrix give an indication of the presence of stochastic effects. Using a 3500 instrument results in a much higher RFU dynamic range than what is seen with a 3130. Most of the peak with sister alleles dropping out are in the 100-200 RFU range. However, there are few instances where a peak in the 400-500 RFU range has known sister allele that is not detected with a threshold of 100 RFU. A table of the highest peaks associated with dropout is shown in Table 2. This data is summarized from the data in Table 1, and also includes a comparison of the two 3500 instruments.

Table 2 – The top 10 highest allele peaks associated with dropout (RFU) on 3500A and 3500B

3500A	3500B
596	570
478	487
419	481
405	416
400	410
394	399
388	398
381	397
370	390
364	382

Another stochastic effect that is known to occur is peak height imbalance. As the input amount of DNA drops, and therefore the overall RFU levels, heterozygous peaks can become more imbalanced due to stochastic amplification effects. See more details about peak height imbalance and peak height ratios in the Peak Height Study. As heterozygous peaks become imbalanced, dropout starts to occur.

The majority of sister alleles with known associated dropout were on the lower end of the stochastic range. An example of dropout is displayed in Figure 1, where the sister allele is visible below the analytical threshold. There are other instances, where due to the stochastic nature of amplifying a very low amount of template DNA, the known sister allele is indistinguishable from noise (Figure 2). An important thing to note is that these two scenarios are not necessarily correlated with the height of the "false homozygote" peak.

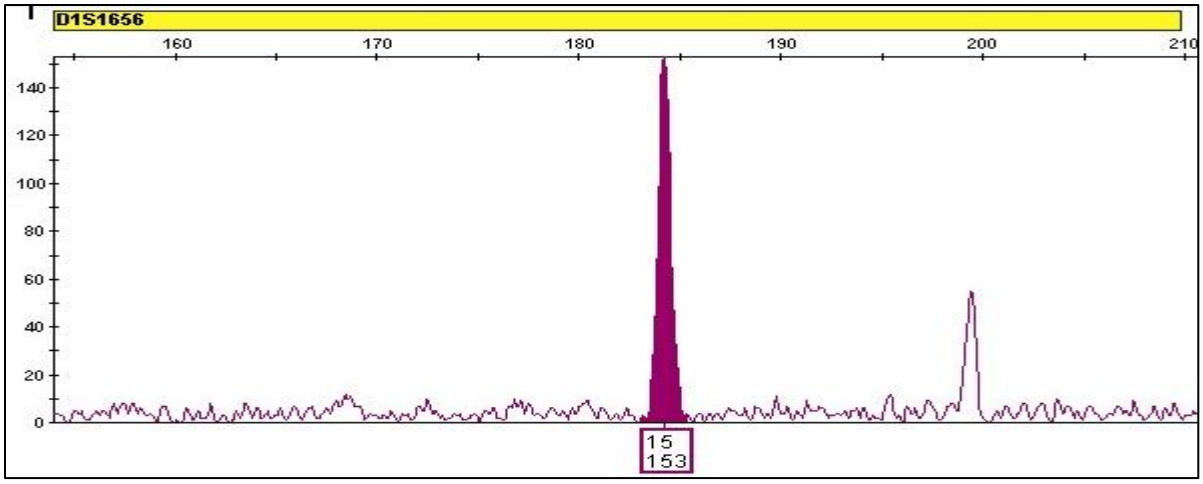


Figure 1 – The 15 allele at D1S1656 with a peak height of 153 RFU was a typical peak observed for which the known sister allele was below 100 RFU.

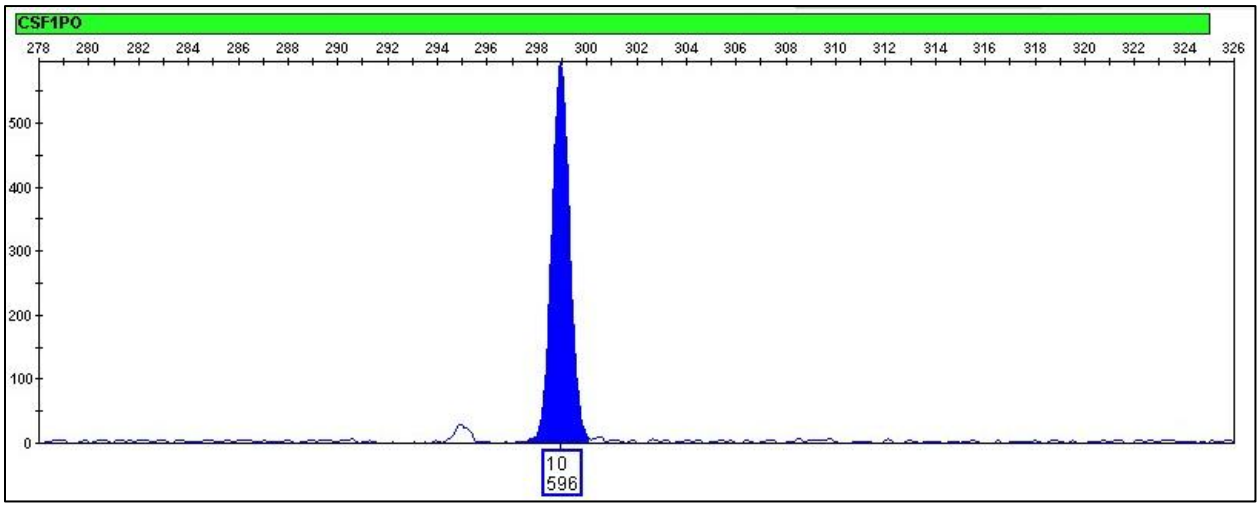


Figure 2 – The 10 allele at CSF1PO with a peak height of 596 RFU was the tallest peak observed for which the known sister allele was below 100 RFU. Its sister allele is also indistinguishable from noise in this instance, demonstrating the stochastic nature of amplification.

An important consideration of this study is the occurrence of dropout. The dropout matrix above displays this information in heat map form. Another way to analyze this data is by calculating both the incidence and probability of dropout. The data presented in Figure 3 takes into account all dropout and what range the remaining sister allele falls into. The majority of dropout (56%) is associated with a detected peak (false homozygote) in the range of 100-200 RFU. As the remaining peaks known to be associated with a dropout peak get higher, the occurrence of dropout decreases.

With a probabilistic interpretation approach in mind, another way of looking at this data is shown in Figure 4. It provides a guide to an analyst about how likely dropout is associated with a single peak at any given locus. In other words, it is the probability of dropout. More details about the probability of dropout can be found in the STRmix ModelMaker Study of the validation.

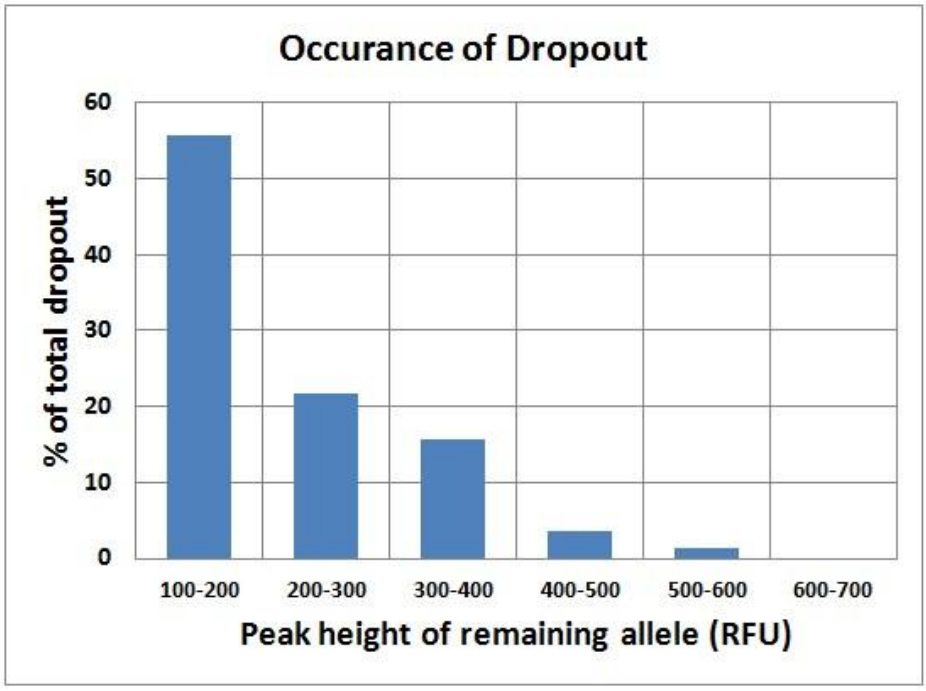


Figure 3 – A histogram of the incidence of dropout. The majority of dropout is associated with a detected sister allele peak height of 100-200 RFU. Data shown from 3500A, and is similar using 3500B.

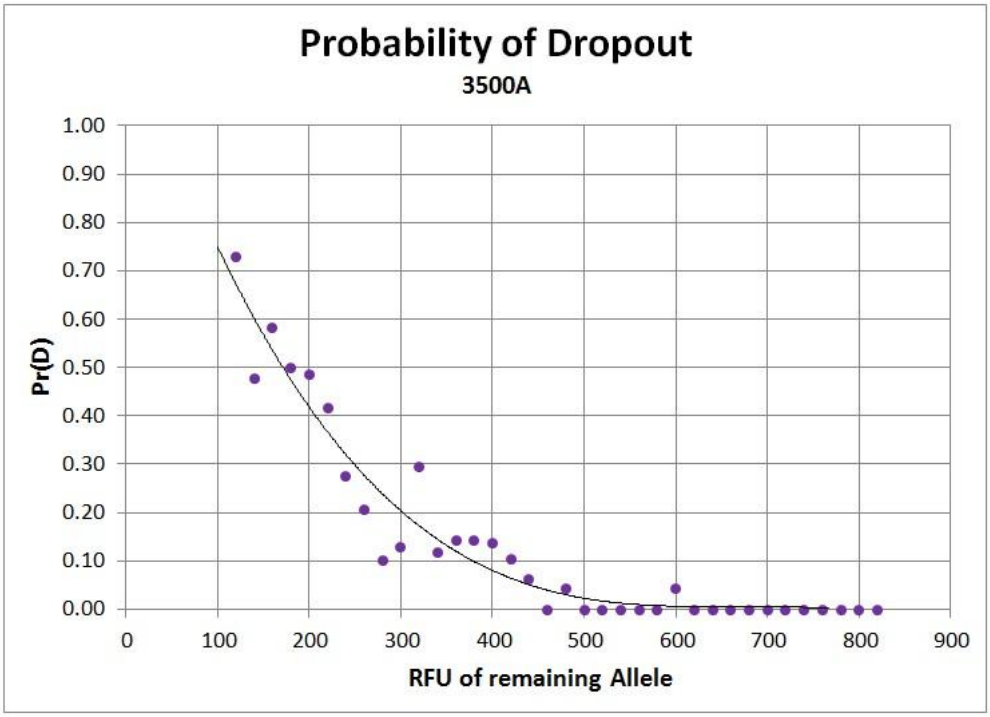


Figure 4 – An estimate of the probability of dropout (Pr(D)). This graph is also a histogram of the proportion of sister alleles with known dropout to total alleles within a 20 RFU range (bin). Data shown from 3500A, and is similar using 3500B.

Conclusions

All of this data taken together provides some idea of the stochastic range of genetic analysis using GlobalFiler on the 3500 instrument. Based on these validation studies, peaks occurring between 100 and 200 RFU have a higher likelihood of pairing with an undetected allele. This likelihood decreases until approximately600 RFU where it becomes exceedingly unlikely that a peak will be paired with an undetected allele.

In addition, the absence of visual evidence of an allele peak below the 100 RFU threshold (so-called "blips") is not sufficient to assume that no data is missing. Analysts should not attempt to look below the detection threshold due to the unreliability of this low level data.

Conclusions from this study are consistent with conclusions drawn in the Peak Height and Sensitivity Studies.

References

Butler, J. Advanced Topics in Forensic DNA Typing: Methodology. 2012 Academic Press.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Applied Biosystems

Peak Height Study

Purpose

The goal of this study is to evaluate peak height ratios, the inter-locus dye balance and intra-locus balance by examining peak heights following amplification with the GlobalFiler PCR Amplification Kit.

Materials and Methods

This study was conducted in two parts. The first part was done using the original GlobalFiler master mix formulation. Over the course of the validation of this kit, Applied Biosystems released a new master mix formulation. The second part of this study was done using the reformulated GlobalFiler master mix. Samples amplified at both optimal and low target inputs were used in this study in order to evaluate peak height ratios across a range of input amounts.

Part 1: 67 reference samples were amplified at quantities within the DNA input range (0.85 ng-2 ng), and a dilution series of 10 samples were amplified 10 different times at a range of target amounts from >2ng to ~20pg for 29 cycles using the original formulation of the GlobalFiler kit.

Part 2: 56 reference samples amplified at quantities within the optimal DNA input range (600pg). The optimal target input amount decreased with the new master mix formulation; see Sensitivity Target Input Study. In addition, 13 samples were amplified 6 or 7 different times within a target range of ~800 to ~12pg. Samples were amplified using the standard cycling parameters and 29 cycles.

Capillary electrophoresis was performed on the amplified product using both 3500 genetic analyzers and the resultant data was analyzed in GMID-X (v1.4) using a 100 RFU detection threshold (normalization off). Results presented in this study will focus on data collected using only the reformulated GlobalFiler amplification kit, because the original formulation has been discontinued and the reformulation is the only product commercially available. Details from both studies are available in the validation files on the FB network.

Peak height ratios were determined from all loci that have two sister alleles (heterozygotes). If the two heights are equal, the peak height ratio is 100%. If one is smaller than the other, the height of the smaller one is divided by the larger one to get the ratio, and this is converted to a percentage. Additionally, peak height ratios from adjacent sister alleles were excluded so that possible stutter from one of the alleles did not inflate peak height of the other allele.

Results and Discussion

<u>Inter and Intra-locus Dye balance</u>

In order to get an idea of how the loci and dye sets differed from each other, the average peak height of all heterozygous peaks for each locus was calculated in every sample, and then averaged for each locus. Overall, there is very good balance between dye sets and the individual markers were relatively well balanced. There are a few loci that have consistently higher peak heights than others, and that is reflected in Figure 1. D3S1358 (the smallest locus using the 6-FAM fluorophore) had consistently higher peaks than all the other makers using 6-FAM. Amelogenin and Yindel were higher than other loci using the VIC fluorophore. After the master mix reformulation, the yellow dye (NED) was a little lower than the other ones. The markers using both the NED, TAZ, and SID fluorophores are very well balanced. Data shown in Figure 1 is collected from samples only in the optimal DNA input range (400-800pg). One thing of note is that this peak height pattern between loci doesn't seem to break down at low input levels, because it remains consistent when data is compiled from all input amounts.

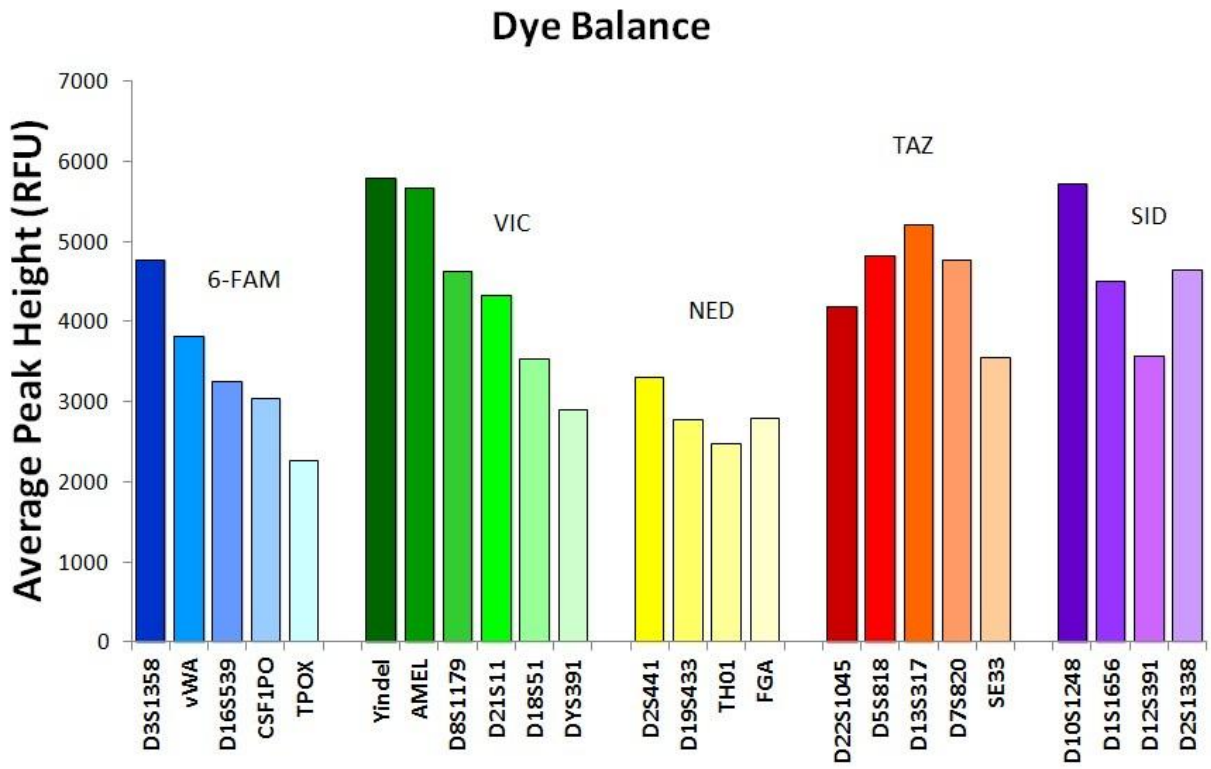


Figure 1 - Average peak height (per locus) for samples amplified within the optimal DNA input range separated by locus and dye channel.

Peak Height Ratios

The global average peak height ratio for samples in the optimal DNA input range was 86.8% (dotted horizontal line in Figure 2). The average peak height ratio did not vary substantially between loci or dye channels. Data shown here is compiled from instrument 3500B, and was similar on 3500A. Furthermore, the master mix formulation did not appear to have an effect on peak height ratio at optimal input amounts, because the pattern and average peak height ratio was also similar with the original formulation.

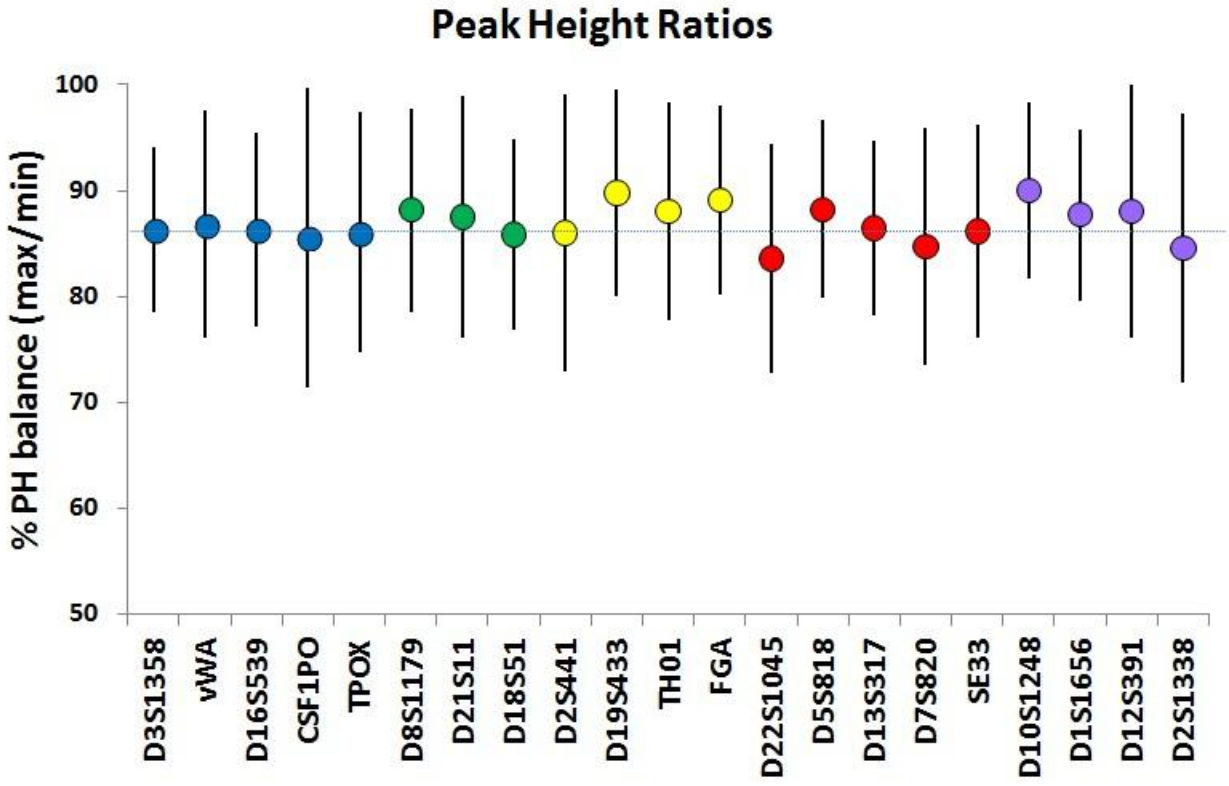


Figure 2 - Average peak height ratios with standard deviation bars (1SD) for samples amplified within the optimal DNA input range separated by locus and dye channel.

Peak height ratios do begin to decrease as DNA input amount decreases. When peak height is low, it raises a question about whether a sister allele has dropped below the analytical threshold. More data on the stochastic range of samples amplified with GlobalFiler can be found in the Stochastic Study.

Figure 3 plots peak height ratios from all samples and all DNA inputs 800pg and below. It plots the relationship between the smaller and taller peaks regardless of molecular weight. Only peak height ratios from heterozygous genotypes with alleles more than 1 repeat unit apart are displayed in Figure 3. The peak height ratio remains at about 50% or greater until peaks start to drop below 2,000 RFU. Considering that the target peak height range on a 3500 is between 3,000

and 10,000 RFU, this is consistent with the detection limits and where variability starts to occur. Data shown here are collected with 3500B. Data is similar on 3500A.

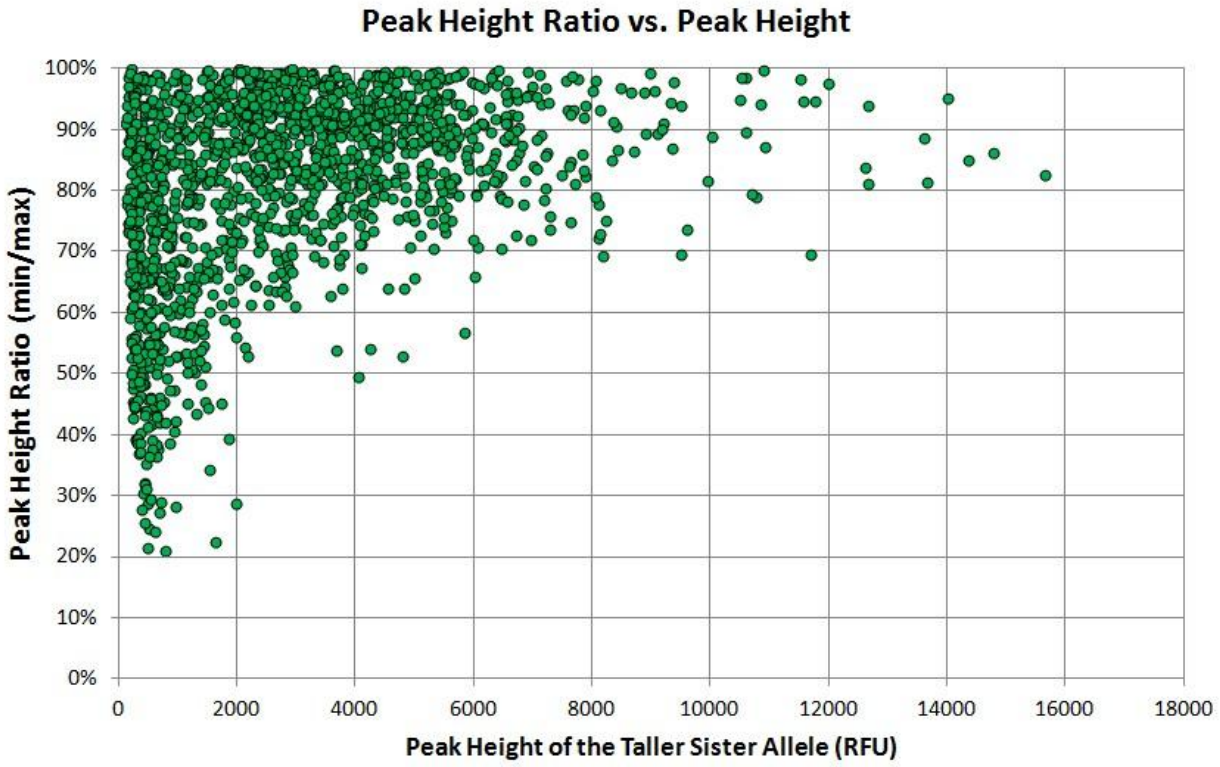


Figure 3 – Plot of peak height ratio versus height of tallest allele for single source heterozygotes.

Identifying the outliers in Figure 3 above can be informative because it displays the limits detected during internal validation. It is also important to look at the variance in this data, and determine how often an extreme peak height ratio occurs. Variance in peak height balance can be visualized in a different way in Figure 4, below, and becomes an important parameter in probabilistic genotyping.

In Figure 4, the higher molecular weight (hmw) peak was divided by the lower molecular weight peak (lmw) at a locus, and a log of that ratio was calculated (plotting it in log form give a more representative visual display of variation). STRmix Model Maker calculated the variance constant that was used to determine the 95% confidence interval in the graph below. See STRmix ModelMaker study for more details.

Like Figure 3, Figure 4 displays the relationship between the shorter and taller peaks, but also keeps track of molecular weight. Sometimes the larger sister allele is taller than the smaller one, and these occurrences are plotted above zero on the horizontal axis. Sometimes the smaller sister allele is taller than the larger one, and these occurrences are plotted below zero.

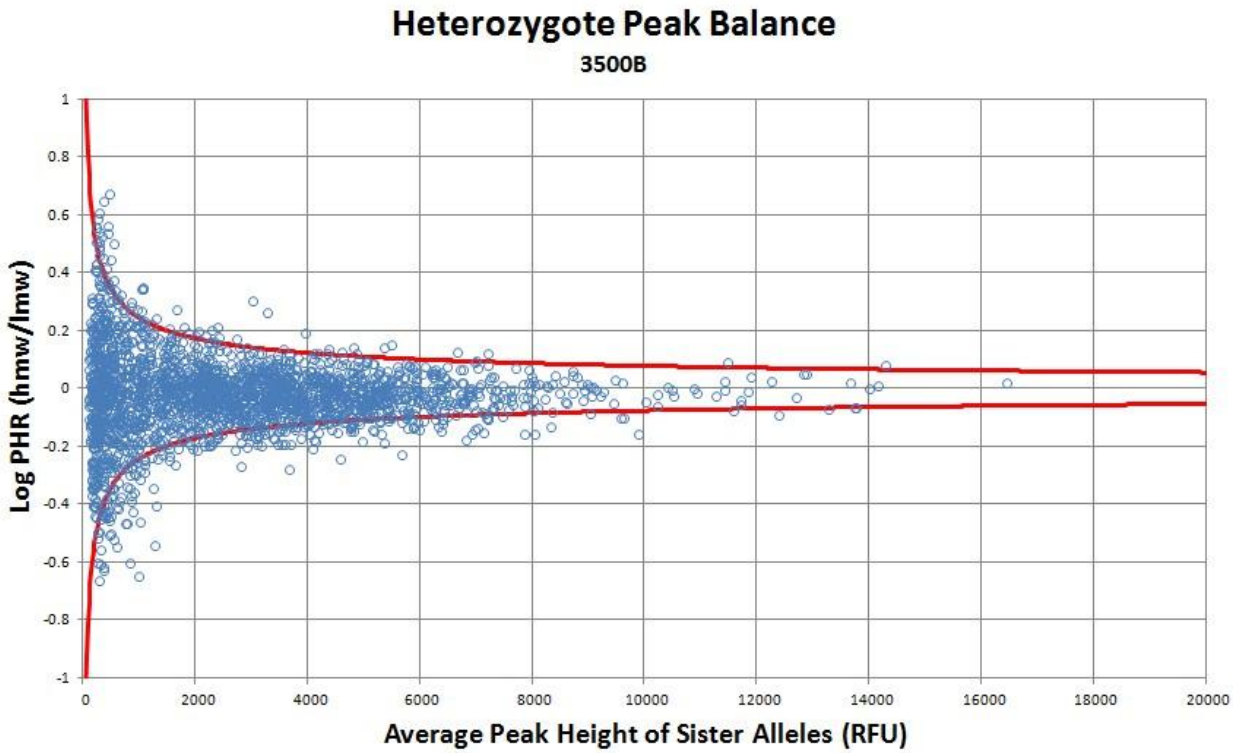


Figure 4 – Observed log of peak height ratio (PHR; hmw/lmw) with 95% boundaries (red dotted line) plotted against the average height of heterozygote alleles for data from all input amounts.

Samples amplified with quantities below the optimal DNA input range demonstrate more peak height imbalance than those within the optimal range. While the majority of peak height ratios are greater than 60% (0.22 on the log scale), this is not true of all observed data. There are a few observed heterozygous peaks where the shorter peak is less than 25% of the taller peak (0.6 on the log scale). Generally, the lower the input DNA amount, the greater the imbalance. This is demonstrated in Figure 5 where the same sample is amplified at decreasing input amounts (800-100pg).

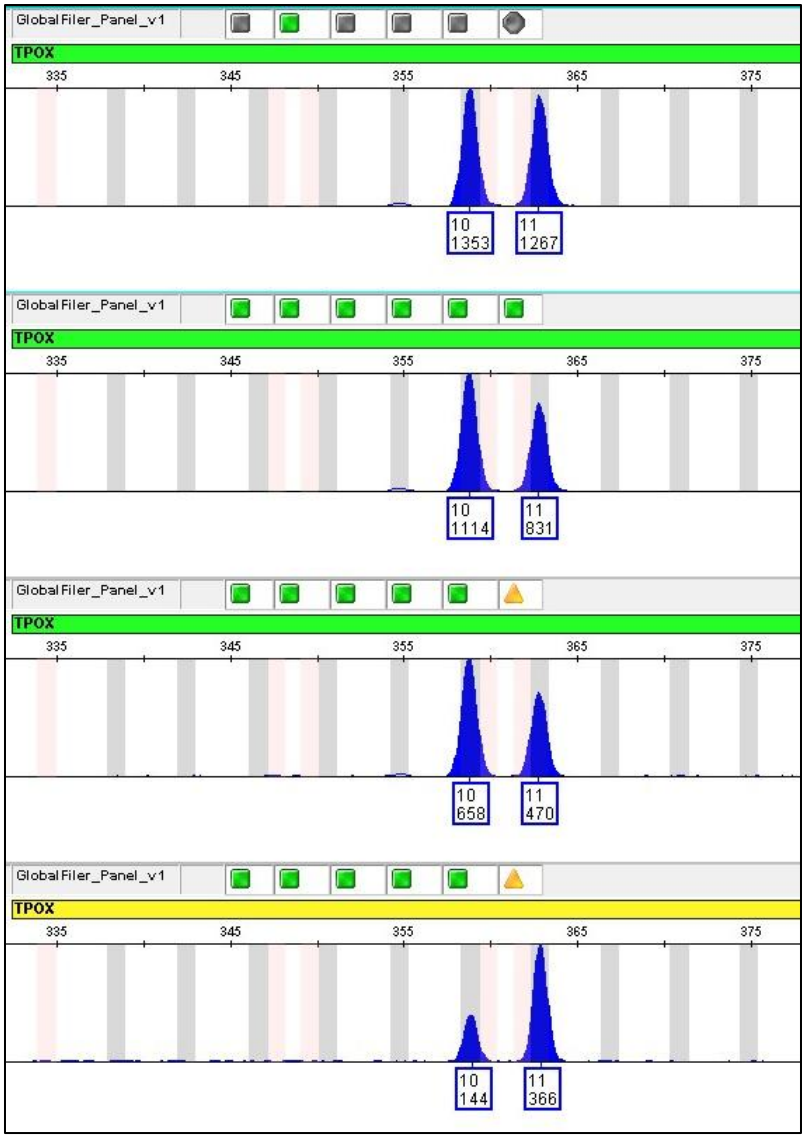


Figure 5 – One example of decreasing peak height balance with lower DNA input amounts and peak heights.

Conclusion

Samples amplified within the optimal DNA input range with GlobalFiler demonstrated well balanced loci, both within and between dyes. The optimal DNA input range of ~600pg (as determined by the sensitivity study) has been verified to be appropriate since samples within this range produced peak height ratios of 80% or greater. Robust samples can be expected to routinely generate peak height ratios of greater than 80% with imbalance only occasionally reaching as low as ~50%. Samples amplified with lower DNA input amounts produce lower DNA peaks, and there is more imbalance associated with smaller allele peaks. As input amounts approach the stochastic range, peak heights start to drop below 2000 RFU and peak height ratios of less than 30% may be observed. Therefore, low level samples should be interpreted with caution.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Applied Biosystems

Stutter Study

Purpose

The goal of this study is to evaluate the levels of stutter that are observed using the GlobalFiler PCR Amplification Kit on the 3500 genetic analyzers. The GlobalFiler kit contains the same loci used in the AmpFℓSTR® Identifiler® Plus PCR Amplification kit, as well as additional loci, some of which have different stutter patterns. Both locus and allele specific stutter will be used to guide interpretation of DNA profiles amplified with GlobalFiler.

Materials and Methods

This study was conducted in two parts. The first part was done using the original GlobalFiler master mix formulation (29 cycles). During the validation, Applied Biosystems released a new master mix formulation. The second part of this study was done using the reformulated GlobalFiler master mix (29 cycles). For the first part, 67 reference samples were amplified within the optimal input range. A dilution series of an additional 10 samples were amplified at different target amounts from a range of >2ng to ~20pg. For the second part using the reformulated master mix, 56 reference samples were amplified within the optimal target range, and an additional 13 samples were amplified 6 or 7 times each within a target range of 800 to 12.5pg. A total of 156 samples were used for the stutter study.

Capillary electrophoresis was performed on the amplified product using both 3500 genetic analyzers and the resultant data was analyzed in GMID-X (v1.4) using a 100 RFU detection threshold (normalization off). Data was first analyzed in the absence of stutter filters in order to collect all stutter heights. Stutter percentage was calculated using the height of the stutter peak with regard to the parent peak.

Results and Discussion

Multiple types of stutter were observed in profiles amplified with GlobalFiler. Historically, stutter has been referred to based on number of base pairs less than the parent peak, but with the incorporation of loci other than tetranucleotide repeats (trinucleotide repeat D22S1045), the number of *repeats* to designate stutter will be used, i.e. n-1 repeat stutter. Several of the new loci in GlobalFiler exhibited stutter half a repeat (2 base pairs) smaller than the parent peak and also 1 repeat unit larger than the parent peak. These will be referred to as n-0.5 repeat and n+1 repeat stutter. Data shown here is only from part two, the reformulated GlobalFiler master mix. Data from the original formulation and how this compares to the reformulated master mix can be found in the validation folder for the lab. Default GMID-X filters were examined with reference to the internal validation, and adjusted accordingly at some loci in order to create a balance

between filtering most stutter without losing allelic peaks in a mixture where DNA contribution is imbalanced.

A summary of the range of n-1 stutter is show in Table 1 alongside the corresponding GMID-X stutter filter. Since GMID-X employs a binary stutter filter, there is no one value that perfectly describes the stutter patterns. It is important to note that the stutter filter is much higher than the mean + 2 standard deviations for the smaller alleles in the locus, and occasionally, an outlier or larger alleles have an acceptable level of observable stutter above this filter (based on the mean and standard deviation for the allele). Larger alleles typically exhibit the highest stutter peaks. Refer to the allele specific stutter in Figure 2 at the individual loci to help assess stutter on an allele specific basis.

Table 1. Summary of stutter filters used by GMID-X for the GlobalFiler loci.

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

Table 1 also lists the stutter filters for n+1 and n-0.5 stutter. Examples of this type of stutter are shown in Figure 1. n-0.5 stutter was observed much more at two loci compared to all other loci, both of which are new STR markers with the GlobalFiler kit: SE33 and D1S1656. With STRmix in mind during this validation, it was ideal to create a separate analysis method that optimally filtered these two types of stutter artifacts, but allowing n-1 stutter to be labeled as alleles and incorporated into the STRmix model (see STRmix ModelMaker Study for more details). Not every locus required this type of filtering.

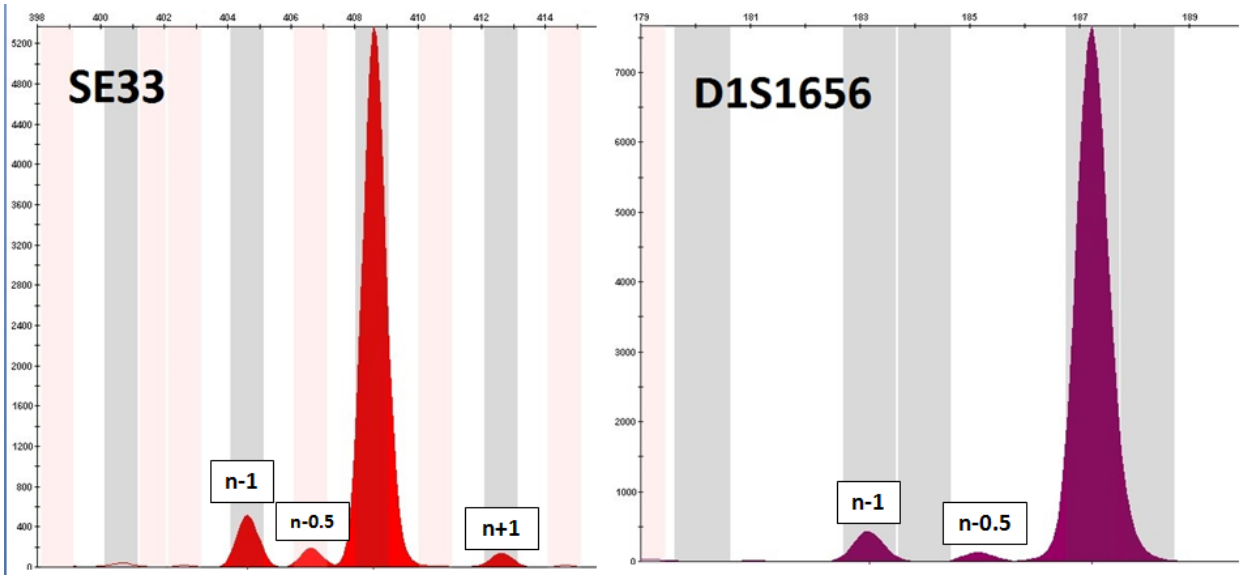
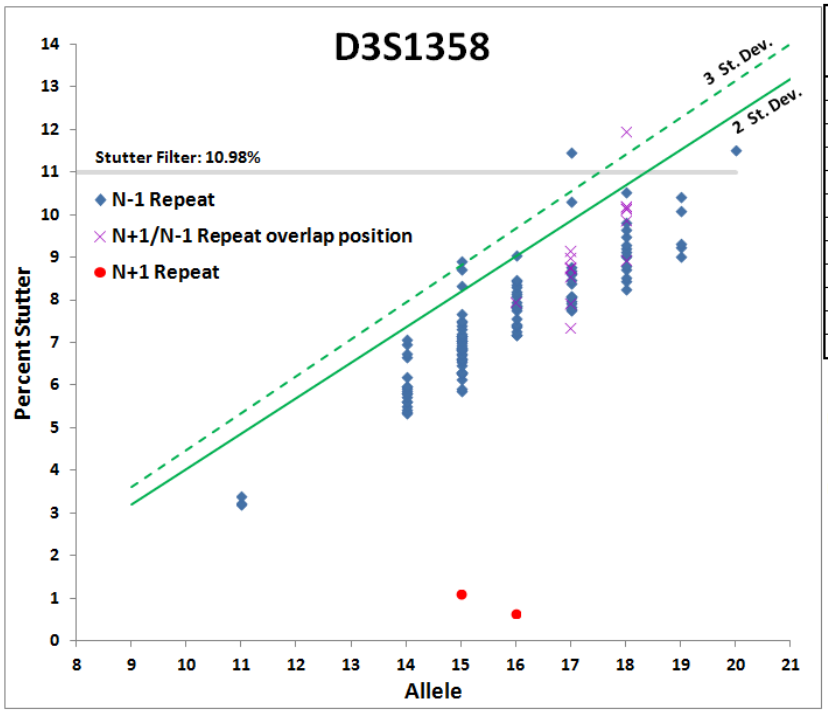


Figure 1. Multiple stutter artifacts originating from one allele peak.

The amount of stutter expected depends on several things: 1 – where the stutter artifact falls in relation to the allele peak (n-1, n+1, etc), 2 – the allele position within the locus, and 3 – the nature of the STR repeat (simple, compound, etc), more specifically, the longest uninterrupted sequence within the STR repeat region that is being amplified. Several of the new loci included in the GlobalFiler kit have complex repeat patterns. For example, at D2S441, a 14 allele is not 14 consecutive repeats, but could be broken into several repeat blocks of variable unit length (See Reference 1 for more information on this topic). This drastically lowers the expected stutter at that allele. It can also be used to better estimate stutter, which varies from allele to allele when there appears to be a second population of stutter. More information about estimating expected stutter based on the longest uninterrupted sequence is presented in the STRmix ModelMaker validation write-up.

The graphs below represent all loci that displayed stutter above the analytical threshold. Stutter varies from allele to allele. Because this validation data is base only on ~150 samples in the internal validation, it does not contain all possible alleles that may be observed in casework. The green lines on this graph are the best fit line of the mean and 2 (solid green line) or 3 standard

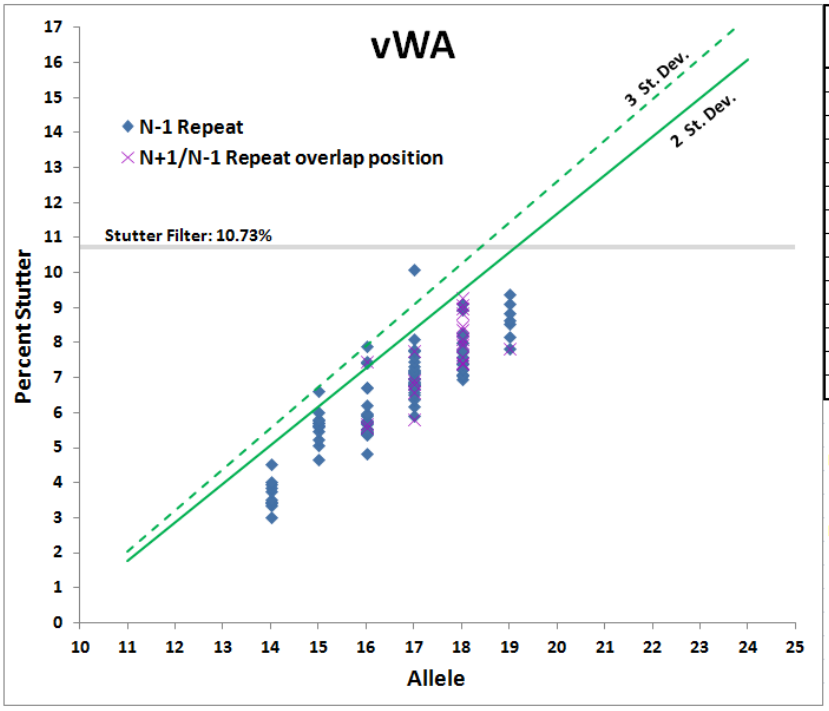
deviations (dashed green line). They model stutter in an allele specific manner, unlike the stutter filter utilized by GMID-X (grey line).



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
9	0	3.21	3.61		
10	0	4.04	4.47		
11	3	4.87	5.34		
12	0	5.70	6.21		
13	0	6.53	7.07		
14	18	7.08	7.62	6.00	0.54
15	39	8.29	8.95	6.97	0.66
16	21	8.87	9.35	7.91	0.48
17	16	10.32	11.21	8.55	0.89
18	14	11.23	12.08	9.51	0.86
19	5	10.84	11.44	9.63	0.60
20	1	12.34	13.14		

Minimum Stutter:
3.22%

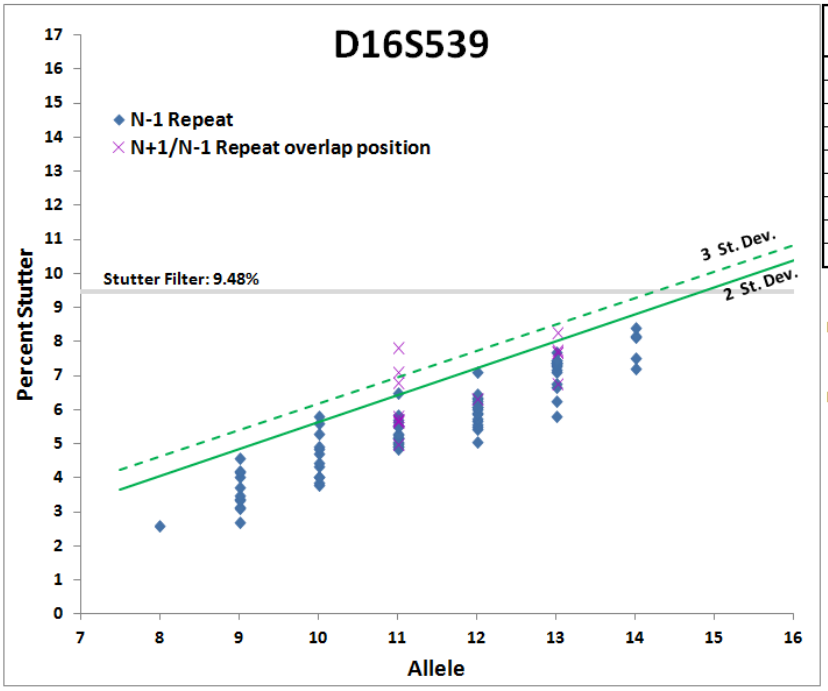
Maximum Stutter:
11.52%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
11	0	1.78	2.05		
12	0	2.88	3.22		
13	0	3.98	4.39		
14	10	4.55	4.98	3.68	0.43
15	13	6.54	7.01	5.61	0.47
16	24	7.65	8.46	6.03	0.81
17	25	8.71	9.51	7.12	0.80
18	18	8.96	9.59	7.70	0.63
19	7	9.72	10.25	8.65	0.53
20	0	11.68	12.60		
21	0	12.78	13.77		
22	0	13.88	14.95		
23	0	14.98	16.12		
24	0	16.08	17.29		

Minimum Stutter:
3.01%

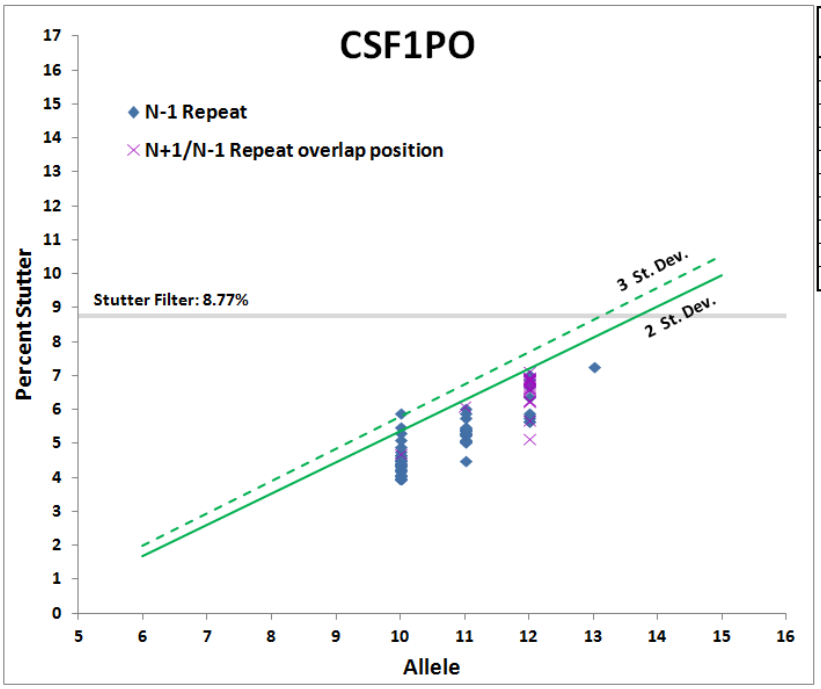
Maximum Stutter:
10.08%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
7	0	3.24	3.86		
8	1	4.03	4.64		
9	11	4.77	5.34	3.64	0.57
10	12	6.00	6.67	4.65	0.67
11	18	6.19	6.61	5.36	0.42
12	17	6.94	7.41	5.99	0.47
13	15	8.14	8.65	7.11	0.51
14	5	8.91	9.41	7.90	0.50
15	0	9.60	10.06		

Minimum Stutter:
2.62%

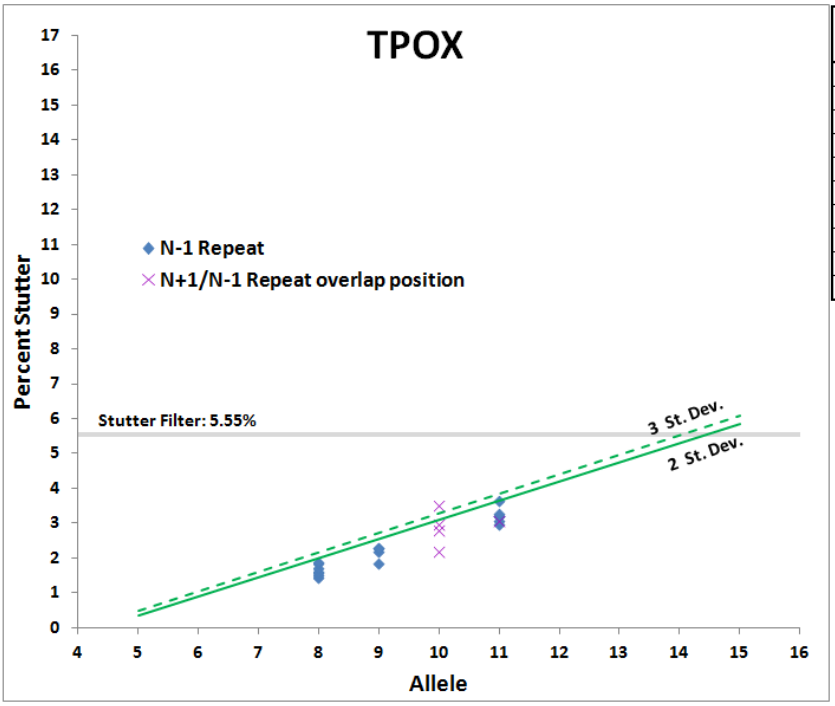
Maximum Stutter:
8.42%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
6	0	1.70	2.00		
7	0	2.62	2.95		
8	0	3.53	3.90		
9	0	4.45	4.85		
10	27	5.51	5.99	4.55	0.48
11	19	6.02	6.36	5.34	0.34
12	10	7.35	7.89	6.26	0.54
13	1	8.13	8.64		
14	0	9.05	9.59		
15	0	9.97	10.54		

Minimum Stutter:
3.98%

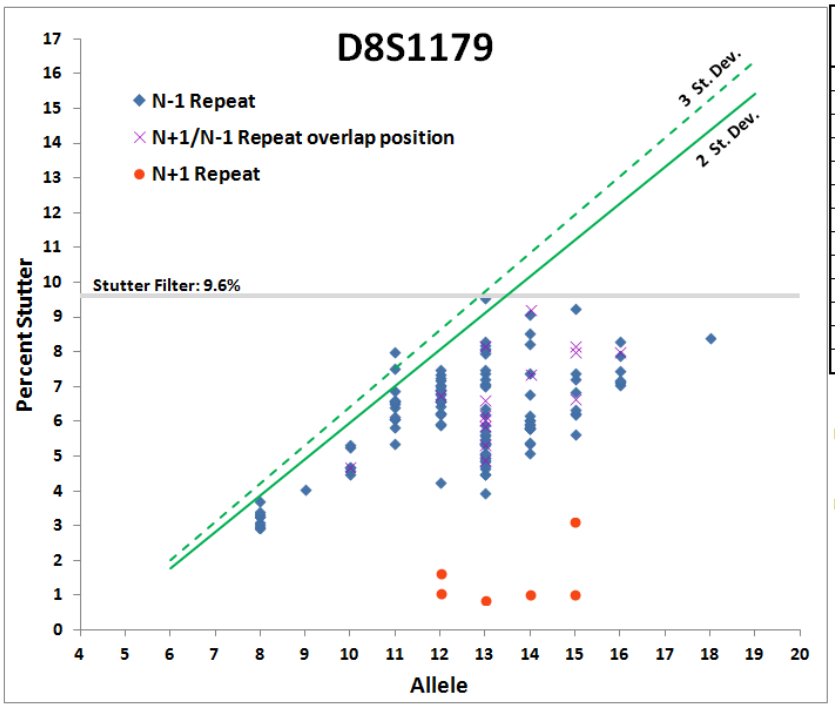
Maximum Stutter:
7.26%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
6	0	0.90	1.05		
7	0	1.45	1.61		
8	7	1.98	2.14	1.66	0.16
9	4	2.56	2.76	2.15	0.20
10	0	3.09	3.29		
11	7	3.63	3.84	3.22	0.21
12	0	4.19	4.41		
13	0	4.74	4.97		
14	0	5.28	5.53		
15	0	5.83	6.09		

Minimum Stutter:
1.44%

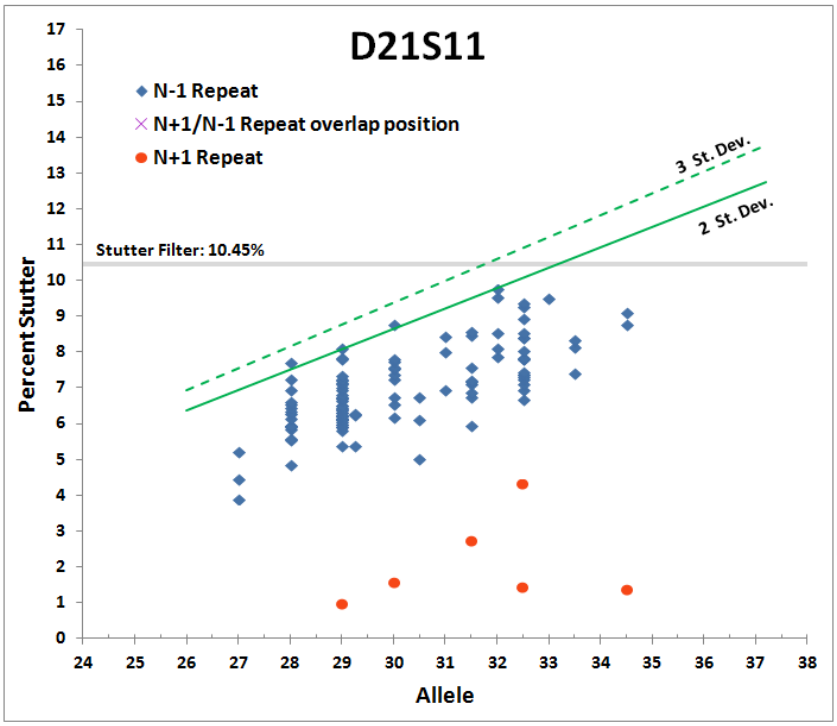
Maximum Stutter:
3.63%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean	SD
6	0	1.79	2.03		
7	0	2.84	3.13		
8	9	3.89	4.23	3.22	0.25
9	1	4.94	5.33		
10	6	5.55	5.91	4.83	0.36
11	13	7.89	8.58	6.51	0.69
12	20	7.65	8.12	6.71	0.47
13	35	9.13	9.74	6.09	1.39
14	17	10.18	10.85	6.43	1.17
15	8	11.22	11.95	6.89	1.11
16	7	12.27	13.05	7.45	0.47
17	0	13.32	14.15		
18	1	14.37	15.25		

Minimum Stutter:
2.94%

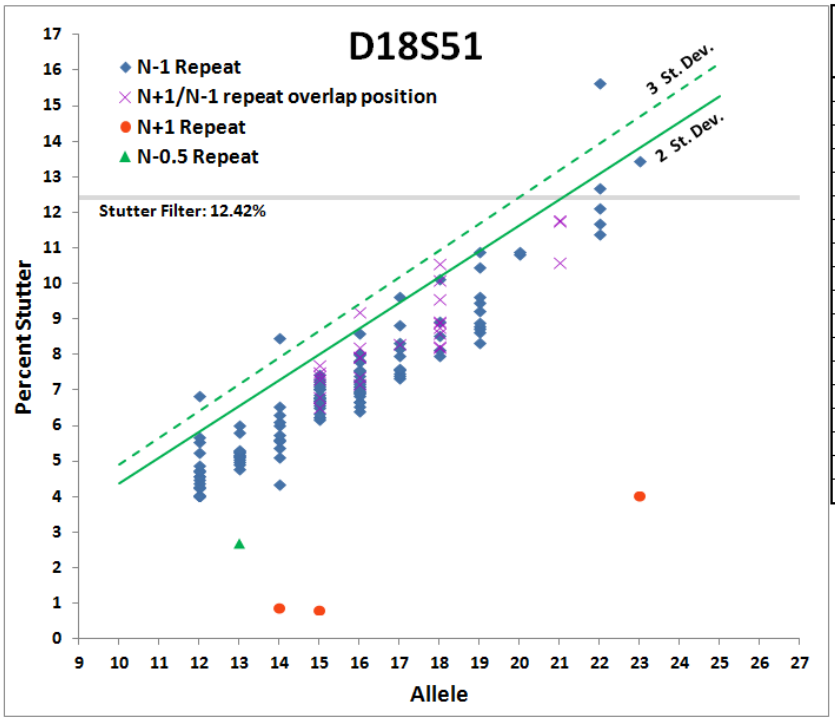
Maximum Stutter:
9.53%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
24	0	5.23	5.70		
25	0	5.79	6.31		
26	0	6.36	6.93		
27	3	6.93	7.54		
28	16	7.65	8.34	6.26	0.69
29	36	7.82	8.44	6.59	0.62
29.1	3	8.13	8.82		
30	10	8.83	9.56	7.36	0.73
30.2	3	8.75	9.49		
31	3	9.21	9.98		
31.2	9	8.96	9.78	7.32	0.82
32	5	10.46	11.30	8.77	0.84
32.2	18	9.49	10.28	7.90	0.79
33	1	10.34	11.20		
33.2	3	10.46	11.33		
34	0	10.91	11.81		
34.2	2	11.02	11.94		
35.2	0	11.59	12.55		
36.2	0	12.16	13.16		
37.2	0	12.73	13.77		

Minimum Stutter:
3.91%

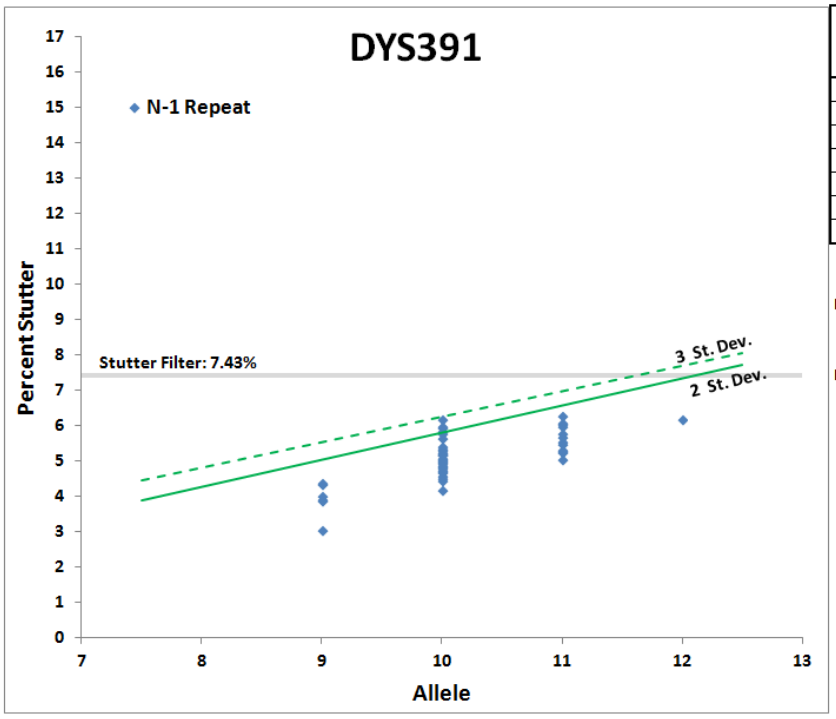
Maximum Stutter:
9.77%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
9	0	3.66	4.15		
10	0	4.38	4.91		
11	0	5.11	5.66		
12	16	6.26	7.00	4.77	0.74
13	12	5.94	6.29	5.23	0.35
14	11	7.99	9.03	5.92	1.03
15	22	7.51	7.87	6.79	0.36
16	22	8.36	8.90	7.26	0.55
17	10	9.50	10.23	8.03	0.73
18	5	10.46	11.32	8.73	0.86
19	10	10.96	11.79	9.31	0.83
20	2	11.63	12.44		
21	0	12.36	13.20		
22	5	13.08	13.95	12.70	1.71
23	1	13.81	14.70		
24	0	14.53	15.46		
25	0	15.26	16.21		
26	0	15.98	16.96		

Minimum Stutter:
4.03%

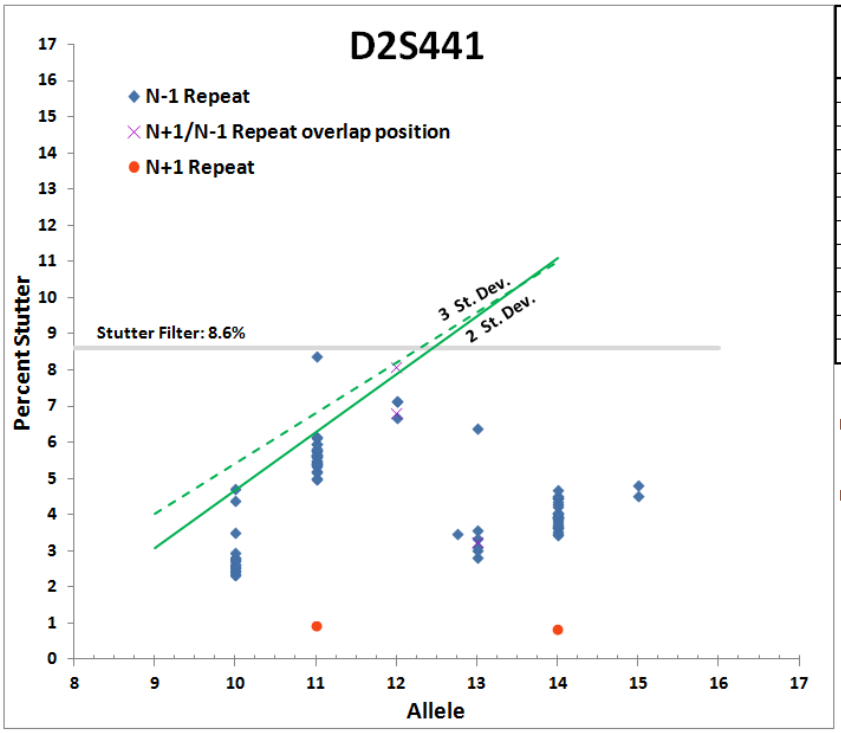
Maximum Stutter:
15.63%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
7	0	3.49	4.08		
8	0	4.26	4.80		
9	6	4.86	5.34	3.91	0.48
10	28	6.12	6.62	5.12	0.50
11	12	6.41	6.79	5.63	0.39
12	1	7.34	7.70		
13	0	8.11	8.43		

Minimum Stutter:
3.02%

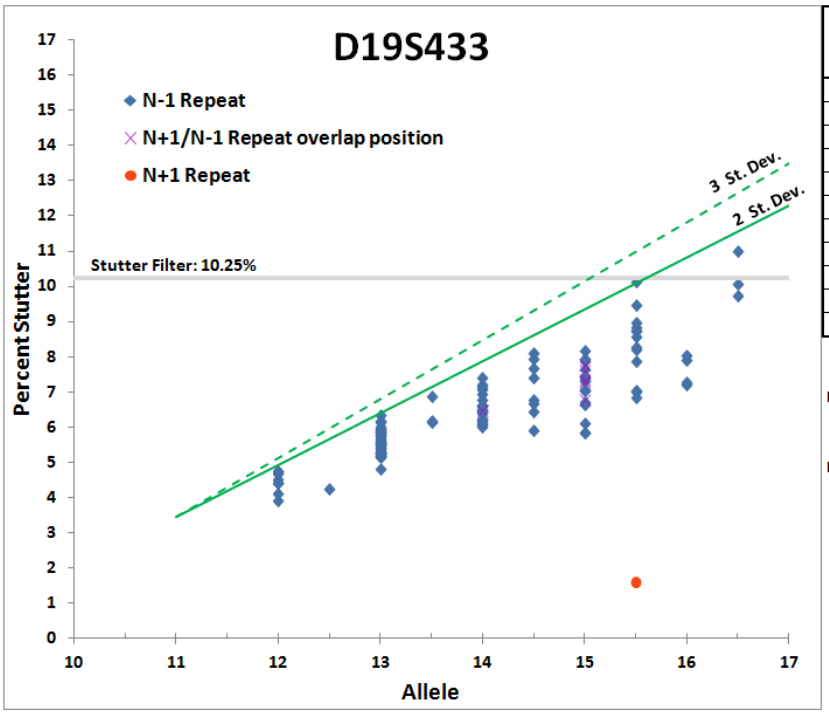
Maximum Stutter:
6.18%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
8	0	1.45	2.62		
9	0	3.06	4.02		
10	16	4.33	5.03	2.93	0.70
11	26	6.95	7.58	5.70	0.63
11.3	0	6.76	7.23		
12	3	7.55	7.82	7.00	0.28
12.3	1	8.37	8.63		
13	6	9.50	9.60	3.72	1.33
14	26	11.10	11.00	4.01	0.35
15	3	12.71	12.40		
16	0	14.32	13.79		
17	0	15.93	15.19		

Minimum Stutter:
2.36%

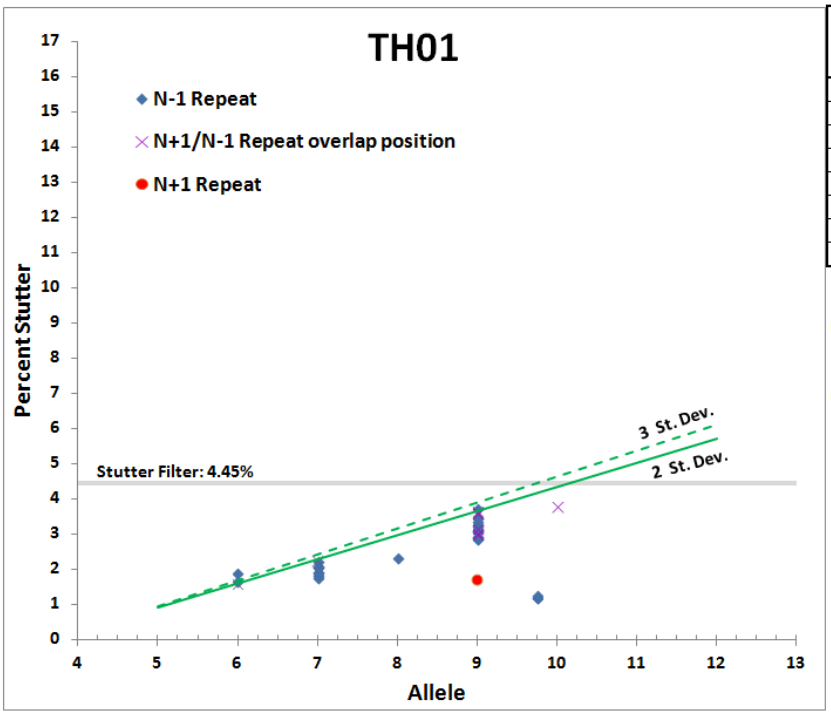
Maximum Stutter:
8.38%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
10	0	1.97	1.79		
11	0	3.44	3.46		
12	8	5.06	5.36	4.44	0.31
13	34	6.28	6.63	5.59	0.35
14	20	7.38	7.78	6.58	0.40
14.2	8	8.68	9.47	7.11	0.78
15	14	8.58	9.34	7.07	0.76
15.2	13	10.32	11.30	8.36	0.98
16	4	10.81	11.83		
16.2	3	11.11	12.16		
17	0	12.28	13.50		

Minimum Stutter:
3.91%

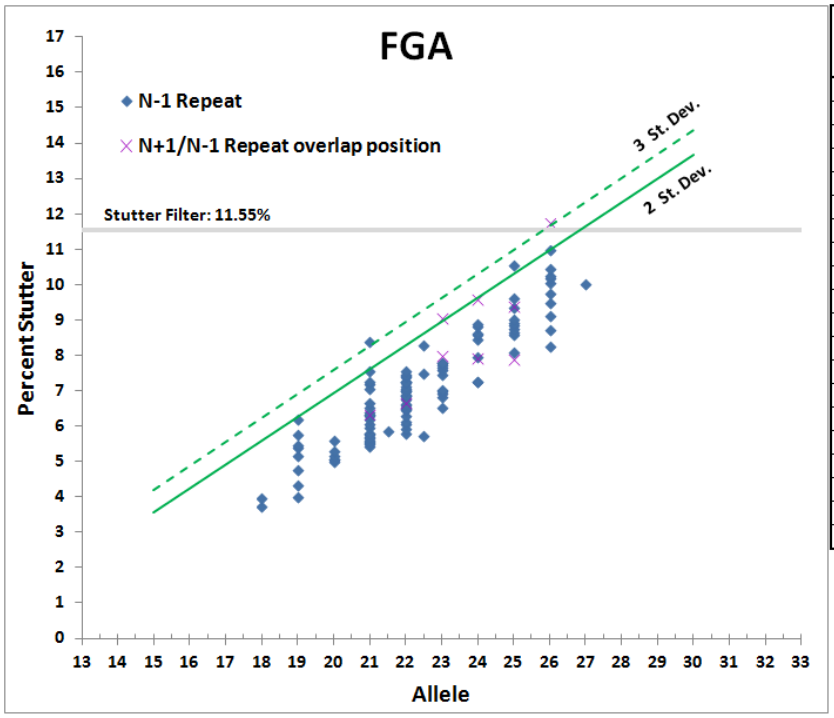
Maximum Stutter:
11.00%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
5	0	0.89	0.94		
6	2	1.58	1.67		
7	8	2.26	2.41	1.96	0.15
8	1	2.95	3.15		
9	13	3.63	3.88	3.14	0.25
9.3	3	3.84	4.10		
10	0	4.32	4.62		
11	0	5.01	5.35		

Minimum Stutter:
1.16%

Maximum Stutter:
3.69%

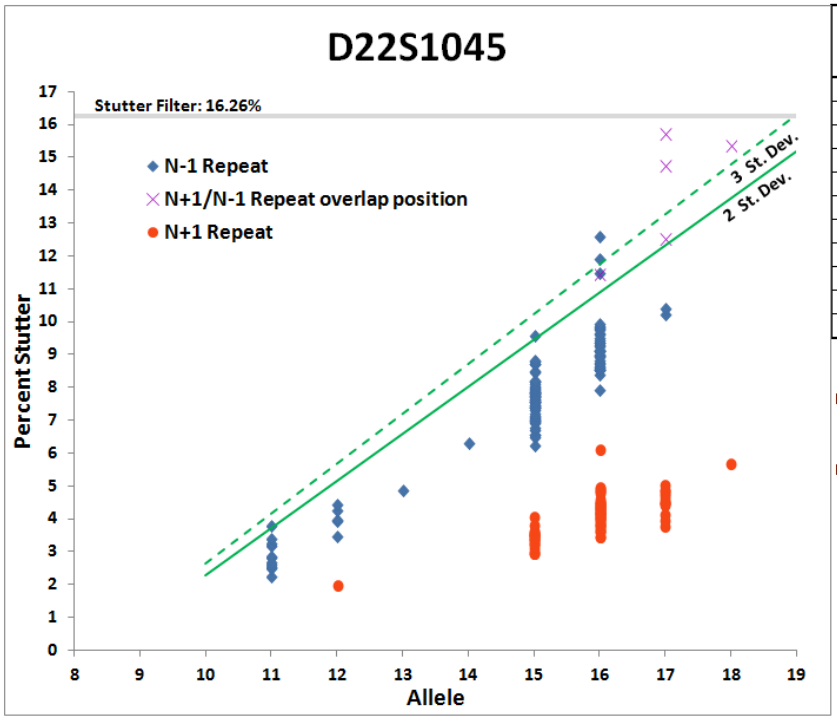


		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
15	0	3.56	4.19		
16	0	4.23	4.86		
17	0	4.90	5.54		
18	2	5.58	6.22		
19	8	6.61	7.34	5.16	0.72
20	9	6.95	7.64	5.57	0.69
21	21	7.84	8.61	6.30	0.77
21.2	1	7.74	8.39		
22	27	7.83	8.32	6.85	0.49
22.2	3	8.41	9.07		
23	11	8.23	8.68	7.33	0.45
24	8	9.60	10.27	8.26	0.67
25	10	10.42	11.09	9.07	0.67
26	10	11.41	12.24	9.75	0.83
27	1	11.65	12.32		
28	0	12.32	13.00		
29	0	12.99	13.68		
30	0	13.67	14.36		
31	0	14.34	15.04		
32	0	15.02	15.72		

Minimum Stutter:
3.75%

Maximum Stutter:
10.99%

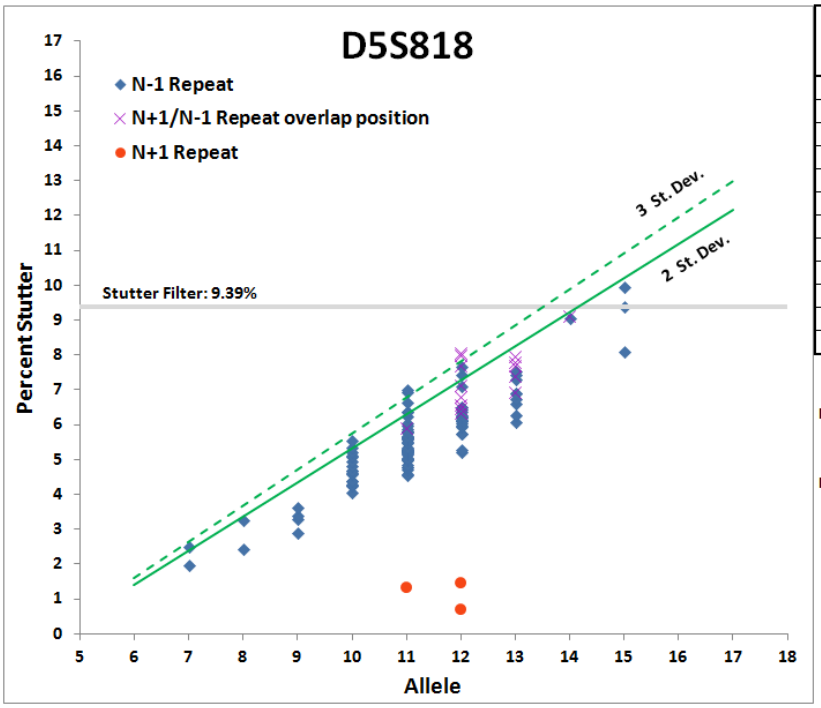
* trinucleotide repeat



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
9	0	0.85	1.11		
10	0	2.28	2.63		
11	11	3.81	4.27	2.89	0.46
12	5	4.76	5.13	4.01	0.37
13	1	6.57	7.18		
14	1	8.00	8.70		
15	55	8.93	9.58	7.62	0.65
16	35	11.27	12.23	9.34	0.96
17	2	12.30	13.24		
18	0	13.73	14.76		
19	0	15.16	16.28		

Minimum Stutter:
2.24%

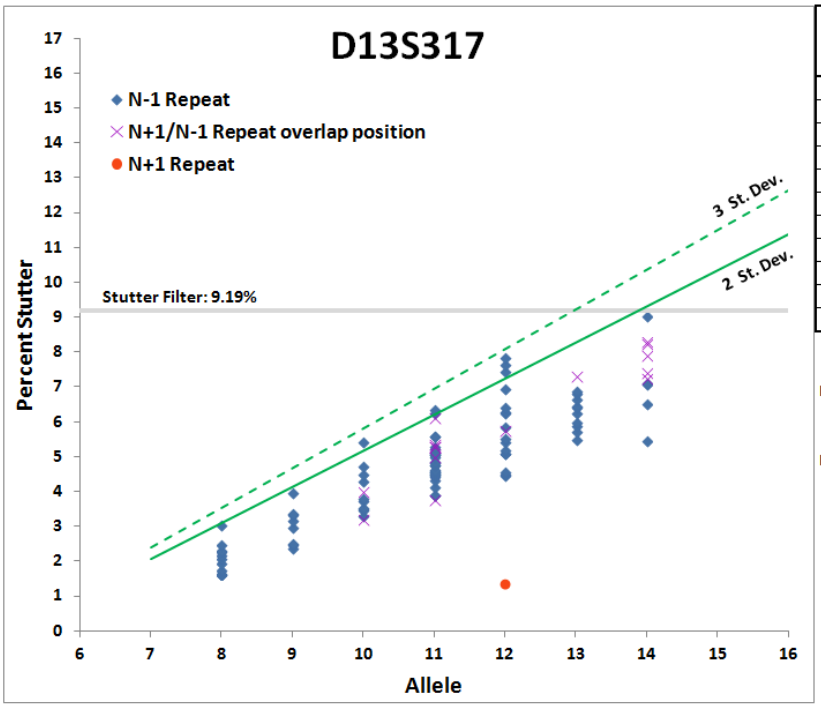
Maximum Stutter:
12.59%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
7	2	2.38	2.64		
8	2	3.36	3.67		
9	4	3.89	4.19	3.30	0.30
10	17	5.58	6.02	4.72	0.43
11	41	6.70	7.30	5.51	0.59
12	21	7.44	8.04	6.25	0.60
13	9	7.86	8.36	6.87	0.50
14	1	9.24	9.89		
15	3	10.22	10.93		
16	0	11.20	11.96		
17	0	12.18	13.00		
18	0	13.16	14.03		

Minimum Stutter:
1.95%

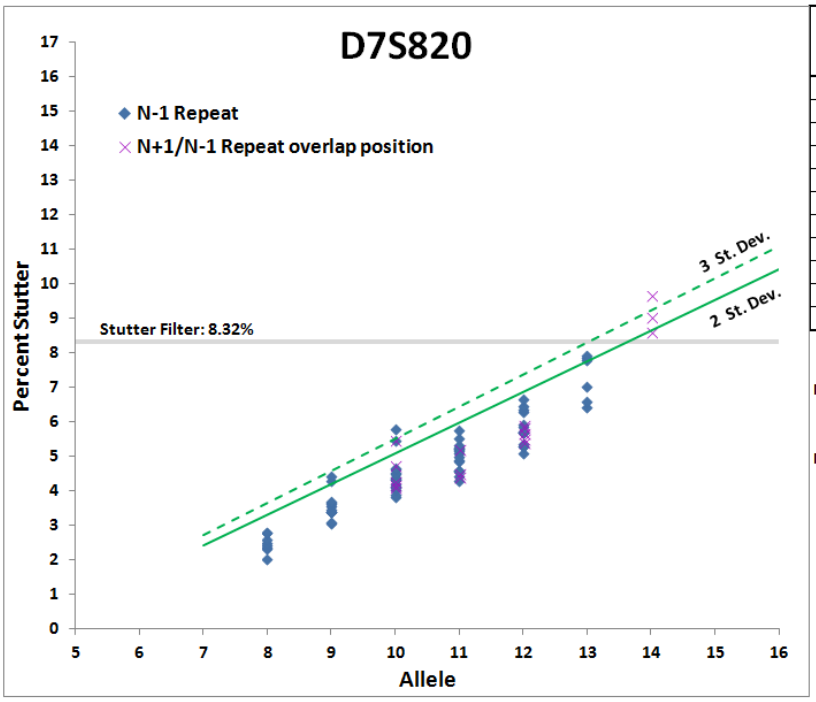
Maximum Stutter:
9.95%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
6	0	1.03	1.26		
7	0	2.07	2.39		
8	11	2.95	3.38	2.08	0.44
9	8	4.10	4.65	3.01	0.55
10	10	5.39	6.08	4.02	0.69
11	28	6.02	6.61	4.84	0.59
12	16	8.14	9.26	5.90	1.12
13	10	7.20	7.67	6.25	0.48
14	5	9.62	10.91	7.03	1.29
15	0	10.34	11.48		
16	0	11.37	12.62		

Minimum Stutter:
1.61%

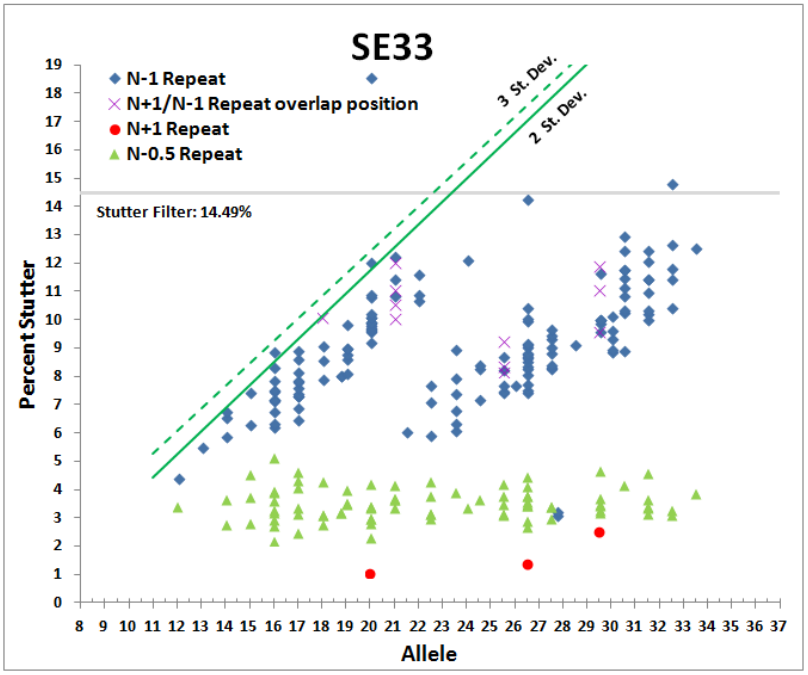
Maximum Stutter:
9.01%



Allele	N	Uses equation of line: Mean + 2 St. Dev. (%)	Uses equation of line: Mean + 3 St. Dev. (%)	Mean (%)	SD
6	0	1.55	1.79		
7	0	2.44	2.72		
8	8	3.02	3.27	2.50	0.26
9	12	4.43	4.84	3.61	0.41
10	17	5.47	5.98	4.45	0.51
11	17	5.82	6.22	5.03	0.40
12	15	6.76	7.23	5.83	0.46
13	6	8.64	9.31	7.30	0.67
14	0	8.65	9.22		
15	0	9.54	10.15		
16	0	10.43	11.08		

Minimum Stutter:
2.04%

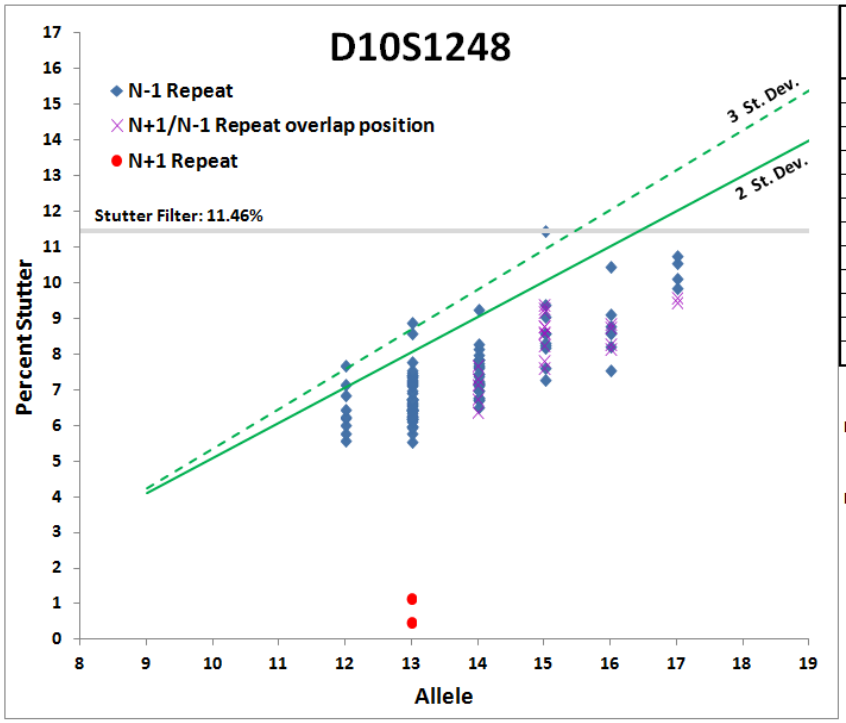
Maximum Stutter:
7.93%



Minimum Stutter:
4.43%

Maximum Stutter:
18.60%

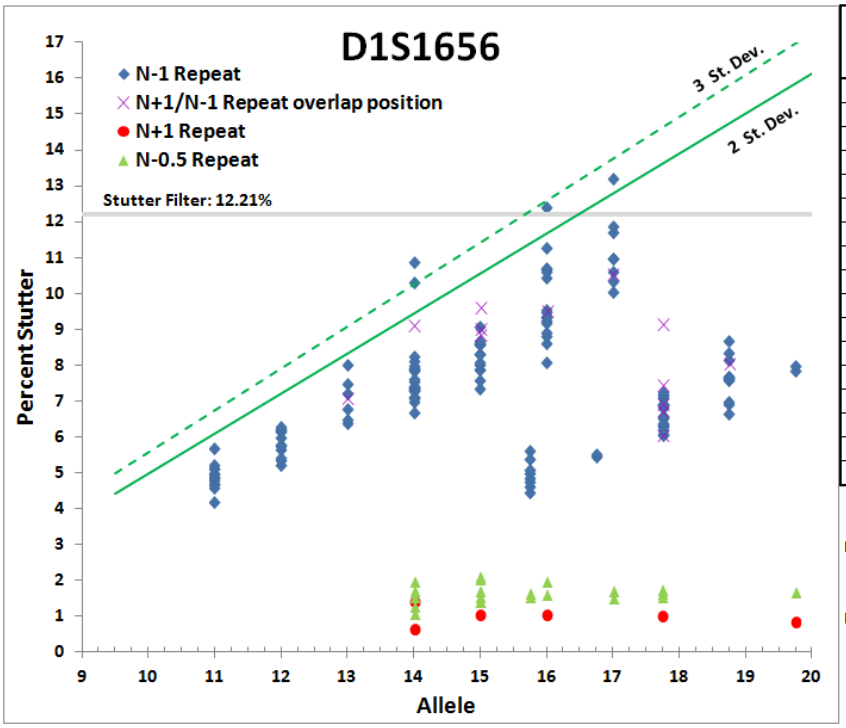
Allele	N	Uses equation of line: Mean + 2 St. Dev. (%)	Uses equation of line: Mean + 3 St. Dev. (%)	Mean (%)	SD
8	0	1.98	2.89		
9	0	2.80	3.68		
10	0	3.61	4.48		
11	0	4.43	5.27		
12	1	5.24	6.06		
13	1	6.05	6.86		
14	3	6.87	7.65		
15	2	7.68	8.45		
16	11	8.96	9.75	7.37	0.80
17	11	9.10	9.82	7.67	0.71
18	3	10.12	10.83		
19	6	10.05	10.62	8.90	0.58
20	16	12.02	12.71	10.64	0.69
21	3	12.92	13.61	11.52	0.70
21.2	1	12.72	13.37		
22	3	13.37	14.00		
22.2	3	13.54	14.16		
23.2	6	14.35	14.96	7.26	1.08
24	1	15.00	15.59		
24.2	3	15.16	15.75		
25.2	5	15.98	16.54	7.92	0.55
26	1	16.63	17.18		
26.2	18	16.79	17.34	9.06	1.54
27.2	8	17.60	18.13	8.93	0.55
28.2	1	18.42	18.93		
29.2	6	19.23	19.72	10.20	0.73
30	5	19.88	20.35	9.39	0.52
30.2	10	20.04	20.51	11.21	1.17
31.2	8	20.86	21.31	11.14	0.90
32.2	5	21.67	22.10	12.25	1.65
33.2	1	22.49	22.90		
34	0	23.14	23.53		
35	0	23.95	24.32		
36	0	24.76	25.12		
37	0	25.58	25.91		



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
8	0	3.10	3.11		
9	0	4.09	4.22		
10	0	5.08	5.34		
11	0	6.07	6.46		
12	9	8.08	8.75	6.45	0.68
13	48	8.08	8.75	6.75	0.67
14	29	8.52	9.11	7.35	0.59
15	10	10.99	12.15	8.68	1.15
16	7	10.57	11.46	8.77	0.90
17	4	12.02	13.16		
18	0	13.01	14.27		
19	0	14.00	15.39		

Minimum Stutter: 5.54%

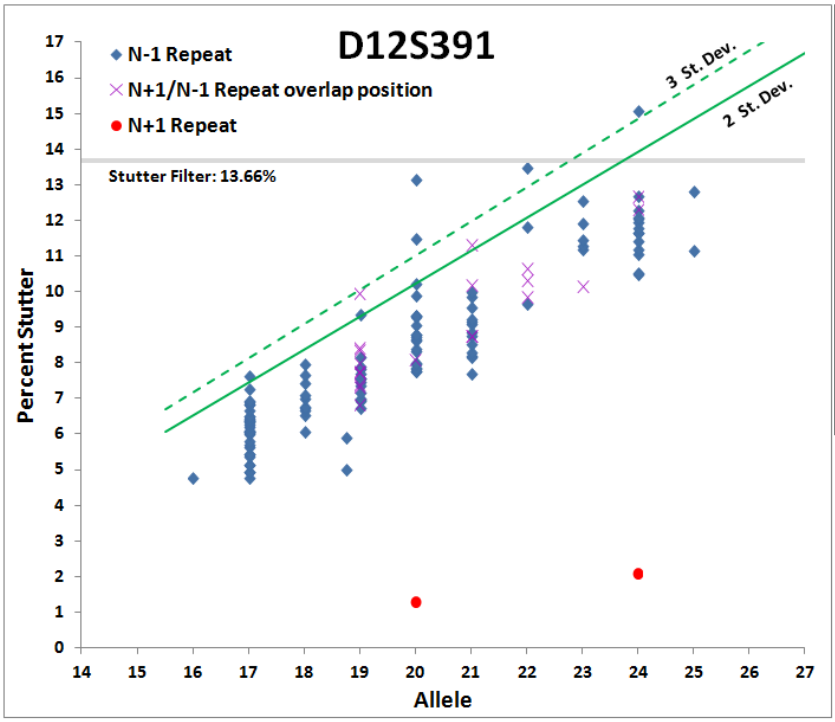
Maximum Stutter: 11.47%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
9	0	3.86	4.38		
10	0	4.97	5.55		
11	13	6.79	7.73	4.89	0.95
12	11	6.56	6.94	5.81	0.38
13	6	8.32	8.95	7.05	0.63
14	20	9.87	10.91	7.80	1.04
15	14	9.19	9.67	8.24	0.48
15.3	8	10.88	11.76	4.96	0.38
16	16	12.04	13.17	9.78	1.13
16.3	2	12.00	12.93		
17	8	13.25	14.27	11.21	1.02
17.3	17	13.11	14.10	6.63	0.41
18	11	14.22	15.27		
18.3	11	14.22	15.27	7.62	0.62
19	0	15.01	16.09		
19.3	7	15.34	16.44	9.17	1.32
20	0	16.12	17.26		

Minimum Stutter: 4.16%

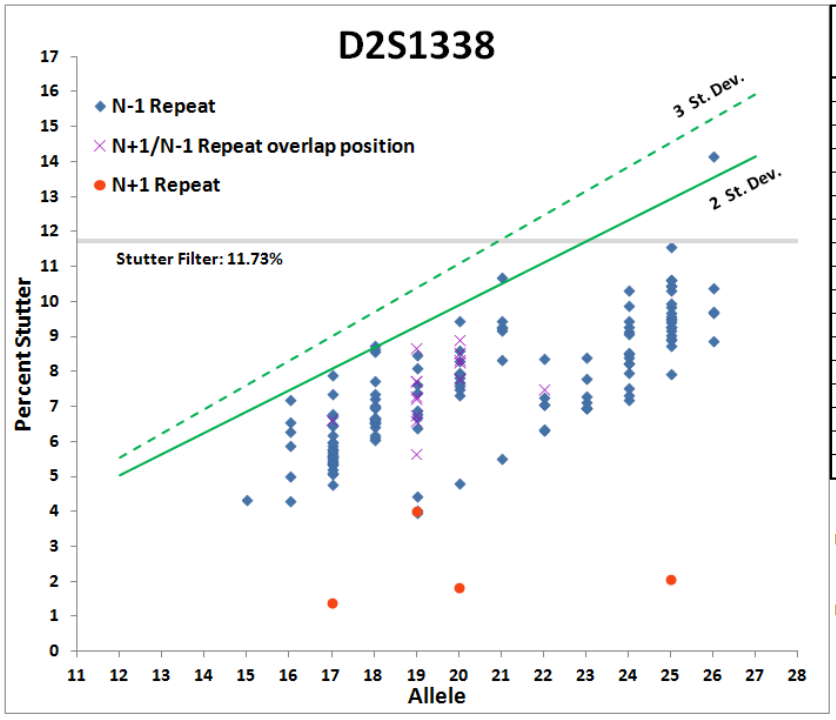
Maximum Stutter: 13.20%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
14	0	4.67	5.25		
15	0	5.59	6.21		
16	1	6.52	7.17		
17	33	7.50	8.19	6.11	0.69
18	10	8.15	8.71	7.01	0.57
18.3	2	8.65	9.38		
19	14	8.90	9.56	7.57	0.66
20	21	11.67	12.99	9.03	1.32
21	13	10.28	10.97	8.89	0.69
22	3	12.07	12.94		
23	5	12.81	13.37	11.69	0.56
24	14	14.11	15.23	11.87	1.12
25	2	14.85	15.82		
26	0	15.78	16.78		
27	0	16.70	17.74		

Minimum Stutter:
4.78%

Maximum Stutter:
15.09%



		Uses equation of line			
Allele	N	Mean + 2 St. Dev. (%)	Mean + 3 St. Dev. (%)	Mean (%)	SD
12	0	5.02	5.55		
13	0	5.63	6.24		
14	0	6.23	6.93		
15	1	6.84	7.62		
16	6	8.00	9.06	5.88	1.06
17	28	7.26	7.97	5.86	0.70
18	22	8.56	9.35	6.99	0.79
19	11	9.65	11.25	6.44	1.60
20	13	9.85	10.89	7.77	1.04
21	7	10.49	11.78	8.83	1.62
22	6	11.10	12.48	7.08	0.74
23	7	11.70	13.17	7.37	0.56
24	15	12.31	13.86	8.62	0.92
25	22	12.92	14.56	10.03	1.84
26	5	13.53	15.25	10.57	2.07
27	0	14.13	15.94		
28	0	14.74	16.64		

Minimum Stutter:
3.96%

Maximum Stutter:
14.15%

Figure 2. Allele specific stutter artifacts in single source samples amplified with GlobalFiler

Conclusions

The GeneMapper ID-X marker specific stutter ratios listed in Table 1 are capable of filtering out the majority of observed stutter peaks without being excessively high to the point where

substantial filtering of true alleles may occur. For interpretation of DNA samples, stutter should be assessed in an allele specific manner. Most of the stutter will be filtered by the binary GMID-X filters, but as described above, and as the grey lines in the graph indicate, it does not accurately model the biological pattern of stutter that is observed. The tables of allele specific stutter at loci in Figure 2 can assist with this assessment.

References

1. Butler, J. Advanced Topics in Forensic DNA Typing: Methodology. 2012 Academic Press.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Life Technologies

Challenged Samples Study

Purpose

The goal of this study is to demonstrate the utility of the GlobalFiler PCR Amplification Kit on challenging adjudicated casework and mock casework samples.

Materials and Methods

Three challenging adjudicated casework and mock casework samples were previously amplified and analyzed using the Minifiler, Identifiler and Identifiler Plus (ID Plus) amplification kits (Table 1). The current study examined results after amplification with GlobalFiler and compared these results to what was previously obtained. The DNA extracts were amplified for 29 cycles using GlobalFiler. Capillary electrophoresis was performed on the 3500 Genetic Analyzers and the resultant data was analyzed in GMID-X (v1.4) with a 100 RFU analytical threshold.

Table 1 - Description of samples used for the Challenged Samples Study.

Sample	Description
846	Bloodstain from a t-shirt in a 1977 homicide (inhibited and degraded).
(2-1)	Bone sample (degraded).
(3-1)	Bone sample from a 1992 helicopter crash (degraded).

Results and Discussion

Figure 1 presents the DNA typing results obtained using Minifiler, Identifiler, ID Plus, and GlobalFiler for comparison. All three of the samples amplified with GlobalFiler (GF) resulted in more genetic information overall than was previously obtained with Minifiler, Identifiler, ID Plus, although a full profile was not obtained.

Because GlobalFiler is the only kit shown here with a 6^{th} dye channel, it has more mini-STRs than the other 3 kits. Because of this, there is a clear advantage of GlobalFiler over Identifiler and Identifiler Plus. And although GlobalFiler provides more information than Minifiler does, it is not always the same information. Minifiler still amplifies smaller amplicons than GlobalFiler so it can supplement the genetic information that GlobalFiler Provides. Table 2 lists the genetic information obtained with GlobalFiler compared to Minifiler.

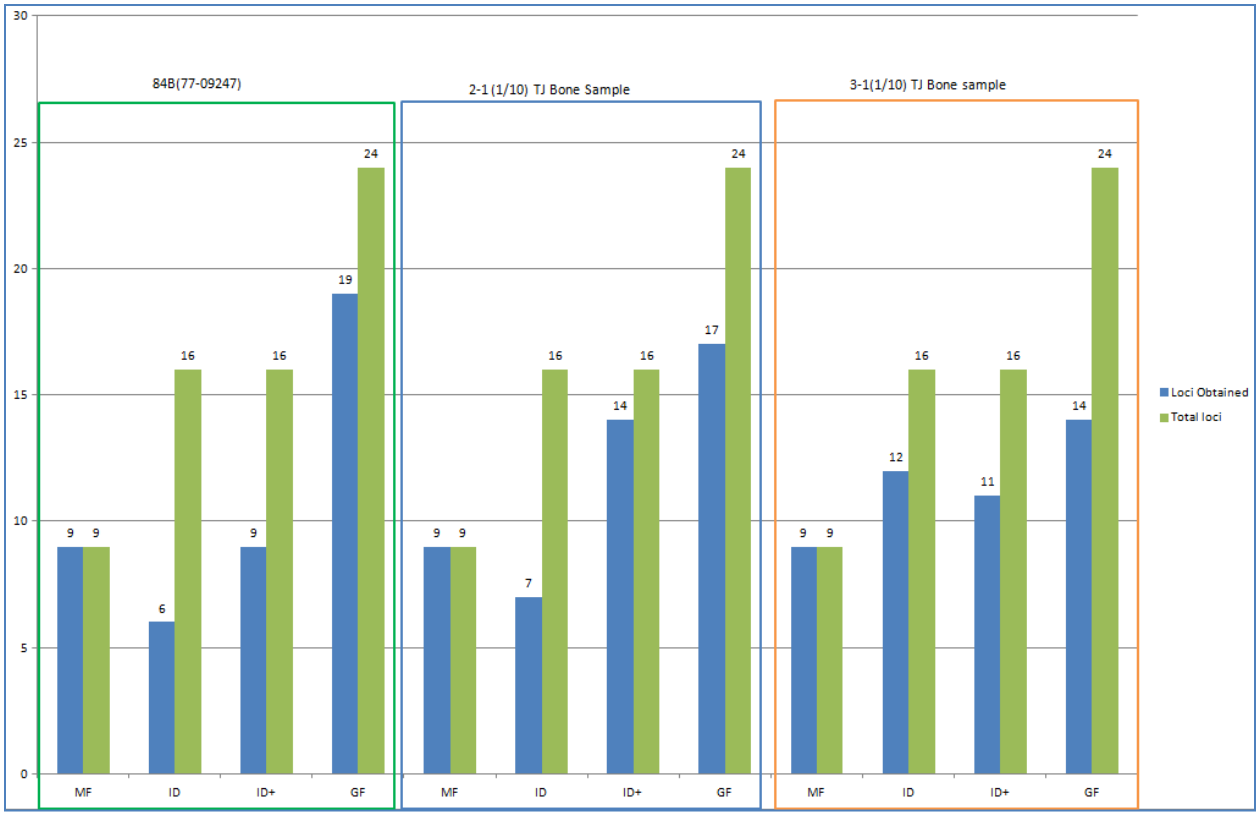


Figure 1 – Summary of genetic information tested and obtained with each of the kits.

Kit	Sample	D3S1358	vWA	D16S539	CSF1PO	TPOX	Yindel	Amel	D8S1179	D21S11	D18S51	DYS391	D2S441	D19S433
MF	846	ND	ND	12	10,12	ND	-	X	ND	28,30	12,17	-	-	-
GF	846	15,19	16,19	12*	ND	ND	NR	X	12,15	28,30	17*	NR	10,11.3	12,14
MF	2-1	ND	ND	11,12,13,14	11,12	ND	-	X,Y	ND	29,30,31	13,15	-	-	-
GF	2-1	15,[16],[18]	14,17*,18	NR	NR	NR	2	X,Y	13,[14]*,[15]*	29*,30*	15*	NR	10,[11],14	[12],13,14
MF	3-1			12,13	10			X,Y		31.2,35.2	14,17			
GF	3-1	15,18	18	NR	NR	NR	2	X,Y	13,15	NR	NR	NR	11,12	13

loci, continued below.....

Kit	Sample	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338	PentaD	PentaE
MF	846	ND	23,24	-	ND	11,12	11,12	-	-	-	-	17,26	-	-
GF	846	8,9	23,24	11,17	11,13	11*,12*	NR	15*	13,14	15	18*,19*	NR	-	-
MF	2-1	ND	19,23,25	-	ND	9,11,12,13	11,12	-	-	-	-	17,20,24	-	-
GF	2-1	6,[9.3]	19,23,25*	15,16	9,10,12,13	9*,11*	NR	NR	13,[14]	16	15*,20*	NR	-	-
MF	3-1		19,24			8,12	9,10					21,23		
GF	3-1	6,9	19,24	16	11,12	NR	NR	NR	13,16	12,16	18	NR	-	-

Table 2 – Minifiler (MF) and GlobalFiler (GF) DNA typing results for Challenged Samples Study. ND = No DNA types obtained.

An electropherograms from sample 2-1 is shown in Figure 2.

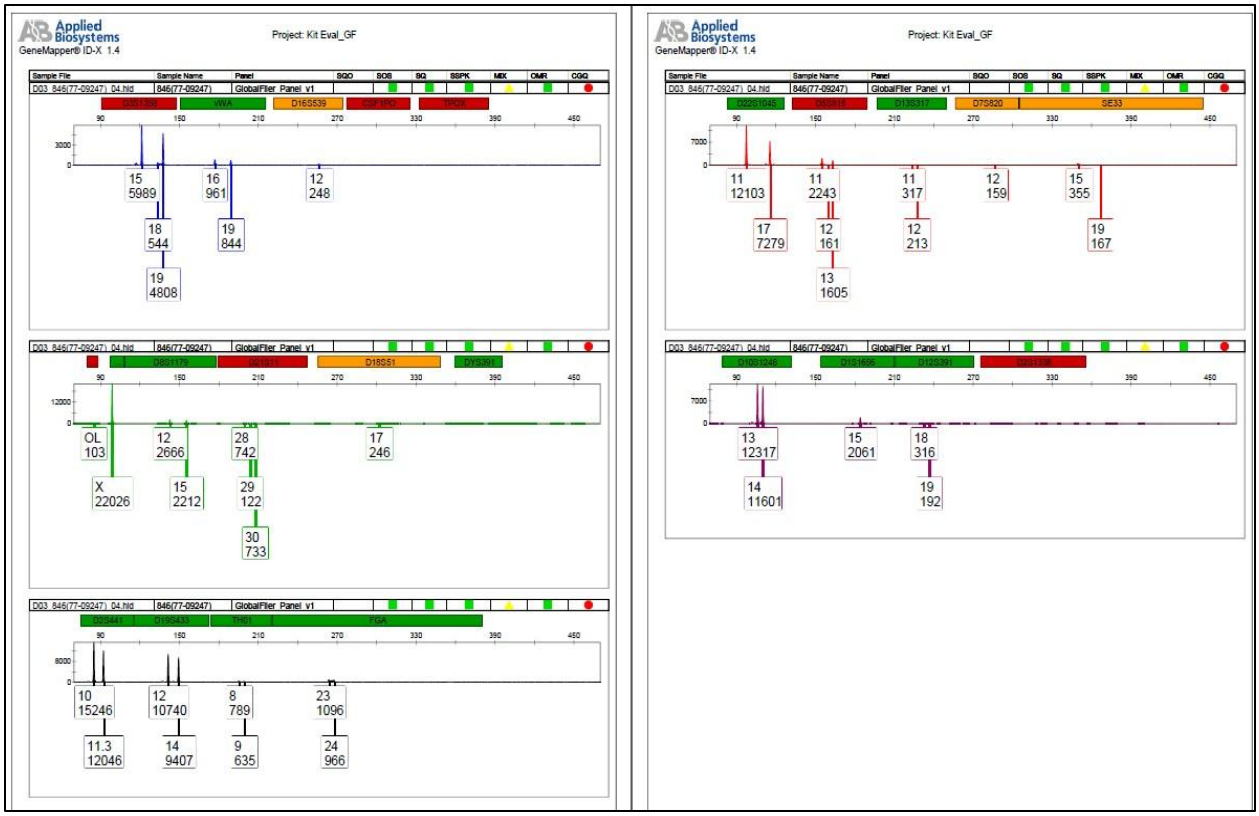


Figure 2: GlobalFiler profile for sample 846.

Conclusions

At least a partial DNA profile was obtained for all of the challenged samples amplified using GlobalFiler. Amplifying these samples with GlobalFiler provided the most genetic information. GlobalFiler has been shown here to be appropriate for STR PCR amplification of challenged samples from adjudicated casework and mock casework.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Applied Biosystems

Mixture Study

Introduction

Mixtures are an important component of an STR typing kit validation. Mixture studies should inform DNA interpretation guidelines for the laboratory. In contrast to previous mixture interpretation procedures, the lab is incorporating the probabilistic genotyping software STRmix for interpretation of DNA results. The validation of GlobalFiler is tightly intertwined with the validation of STRmix.

Purpose

The main goal of this study was to create a very wide variety of mixtures in order to assess the GlobalFiler kit performance, but more importantly to assess the deconvolution of mixed DNA samples by STRmix. These mixtures were constructed using DNA from contributors whose profiles were previously typed with GlobalFiler, and are referred to as "ground truth" mixtures, because the genotypes and amounts of each contributor are known. The range of contributor combinations and input amounts was designed to be very large in order to measure consistency and accuracy of STRmix, and to help determine where its limits are. Deconvolution results of these mixtures are further described in the STRmix MCMC study. The purpose of this study is to verify that the contributor amounts reflect the mixture ratio design, verify that only the correct genotypes are included in the mixtures, and to quantify the amount of dropout occurring, if any.

Materials and Methods

Two, three, four, and five person DNA mixtures were created in a variety of ratios (Table 1) using reference samples previously genotyped with GlobalFiler. Target ratios were designed similar to the previous mixtures generated in the Identifiler Plus mixture study. In order to examine the affects of allele stacking, two different combinations of people were used for every different mixture ratio combination. Because electropherograms of single source reference samples were examined closely before mixture design, average peak height of the alleles from the single source samples was taken into account when combining them to make mixtures. For example, if two samples were amplified with the same DNA target input amount and one single source sample had approximately double the peak heights of another sample, that would be factored into the construction of the mixtures so that approximately double the amount of DNA would be used for the one with shorter peak heights in a 1:1 mixture (see Table 2 for details and amounts used for each sample). Each set of mixtures was amplified at three different overall target levels to examine the effect of peak height on mixture deconvolution.

Table 1 – Summary list of all samples with target ratios

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

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

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

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

Table 2 – List of each mixture with target ratio, target input amount, and adjusted input amount

	Sample ID	Ratios	Target pg	Target pg	Adjusted for PH (pg)	Adjusted for PH (pg)
High (highest at 3-10K RFU)	2-1	8:1	800	100	**800**	125
	2-2	5:1	800	160	**800**	200
	2-3	2:1	700	350	**700**	425
	2-4	1:1	700	700	**700**	850
	2-5	1:2	400	800	**400**	975
	2-6	1:5	175	875	**175**	1000
	2-7	1:8	125	1000	**125**	1200
	2-8	8:1	800	100	**800**	175
	2-9	5:1	800	160	**800**	275
	2-10	2:1	700	350	**700**	600
	2-11	1:1	700	700	**700**	1225
	2-12	1:2	400	800	**400**	1400
	2-13	1:5	175	875	**175**	1500
	2-14	1:8	125	1000	**125**	1750
Mid (highest at 1-3K RFU)	2-15	8:1	600	75	**600**	90
	2-16	5:1	600	100	**600**	125
	2-17	2:1	600	200	**600**	240
	2-18	1:1	600	600	**600**	720
	2-19	1:2	200	600	**200**	720
	2-20	1:5	100	600	**100**	720
	2-21	1:8	75	600	**75**	720
	2-22	8:1	600	75	**600**	130
	2-23	5:1	600	100	**600**	175
	2-24	2:1	600	200	**600**	350
	2-25	1:1	600	600	**600**	1050
	2-26	1:2	200	600	**200**	1050
	2-27	1:5	100	600	**100**	1050
	2-28	1:8	75	600	**75**	1050
Low (lowest samples 200-500 RFU)	2-29	8:1	320	40	325	**50**
	2-30	5:1	250	50	250	**60**
	2-31	2:1	150	75	150	**90**
	2-32	1:1	75	75	**75**	**90**
	2-33	1:2	75	150	**75**	180
	2-34	1:5	50	250	**50**	300
	2-35	1:8	40	320	**40**	380
	2-36	8:1	320	40	375	**80**
	2-37	5:1	250	50	300	**100**
	2-38	2:1	150	75	150	**130**
	2-39	1:1	75	75	**75**	130
	2-40	1:2	75	150	**75**	250
	2-41	1:5	50	250	**50**	440
	2-42	1:8	40	320	**40**	550

	Sample ID	Ratios (in %)	Target pg	Target pg	Target pg	Adjusted for PH (pg)	Adjusted for PH (pg)	Adjusted for PH (pg)
High (highest at 3-5K RFU)	3-1	33.3:33.3:33.3	**650**	650	650	**650**	780	810
	3-2	33.3:33.3:33.3	**650**	650	650	**650**	650	650
	3-3	70:20:10	**700**	200	100	**700**	200	80
	3-4	70:20:10	**900**	250	125	**900**	250	125
	3-5	60:30:10	**750**	375	125	**750**	1900	125
	3-6	60:30:10	**900**	450	150	**900**	275	100
	3-7	50:40:10	**700**	560	140	**700**	850	230
	3-8	50:40:10	**700**	560	140	**700**	728	180
	3-9	50:30:20	**750**	450	300	**750**	450	200
	3-10	50:30:20	**700**	425	280	**700**	550	330
	3-11	45:45:10	**700**	700	150	**700**	700	150
	3-12	45:45:10	**500**	500	110	**500**	500	120
	3-13	40:40:20	**600**	600	300	**600**	900	300
	3-14	40:40:20	**600**	600	300	**600**	525	400
	3-15	35:35:30	**650**	650	550	**650**	700	650
	3-16	35:35:30	**750**	750	650	**750**	750	550
	3-17	60:20:20	**700**	125	125	**700**	175	150
	3-18	60:20:20	**700**	125	125	**700**	130	130
	3-19	50:25:25	**700**	350	350	**700**	350	250
	3-20	50:25:25	**700**	350	350	**700**	350	490
	3-21	40:30:30	**650**	475	475	**650**	675	775
	3-22	40:30:30	**650**	475	475	**650**	500	475
Mid (highest at 1-1.5K RFU)	3-23	33.3:33.3:33.3	**300**	300	300	**300**	360	375
	3-24	33.3:33.3:33.3	**300**	300	300	**300**	300	300
	3-25	70:20:10	**475**	130	65	**475**	115	50
	3-26	70:20:10	**650**	180	90	**650**	170	90
	3-27	60:30:10	**600**	300	100	**600**	1500	100
	3-28	60:30:10	**600**	300	100	**600**	180	70
	3-29	50:40:10	**500**	400	100	**500**	625	175
	3-30	50:40:10	**450**	360	90	**450**	475	125
	3-31	50:30:20	**400**	240	160	**400**	240	100
	3-32	50:30:20	**400**	240	160	**400**	300	200
	3-33	45:45:10	**450**	450	100	**450**	450	100
	3-34	45:45:10	**300**	300	65	**300**	300	75
	3-35	40:40:20	**350**	350	175	**350**	525	175
	3-36	40:40:20	**350**	350	175	**350**	300	250
	3-37	35:35:30	**350**	350	300	**350**	375	350
	3-38	35:35:30	**400**	400	350	**400**	375	300
	3-39	60:20:20	**425**	150	150	**425**	200	175
	3-40	60:20:20	**400**	130	130	**400**	145	145
	3-41	50:25:25	**375**	175	175	**375**	175	126
	3-42	50:25:25	**400**	200	200	**400**	200	280
	3-43	40:30:30	**350**	250	250	**350**	350	400
	3-44	40:30:30	**350**	250	250	**350**	275	250
Low (lowest samples 200-500 RFU)	3-45	33.3:33.3:33.3	50	50	**50**	40	50	**50**
	3-46	33.3:33.3:33.3	30	30	**30**	30	30	**30**
	3-47	70:20:10	280	80	**40**	350	90	**40**
	3-48	70:20:10	420	120	**60**	420	110	**60**
	3-49	60:30:10	490	210	**70**	490	1000	**70**
	3-50	60:30:10	240	120	**40**	360	120	**40**
	3-51	50:40:10	300	240	**60**	174	240	**60**
	3-52	50:40:10	200	160	**40**	150	160	**40**
	3-53	50:30:20	75	45	**30**	110	70	**30**
	3-54	50:30:20	125	75	**50**	105	75	**50**
	3-55	45:45:10	225	225	**50**	225	225	**50**
	3-56	45:45:10	135	135	**30**	125	125	**30**
	3-57	40:40:20	100	100	**50**	100	145	**50**
	3-58	40:40:20	100	100	**50**	75	68	**50**
	3-59	35:35:30	70	70	**60**	60	65	**60**
	3-60	35:35:30	50	50	**40**	60	60	**40**
	3-61	60:20:20	150	50	**50**	135	60	**50**
	3-62	60:20:20	90	30	**30**	80	30	**30**
	3-63	50:25:25	60	30	**30**	85	40	**30**
	3-64	50:25:25	140	70	**70**	100	55	**70**
	3-65	40:30:30	50	40	**40**	30	35	**40**
	3-66	40:30:30	40	30	**30**	40	32	**30**

	Sample ID	Ratios (in %)	Target pg	Target pg	Target pg	Target pg
High (highest at 3-5K RFU)	4-1	25:25:25:25	350	350	350	350
	4-2	25:25:25:25	600	600	600	600
	4-3	60:20:10:10	900	300	150	150
	4-4	60:20:10:10	900	300	150	150
	4-5	50:20:20:10	1000	400	400	100
	4-6	50:20:20:10	1200	480	480	240
	4-7	70:10:10:10	1050	150	150	150
	4-8	70:10:10:10	1400	200	200	200
	4-9	40:20:20:20	1000	500	500	500
	4-10	40:20:20:20	1400	700	700	700
	4-11	40:40:15:5	800	800	300	100
	4-12	40:40:15:5	1000	1000	375	125
	4-13	35:35:20:10	650	650	375	185
	4-14	35:35:20:10	800	800	450	225
	4-15	40:40:10:10	800	800	200	200
	4-16	40:40:10:10	1000	1000	250	250
	4-17	35:35:25:5	1000	1000	700	150
	4-18	35:35:25:5	500	500	350	75
	4-19	30:30:20:20	525	525	350	350
	4-20	30:30:20:20	***2400***	1200	800	800
	4-21	30:30:30:10	600	600	600	200
	4-22	30:30:30:10	1000	1000	1000	325
Mid (highest at 1-1.5K RFU)	4-23	25:25:25:25	175	175	175	175
	4-24	25:25:25:25	250	250	250	250
	4-25	60:20:10:10	600	200	100	100
	4-26	60:20:10:10	600	200	100	100
	4-27	50:20:20:10	600	200	200	100
	4-28	50:20:20:10	800	320	320	160
	4-29	70:10:10:10	700	100	100	100
	4-30	70:10:10:10	1050	150	150	150
	4-31	40:20:20:20	550	275	275	275
	4-32	40:20:20:20	800	400	400	400
	4-33	40:40:15:5	600	600	225	75
	4-34	40:40:15:5	700	700	250	85
	4-35	35:35:20:10	350	350	200	100
	4-36	35:35:20:10	525	525	300	150
	4-37	40:40:10:10	500	500	125	125
	4-38	40:40:10:10	600	600	150	150
	4-39	35:35:25:5	700	700	500	100
	4-40	35:35:25:5	420	420	300	60
	4-41	30:30:20:20	300	300	200	200
	4-42	30:30:20:20	675	675	450	450
	4-43	30:30:30:10	400	400	400	130
	4-44	30:30:30:10	600	600	600	200

Low (lowest samples 200-500 RFU)	4-45	25:25:25:25	30	30	30	30
	4-46	25:25:25:25	75	75	75	75
	4-47	60:20:10:10	250	100	50	50
	4-48	60:20:10:10	450	150	75	75
	4-49	50:20:20:10	250	100	100	50
	4-50	50:20:20:10	375	150	150	75
	4-51	70:10:10:10	350	50	50	50
	4-52	70:10:10:10	630	90	90	90
	4-53	40:20:20:20	100	50	50	50
	4-54	40:20:20:20	200	100	100	100
	4-55	40:40:15:5	400	400	150	50
	4-56	40:40:15:5	400	400	150	50
	4-57	35:35:20:10	175	175	100	50
	4-58	35:35:20:10	250	250	150	75
	4-59	40:40:10:10	200	200	50	50
	4-60	40:40:10:10	300	300	75	75
	4-61	35:35:25:5	350	350	250	50
	4-62	35:35:25:5	300	300	225	45
	4-63	30:30:20:20	120	120	80	80
	4-64	30:30:20:20	***300***	150	100	100
	4-65	30:30:30:10	180	180	180	60
	4-66	30:30:30:10	225	225	225	75

Five person mixtures continued below….

	Sample ID	Ratios (in %)	Target pg	Target pg	Target pg	Target pg	Target pg
High (highest at 3-5K RFU)	5-1	20:20:20:20:20	350	350	350	350	350
	5-2	20:20:20:20:20	600	600	600	600	600
	5-3	60:10:10:10:10	900	150	150	150	150
	5-4	60:10:10:10:10	900	150	150	150	150
Mid (highest at 1-1.5K RFU)	5-5	20:20:20:20:20	175	175	175	175	175
	5-6	20:20:20:20:20	250	250	250	250	250
	5-7	60:10:10:10:10	600	100	100	100	100
	5-8	60:10:10:10:10	600	100	100	100	100
Low (lowest at 200-500 RFU)	5-9	20:20:20:20:20	30	30	30	30	30
	5-10	20:20:20:20:20	75	75	75	75	75
	5-11	60:10:10:10:10	250	50	50	50	50
	5-12	60:10:10:10:10	450	75	75	75	75

The mixtures were amplified for 29 cycles using GlobalFiler followed by capillary electrophoresis on the 3500. The resultant data was analyzed in GMID-X with an analytical threshold of 100 RFU and stutter filtered.

Genotypes for each electropherogram were accounted for in order to identify any extraneous alleles. At loci where the contributors did not overlap genotypes or at least one unique allele from each person was represented, contributor ratio was manually calculated. The results across loci were averaged for comparison to the target ratio, and how much that varied from locus to locus was also recorded. Unlike STRmix assessment of contributor ratio, stutter was not taken into account for these calculations.

Because all 16 previously characterized Identifiler Plus loci are included in the GlobalFiler kit, peak height balance and mixture ratio balance between the contributors for those loci could be generally assessed. Any extreme peak height balance or contributor ratio imbalance from locus to locus was noted for each of the mixtures.

Results and Discussion

The mixture ratio results and peak heights reflected the target to a large degree. Stutter was not taken into account when determining these ratios manually, so there is some variation from locus to locus. The mixtures range from being very robust (peak heights above the target range creating spectral artifacts) to very low (considerable dropout occuring in one or more of the contributors). Sixty-eight of the mixtures had at least one instance of dropout (when stutter was filtered).The correlation between the target % contribution and the manual calculation of contributor percentage is shown in Figure 1. Contributors from 6 of the mixtures are not shown, because there was not enough reliable data to get an accurate contributor breakdown. See the STRmix MCMC study for more details on how the STRmix estimation of contributor ratio

compares to the target ratio. A representative set graphs demonstrating contributor ratio by locus is included at the end of this study.

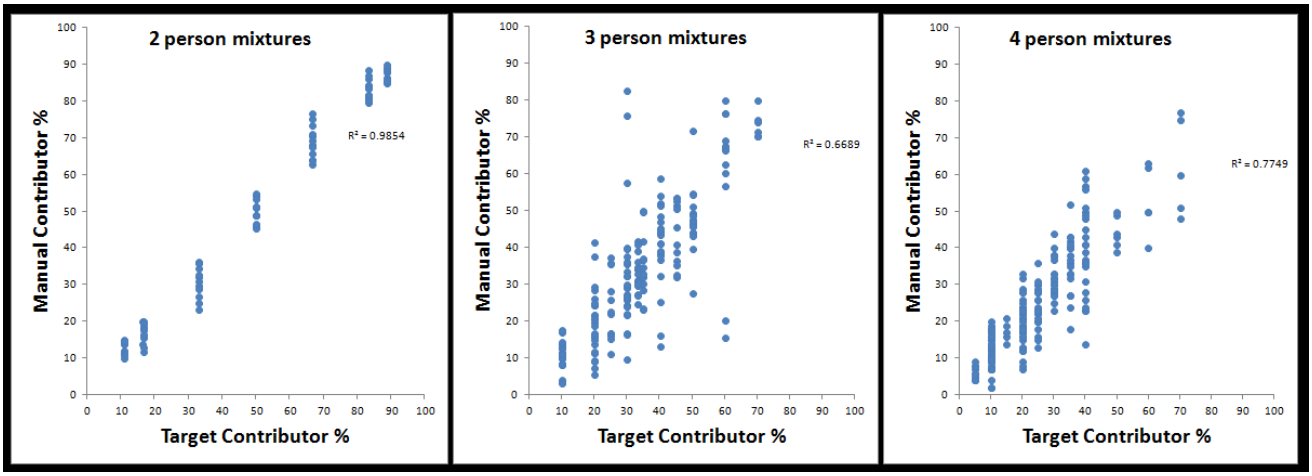
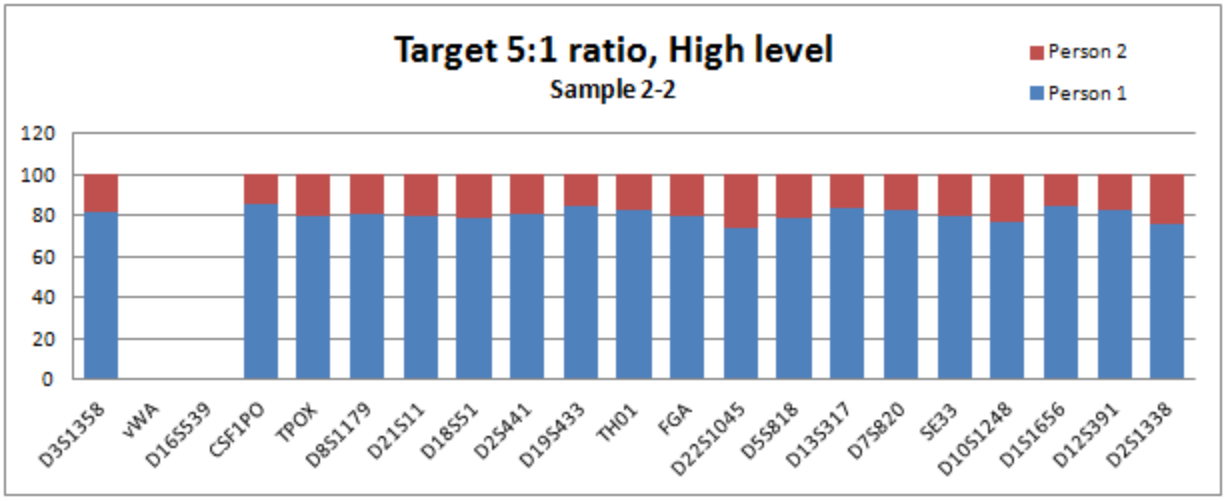
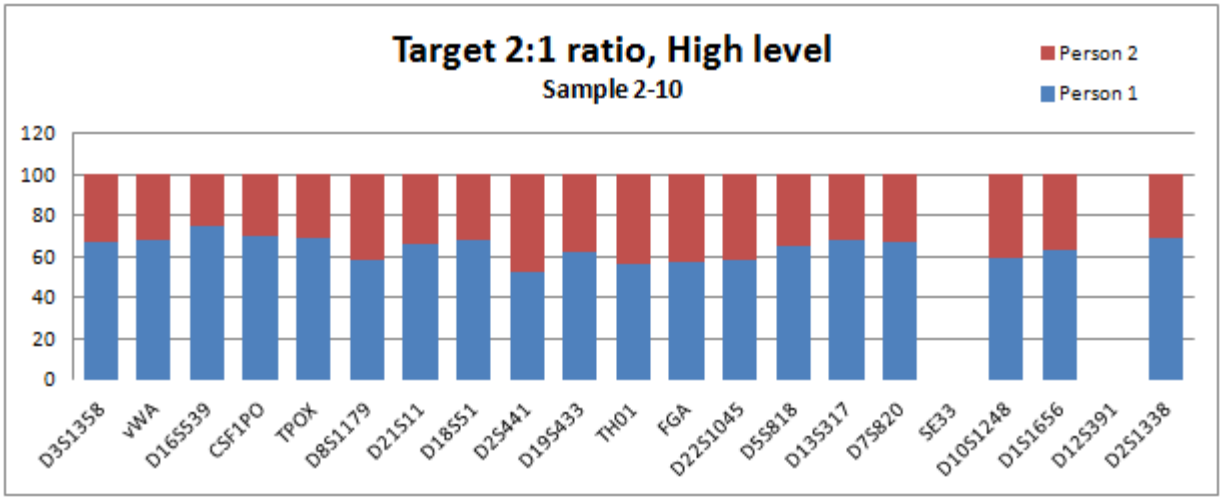


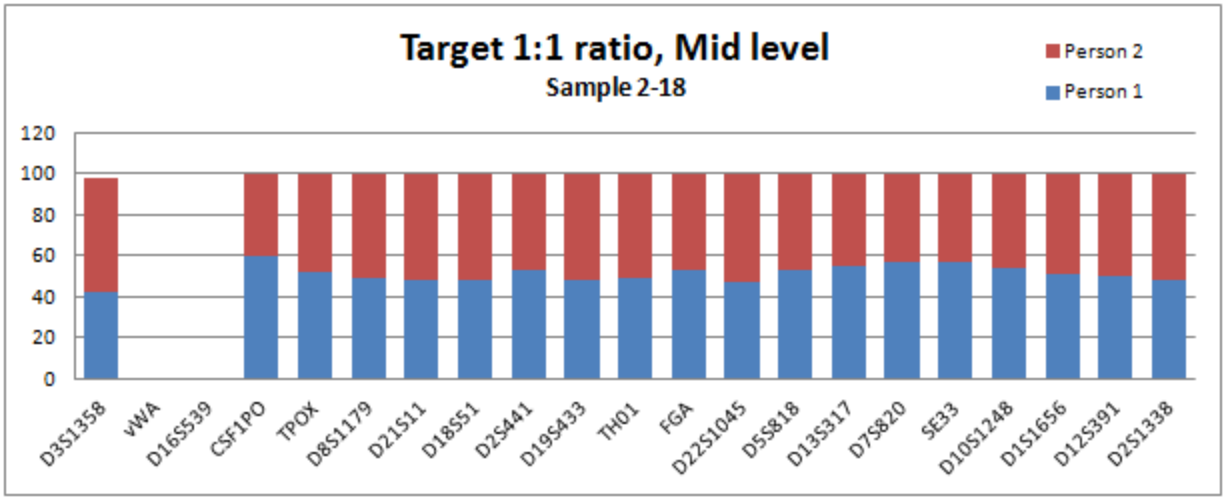
Figure 1 – Correlation between target mixture ratio design and a manual estimation of mixture results by contributor.

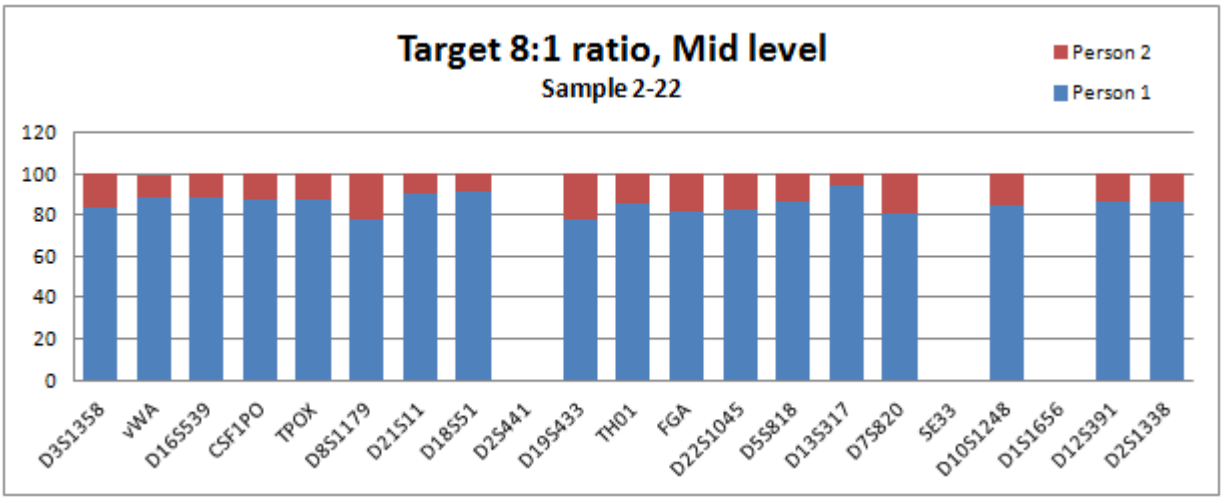
Conclusions

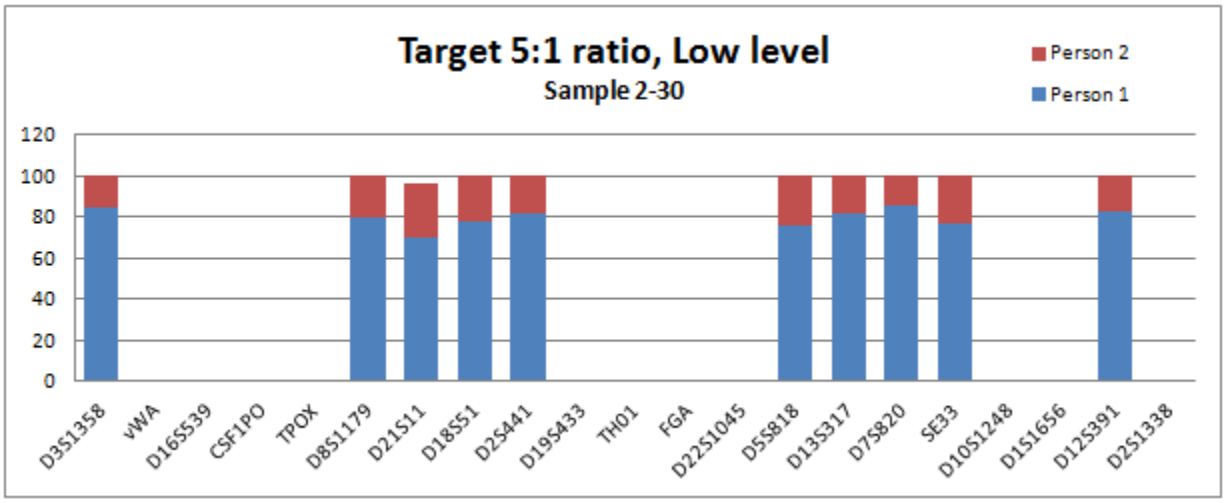
The results of this mixture set provide a sufficiently wide range of contributor ratio and input amount to extensively test the deconvolution ability, accuracy and limits of STRmix.

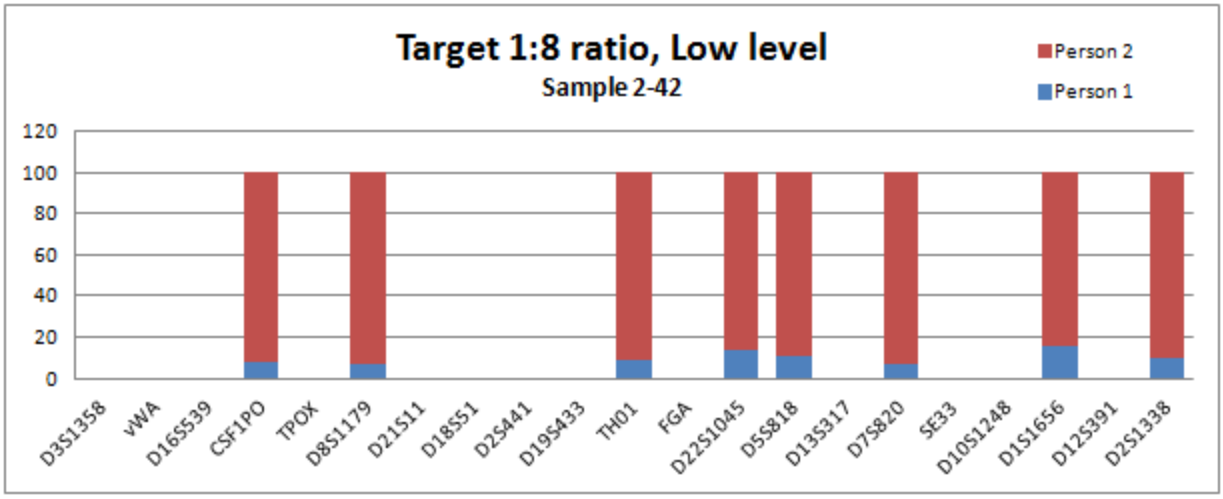


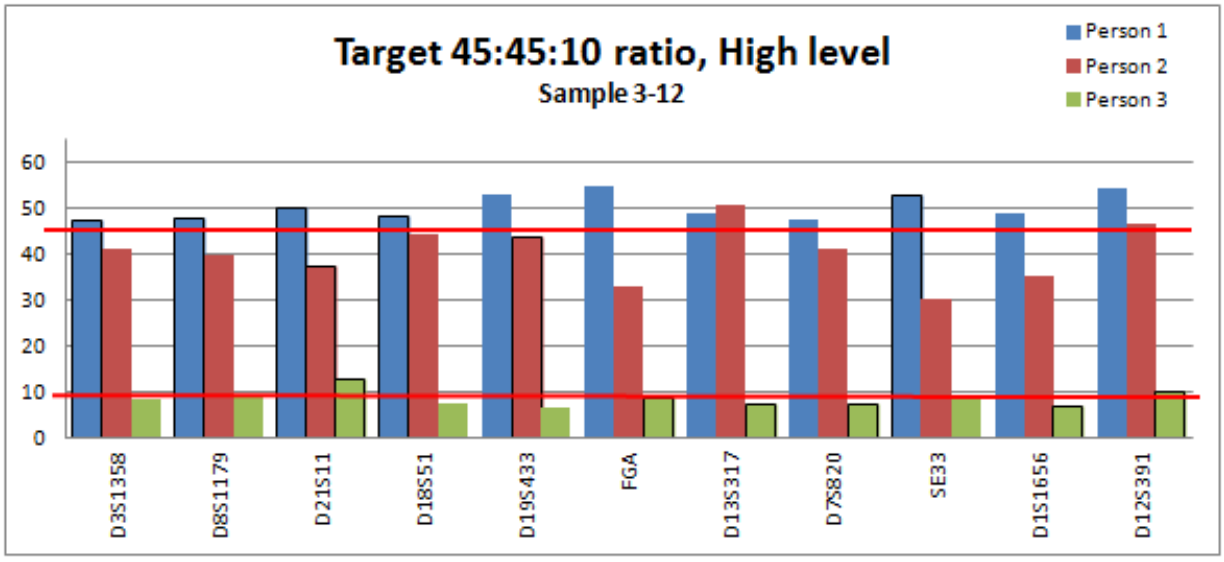

Target 2:1 ratio, High level
Sample 2-10
Person 2
Person 1
120
100
80
60
40
20
0
D3S1358
vWA
D16S539
CSF1PO
TPOX
D8S1179
D21S11
D18S51
D2S441
D19S433
TH01
FGA
D22S1045
D5S818
D13S317
D7S820
SE33
D10S1248
D1S1656
D12S391
D2S1338

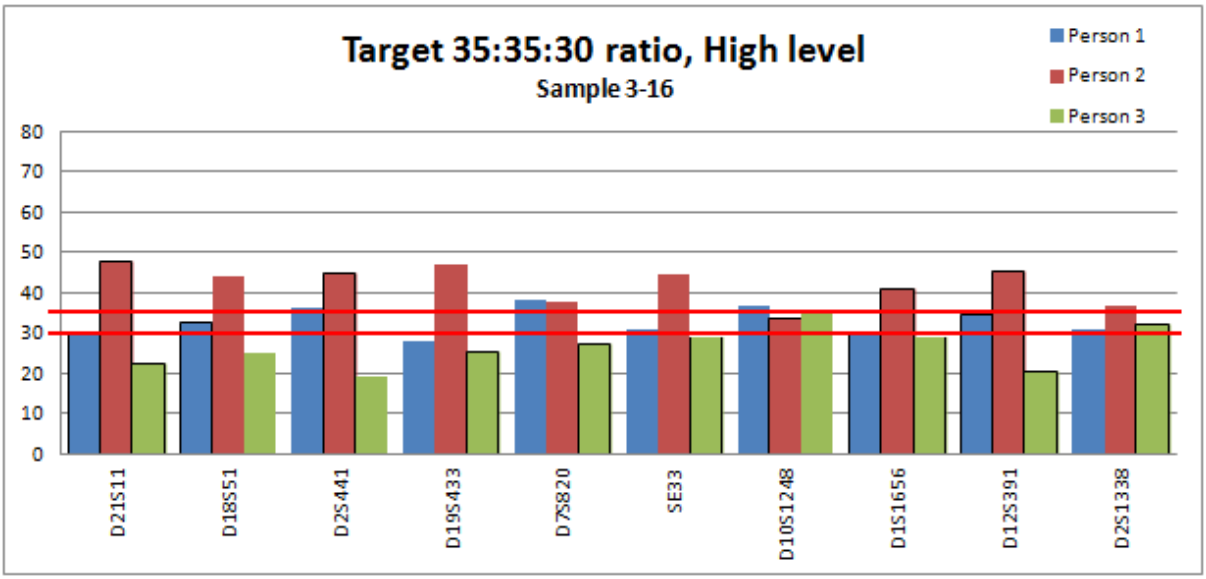

Target 1:1 ratio, Mid level
Sample 2-18
Person 2
Person 1
120
100
80
60
40
20
0
D3S1358
vWA
D16S539
CSF1PO
TPOX
D8S1179
D21S11
D18S51
D2S441
D19S433
TH01
FGA
D22S1045
D5S818
D13S317
D7S820
SE33
D10S1248
D1S1656
D12S391
D2S1338

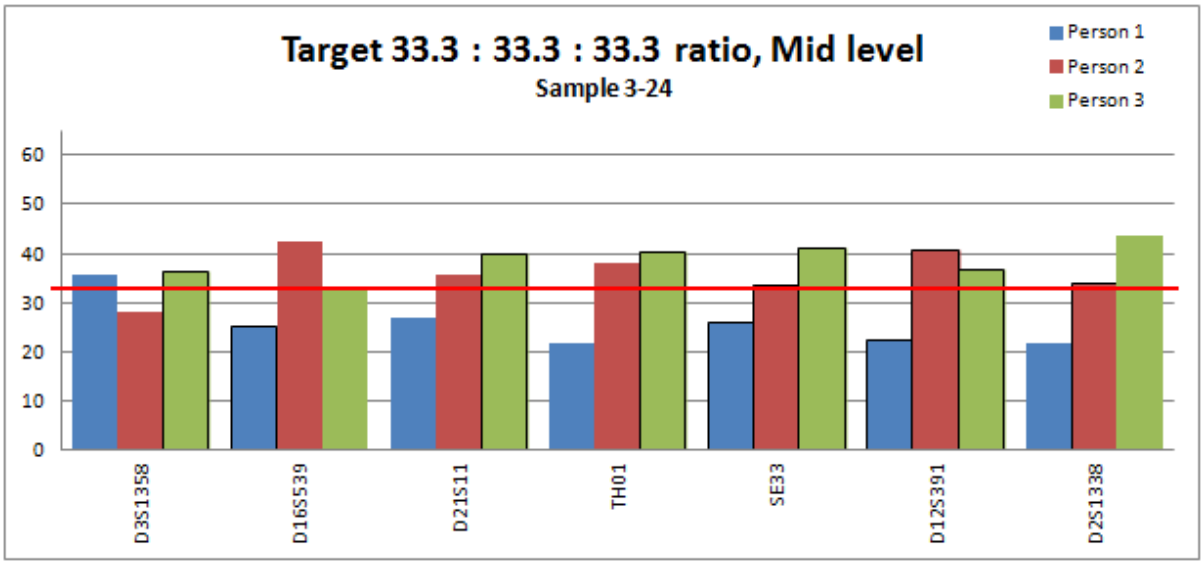

Target 8:1 ratio, Mid level
Sample 2-22
Person 2
Person 1
120
100
80
60
40
20
0
D3S1358
vWA
D16S539
CSF1PO
TPOX
D8S1179
D21S11
D18S51
D2S441
D19S433
TH01
FGA
D22S1045
D5S818
D13S317
D7S820
SE33
D10S1248
D1S1656
D12S391
D2S1338

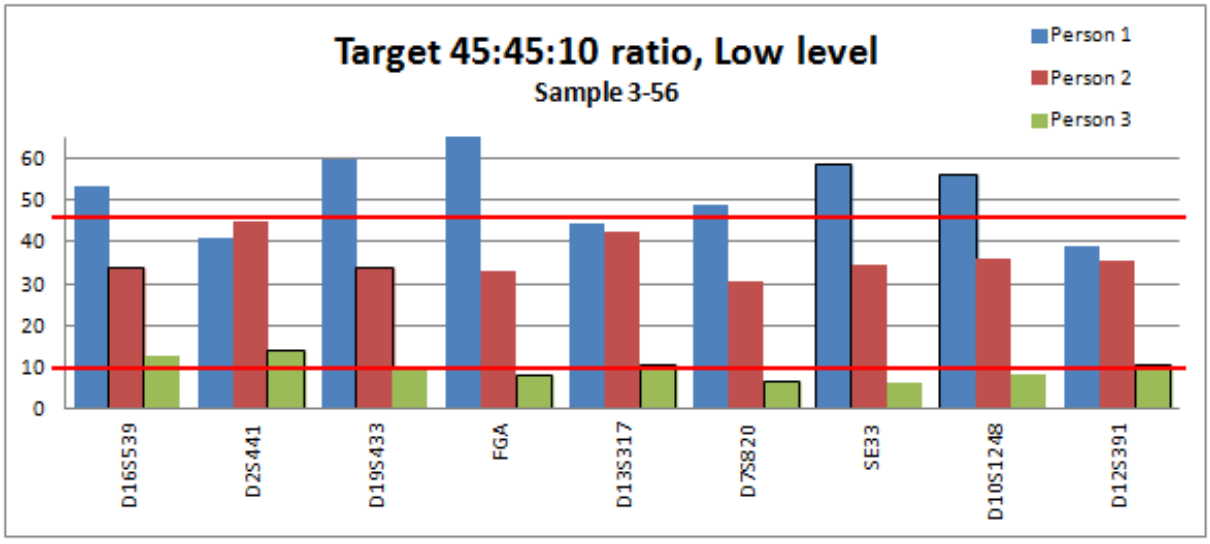

Target 5:1 ratio, Low level
Sample 2-30
Person 2
Person 1
120
100
80
60
40
20
0
D3S1358
vWA
D16S539
CSF1PO
TPOX
D8S1179
D21S11
D18S51
D2S441
D19S433
TH01
FGA
D22S1045
D5S818
D13S317
D7S820
SE33
D10S1248
D1S1656
D12S391
D2S1338

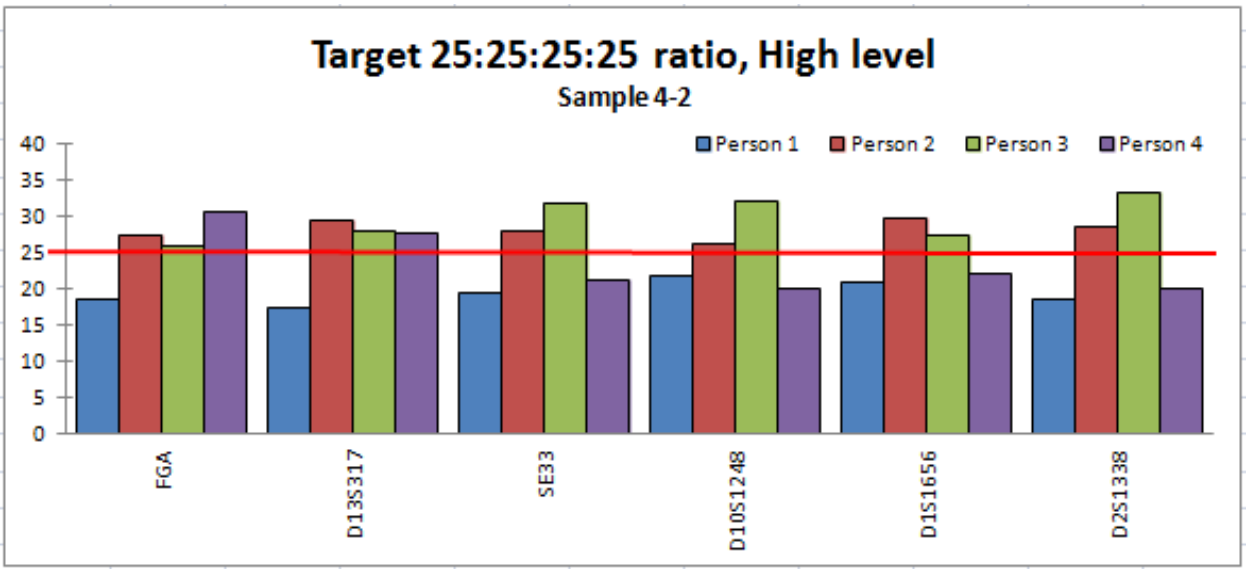

Target 1:8 ratio, Low level
Sample 2-42
Person 2
Person 1
120
100
80
60
40
20
0
D3S1358
vWA
D16S539
CSF1PO
TPOX
D8S1179
D21S11
D18S51
D2S441
D19S433
TH01
FGA
D22S1045
D5S818
D13S317
D7S820
SE33
D10S1248
D1S1656
D12S391
D2S1338


Target 45:45:10 ratio, High level
Sample 3-12
Person 1
Person 2
Person 3
60
50
40
30
20
10
0
D3S1358
D8S1179
D21S11
D18S51
D19S433
FGA
D13S317
D7S820
SE33
D1S1656
D12S391


Target 35:35:30 ratio, High level
Sample 3-16
Person 1
Person 2
Person 3
80
70
60
50
40
30
20
10
0
D21S11
D18S51
D2S441
D19S433
D7S820
SE33
D10S1248
D1S1656
D12S391
D2S1338


Target 33.3 : 33.3 : 33.3 ratio, Mid level
Sample 3-24
Person 1
Person 2
Person 3
60
50
40
30
20
10
0
D3S1358
D16S539
D21S11
TH01
SE33
D12S391
D2S1338


Target 45:45:10 ratio, Low level
Sample 3-56
Person 1
Person 2
Person 3
60
50
40
30
20
10
0
D16S539
D2S441
D19S433
FGA
D13S317
D7S820
SE33
D10S1248
D12S391


Target 25:25:25:25 ratio, High level
Sample 4-2
Person 1
Person 2
Person 3
Person 4
40
35
30
25
20
15
10
5
0
FGA
D13S317
SE33
D10S1248
D1S1656
D2S1338

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Life Technology

Injection Time Study

Purpose

The goal of this study is to evaluate the use of increased and decreased injection times during capillary electrophoresis (on the 3500 Genetic Analyzer) of samples amplified with the GlobalFiler Amplification Kit and to determine the appropriate analysis parameters to use for these samples. The default injection time on a 3500 is 15 seconds. This study evaluates the decreased injection times of 5 and 10 seconds and an increased injection time of 24 seconds (the manufacturer recommends not exceeding 24 seconds). A secondary purpose of this study is to generate a sufficient amount of samples injected for 24 seconds to use in STRmix ModelMaker. Increasing the injection time has the potential to increase the peak height range in which pronounced stochastic events occur (like dropout and extreme peak height imbalance). This will allow comparison between the 15 and 24 second allele and stutter variance parameters. Having these variance parameters will allow interpretation of mixtures injected for 24 seconds. See STRmix ModelMaker study for description of 24 second injection ModelMaker results.

Materials and Methods

Decreased injection time

Seven of the 10 reference samples previously amplified for the sensitivity study were used for this study. This was done with the original GlobalFiler formulation first. When the new product with the reformulated master mix was released, it was done again. The second time, seven different single source reference samples (previously amplified with the reformulated GlobalFiler Master Mix) were used. Three different input amounts for each sample (200, 400, and 800pg) were injected for 15 seconds, 10 seconds, and 5 seconds from the same CE plate, within the same day. The negative TE amplification control was also examined for artifacts at each of these injection times. Results presented in this study will focus on data collected using the reformulated GlobalFiler amplification kit, because that is the only product commercially available now, however all data is accessible in the lab validation files on the FB network.

The amplification product from these samples was injected for capillary electrophoresis on both 3500 instruments according to the manufacturer's recommended parameters. A detection threshold of 100 RFU was applied during GeneMapper ID-X (v1.4) analysis. Peak heights were observed across the dilution series to determine the DNA target amounts that resulted in peak heights between 3,000 and 10,000 RFU (the ideal peak height range for data processed with the 3500 instrument).

Fold changes for all peak heights detected in these samples were calculated between the 15 second injection and the 10 second injection, as well as between the 10 second injection and the 5 second injection. Fold change was averaged across all peaks in all samples.

Increased injection time

Seven of the 10 reference samples previously amplified for the sensitivity study (lowest input amounts for each sample) were used for this study. This was done with the original GlobalFiler formulation first. When the new product with the reformulated master mix was released, it was done again. The second time (with the reformulated master mix), many more samples were used in order to collect enough data for STRmix ModelMaker (24 second kit). The data presented here was compiled from 47 of the samples injected at 24 seconds. These are made up of 7 different dilutions (target input ~12pg-200pg) of 7 different samples (previously amplified for the sensitivity study. The 15 second and 24 second injections were done from the same CE plate, within the same day. The negative TE amplification control was also examined for artifacts at each of these injection times for artifacts and potential baseline elevation. Results presented in this study will focus on data collected using the reformulated GlobalFiler amplification kit, because that is the only product commercially available now, however all data is accessible in the lab validation files on the FB network.

The amplification product from these samples was injected for capillary electrophoresis on both 3500 instruments. A detection threshold of 100 RFU was applied during GeneMapper ID-X (v1.4) analysis. Peak heights were observed across the dilution series to determine the DNA target amounts that resulted in peak heights between 3,000 and 10,000 RFU (the ideal peak height range for data processed with the 3500 instrument).

Fold changes for all peak heights detected in these samples were calculated between the 24 second injection and the 15 second injection. Fold change was averaged across all peaks in all samples.

Results and Discussion

Decreased injection time

Peak height decreases linearly when injection time decreases. Figure 1 shows an example of one locus in one sample injected for three different durations. Decreasing injection time is useful for samples with peak heights about ~10,000 RFU. Peaks around and above 10,000 RFU almost always have spectral artifacts associated with them because of the large dynamic range of the 3500 instrument.

Peak heights from 15 second injections are, on average, 1.59 times higher than peak heights from 10 second injections. Peak heights from 10 second injections are, on average, 2.06 times higher than peak heights from 5 second injections.

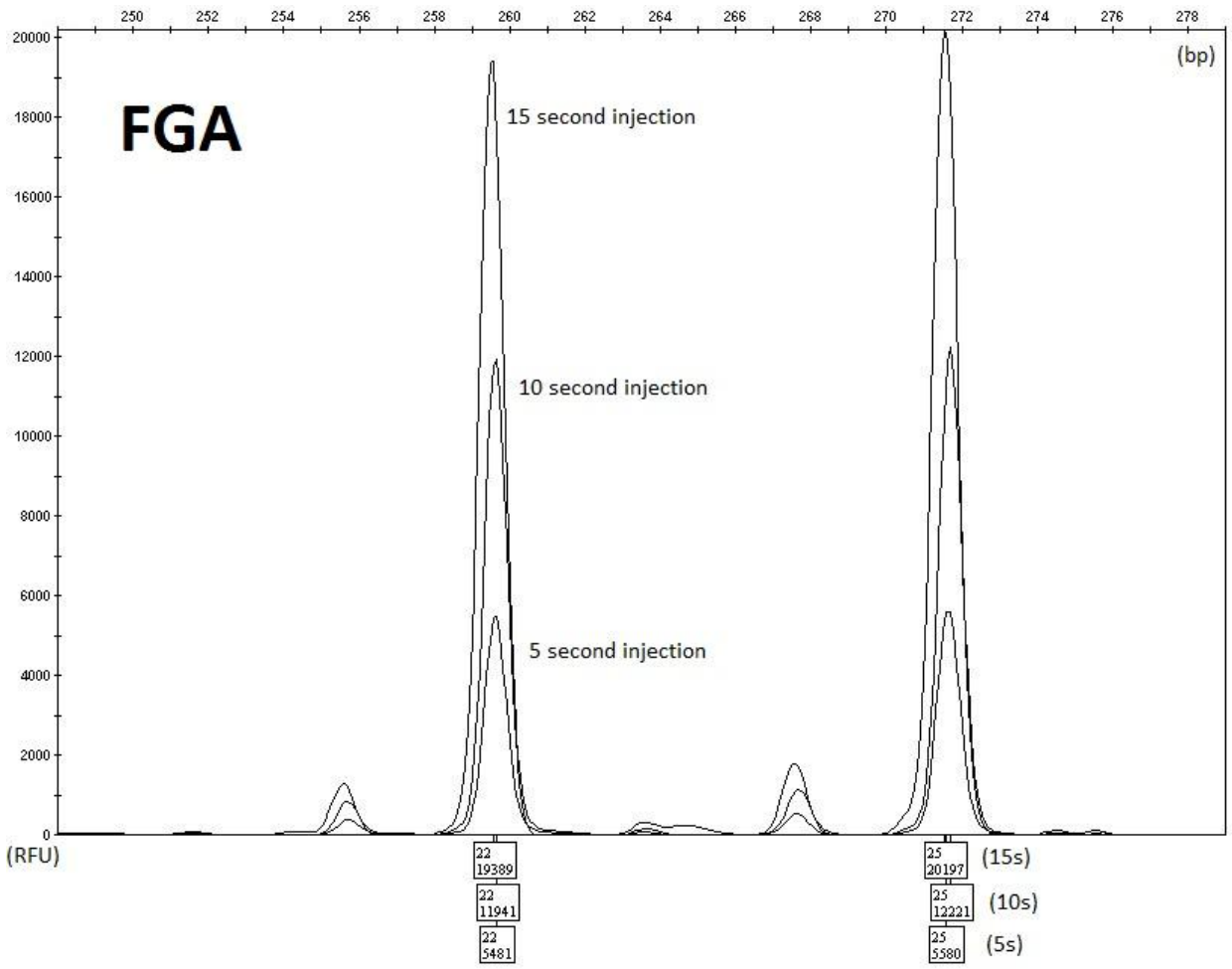


Figure 1 – An example of peak heights when the same sample is injected for 15, 10 and 5 seconds.

Table 1 – Peak height changes associated with decreased injection time.

Fold Change →	Average 15 to 10	Average 10 to 5	Max 15 to 10	Min 15 to	Max 10 to 5	Min 10 to 5
3500A	1.49	2.13	1.84	1.24	2.77	1.71
3500B	1.68	2.00	3.36	1.40	2.47	1.36
Both Instruments	1.59	2.07				

Increased injection time

The baseline noise for the amplification blank was evaluated for the 24 second injection (Figure 2). On both instruments, the average plus 10 standard deviations was less than 100 RFU (Table 2), and the max peak also didn't indicate any need to increase the analytical threshold for samples injected for 24 seconds.

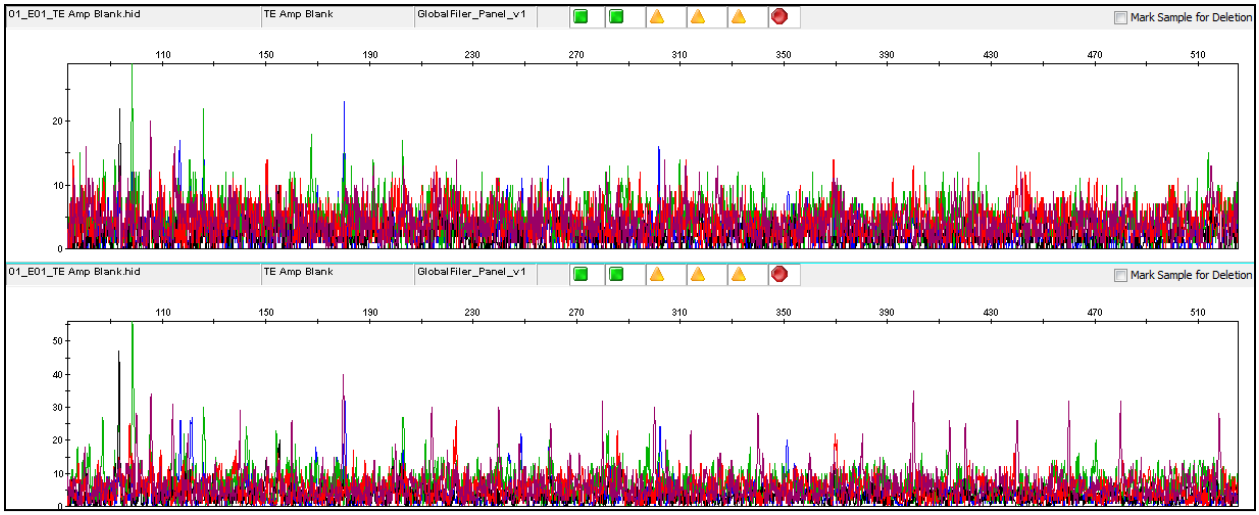


Figure 2 - Electropherograms for amplification blank injected for 24 sec. on 3500A and 3500B, respectively.

Table 2: Baseline levels for amplification blank.

Average (RFU)	Standard Deviation	Avg +3 SD	Avg + 10SD	MAX	MIN
7.58	4.31	20.50	50.64	56	1

Injecting samples for 24 seconds increased peak heights. This is especially useful for samples with limited input amounts or profiles with maximum peak heights of less than ~1000 RFU. Peak heights from 24 second injections are, on average, 1.7 times higher than peak heights from 15 second injections.

Table 3 – Peak height changes associated with increased injection time.

Fold Change →	Average 24 to 15	Max 24 to 15	Min 24 to 15
3500A	1.79	2.34	1.02
3500B	1.62	2.08	0.94
Both Instruments	1.70		

The lower the peak height, the more likely dropout is to be associated with that peak height. In this study, only single source samples with known genotypes were used, so occurrences of dropout were recorded. Sometimes, partial dropout is rescued by an increased injection time. This would be the case if there was a visible allele just below the analytical threshold. Often, increasing the injection time increases the number of alleles detected for a sample. It is also possible to have partial dropout at 15 seconds, increase the injection time to 24 second, and still have dropout (Figure 3). The effect this would have would be to increase the stochastic range. As a result, different interpretation guidelines should be imposed on evidence samples that have been injected for 24 seconds as opposed to 15 seconds.

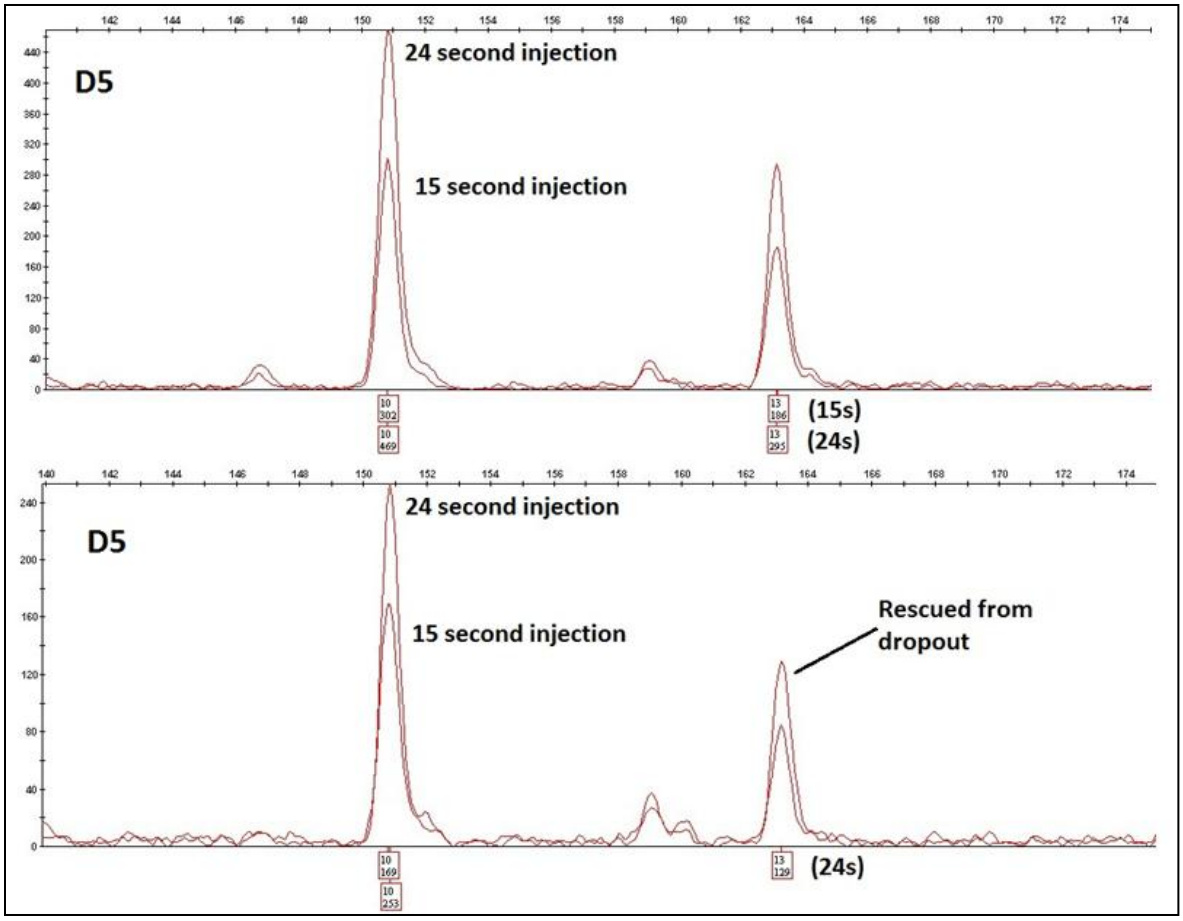


Figure 3 – The effect of a 24 second injection on dropout in two different samples.

Table 4 - Top 10 highest allele peaks associated with dropout (24 sec)

844
742
656
621
601
546
532
514
502
482

Table 4 contains peak heights of alleles whose known sister allele has dropped out. This list is higher than the peak heights for samples injected at 15 seconds. For this particular set of samples, the highest peak associated with dropout is 495 RFU when injected for 15 seconds. A more in-depth analysis about the allelic variance is described in the STRmix ModelMaker study.

Two samples from the 24 second injection study had two extra peaks detected. The peaks were not originally detected when the samples were injected for 15 seconds (see Contamination Assessment Study, Figure 6). Additional samples were amplified with a GlobalFiler Primer Mix that was purified from residual Identifiler Plus Primers, and this problem was not detected when the samples were injected for 24 seconds.

Conclusions

Decreasing injection time decreases peak heights in a linear way. Peak heights from samples injected for 10 seconds are only about 2/3 as high as corresponding peak heights from samples

injected for 15 seconds. Peak heights from samples injected for 5 seconds are half as high as corresponding peak heights from the same samples injected for 10 seconds and 1/3 as high as corresponding peak heights from the same samples injected for 15 seconds. This should be used as a guide when a CE injection results in peak heights greater than ~10,000 RFU.

Evaluation of the baseline noise levels from the amplification blank verifies that the 100 RFU detection threshold is suitable for use with 24 second injection times. Increasing injection time to 24 seconds results in peak heights that are ~1.7 times higher than those from the same sample injected for 15 seconds. These increases did not appear to be locus dependent. The tallest peaks detected when there was known to be allelic dropout of the sister peak were higher for 24 second injection times than they were for 15 second injections. For increased injection times it is necessary to elevate the stochastic threshold in order to account for this increase in peak heights. This will be done using STRmix ModelMaker, and a new STRmix 24 second kit will be used to analyze samples that have been injected for 24 seconds.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Life Technology

Contamination Assessment Study

Purpose

The goal of this study is to verify that contamination is not a concern for samples amplified with the GlobalFiler PCR Amplification Kit.

Materials and Methods

Over the course of the validation a total of 555 samples have been amplified with 29 cycles using GlobalFiler (both the original and reformulated master mix). Fourteen of these samples were amplification blanks containing only TE buffer. The remaining samples were from known sources (007 positive control and reference samples). Because there are new loci in the GlobalFiler kit, only the ones that overlapped with Identifiler Plus could be used for genotype references. The additional loci were scrutinized for possible contamination or artifacts. In addition to multiple amplifications, the amplified samples were injected multiple times, either to compare 3500A and 3500B, or to examine the effect of different injection times on the same samples. Each of these injections was examined for the known genotype

Results and Discussion

No DNA types were detected in any of the 14 TE amplification blanks. No contamination was detected in any of the samples amplified with GlobalFiler and analyzed on the 3500 instruments. No carryover from the CE injection process was observed.

There were, however, multiple extraneous alleles and one out of marker range (OMR) peak that was called in several instances. All of these can be attributed to artifacts. These are listed below.

The first set of peaks are artifacts identified by the GlobalFiler manufacturer to be caused by the source DNA in the sample. All lots of GlobalFiler primer manufactured in 2014 and the first half of 2015 were known to be contaminated with residual Identifiler Plus primers (Figure 1). See the communication from ThermoFisher (7/13/15, GlobalFiler D18 artifacts) for more details on this artifact.

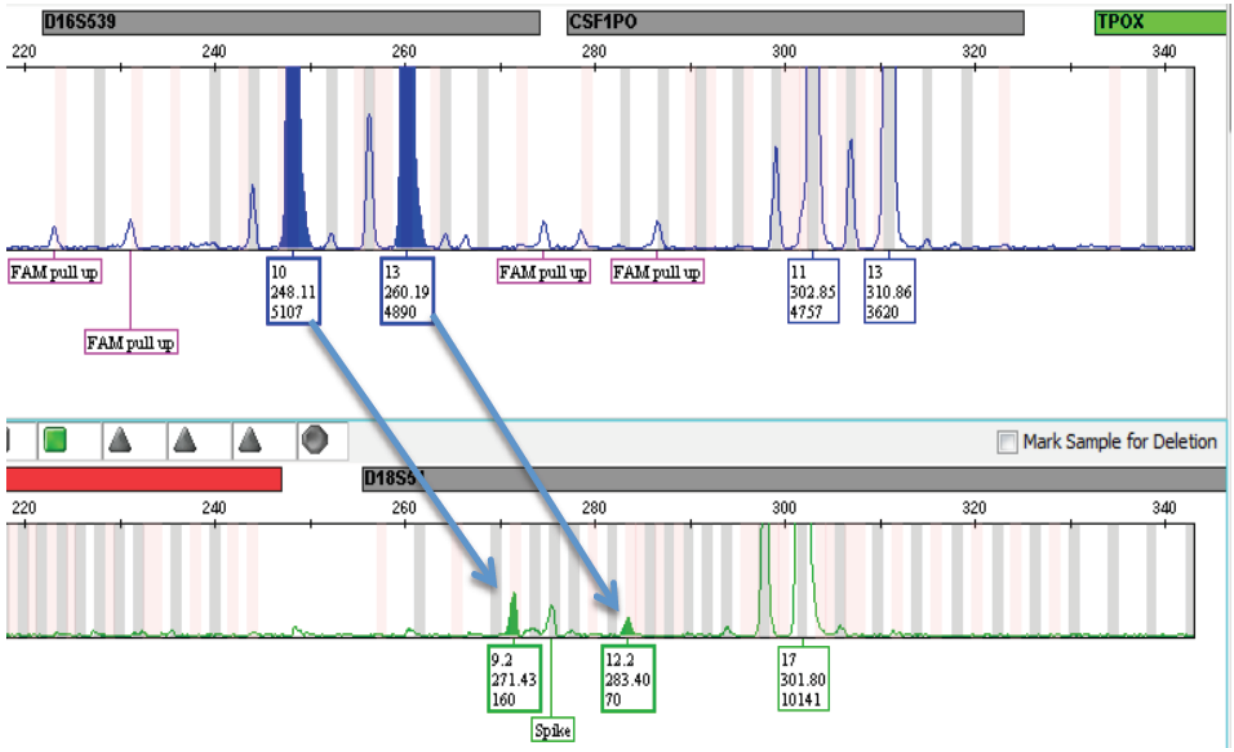


Figure 1 – Figure from the ThermoFisher communication showing VIC-D16S539 artifacts shifted +23 base pairs from FAM-labeled D16S539 peaks in the GlobalFiler kit.

These artifacts were also seen in the GlobalFiler validation study at SDPD. These are shown in Figure 2 below. Only the artifacts in D18S51 are shown in Figure 2. Each of these peaks were shifted 23 base pairs from a corresponding peak in D16S539. These are only 5 representative samples where this artifact was seen. More were detected over the course of the validation, but all were confirmed to be artifacts from D16S539 primers based on the base pair shift. Therefore, this is not contamination of genomic DNA, as the residual VIC-labeled primers are amplifying DNA from the reference sample. This primer contamination has been reported by the manufacturer, and will be corrected in future lots of GlobalFiler primer mix.

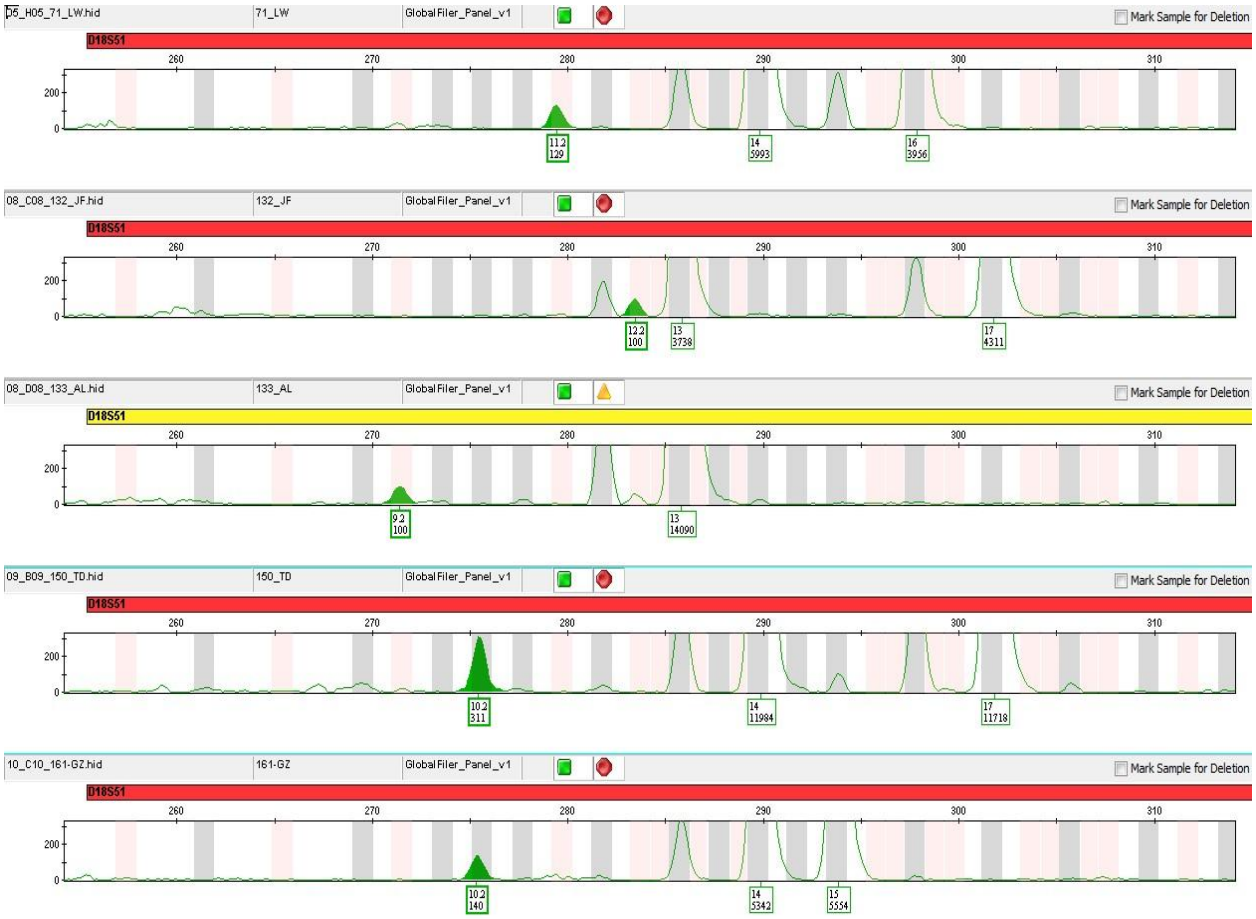


Figure 2 – Samples from the SDPD GlobalFiler validation showing VIC-D16S539 artifacts shifted +23 base pairs from FAM-labeled D16S539 peaks in the GlobalFiler kit.

Another type of artifact that was detected was consistently ~5.3 base pairs smaller than the X allele peak at Amelogenin (Figure 3). This artifact was observed 10 times in the 555 samples amplified with GlobalFiler, and it was only seen when very high template amounts were used. In Figure 3, the height of the X allele at Amelogenin is over 20,000 RFU, which is much higher than the ideal peak height range.

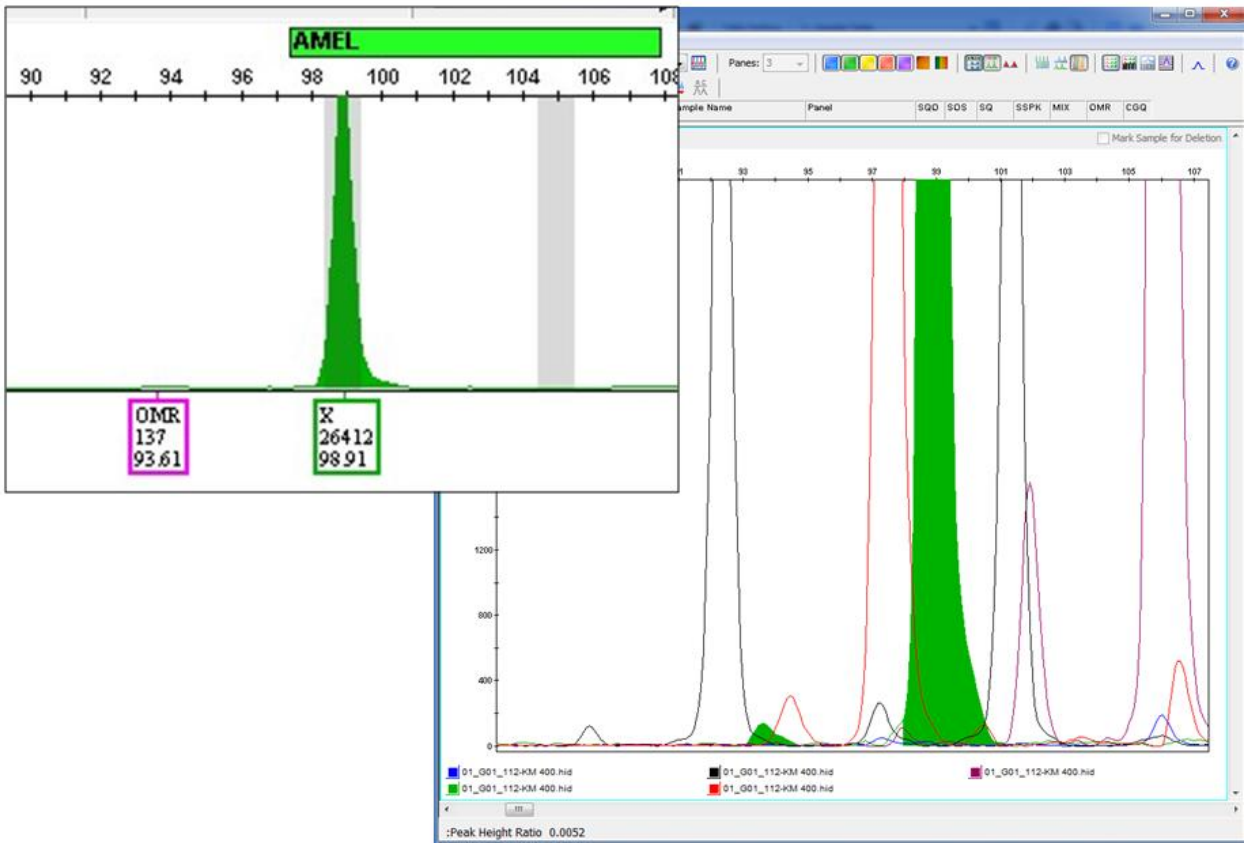


Figure 3 – Artifact in the green channel at ~93.5 base pairs.

Another artifact was observed in three samples using the reformulated GlobalFiler master mix (Figure 4). One sample was the 007 positive control, and the other two were in a reference sample from the sensitivity study that had been previously amplified with no artifact, and also was not present in 7 other dilutions of the same sample prepared at the same time. These three samples were in the same plate for amplification preparation, and in the same plate for capillary electrophoresis. It was injected on both 3500 instruments several times, and was present in every CE run of the sample. It primarily affected the yellow channel from 75 to 110 basepairs. Results of analyzed data are shown in Figure 4, and raw data traces for these samples are shown in Figure 5.

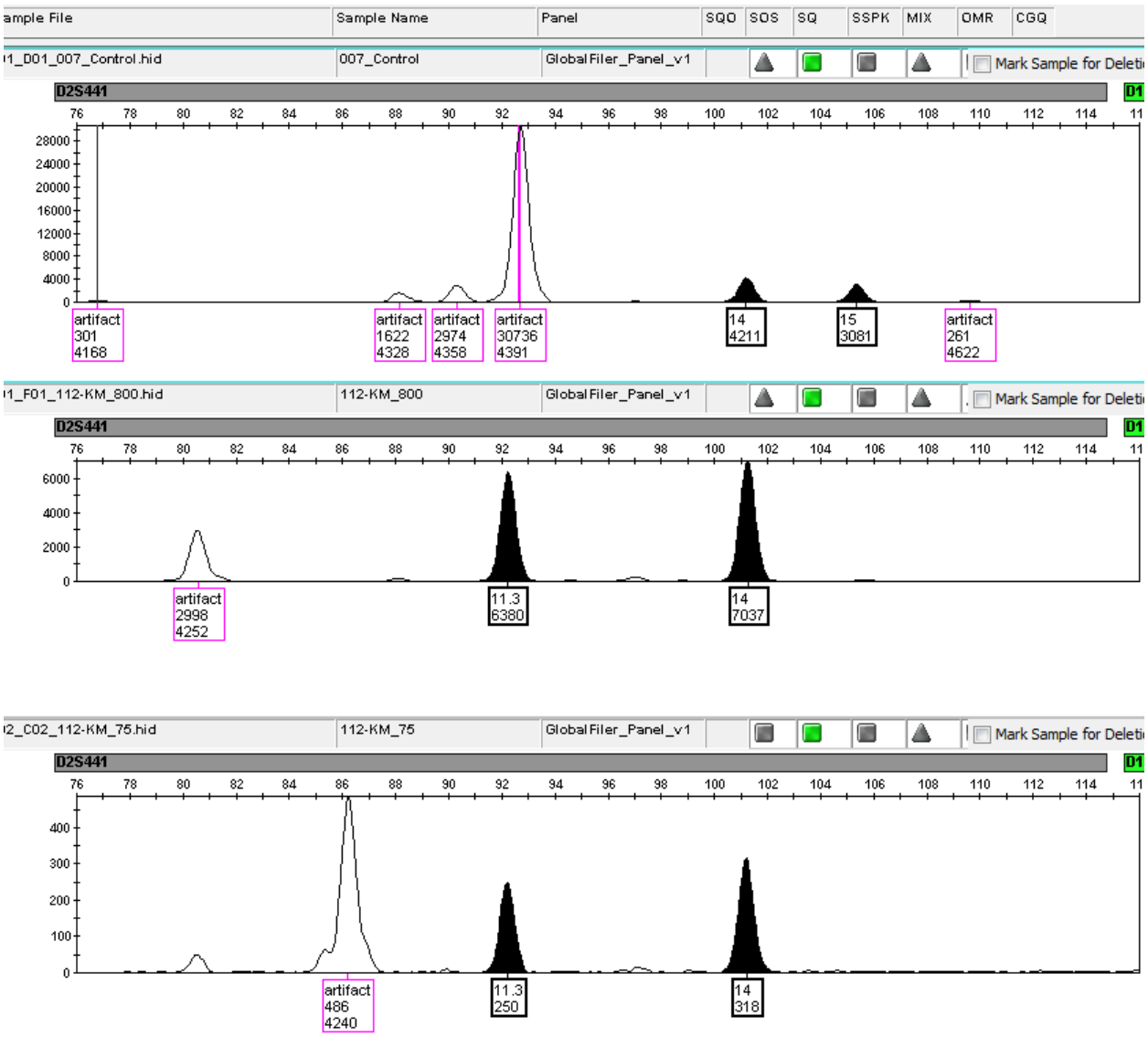


Figure 4 – Artifact observed in 3 samples. Known genotypes of the sample are highlighted in black.

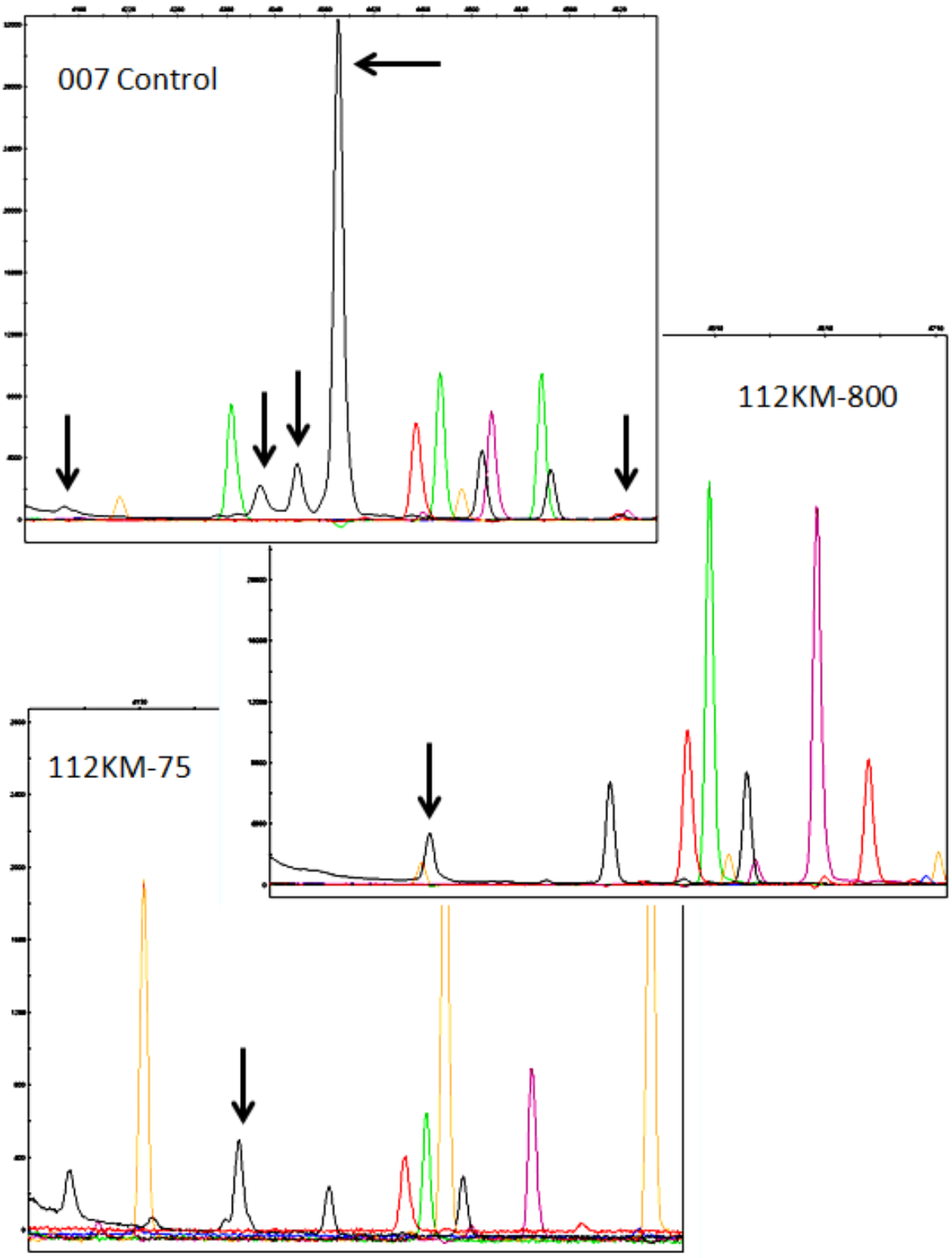


Figure 5 – Raw data of samples shown in Figure 4.

Two samples from the 24 second injection study had two extra peaks detected. The peaks were not originally detected when the samples were originally injected for 15 seconds (Figure 6).

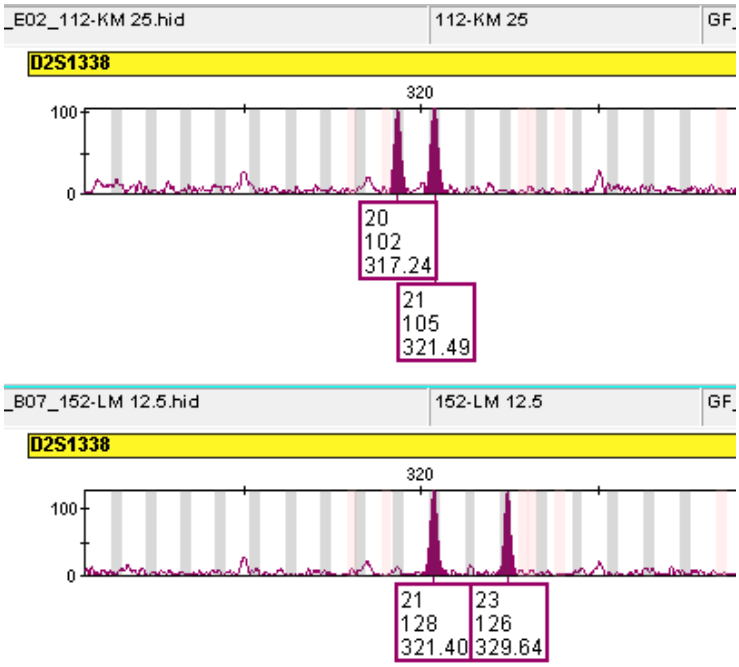


Figure 6 – Extraneous peaks detected in samples injected for 24 seconds.

It is possible that these peaks are a result of the primer mix not being purified in this lot. These peaks are closely associated and match the pattern of alleles at D18S51. Furthermore, D18S51 created amplicons of this size in the Identifiler Plus kit. Without further investigation of this primer mix and two samples, it is impossible to determine the cause of these extraneous peaks. When additional samples were amplified with the purified primer mix, these extraneous peaks were not detected.

Conclusions

All extraneous peaks investigated and documented over the course of this validation. These events appear to be sample or GlobalFiler kit lot number specific and not a systematic problem.

Contamination is not an inherent problem associated with the GlobalFiler kit. Each new lot # of GlobalFiler kits received into the laboratory will have a quality control ("QC") run performed prior to use in casework. This QC and the TE amplification blank included in each amplification will serve as ongoing verification that the components of the GlobalFiler kits do not exhibit inherent contamination.

References

Communication by ThermoFisher Scientific, GlobalFiler CN D18 artifacts_final_07132015 (460015_2), Ref: CN_07_15_01 (see user manuals and bulletins folder)

Adjudicated Case Study

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

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

Analytical record AND the write up of these cases is included in the STRmix Validation, Adjudicated case study. Please refer to this document for details on the analysis and interpretation of the data.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler® PCR Amplification Kit from Applied Biosystems

Normalization Study

Purpose

Normalization is a new feature included in the 3500 data collection software. It is designed to attenuate signal variations associated primarily between instruments, but also associated with capillary array and sample salt load. The purpose of this study is to determine if normalization would be beneficial for use on the two 3500 Genetic Analyzers. This feature was assessed with samples amplified with the GlobalFiler PCR Amplification Kit.

Materials and Methods

In order for the normalization feature to be utilized in the SDPD Crime Lab, three things must occur: 1 - samples must be injected for capillary electrophoresis with the GS600LIZ v.2 Size Standard, 2- in the 3500 data collection software, data must be collected with a normalization protocol, and 3- Data must be analyzed in GMID-X (v1.4) with an analysis method that has normalization enabled. Over the course of the internal validation of GlobalFiler on the 3500, the first two requirements were always met. The use of this feature was controlled in GMID-X with the analysis method. Data could be compared normalized vs. not-normalized by analyzing with two different methods.

In the 3500 data collection software, an injection assay must be specified before capillary electrophoresis can begin. Within this injection assay, the injection protocol is specified. The injection protocol defines parameters such as injection time, voltage, and matrix standard. It also defines whether or not normalization will be allowed, and if so, what the expected *normalization target* (in RFU) should be.

Normalization works by averaging the peak heights of 11 specific size standard peaks (200, 220, 240, 260, 280, 300, 314, 320, 340, 360, and 400 bp) in each sample. This average is used to divide the *normalization target* pre-specified in the injection protocol to get a *normalization factor*. This *normalization factor* is then used as a multiplier to adjust peak height of the sample peaks relative to the size standard peaks. For example, if the *normalization target* is 3200 RFU (default provided by Applied Biosystems), and the average of those 11 peak heights for one sample is 1600 RFU, the *normalization factor* is 2 in order to bring the average of those peaks up to 3200 RFU. Then, if the data is analyzed in GMID-X with normalization enabled, each of the peaks in the electropherogram for that sample are multiplied by 2.

To study how normalization would affect the data collected at SDPD, the normalization factor was tracked over the course of the validation for all GlobalFiler amplified samples injected on both 3500 instruments. Variability was compared between instruments, between samples, between size standard lot numbers, and between 3500 plate prep batches. To study the sensitivity of each instrument, samples amplified at a range of target amounts were used to assess average peak height on instrument 3500A vs. 3500B. The resultant data was analyzed in GMID-X (v1.4) using a 100 RFU detection threshold. The normalization target was reduced from the manufacturer recommended 3200 a few times over the course of the validation to obtain a more accurate value for the SDPD crime lab.

Results and Discussion

Normalization Target

Four different values for the normalization targets were tested. The first several plates run on the instrument used the recommended target of 3200 RFU. This recommendation is for samples amplified with GlobalFiler (29 cycles, manufacturer recommended target amount) and set up in a CE plate with GS600 LIZ v.2. As more samples were run, it became apparent that the size standard peaks were not reaching an average value of 3200 RFU in either 3500A or B. As a result, most normalization factors were over 1, causing an increase in sample peak heights (if normalization was turned on for analysis). Some samples even reached the maximum normalization factor of 3. 3000 RFU was the normalization target that was tested next with the same result. The normalization target was lowered to 2800, and then to 2700 to reflect an average size standard peak height for both instruments with the SDPD plate setup protocols.

With a normalization target of 2700, the average size standard peak heights for a sample ranged from ~600 RFU to over 3000 RFU, and as a result, normalization factors for the samples ranged from below 1 to a maximum value of 3. This was observed on 3500A and 3500B. The variability from sample to sample made it difficult to choose a normalization target, but the data indicated an average size standard peak height of 2700 RFU for both instruments to achieve the best balance of sample peak height increases and decreases.

Sensitivity between Instruments

To measure the general sensitivity of each instrument, capillary electrophoresis was performed with both instruments using the same two setup plates (70 samples, total). Each point on the graph represents the average allele peak height (not normalized) across the entire sample (Figure 1). The correlation between the two instruments is very good ($R^2 = 0.9846$), suggesting low variability between injections on 3500A and 3500B. The slope of the line also indicates very similar sensitivity between 3500A and 3500B.

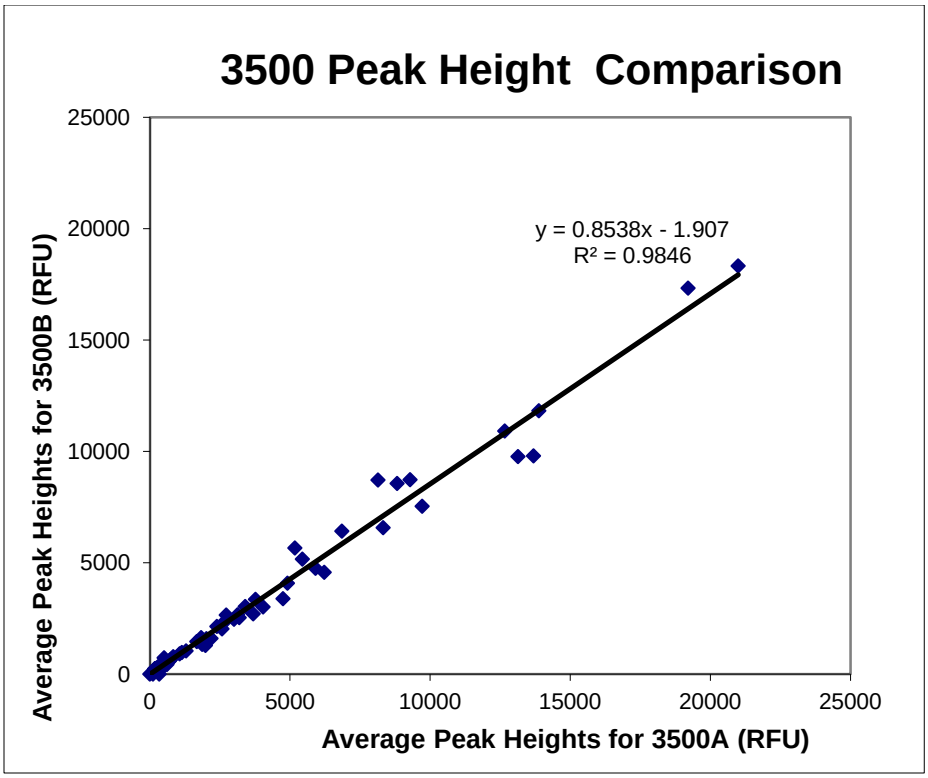


Figure 1: Correlation of sample peak heights between 3500A and 3500B from two different plate setups. Sensitivity is equal between the two instrument

To further characterize the differences between the two instruments, 17 different plate setups on 3500A and 15 different plate setups on 3500B were used. The average values of the 11 size standard peaks were averaged for all the samples in each plate (Table 1). A t-test was performed to assess internal size standard peak height differences between the two instruments. No significant differences in size standard peak height between 3500A and 3500B were found (p=0.162).

Table 1

Average Size Standard PH	
3500A	3500B
2813	2989
2133	2293
2507	2225
2364	2508
2162	2658
3811	3183
2139	2778
1782	2378
1647	2120
1203	1759
1909	1391
1716	1671
2608	3316
1637	2255
1758	2981
1453	
2591	

Normalization was primarily designed to minimize signal variation between 3500 instruments. The results of this study show that the signal variation between 3500A and 3500B is very well correlated and that these instruments are similarly sensitive.

Variability of normalization factor

The normalization factor for each sample is determined by the average of 11 size standard peaks in that well of the CE plate, and how that average compares to the pre-determined normalization target. As discovered over the course of the validation studies – there is a large amount of variability from

sample to sample in the average size standard peak height. This is most likely due to pipetting variability, either by the Corbett instrument or an analyst performing a manual plate setup. Within one plate setup, the average peak height of the 11 peaks varied as much as 3.53 fold from the lowest to the highest size standard peak height. In one sample from this plate, the average size standard peak height was 5042 (normalization factor of 0.63). In another sample on this plate, the average size standard peak height was 1428 (normalization factor of 2.24). All other plate setups had sample to sample to variability, as well, with fold changes of the maximum average to the minimum average of 1.39 to 3.53. This sample to sample variability resulted in normalization factors that brought the sample peak heights down in some samples, and peak heights up in adjacent samples on the same plate in a random fashion.

There is also some variability between lot number of size standard, but not as much as sample to sample variability. Five different lot numbers were tested over the course of the validation. Most of the plates in this study were prepared with one lot number of GS600 (1404044). The average size standard peak height did differ with the use of other lot numbers (t-test, p=0.112). The normalization factors for the same samples prepared with the four other lot numbers still varied from well to well (from 0.6 to 3).

Plate to plate variability was also observed. Across all samples in all plates the average of the size standard peak heights was 2273 RFU with a standard deviation of 606 RFU. This variability was also smaller than the sample to sample variability.

The observation that sample to sample variability is by far the largest source of variability creates more consideration for sample interpretation. Sample peak heights are potentially increasing by a factor of 3. The results of normalizing (increasing sample peak heights) could have drastic effects on the peak height ratio expectations and dropout. This would be less a concern if one of the instruments was significantly less sensitive than the other, and if this factor of change was being applied to all of the samples in a plate more equally than it is now. But this study indicates that it is primarily the internal size standard variability (presumably due to pipetting) that has the largest effect on normalization factor at this time.

Conclusions

Normalization is a feature that is not recommended for analysis in the SDPD lab at this time. The sensitivity difference and variability between the two instruments is so minimal when compared with the variability associated with the other analytical procedures that allowing for normalization would potentially create unwanted variability when it comes to interpretation. Using the normalization feature is not recommended until a time where the sensitivity of the two instruments differs more than sample to sample variability.

References

Applied Biosystems 3500/3500xL Genetic Analyzer User Guide. Part Number 4401661C

Applied Biosystems 3500/3500xL Genetic Analyzer. Part Number 4469192A

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Life Technology

Concordance Study

Purpose

The goal of this study is to verify that DNA typing results obtained with the GlobalFiler PCR Amplification Kit are concordant with those obtained with the AmpFℓSTR® Identifiler® Plus (ID Plus) PCR Amplification Kit. While many of the primer sequences remain the same, there were several modifications (primer redesign, new SNP specific primers, and linkers added) in moving from a 5-dye kit to a 6-dye kit.

Materials and Methods

Seventy reference samples that have documented ID Plus profiles and NIST standard reference samples were amplified at quantities within the optimal input range for 29 cycles using the GlobalFiler kit. Capillary electrophoresis was performed using both 3500 genetic analyzers and the resultant data was analyzed in GMID-X (v1.4). The results from the reference samples amplified with GlobalFiler were compared to the results at the 16 loci that are shared with ID Plus.

Results and Discussion

Genotypes for the NIST standard reference material (NIST SRM2391c Component A, B, and C) were 100% concordant at all loci after being amplified with the GlobalFiler kit.

The genotypes obtained following GlobalFiler amplification of the 70 reference samples were 100% concordant at the loci previously analyzed using ID Plus. A total of 1991 concordant alleles were observed. In addition, the positive control (DNA control 007) was concordant with the published profile (at all loci) in every amplification done with the GlobalFiler kit.

Since there are additional loci (D2S441, D22S1045, SE33, D10S1248, D1S1656, D12S391, and two Y chromosome markers Yindel and DYS391) associated with the GlobalFiler kit, these loci could not be tested for concordance with the previous kit. However, concordance did occur with respect to gender. Only male profiles (and every male profile) had alleles at the two additional Y markers (Yindel and DYS391).

Interestingly, SE33 had some interesting features. One tri-allele pattern was detected (Figure 1), and one apparent partial null allele (30.2) was detected at this locus (Figure 2). The tri-allele was amplifed three separate times to verify the presence of three alleles at SE33 in this sample. On in all three eletropherograms, the 29.2 and 30.2 alleles were approximately half the 21 allele.

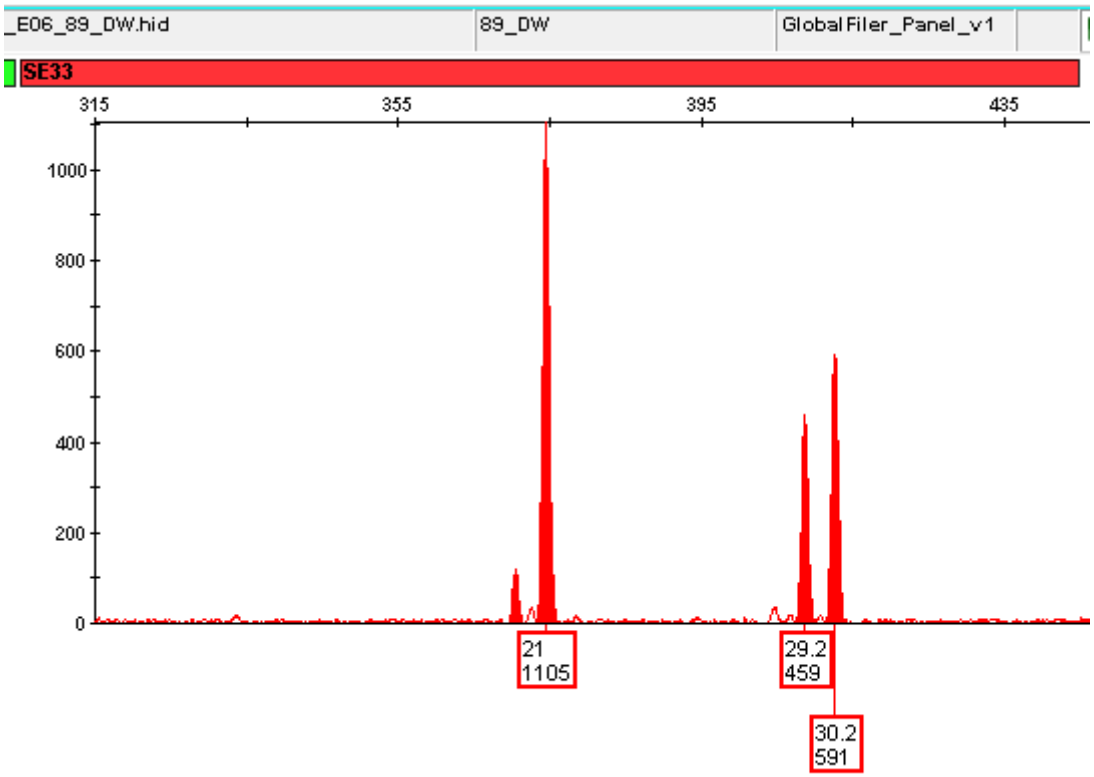


Figure 1 – A triallele pattern at SE33

The sample with the partial null allele was included in the sensitivity study, and was amplified with target amounts ranging from 800pg to 12.5pg. The partial null allele was no longer detectable above the analytical threshold at 50pg. The 30.2 allele ranged from 17% of the 26.2 allele to 37% of the 26.2 allele.

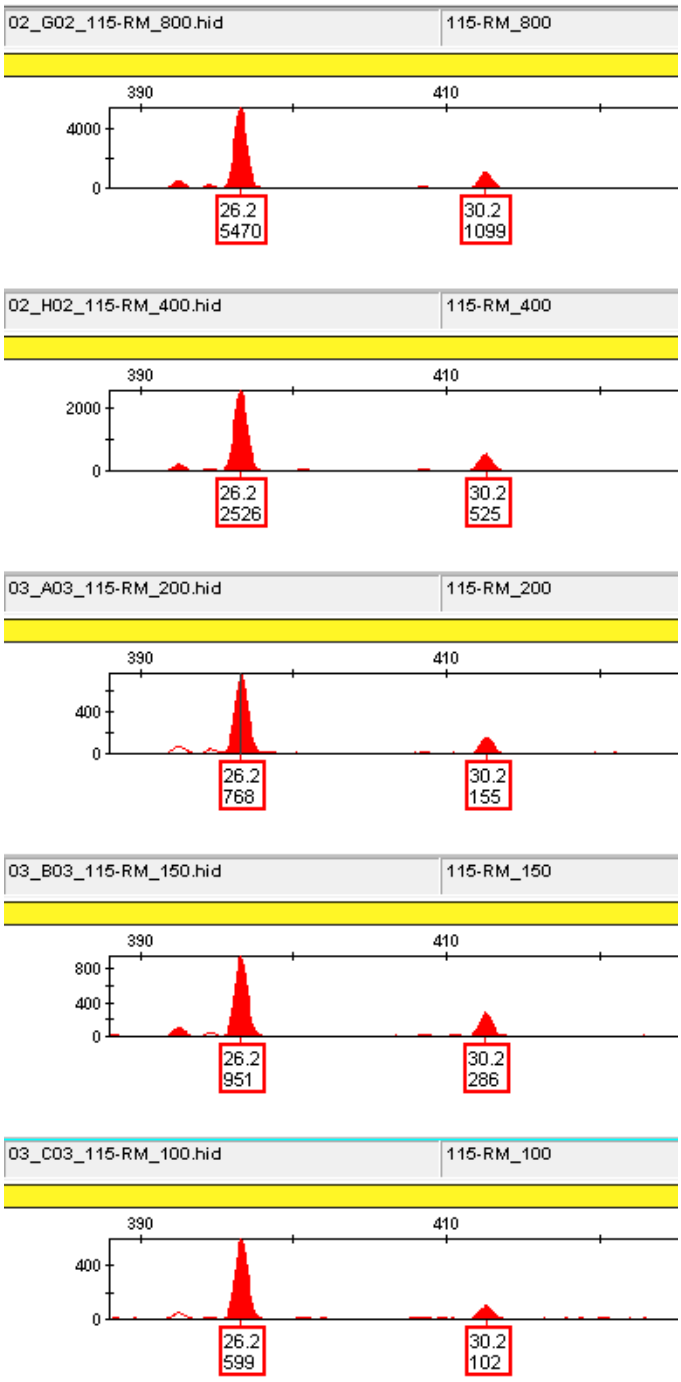


Figure 2 – An apparent partial null genotype at SE33 for one sample amplified at multiple target amounts.

Conclusions

DNA typing results obtained using GlobalFiler are fully concordant with those obtained using ID Plus. This confirms the expected results, as well as providing additional genotype characterization.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Life Technology

Reproducibility Study

Purpose

The goal of this study is to verify that the results obtained from samples amplified with the GlobalFiler PCR Amplification Kit are reproducible.

Materials and Methods

Over the course of the validation, a total of 167 single source DNA samples and 174 mixtures samples containing DNA from a combination of 2-4 people have been amplified for 29 cycles using GlobalFiler. Included in each amplification was at least one 007 positive control. This data was sourced from 77 reference samples used throughout the validation, therefore, many of these samples were amplified multiple times by several analysts. This included amplification reaction with the original GlobalFiler master mix formulation as well as the reformulated master mix released during the course of the validation studies. Ten of the samples were amplified with the original master mix formulation at 10 different dilutions each. Thirteen different samples were amplified with the reformulated master mix at 6 different dilutions each. Aside from the instances of dropout, the same genotype was obtained every time. Likewise, for robust samples in the mixture study, every allele could be attributed to the known contributors to the mixture.

Results and Discussion

The results from the 007 positive control were consistent between each amplification, and were consistent with the known profile published by the manufacturer. Replicates of the same employee reference sample produced concordant results between multiple GlobalFiler amplifications and with previously obtained Identifiler Plus results (also see concordance study).

Conclusions

Results obtained with GlobalFiler are reproducible.

SDPD Forensic Science Section – Forensic Biology Unit

Validation of the GlobalFiler PCR Amplification Kit from Life Technology

Precision Study

Purpose

The goal of this study is to evaluate the sizing precision of data generated from samples amplified with the GlobalFiler PCR Amplification Kit using 3500 genetic analyzers.

Materials and Methods

Multiple injections of the GlobalFiler ladder were performed on both 3500 genetic analyzers. The data was analyzed from 25 separate injections of ladder, which were pooled from several studies and both instruments over the course of the validation. The extent of variability across each ladder allele from all 25 injections was determined.

Results and Discussion

Precision was evaluated by assessing the average allele size and standard deviation for each allele in all ladders. A representative example (D16S539, Allele 5) is shown in Figure 1. An overlay view of the 25 ladders at Allele 5 of this locus visually illustrates the small amount of variability between different injections. Table 1 describes the minimum, maximum, average and standard deviation size of Allele 5 at D16S539 for 25 ladder samples.

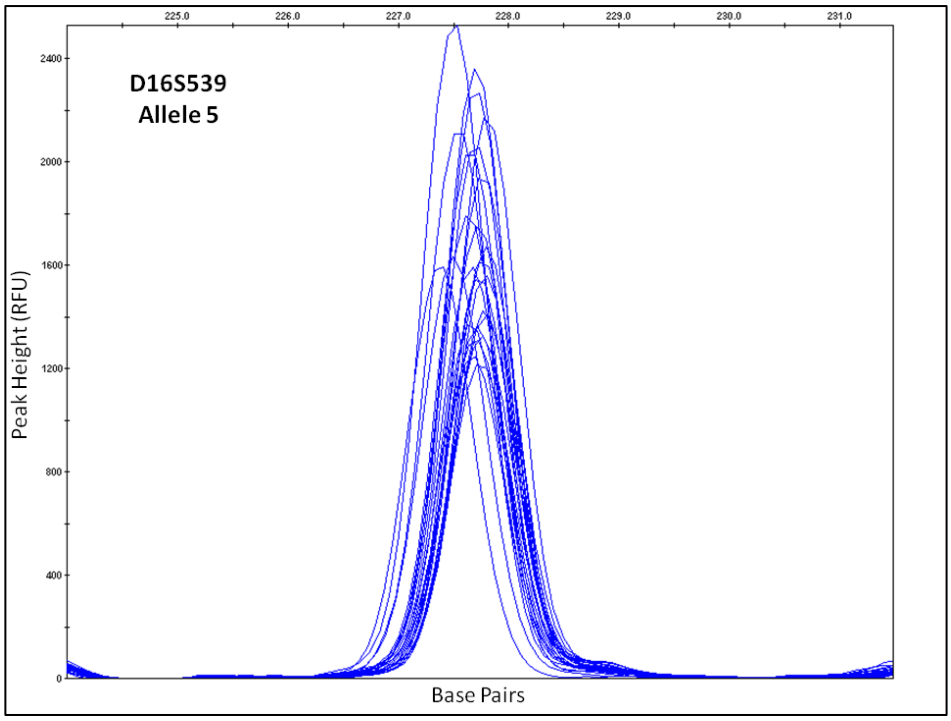


Figure 1. Electropherogram showing an overlay view of 25 ladder samples at Allele 5 of locus D15S539.

Table 1. Variation at Allele 5, D16S539

Description	Size (bp)
Min	227.41
Max	227.71
Average	227.68
Standard Deviation	0.103

A representative example is shown above, but precision across all alleles and loci were examined to determine the degree of consistency. Table 2 lists the smallest and greatest standard deviations from the average peak height. As a general trend, larger amplicons had a higher standard deviation (Table 2). The target value for the standard deviation is less than 0.15 bp. When the standard deviation is below 0.15 bp, the sample sizing error rarely exceeds 0.5 nucleotides, which minimizes the likelihood of an erroneous call (1). The only locus with a standard deviation across the mean peak height exceeding 0.15 bp was TPOX, whose amplicons are located above 330 bp in the GlobalFiler kit.

Table 2. Summary of precision across all loci in the GlobalFiler ladder.

Description	Standard Deviation (bp)	Amplicon size (bp)
Min Standard Deviation	0.022	160
Max Standard Deviation	0.166	374.98
Median Standard Deviation	0.061	

Sizing Accuracy (the difference between the maximum and minimum size for each allele) also varied with size. Ideally, the size range for any given allele remains below 0.5 bp, so that allele calls are called correctly. Across all 343 allele peaks in the ladder, only 18 alleles had a size range greater than 0.5 bp, with a maximum being 0.81 bp range. This may be due to the fact that the ladders were pooled from multiple runs over time. Ladders within the same run are more likely to be consistent with each other than between runs. All of the alleles with a size range above 0.5 bp were from TPOX and DYS391. Table 3 summarizes these data.

Table 3. Summary of sizing accuracy across all loci in the GlobalFiler ladder.

Description	Size Range (bp)	Amplicon size (bp)
Min size range	0.08	160
Max size range	0.81	367
Median size range	0.24	

Conclusions

Based on data collected across all alleles of all loci, miscalled alleles due to sizing imprecision should not be a concern. Precision is more closely associated with the capillary electrophoresis instrument than the DNA typing kit. This study reveals that reliable results will be obtained by performing capillary electrophoresis on both genetic analyzers. The results obtained following amplification with GlobalFiler demonstrated acceptable levels of precision. These results also

allow us an average size for every allele detected in the ladder. This may be of use when creating a reference profile for use in STRmix. If the only reference available was not typed with GlobalFiler, it would still be possible to create a compatible file for STRmix with the use of these ladder sizes.

References

1. Smith, et al. 1995, Biotechniques 18(1):122-8.