STRmix™
RESOLVE MORE DNA MIXTURES.

Estimation of STRmix™ V2.6 Parameters for Palm Beach County Sheriff's Office (PowerPlex® Fusion6C, 3500xl CE)

REVIEWED
Date: 6/14/19
Initials: ABM
OK for use on Casework.

STRmix™ Implementation

This document describes the estimation of the STRmix™ parameters for PowerPlex® Fusion 6C DNA profiling data (29 cycles, 3500xl CE) for the Palm Beach County Sheriff's Office Crime Laboratory (hereafter called PBSO), for use in version 2.6 of STRmix™, created in conjunction with the STRmix™ support team.

The STRmix™ validation was conducted in conjunction with PBSO's validation of PowerPlex® Fusion 6C. Supporting data for sample preparation may be found in either the STRmix™ validation companion binder or the PowerPlex® Fusion 6C validation companion binder.

STRmix™ parameters

There are a number of parameters which are not optimized by the MCMC in a STRmix™ analysis. These parameters must be set by the user and are either determined by the analysis of empirical data or modelled within STRmix™ using Model Maker. The laboratory specific parameters that are determined prior to the use of STRmix™ are:

- Analytical/Detection Thresholds,
- Stutter Ratios,
- Drop-in Parameters,
- Saturation,
- Allele and Stutter Variances,
- and Locus Specific Amplification Effects (LSAE).

These parameters need to be defined for each STR kit, each protocol (e.g. cycle number variation), and CE platform (e.g. 3130 or 3500), and potentially each time there is a significant change to the platform (e.g. a camera or laser change). Stutter ratios and saturation were determined for PBSO's 29 cycle Fusion 6C data analyzed on a 3500xl capillary electrophoresis instrument. Peak height variance and locus specific amplification efficiencies are calculated using Model Maker within STRmix™ from the analysis of empirical profile data. The results of these analyses are described within this report.

Analytical Thresholds

The assignment of a signal as allelic product as opposed to baseline or noise is important in DNA profile analysis. This differentiation is usually undertaken using a set threshold above which peaks are deemed to be allelic if they also meet certain morphological requirements, and below which they are ignored, regardless of morphology. The issue is to assign a threshold, often termed the limit of detection (LOD) or analytical threshold (AT), to minimise the detection of artifacts while maximizing the detection of allelic peaks.

Optimum AT values have previously been determined by PBSO for all the Fusion 6C loci and are dye specific as follows:

Blue 75 rfu, Green 101 rfu, Yellow 60 rfu, Red 69 rfu, Purple 56 rfu and Orange 50 rfu.

In order to maximise the recovery of data for this study a reduced AT value of **25 rfu** for all dyes was used for all samples.

Stutter

Within STRmix™ v2.6 there is the ability to model any type of stutter the laboratory might observe in their questioned samples. This is known as *generalized stutter* modelling. To model these different types of stutter, stutter ratios for each stutter type must be determined from the empirical data.

Analysis was undertaken to study the PBSO specific observations of stutter. A total of 108 single-source profiles were analyzed in GeneMapper™ at an AT of 25 rfu, capturing allelic, back, forward, double-back and half back stutter peaks. The lowered AT values which are well below the PBSO casework AT values, were used to capture more stutter information to inform the analyses described below.

It is assumed that all loci are stuttering, however at some loci and especially for less common stutter types some of these peaks are below the analytical threshold and therefore not visible. Missing stutter peaks were inserted at half the analytical threshold (12.5rfu), if the parent peak is 4,000 rfu or above. This is a statistical method to account for data paucity. Data points derived from inserted stutter peaks are indicated by triangles in the plots shown in the Appendices.

Because each type of generalized stutter needs to be defined in relation to the stutter's parent peak, new nomenclature has been adopted. The position of a stutter in relation to its parent allele can be defined by the nomenclature (*i,j*). *i* defines the number of whole STR repeat units the stutter is located from the parent peak, and *j* defines the extra nucleotides required to locate the stutter. For example, (-1,0) describes back stutter one whole repeat less than the parent allele. This is further explained in the STRmix™ v2.6 User's Manual. Below is a summary of stutter observations and data work-up. This will begin with back (-1,0) and forward (1,0) stutter as they are most prevalent in profiles then move on to the additional stutters being considered by PBSO.

Back stutter (-1,0)

Definition: Back stutter is one repeat unit smaller than its allelic component (i.e. O_{a-1}). Tetra-nucleotide loci have repeat units of four nucleotides, tri-nucleotide loci have repeat units of three nucleotides and penta-nucleotide loci have repeat units of five nucleotides.

There are three parameters within STRmix™ that are used to calculate the expected back stutter height and require optimization.

The first is the **maximum allowable stutter ratio**. A maximum allowable stutter ratio reduces run time by only permitting peaks in stutter positions below a certain ratio to be considered as stutter. The highest observed back stutter ratio was 0.2165 at D1S1656. Therefore, this parameter has been set at **0.3**, which is conservatively high, based on inspection of PBSO's stutter ratio data.

The second parameter is a file used to model the expected heights of the stutter peaks based on their parent allele designation. The values used to determine expected stutter heights are 'per allele'. Per allele stutter ratios are calculated using a linear equation $SR = m \times Allele + c$ where the intercept (*c*) and slope (*m*) are determined by regressing stutter ratio against allele.

Values for *m* and *c* were calculated from the PBSO data. A plot of *SR* versus Allele (and *SR* versus LUS, which is discussed below) for each locus is provided in **Appendix 1**. A summary of the STRmix™ stutter by allele file for the PBSO data is given in Table 1.

Back (-1,0) stutter variability will be modelled as being *inversely proportional to the observed height of the parent allele*.

Table 1: PBSO per allele Fusion 6C back (-1,0) stutter regression for STRmix™. The associated text file has been titled: "**Palm Beach_IN_Fusion6C_3500_Back Stutter**.txt "

Marker	Intercept	Slope
D3S1358	-0.07057	0.01046
D1S1656	0.00132	0.00604
D2S441	0.05096	-0.00002
D10S1248	-0.06292	0.01068
D13S317	-0.07044	0.01161
Penta E	-0.01607	0.00398
D16S539	-0.0595	0.0118
D18S51	-0.04708	0.00879
D2S1338	-0.0119	0.0049
CSF1PO	-0.05766	0.01144
Penta D	-0.00468	0.00213
TH01	0.00801	0.00185
vWA	-0.09819	0.01098
D21S11	-0.06858	0.00523
D7S820	-0.05172	0.01048
D5S818	-0.05569	0.01106
TPOX	-0.02772	0.00611
D8S1179	0.01873	0.00461
D12S391	-0.10533	0.01063
D19S433	-0.0604	0.00981
SE33	0.04811	0.00304
D22S1045	-0.1354	0.01528
DYS391	0	0
FGA	-0.06731	0.00659
DYS576	0	0
DYS570	0	0

A better explanatory variable for stutter ratio for some loci with compound and complex repeat structure has been shown to be the longest uninterrupted stretch of common repeats (LUS) within the allele [1-3] and not the allele designation itself. Values for LUS are determined by sequencing alleles. A number of common alleles for forensic loci have been typed. A summary of these appear on STRBase [4,5]. A plot of *SR* versus LUS for the compound and complex loci within Fusion 6C is provided within **Appendix 1**. Some of the plots of *SR* versus LUS are provided for comparison only, as loci where neither

a linear model based on Allele nor LUS describes the data well, the average of the observed *SR* for each allele was applied. Please refer to Table 2 for details.

The third parameter within STRmix™ that is used to calculate the expected stutter peak heights is an exceptions file. It is populated with per allele *SR* values based on either the LUS linear regression or an average observed allelic stutter ratio, if either better explain the data than the allelic regression. A stutter exceptions file based on PBSO's data has been created and will form part of the PBSO Fusion 6C STRmix™ method. Where a 0 appears in a cell for a given allele in this file the expected stutter rates are calculated from the allele file (Table 1).

A summary of the explanatory variables selected to model the expected *SR* for each locus is given in Table 2.

Looking at the plots in Appendix 1, it is important for analysts to be aware of iso-alleles, in particular known instances in vWA at allele 14, and to a lesser degree allele 15. Typically these two alleles exhibit two variants within the population that stutter in different amounts due to their repeat structure[1]. This is observed in the PBSO data as a cluster of vWA 14 SR values <0.025 and a second looser grouping between 0.05 and 0.1. Based on the PBSO data, the explanatory variable chosen for *SR* for vWA is the average for each allele. This should be sufficient for the majority of samples analyzed. Despite this, on occasion this may lead to potential over- or under-estimation of the expected stutter ratio compared to the true[2] value for a given individual's vWA allele. Once STRmix™ is in use in casework, it is recommended a review of the output is undertaken paying particular attention to vWA if samples run through STRmix™ contain either a vWA 14 or 15. **The impact of the iso-alleles at vWA 14 on stutter variance is demonstrated in the variance section below.**

[1] This is a known issue and is further explained within a FAQ on the STRmix™ support website.
[2] The true expected *SR* to apply is unknown, unless sequencing were done of the donors DNA

Table 2: Explanatory variables for the expected back *SR* for Fusion 6C loci for PBSO.

Locus	Explanatory variable chosen
D3S1358	Allele Average
D1S1656	LUS Regression
D2S441	Allele Average
D10S1248	Allele Regression
D13S317	Allele Average
Penta E	Allele Regression
D16S539	Allele Regression
D18S51	Allele Regression
D2S1338	Allele Average
CSF1PO	Allele Regression
Penta D	Allele Average
TH01	LUS Regression
vWA	Allele Average (note vWA 14 issue)
D21S11	Allele Average
D7S820	Allele Regression
D5S818	Allele Average
TPOX	Allele Regression
D8S1179	Allele Average
D12S391	Allele Regression
D19S433	Allele Average
SE33	Allele Average
D22S1045	LUS Regression
DYS391	N/A
FGA	Allele Average
DYS576	N/A
DYS570	N/A

The file format in STRmix™ is .csv and this file has been titled: "**Palm Beach_IN_Fusion6C_3500_ Back Stutter ExceptionsI.csv**"

Forward (1,0) Stutter

Definition: Forward stutter is one repeat unit larger than its allelic component (i.e. O_{a+1}).

There are two parameters within STRmix™ that calculate expected forward stutter ratios and that require optimization. The first is the **maximum allowable forward stutter ratio**.

This parameter has been set conservatively high at **0.15** based on inspection of the laboratory's forward stutter ratio data. The marker where most forward stutter was observed in this data set was D2S441. The highest observed forward (1,0) stutter ratio was observed in D22S1045 at 0.0826. This is not unexpected given this is a tri-nucleotide repeat marker so is known to stutter more.

The second parameter is a file used to model the expected heights of the forward stutter peaks based on their parent allele designation. Results are shown in **Appendix 2**. D22S1045 was the only locus where the relationship between allele designation and FSR was obvious, therefore a linear regression model was created based on observed data for this locus. At the remaining loci an average of the forward stutter data for the entire locus was used. This has been implemented by setting the slope to 0 and the intercept to the average of the observed stutter for the locus within the forward stutter text file. A summary of the STRmix™ forward stutter file for PBSO's data is given in Table 3.

Forward (1,0) stutter variability will be modelled as being *inversely proportional to the expected height of the stutter allele*[3].

[3] As recommended in the STRmix™ v2.6 User's Manual, all but back (1,0) stutter should be modelled with this setting.

Table 3: Expected FSR for Fusion 6C for PBSO. This text file has been titled: "**Palm Beach _IN_Fusion6C_3500_Forward Stutter**.txt".

Marker	Intercept	Slope
D3S1358	0.00751	0
D1S1656	0.01149	0
D2S441	0.00753	0
D10S1248	0.00248	0
D13S317	0.00635	0
Penta E	0.00273	0
D16S539	0.00974	0
D18S51	0.00768	0
D2S1338	0.00315	0
CSF1PO	0.00926	0
Penta D	0.00255	0
TH01	0.00225	0
vWA	0.00463	0
D21S11	0.01001	0
D7S820	0.00476	0
D5S818	0.00888	0
TPOX	0.00241	0
D8S1179	0.00669	0
D12S391	0.00309	0
D19S433	0.0039	0
SE33	0.00847	0
D22S1045	-0.07095	0.0088
DYS391	0	0
FGA	0.00728	0
DYS576	0	0
DYS570	0	0

Additional Stutter

Within STRmix™ version 2.6, PBSO will incorporate the modelling of double back (-2,0) stutter and minus two base pair/half back (0,-2) stutter, at appropriate loci. As per the stutters described above, the same 108 samples were analysed at 25 rfu. Empirical Fusion 6C multiplex validation data from PBSO suggested that double back stutters had been detected above casework AT at the following loci; D3S1358, D1S1656, D10S1248, D16S539, D18S51, D21S11, D8S1179, D19S433, SE33 and D22S1045 and minus two base pair stutters at D1S1656, D21S11, D5S818, D19S433, SE33 and FGA, hence GMID-X data analysis permitted collection of these types of stutter only at these loci.

Double Back (-2,0) Stutter

Definition: Double back stutter is two repeat units smaller than its allelic component (i.e. $O_{\alpha-2}$). For Fusion 6C and PBSO, double back stutter modelling was only considered at D3S1358, D1S1656, D10S1248, D16S539, D18S51, D21S11, D8S1179, D19S433, SE33 and D22S1045.

Plots of the double back (-2,0) stutter ratio versus allele are found in **Appendix 3**. Inspection of the empirical data in Appendix 3 shows that these *SR* are low; often at 1%. It is anticipated that most double back stutters will fall below the PBSO casework AT. However, all listed loci were modelled for added functionality. The highest observed (-2,0) stutter originated from D18S51 at 0.0198. The maximum (-2,0) stutter ratio is set conservatively high at **0.05**.

Inspection of **Appendix 3** shows that linear regression based on allelic designation or LUS is not a great explanatory variable for double back *SR* for the majority of the listed loci. The exception being D1S1656 where the LUS regression shows a good fit to the data. D10S1248 and D18S51 have each been modelled using an average *SR* per allele. To implement these modelling choices a double back stutter exceptions file has been populated for these three loci only and has been titled: "**Palm Beach_IN_Fusion6C_3500_ Double Back Stutter Exceptions.csv**" for use at PBSO.

For the remaining listed loci the double back stutter ratio model will use the average *SR* of the entire locus. This value is used as the intercept for these loci. A summary of the STRmix™ double back stutter file for PBSO's data is given in Table 4.

Double back (-2,0) stutter variability will be modelled as being *inversely proportional to the expected height of the stutter allele*.

Table 4: PBSO Fusion 6C double back (-2,0) stutter values for STRmix™. This text file has been titled: "**Palm Beach_IN_Fusion6C_3500_Double Back Stutter**.txt".

Locus	Intercept	Slope	Enabled
D3S1358	0.00632	0	Y
D1S1656	-0.00025	0.00053	Y
D2S441	0	0	N
D10S1248	-0.00751	0.00113	Y
D13S317	0	0	N
Penta E	0	0	N
D16S539	0.00499	0	Y
D18S51	-0.01166	0.0013	Y
D2S1338	0	0	N
CSF1PO	0	0	N
Penta D	0	0	N
TH01	0	0	N
vWA	0	0	N
D21S11	0.00631	0	Y
D7S820	0	0	N
D5S818	0	0	N

TPOX	0	0	N
D8S1179	0.00491	0	Y
D12S391	0	0	N
D19S433	0.00626	0	Y
SE33	0.00984	0	Y
D22S1045	0.0083	0	Y
DYS391	0	0	N
FGA	0	0	N
DYS576	0	0	N
DYS570	0	0	N

Two Base Pair/Half Back (0,-2) Stutter

Definition: Two base pair or half back stutter is a peak that is two nucleotides smaller than its allelic component. While two base pair back stutter modelling was considered at six lociD1S1656, D21S11, D5S818, D19S433, SE33, and FGA , it was only observed at D1S1656 and SE33 in this Fusion 6C PBSO data set.

Plots of the two base pair back (0,-2) stutter ratio versus allele for D1S1656 and SE33 are found in Appendix 4. On inspection of the empirical data in **Appendix 4**, the highest observed (0,-2) stutter was 0.0629 at SE33. The maximum (0,-2) stutter ratio is set conservatively high at **0.1**.

Inspection of **Appendix 4** shows that the regression relationship of *SR* by allele does not appear to hold well for the modelled loci. Therefore the average *SR* of the entire locus has been used to model this type of stutter at each locus. A summary of the STRmix™ two base pair back stutter file for PBSO's data is given in Table 5. Again, setting the slope to 0 means the intercept is the average of the observed stutter ratios for the locus.

Minus two base pair (0,-2) stutter variability will be modelled as being *inversely proportional to the expected height of the stutter allele*.

Table 5: PBSO Fusion 6C two base pair back (-2,0) stutter values for STRmix™. This text file has been titled: "**Palm Beach_IN_Fusion6C_3500_Half Back Stutter**.txt".

Locus	Intercept	Slope	Enabled
D3S1358	0	0	N
D1S1656	0.01836	0	Y
D2S441	0	0	N
D10S1248	0	0	N
D13S317	0	0	N
Penta E	0	0	N
D16S539	0	0	N
D18S51	0	0	N
D2S1338	0	0	N
CSF1PO	0	0	N

Penta D	0	0	N
TH01	0	0	N
vWA	0	0	N
D21S11	0	0	N
D7S820	0	0	N
D5S818	0	0	N
TPOX	0	0	N
D8S1179	0	0	N
D12S391	0	0	N
D19S433	0	0	N
SE33	0.05286	0	Y
D22S1045	0	0	N
DYS391	0	0	N
FGA	0	0	N
DYS576	0	0	N
DYS570	0	0	N

Drop-in parameters

Drop-in is non-reproducible, unexplained peaks observed within a profile. There are four parameters used for the modelling of drop-in in STRmix™. These are:

1. Z: the detection threshold or analytical threshold
2. A cap on the maximum allowed combined drop-in height per locus
3. The drop-in frequency
4. α,β: two parameters for the gamma model.

Drop-in rates for a laboratory platform (multiplex and instrument combination) should be monitored. This is done by recording counts and corresponding heights of drop-in peaks observed in negative controls and counts of negative controls without drop-in peaks. An on-going review will determine whether the following values are fit for purposes in the future.

PBSO reviewed 128 Fusion 6C reagent blank and negative control samples at AT 25 rfu and observed one putative drop-in peak at 82 rfu. This is insufficient to meaningfully inform the drop-in model in STRmix™. Subsequently PBSO's inaugural Fusion 6C drop-in parameters were provided by the STRmix™ scientific support team based on their experience of similar platforms (kit, cycle number and CE). A uniform prior distribution has been chosen, applying a consistent probability of drop-in between AT and the drop-in cap. Appropriate settings have been provided in the following table:

Table 6: PBSO's inaugural drop-in parameters for STRmix™ for the Fusion 6C kit/platform

Drop-in cap	150
Drop-in frequency	0.0001
Drop-in parameters α,β	0,0

Saturation

The peaks in a DNA profile are measured using fluorescence. The amount of fluorescence is proportional to the amount of DNA present. This fluorescence is captured by a camera. It is expected that as more DNA is added into a PCR the resulting peak height (measured in relative fluorescent units) in an electropherogram will increase. The camera can become saturated when there is too much fluorescence detected. This means we can no longer accurately measure the height of the peaks observed or estimate how much DNA is really represented by this result. Following this we can no longer accurately model over-saturated peak heights using STRmix™. Saturation, like the analytical threshold, is mostly instrument related and not kit or method dependent.

A saturation threshold setting for PBSO's 3500 CE instruments has already been established for the previous Fusion 5C kit as **30,000** rfu.

Peak height variance and LSAE using Model Maker

Empirical observations and experience suggests that profiles differ in variance (hereafter "quality"). Within STRmix™ the variability of peaks within profiles is described using a model containing a variance constant. Allele and stutter peaks (each type of stutter, $a \pm i,\ j$) have separate variances; c^2 and $k^2_{a\pm i,\,j}$, respectively. The c^2 and $k^2_{a\pm i,\,j}$ terms are variables which are determined after sampling from a gamma distribution within the MCMC.

The gamma distribution priors that STRmix™ samples from during an interpretation are optimized in Model Maker, an add-on to the STRmix™ software. Model Maker works by using a component wise MCMC. In component 1 each DNA profile has its mass parameters optimized and uses a stable gamma distribution for allele, stutter and LSAE variance constants. In component 2 the mass parameters for each profile are held constant and the hyperparameters for each gamma distribution are varied. Components are 1000 accepts long and they cycle through a number of times depending on the user input value.

The samples used for this section consisted of a series of three known donors and their DNA extracts diluted, to create a range of input templates from 0.0039ng up to 4ng and a range of 20 year old blood stains that were exhibiting some degradation. These samples were amplified and run on the 3500xl capillary electrophoresis machine using PBSO's standard casework procedures. The resultant CE data was analyzed in GeneMapper™ at an AT of 25 rfu across all dyes and then run in Model Maker.

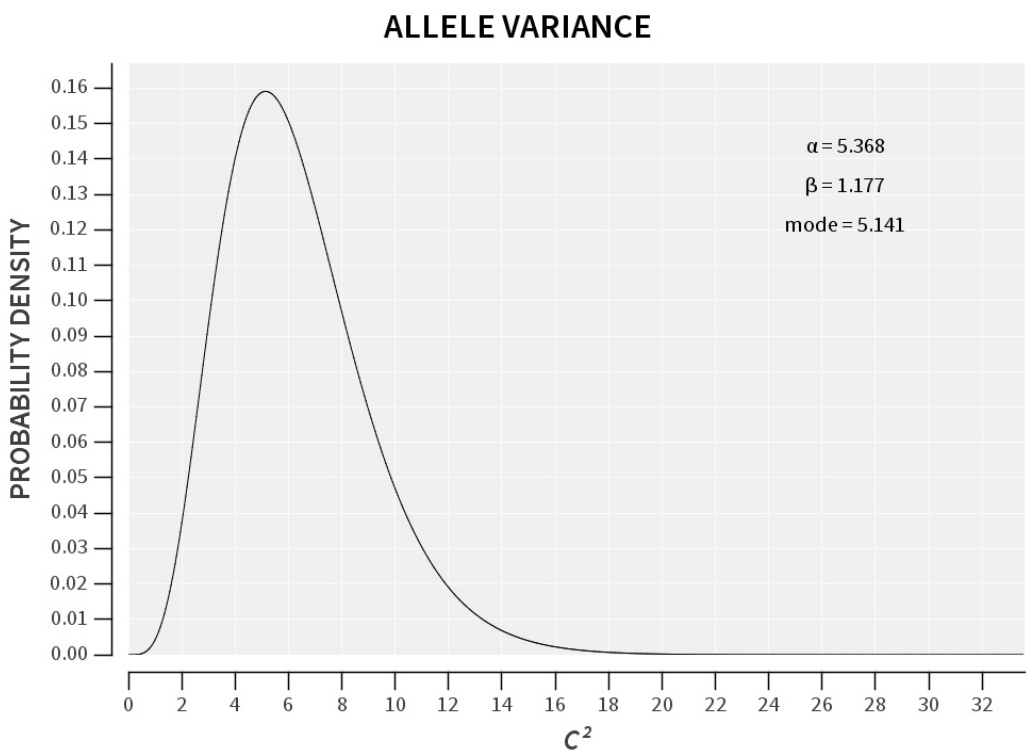
The Model Maker analysis within STRmix™ for the PBSO laboratory was carried out on 128 single source profiles of varying quality (template), each for 100 cycles (100,000 accepts total). One of these samples was not used as it contained less than 11 peaks.

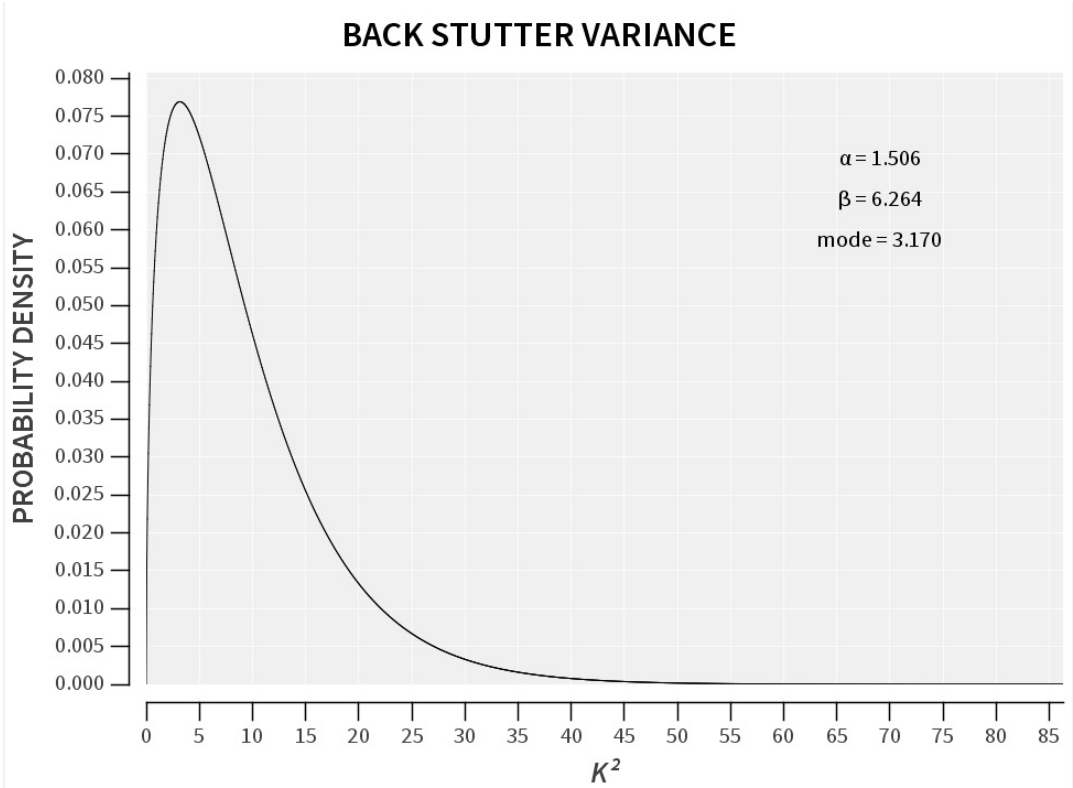
A summary of the results for the dataset is provided in Table 7. The gamma distributions and correlation plots are provided in **Figure** 1.

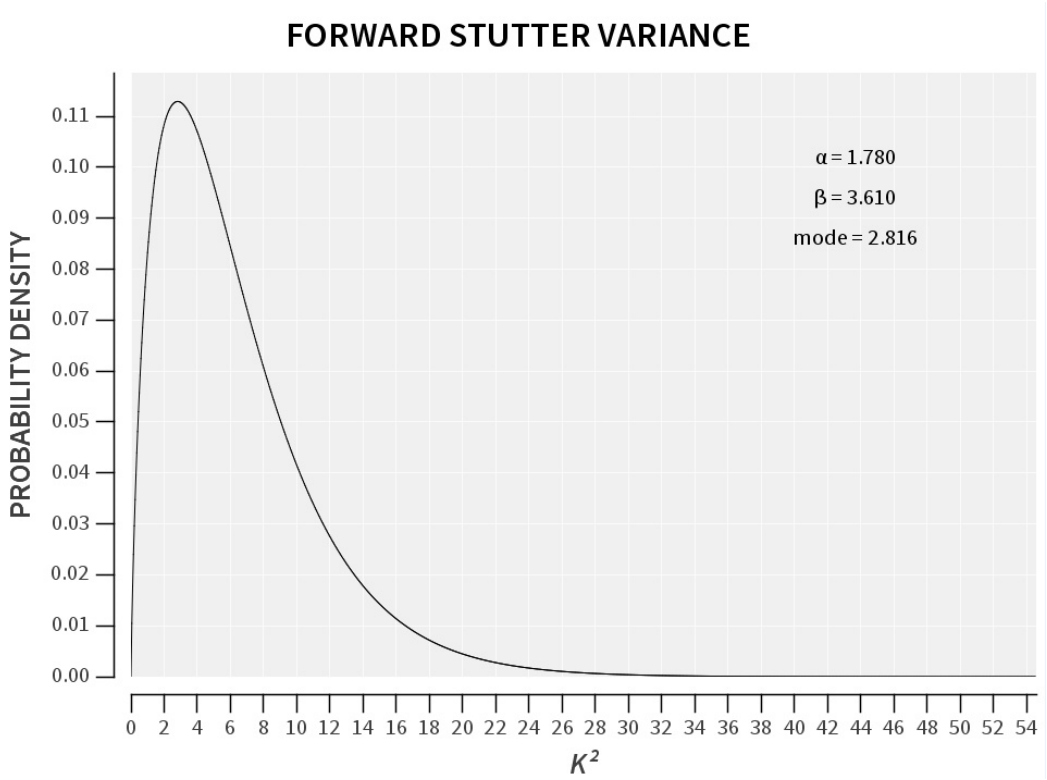
Table 7: Summary of allele and stutter prior variance values for Fusion 6C, 29 cycles, 3500xl CE at PBSO

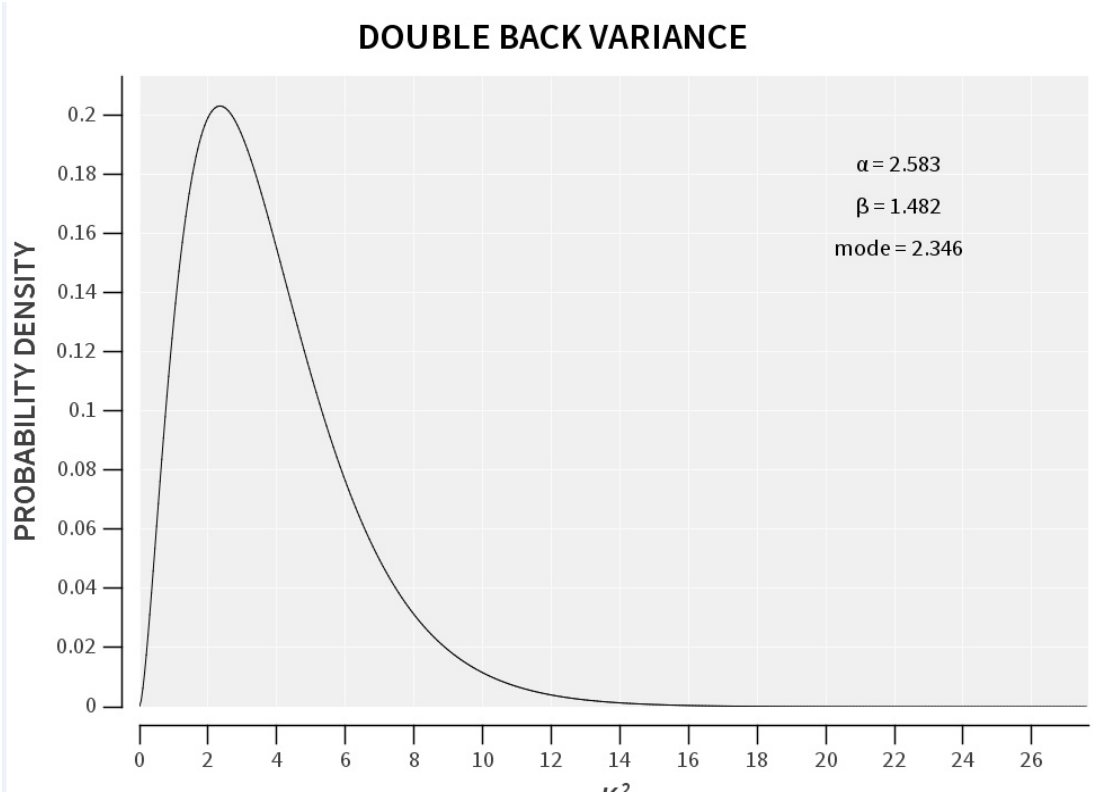
Number profiles analysed	Allele (Mode)	Back stutter (Mode)	Forward stutter (Mode)	Double back stutter (Mode)	Half back stutter (Mode)	Mean LSAE variance
127	5.368,1.177 (5.141)	1.506,6.264 (3.170)	1.780,3.610 (2.816)	2.583,1.482 (2.346)	1.785,1.153 (0.905)	0.014

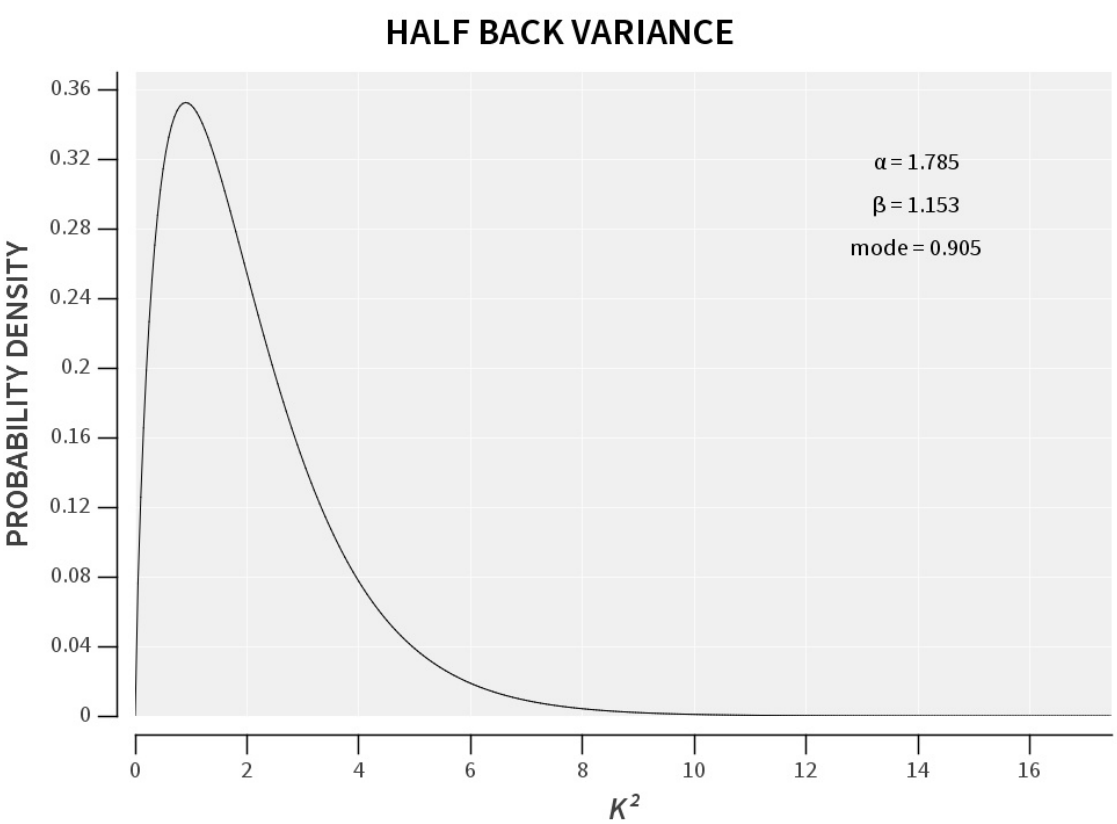
Figure 1: Summary plots of the allele and stutter prior gamma distributions (note the scales differ)




BACK STUTTER VARIANCE
PROBABILITY DENSITY
0.080
0.075
0.070
0.065
0.060
0.055
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
α = 1.506
β = 6.264
mode = 3.170
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
K²


FORWARD STUTTER VARIANCE
PROBABILITY DENSITY
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
α = 1.780
β = 3.610
mode = 2.816
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54
K²


DOUBLE BACK VARIANCE
PROBABILITY DENSITY
α = 2.583
β = 1.482
mode = 2.346
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26


HALF BACK VARIANCE
PROBABILITY DENSITY
α = 1.785
β = 1.153
mode = 0.905
0.36
0.32
0.28
0.24
0.2
0.16
0.12
0.08
0.04
0
0 2 4 6 8 10 12 14 16
K²

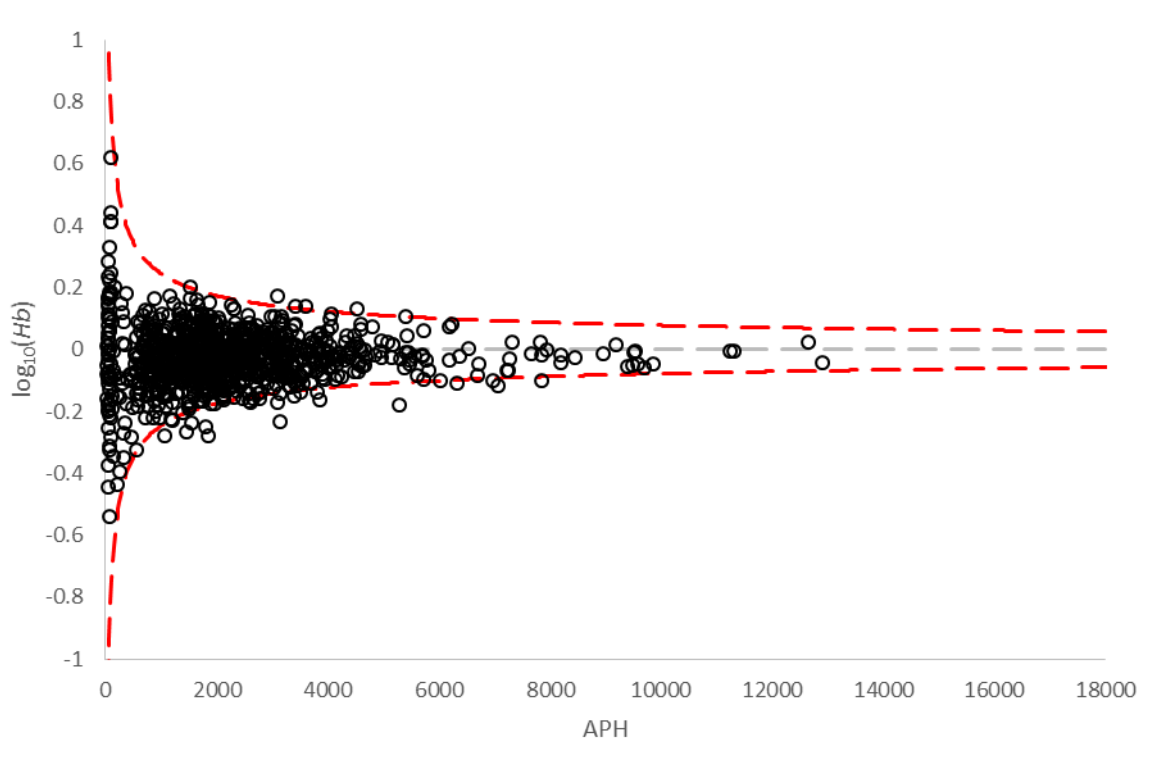
As a further review of these settings, heterozygote balance was calculated for all heterozygote loci for the Model Maker profiles (combined data set). Heterozygote balance (*Hb*) was calculated as:

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

Where O_{HMW} refers to the observed height of the high molecular weight allele and O_{LMW} the observed height of the low molecular weight allele. Previous work has suggested that there is a relationship between the variation in peak height and the variation in *Hb* [6, 7]. In single source profiles, variability in *Hb* reduces as the average peak height (APH) at a locus increases. The variance of *Hb* can be used as a proxy for the variance of individual peaks. This allows an approximate comparison between the variance from the STRmix™ MCMC approach and *Hb*; a readily determined variable from empirical data.

The plot of log*Hb* versus APH, for the dataset described above with expected 95% bounds (plotted as dashed lines) calculated at $\pm\sqrt{2}\times 1.96\times\sqrt{\frac{c^2}{APH}}$ where c^2 = 7.884 (the 75th percentile from the allelic variance prior distribution for this data set), is shown in **Figure** 2. The 95% bounds encapsulate sufficient data as demonstrated in the graphs (coverage = 96.7%) demonstrating that the values for variance are sufficiently optimized (**Figure** 2).

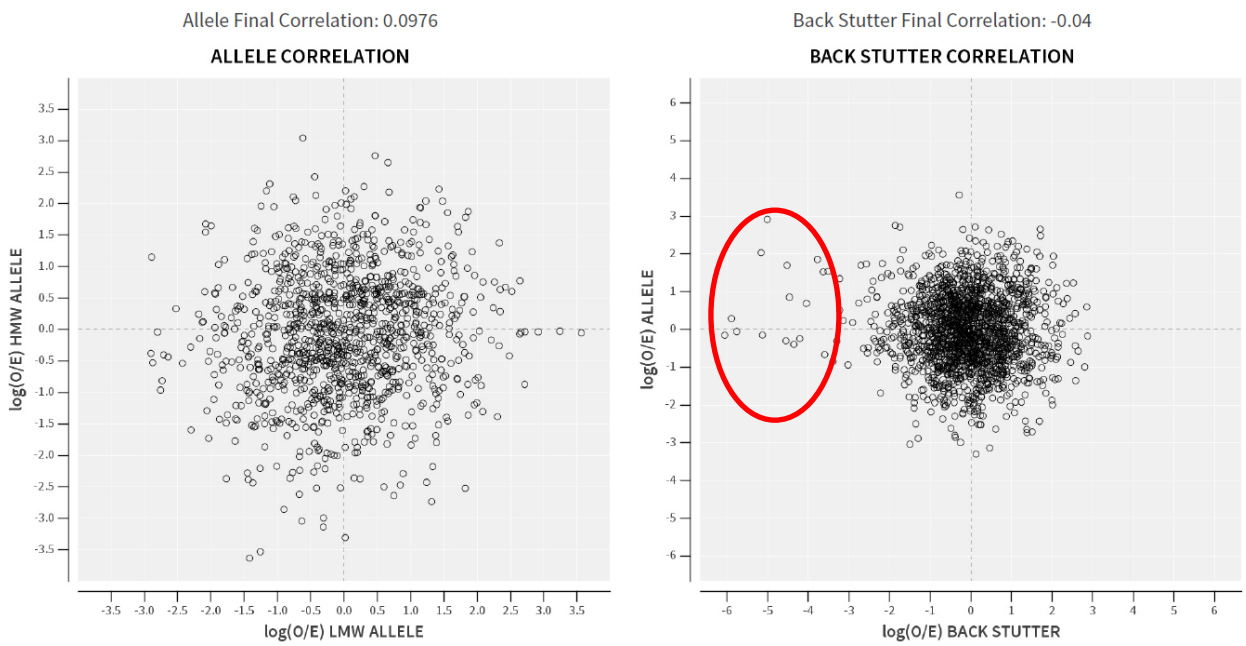
Figure 2: Log(*Hb*) versus APH for the single source profiles used in Model Maker at the PBSO laboratory



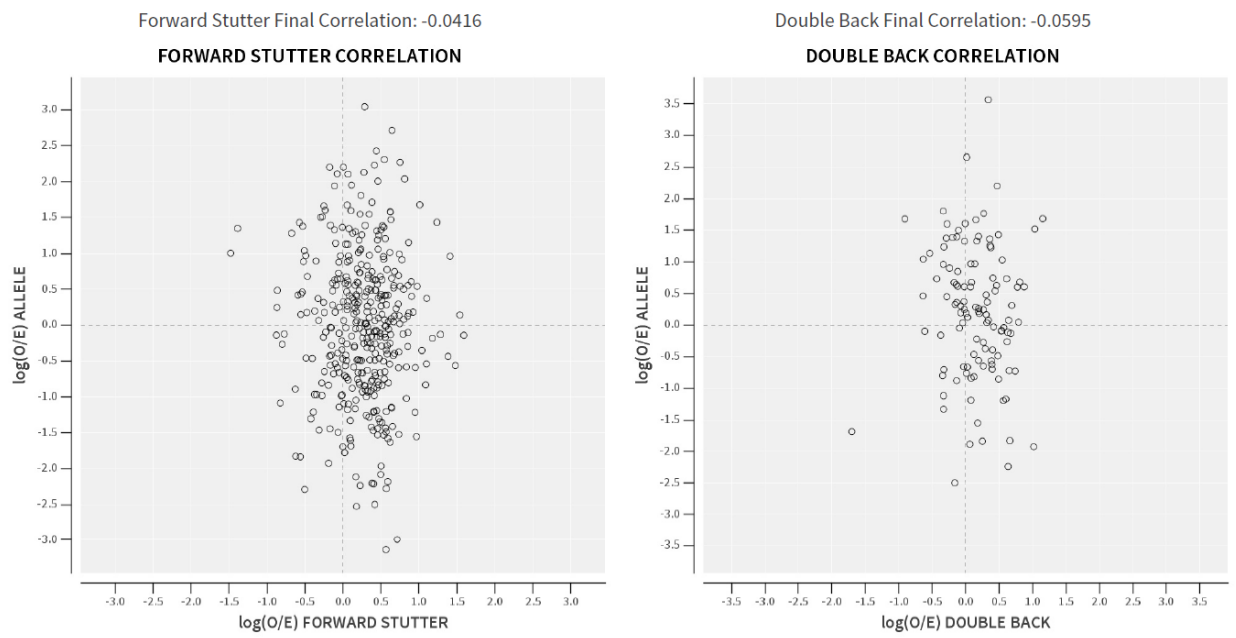
In **Figure** 3 we plot the correlation plots (combined data set) for Low Molecular Weight (LMW) versus High Molecular Weight (HMW) alleles, and stutter versus allele peaks for the Model Maker dataset. The

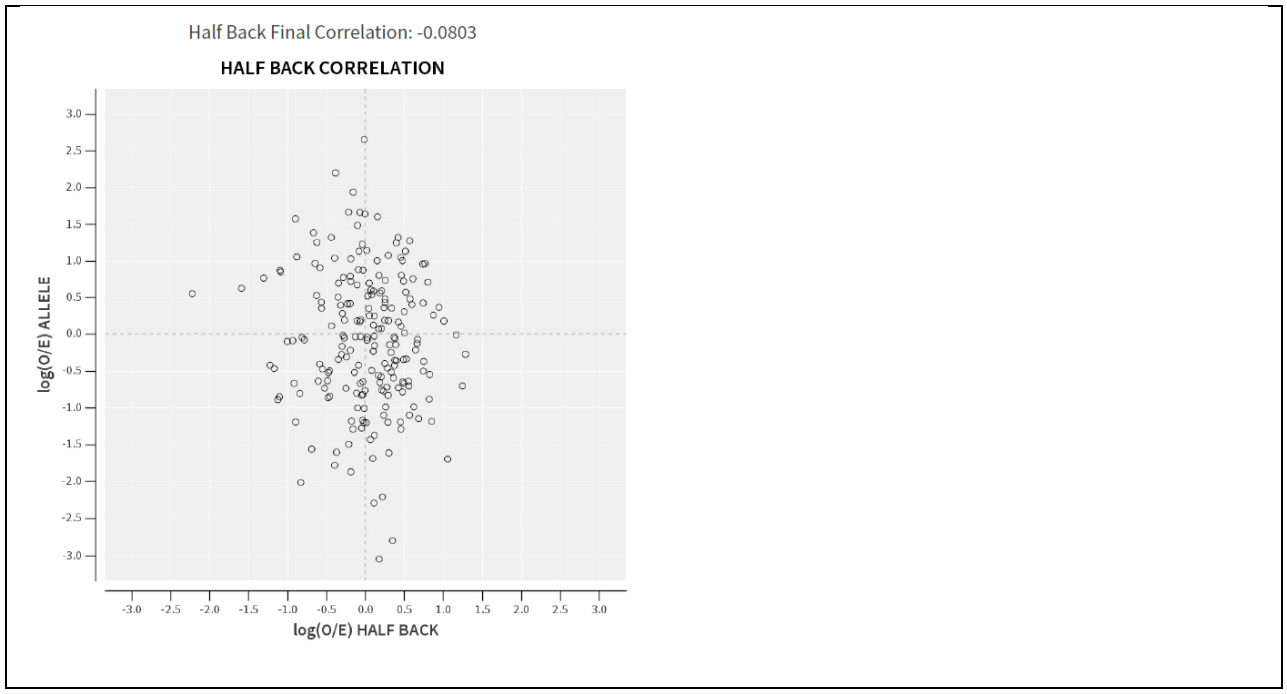
distribution of the points within the figures is as expected, with no observed correlation. No excessive outliers were observed.

Figure 3: PBSO's Fusion 6C correlation plots



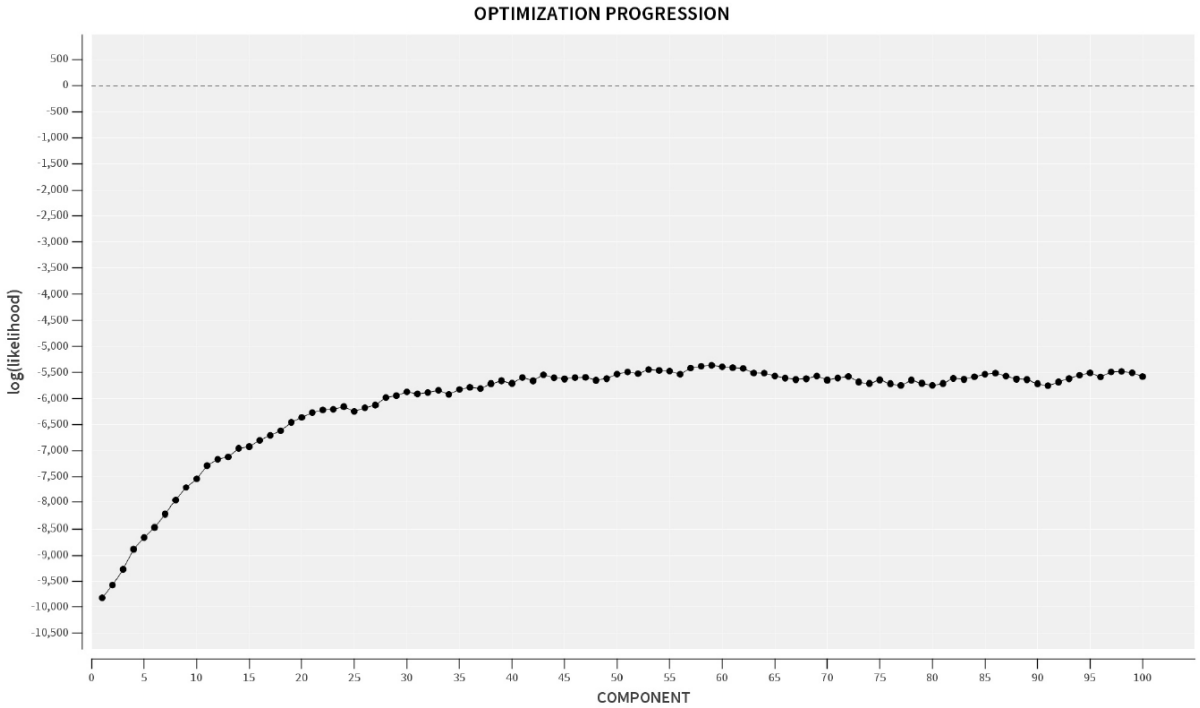
Note: the data points circled in the back stutter correlation plot are all from vWA 14 peaks. See earlier comment about iso-alleles having different stutter ratios.





The optimization progression from the Model Maker report was also as expected (see Figure 4).

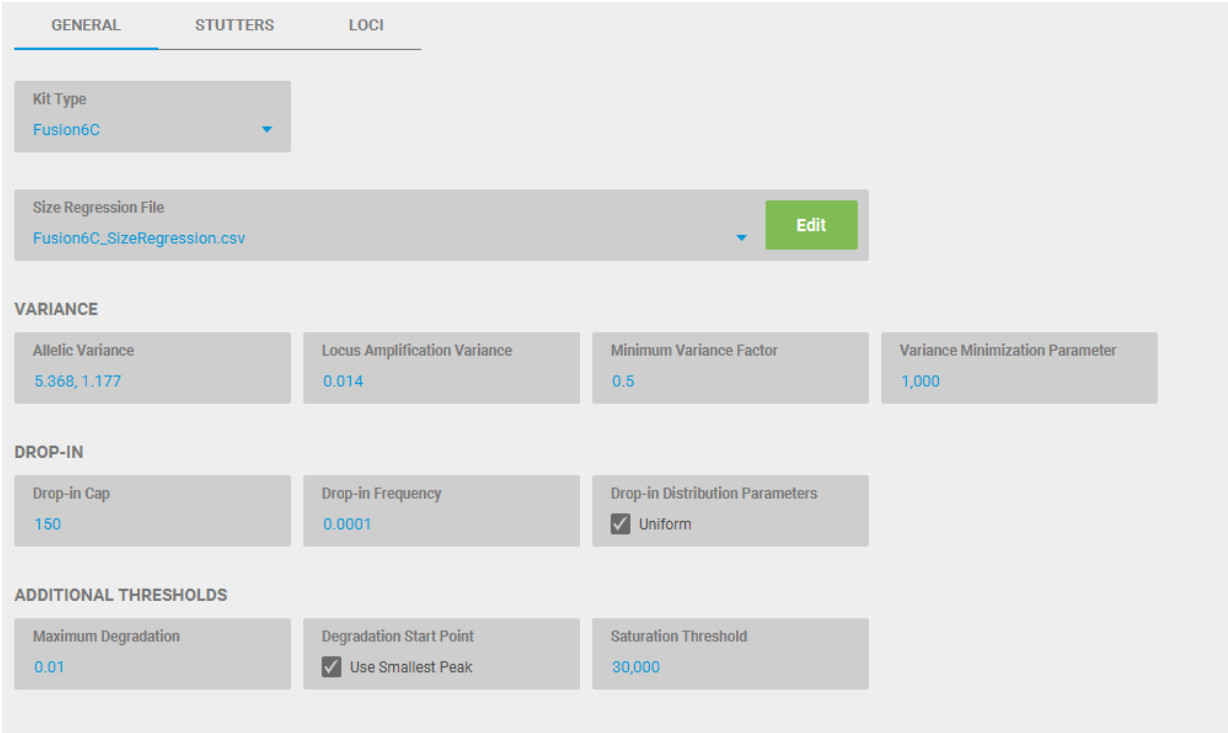
Figure 4: PBSO's Fusion 6C™ progression plot

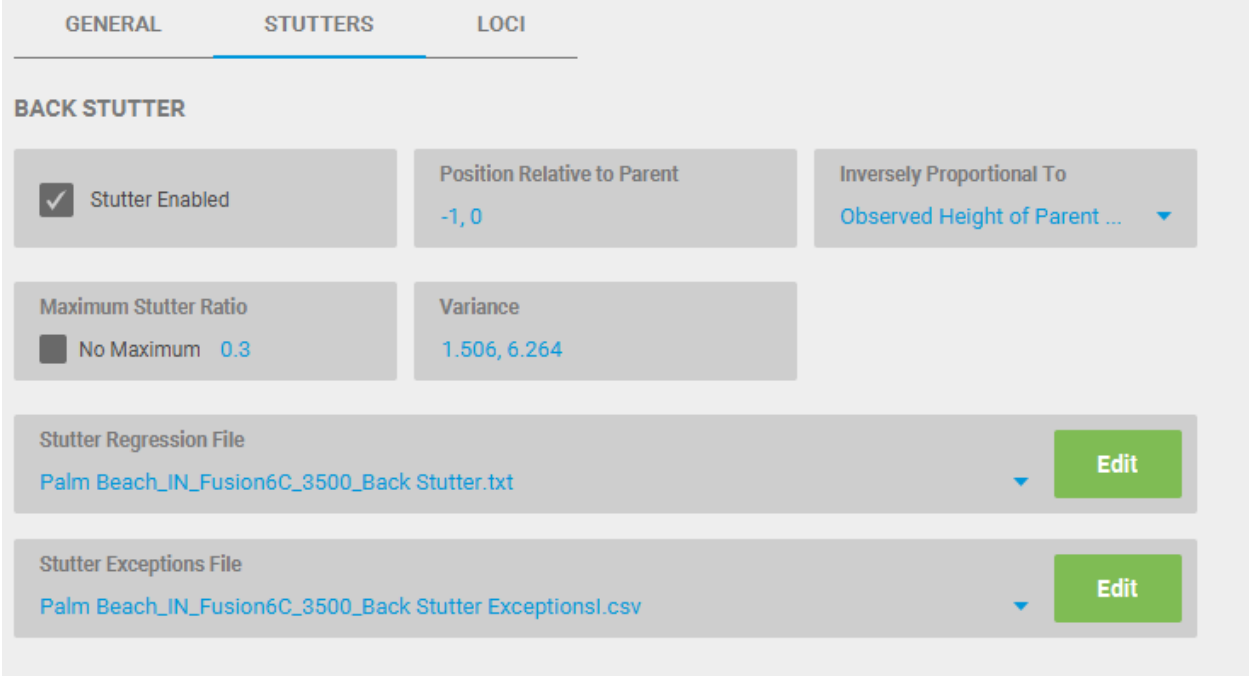


Conclusions

The recommended STRmix™ V2.6 default parameters for the interpretation of PBSO's 29 cycle, Fusion 6C profiles run on a 3500xl CE instrument are given in **Figure 5**. (**Palm Beach_Fusion_3500** with an AT set for each dye to align with PBSOs casework detection thresholds as indicated).

Figure 5: STRmix™ recommended default parameters for PBSO Fusion 6C profile interpretation





FORWARD STUTTER

☑ Stutter Enabled	Position Relative to Parent 1, 0	Inversely Proportional To Expected Height of Stutter ...
Maximum Stutter Ratio ☐ No Maximum 0.15	Variance 1.78, 3.61	

Stutter Regression File
Palm Beach_IN_Fusion6C_3500_Forward Stutter.txt — Edit

Stutter Exceptions File
Select a value — Edit

DOUBLE BACK

☑ Stutter Enabled	Position Relative to Parent -2, 0	Inversely Proportional To Expected Height of Stutter ...
Maximum Stutter Ratio ☐ No Maximum 0.05	Variance 2.583, 1.482	

Stutter Regression File
Palm Beach_IN_Fusion6C_3500_Double Back Stutter.txt — Edit

Stutter Exceptions File
Palm Beach_IN_Fusion6C_3500_Double Back Stutter Exceptions .csv — Edit

HALF BACK

☑ Stutter Enabled	Position Relative to Parent 0, -2	Inversely Proportional To Expected Height of Stutter ...
Maximum Stutter Ratio ☐ No Maximum 0.1	Variance 1.785, 1.153	

Stutter Regression File
Palm Beach_IN_Fusion6C_3500_Half Back Stutter.txt — Edit

Stutter Exceptions File
Select a value — Edit

GENERAL | STUTTERS | LOCI

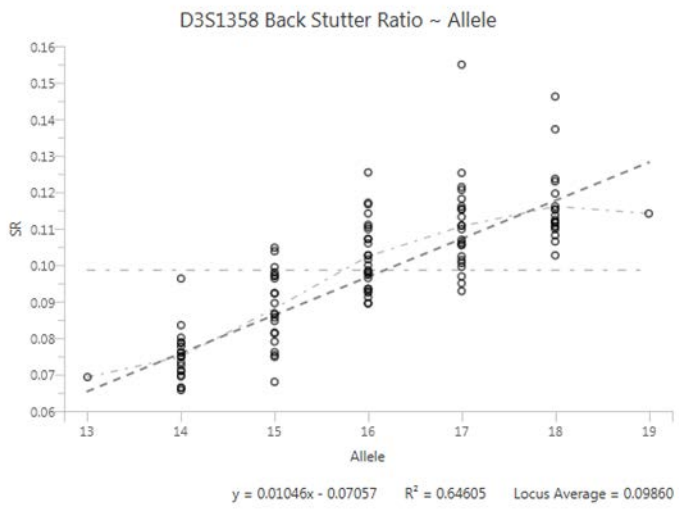
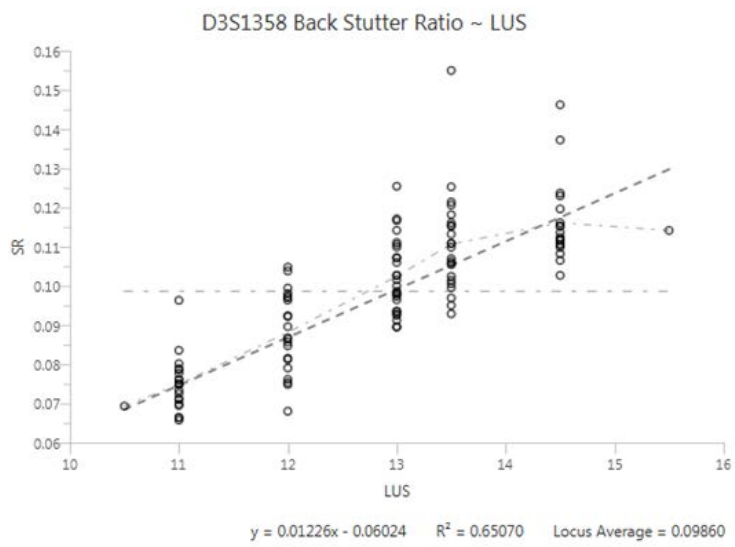
LOCUS NAME	GENDER?	REPEAT LENGTH	IGNORE?	DETECTION THRESHOLD	BACK STUTTER	FORWARD STUTTER	DOUBLE BACK	HALF BACK
AMEL	☑							
D3S1358	☐	4	☐	75	☑	☑	☑	☐
D1S1656	☐	4	☐	75	☑	☑	☑	☑
D2S441	☐	4	☐	75	☑	☑	☐	☐
D10S1248	☐	4	☐	75	☑	☑	☑	☐
D13S317	☐	4	☐	75	☑	☑	☐	☐
Penta E	☐	5	☐	75	☑	☑	☐	☐
D16S539	☐	4	☐	101	☑	☑	☑	☐
D18S51	☐	4	☐	101	☑	☑	☑	☐
D2S1338	☐	4	☐	101	☑	☑	☐	☐
CSF1PO	☐	4	☐	101	☑	☑	☐	☐
Penta D	☐	5	☐	101	☑	☑	☐	☐
TH01	☐	4	☐	60	☑	☑	☐	☐
vWA	☐	4	☐	60	☑	☑	☐	☐

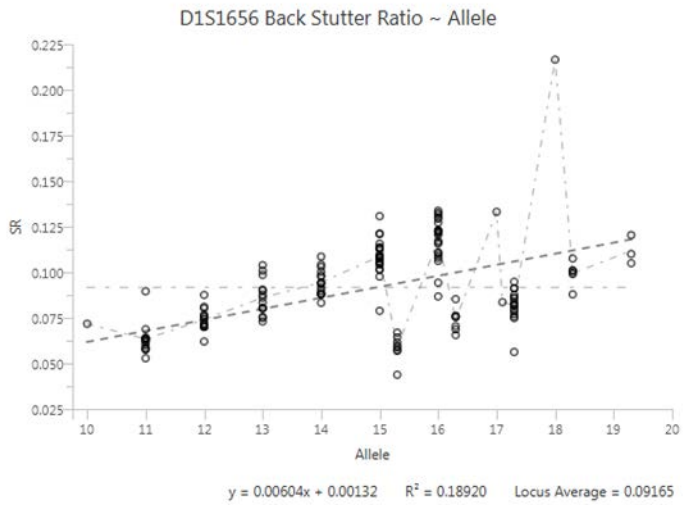
D21S11	☐	4	☐	60	☑	☑	☑	☐
D7S820	☐	4	☐	60	☑	☑	☐	☐
D5S818	☐	4	☐	60	☑	☑	☐	☐
TPOX	☐	4	☐	60	☑	☑	☐	☐
D8S1179	☐	4	☐	69	☑	☑	☑	☐
D12S391	☐	4	☐	69	☑	☑	☐	☐
D19S433	☐	4	☐	69	☑	☑	☑	☐
SE33	☐	4	☐	69	☑	☑	☑	☑
D22S1045	☐	3	☐	69	☑	☑	☑	☐
DYS391	☐	4	☑	56	☑	☑	☐	☐
FGA	☐	4	☐	56	☑	☑	☐	☐
DYS576	☐	4	☑	56	☑	☑	☐	☐
DYS570	☐	4	☑	56	☑	☑	☐	☐

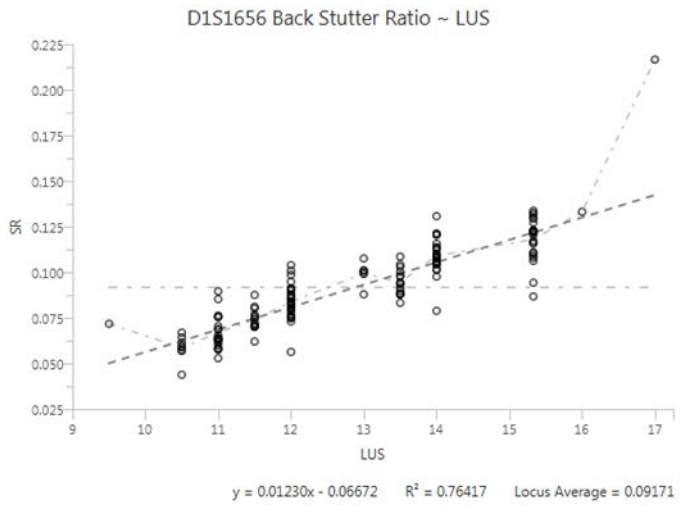
References

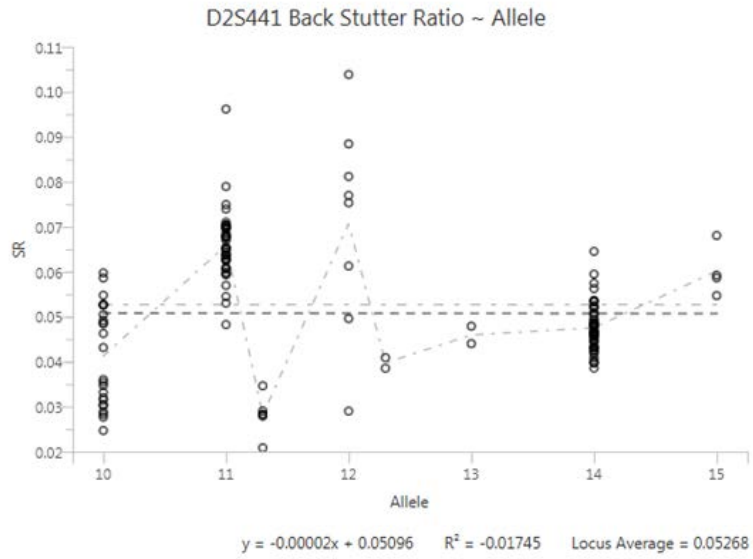
[1] J.-A. Bright, D. Taylor, J.M. Curran, J.S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, Forensic Science International: Genetics 7(2) (2013) 296-304.
[2] C. Brookes, J.-A. Bright, S. Harbison, J. Buckleton, Characterising stutter in forensic STR multiplexes, Forensic Science International: Genetics 6(1) (2012) 58-63.
[3] P.S. Walsh, N.J. Fildes, R. Reynolds, Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA., Nucleic Acids Res. 24 (1996) 2807-2812.
[4] J.M. Butler, D.J. Reeder, Short Tandem Repeat DNA Internet DataBase. (www.cstl.nist.gov/biotech/strbase).
[5] C.M. Ruitberg, D.J. Reeder, J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community., Nucleic Acids Research 29(1) (2001) 320 - 322.
[6] J.-A. Bright, E. Huizing, L. Melia, J. Buckleton, Determination of the variables affecting mixed MiniFiler™ DNA profiles, Forensic Science International: Genetics 5(5) (2011) 381-385.
[7] J.-A. Bright, J. Turkington, J. Buckleton, Examination of the variability in mixed DNA profile parameters for the Identifiler multiplex, Forensic Science International: Genetics 4(2) (2009) 111-114.

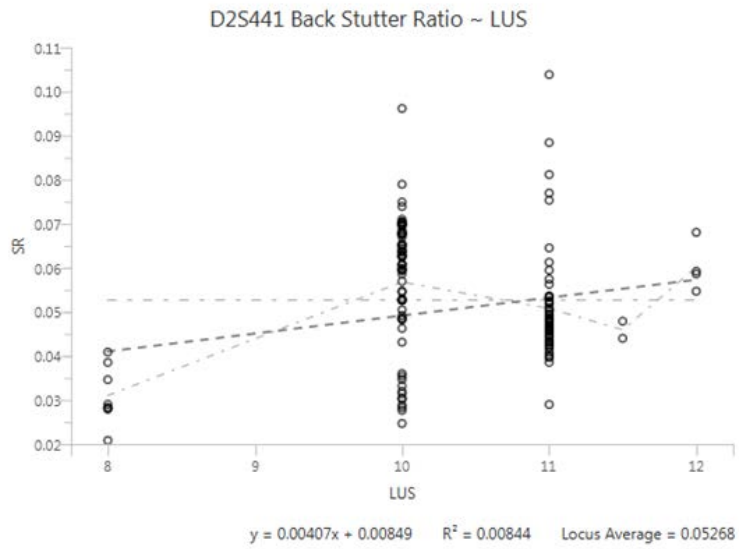
Appendix 1: Back stutter (-1,0) versus allele designation (left pane) and versus LUS (right pane), where appropriate. The dashed line is the regression by allele (or LUS). The dash-dot line is the average SR per allele. The horizontal dashed line is locus average SR.

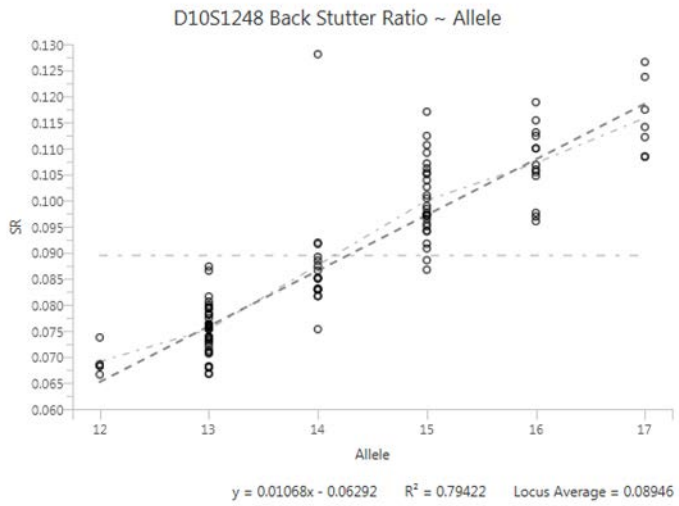
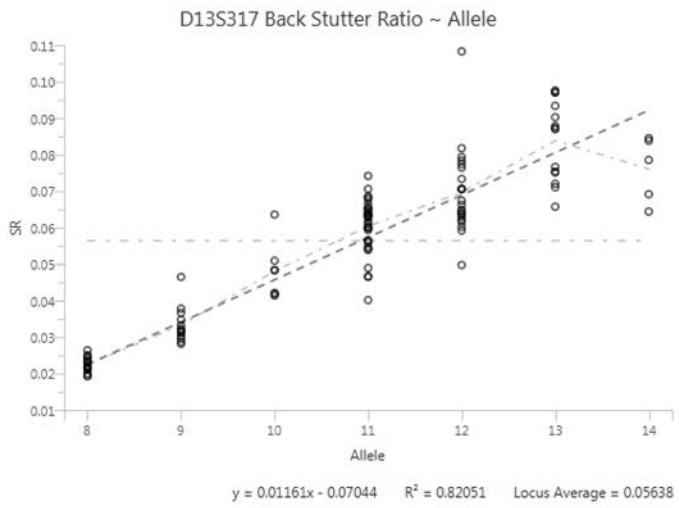
SR ~ Allele	SR ~ LUS
	

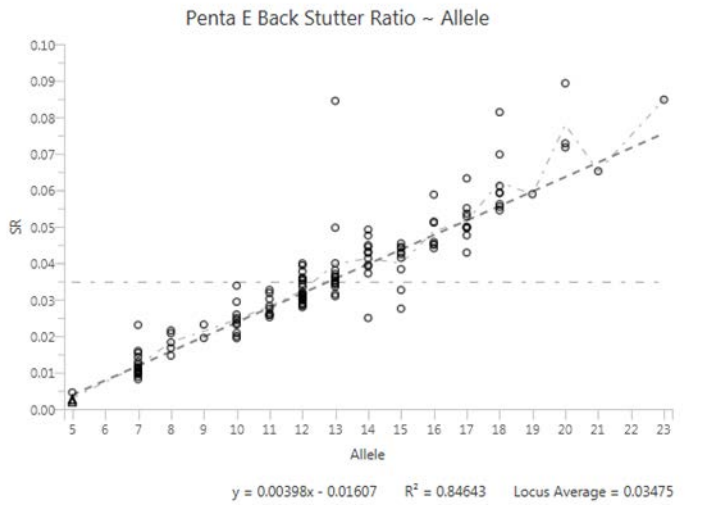
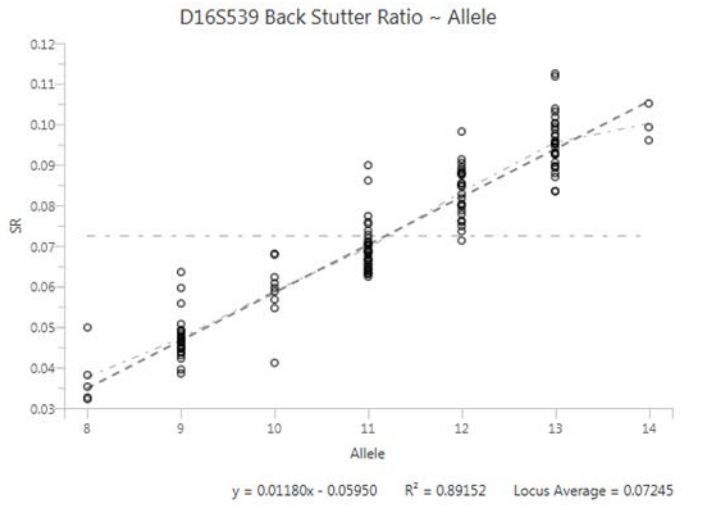

D1S1656 Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00604x + 0.00132 R² = 0.18920 Locus Average = 0.09165

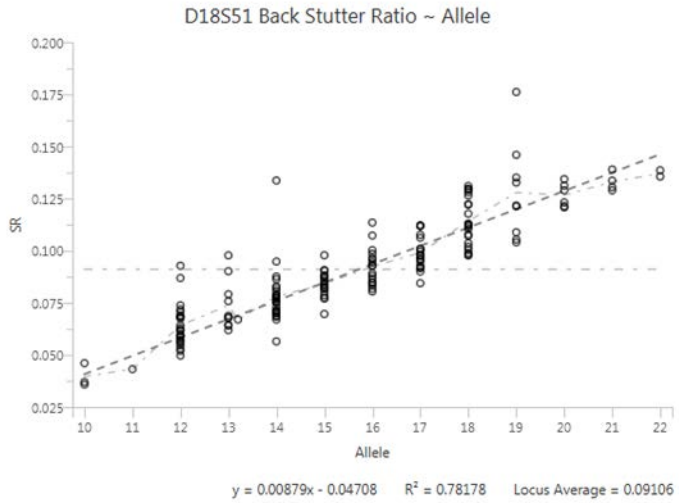

D1S1656 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.01230x - 0.06672 R² = 0.76417 Locus Average = 0.09171

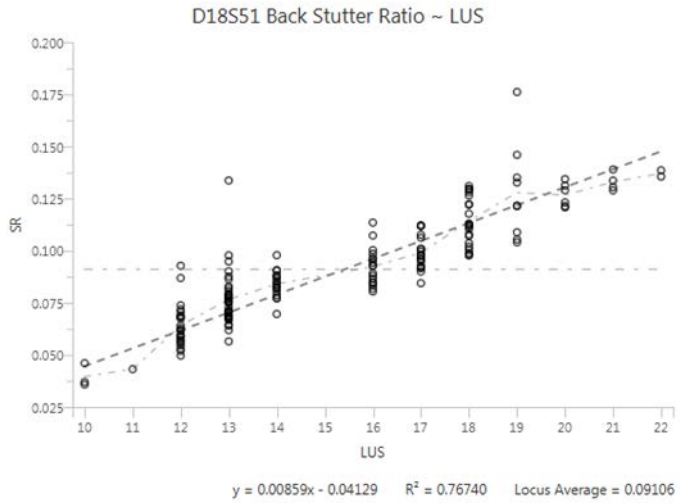

D2S441 Back Stutter Ratio ~ Allele
SR
Allele
y = -0.00002x + 0.05096 R² = -0.01745 Locus Average = 0.05268

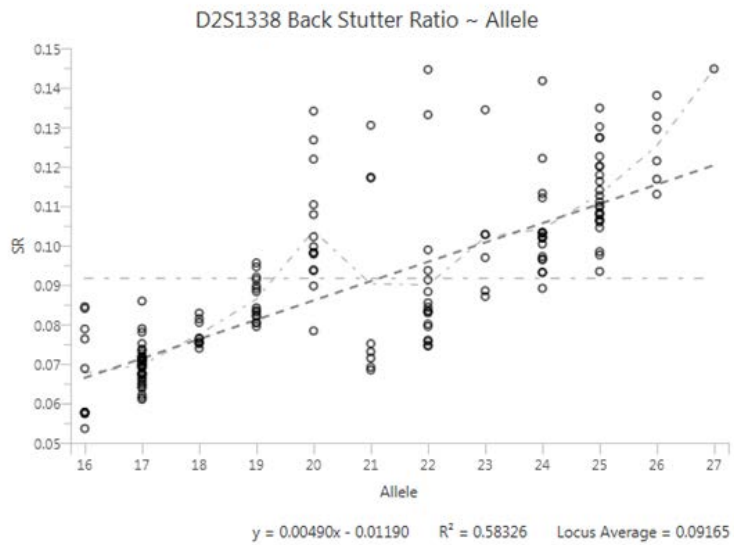

D2S441 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.00407x + 0.00849 R² = 0.00844 Locus Average = 0.05268

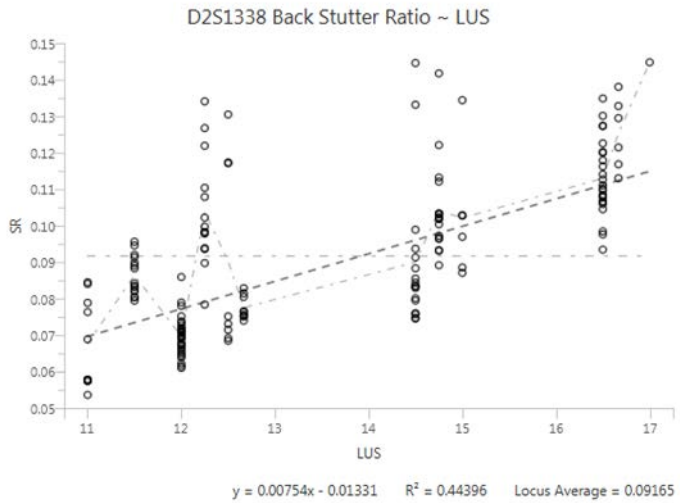
	N/A. Allele and LUS are the same
	N/A. Allele and LUS are the same

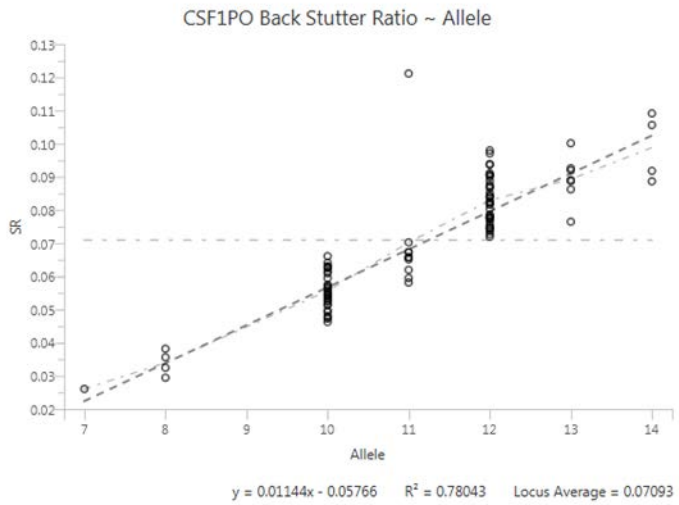
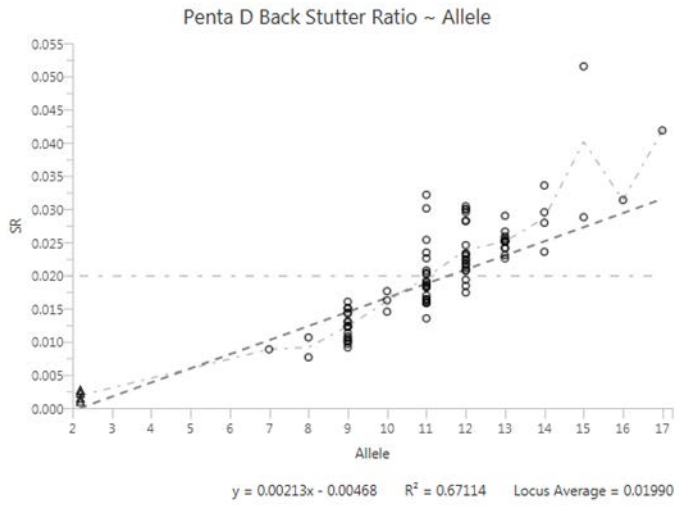
	N/A. Allele and LUS are the same
	N/A. Allele and LUS are the same

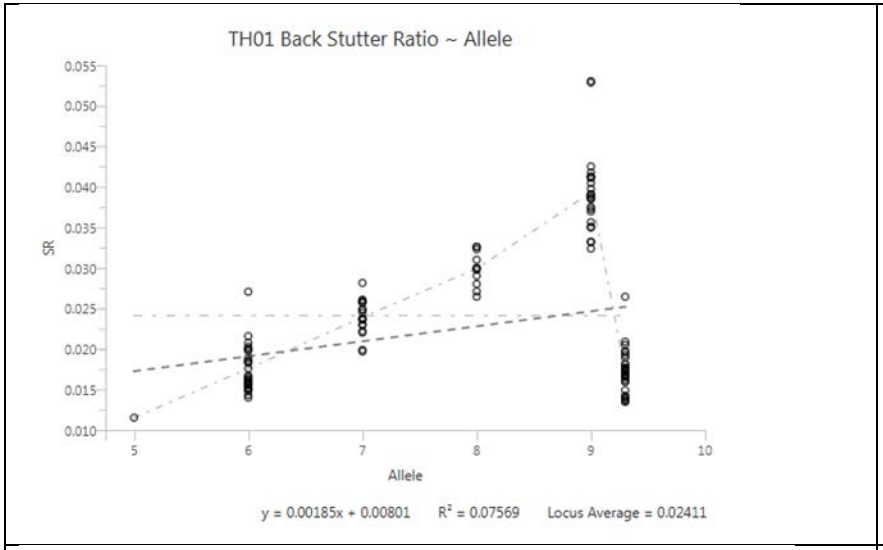

D18S51 Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00879x - 0.04708 R² = 0.78178 Locus Average = 0.09106

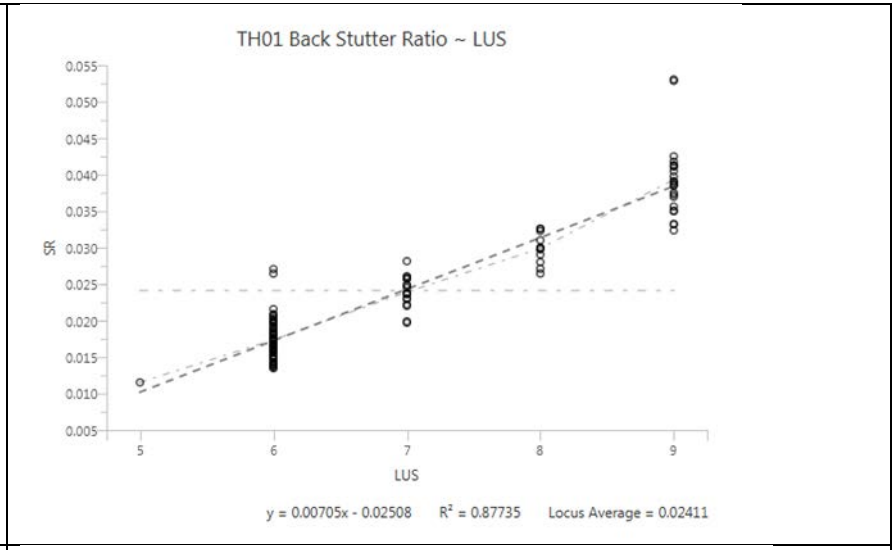

D18S51 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.00859x - 0.04129 R² = 0.76740 Locus Average = 0.09106

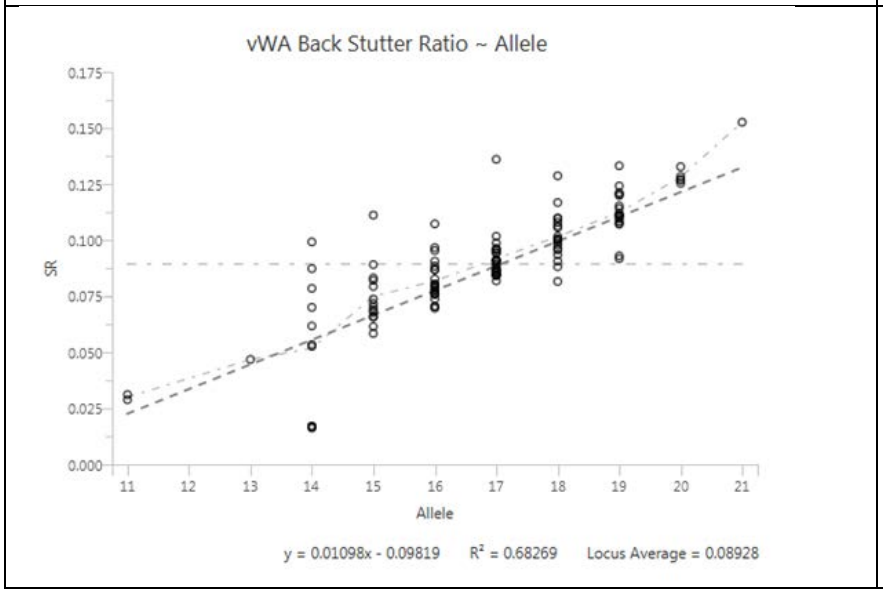

D2S1338 Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00490x - 0.01190 R² = 0.58326 Locus Average = 0.09165

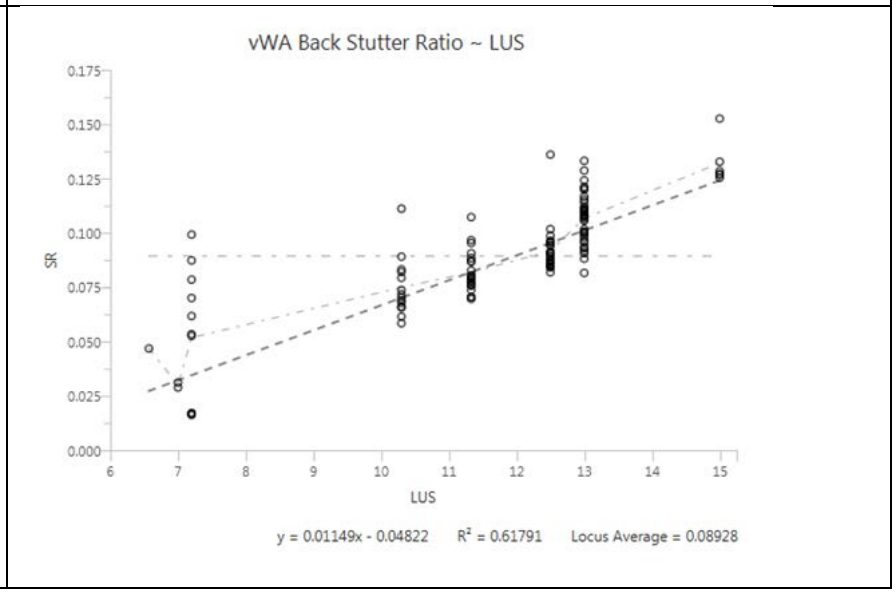

D2S1338 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.00754x - 0.01331 R² = 0.44396 Locus Average = 0.09165

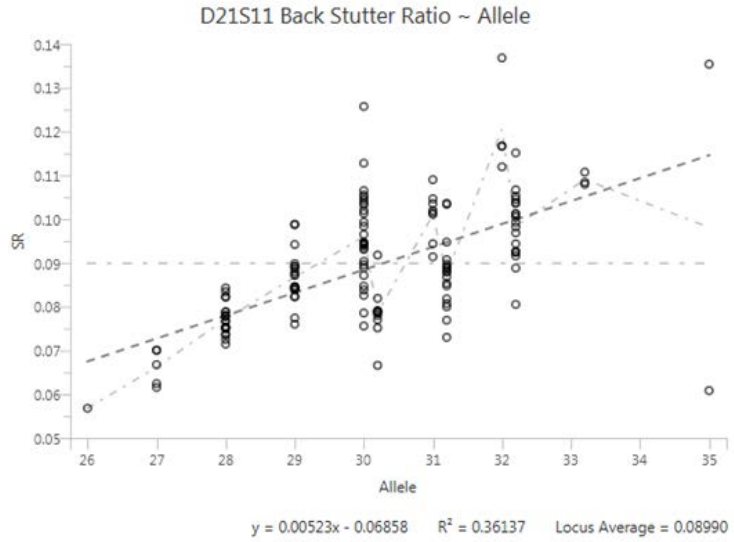
	N/A. Allele and LUS are the same
	N/A. Allele and LUS are the same

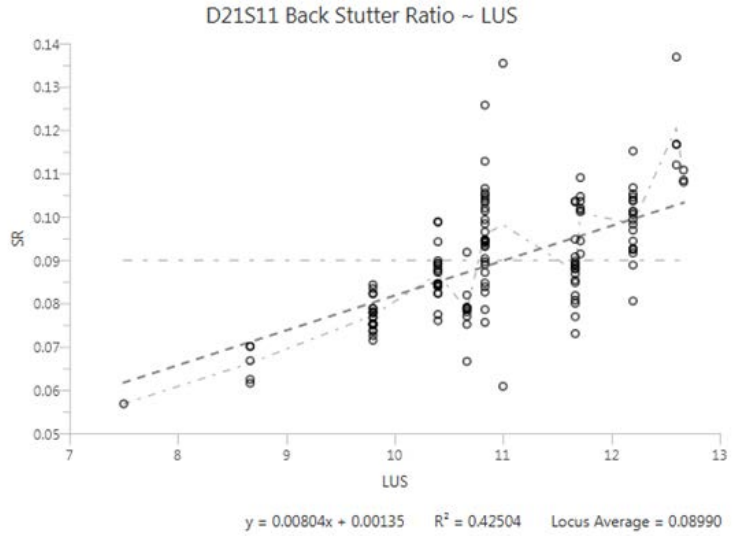

TH01 Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00185x + 0.00801 R² = 0.07569 Locus Average = 0.02411

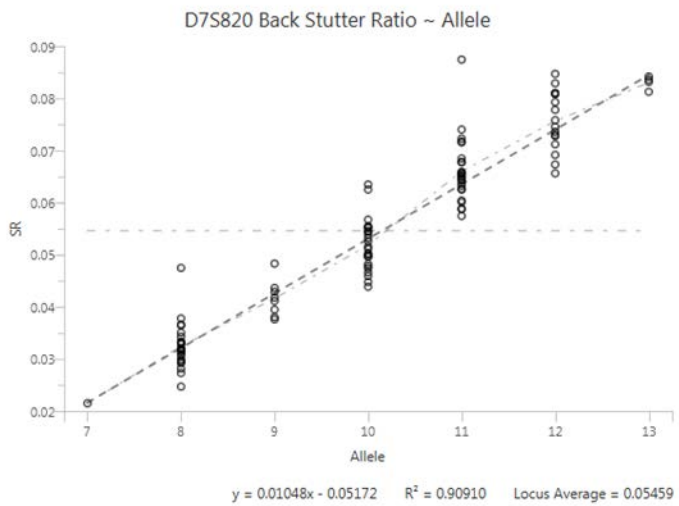

TH01 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.00705x - 0.02508 R² = 0.87735 Locus Average = 0.02411


vWA Back Stutter Ratio ~ Allele
SR
Allele
y = 0.01098x - 0.09819 R² = 0.68269 Locus Average = 0.08928

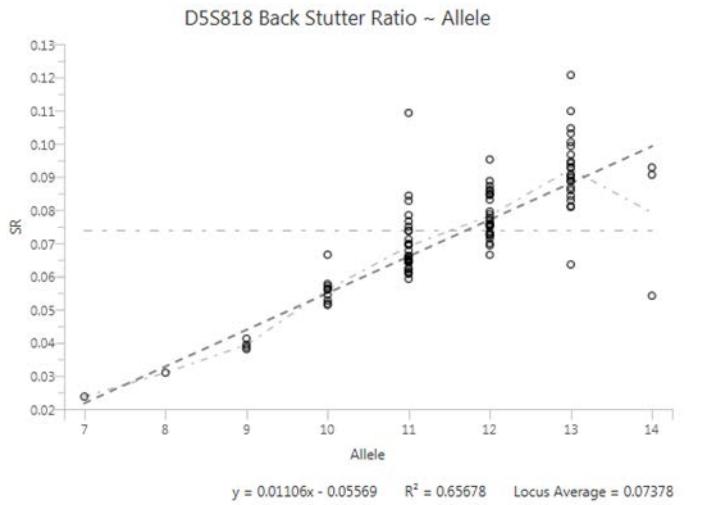
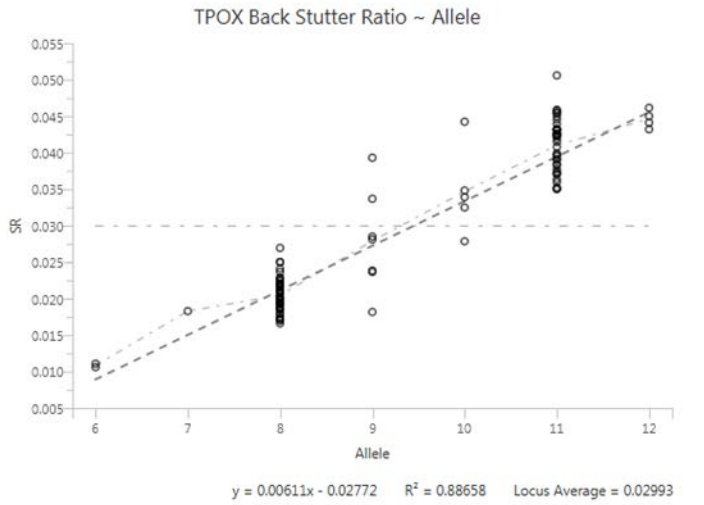

vWA Back Stutter Ratio ~ LUS
SR
LUS
y = 0.01149x - 0.04822 R² = 0.61791 Locus Average = 0.08928

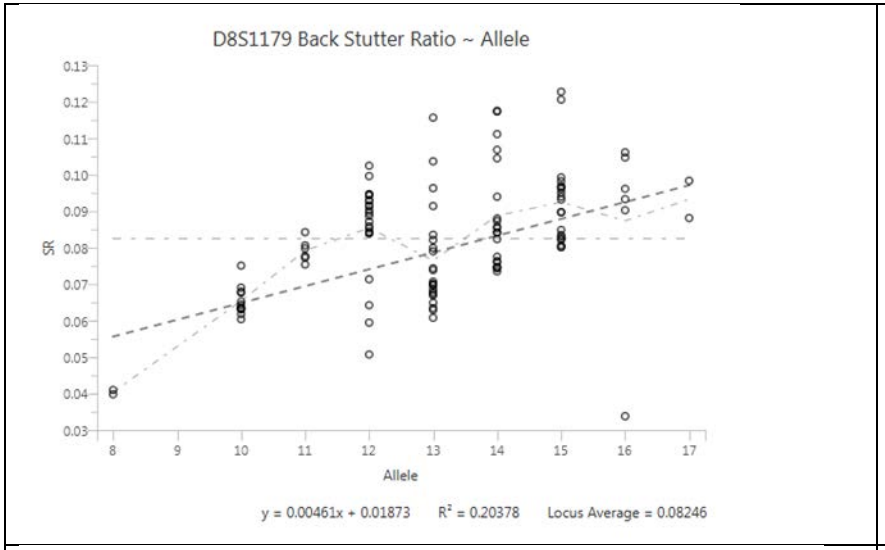


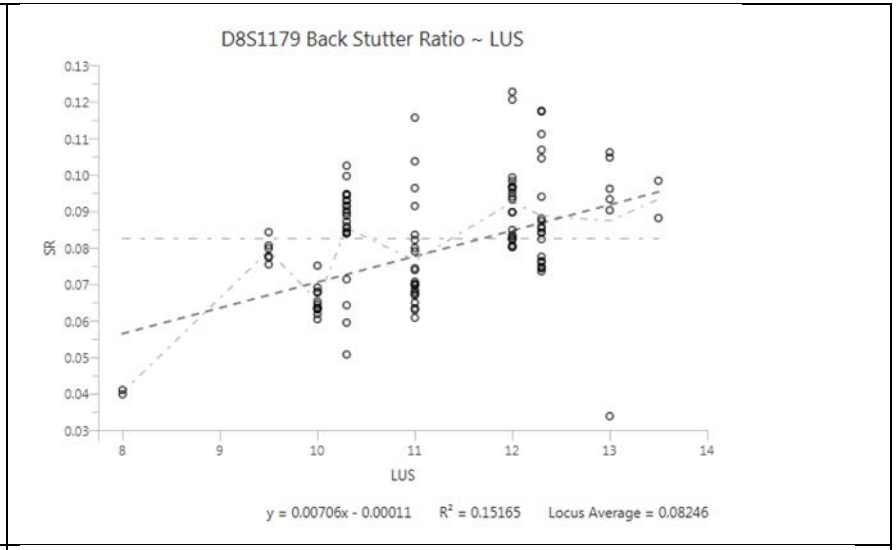


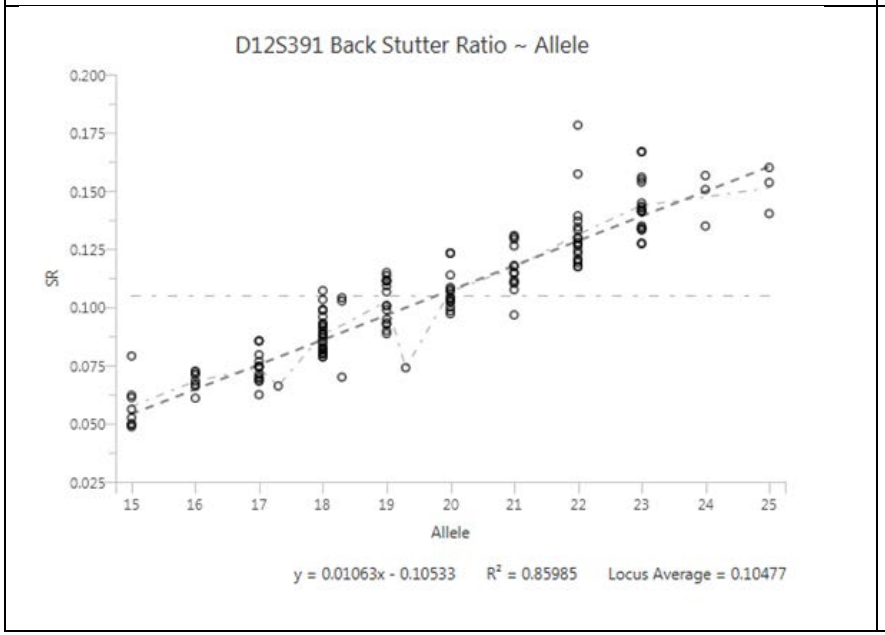


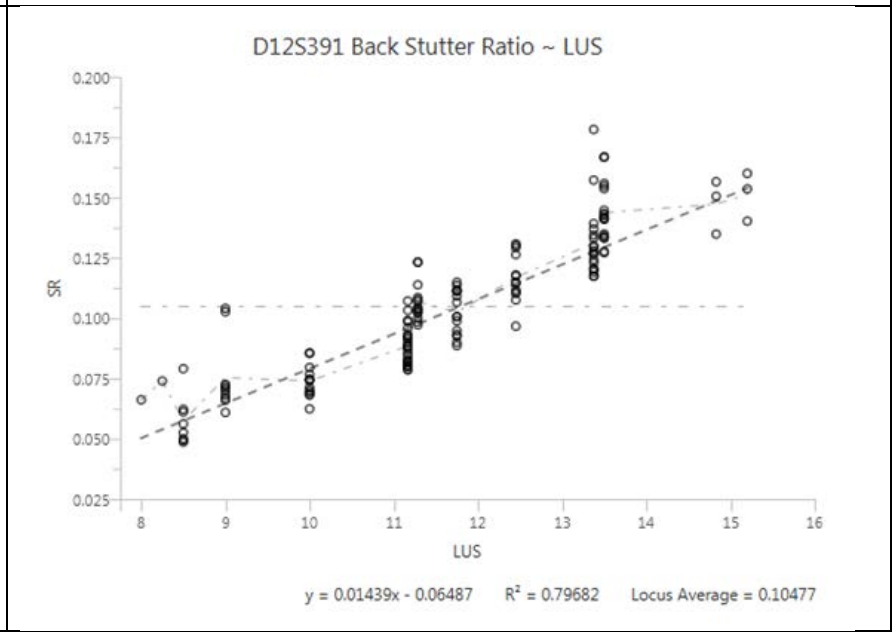
N/A. Allele and LUS are the same

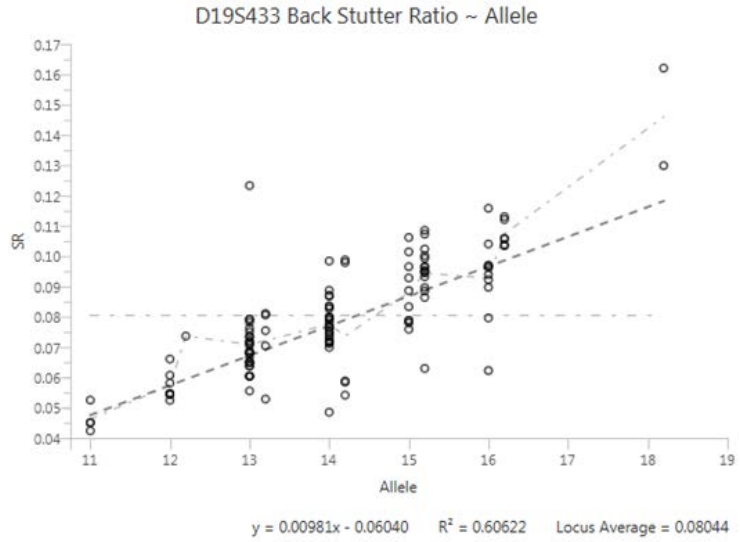
	N/A. Allele and LUS are the same
	N/A. Allele and LUS are the same

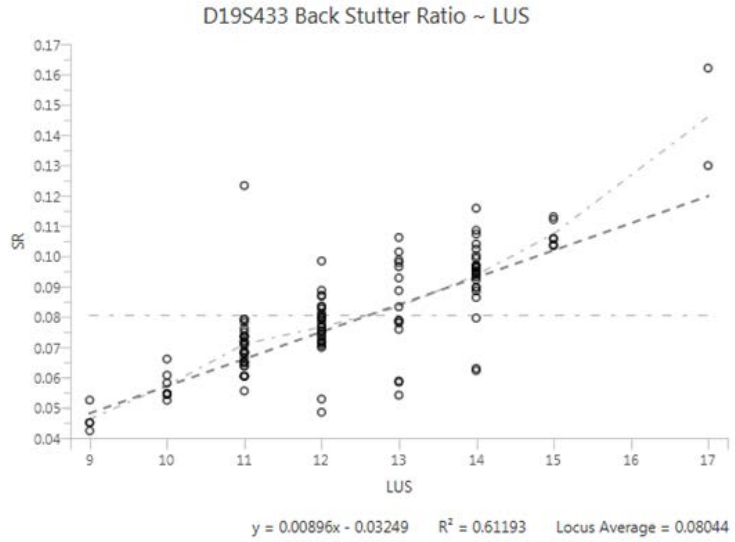

D8S1179 Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00461x + 0.01873 R² = 0.20378 Locus Average = 0.08246

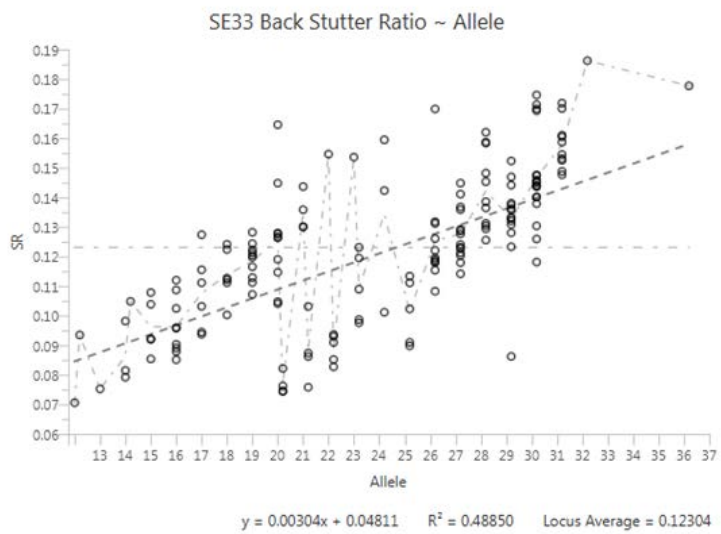

D8S1179 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.00706x - 0.00011 R² = 0.15165 Locus Average = 0.08246

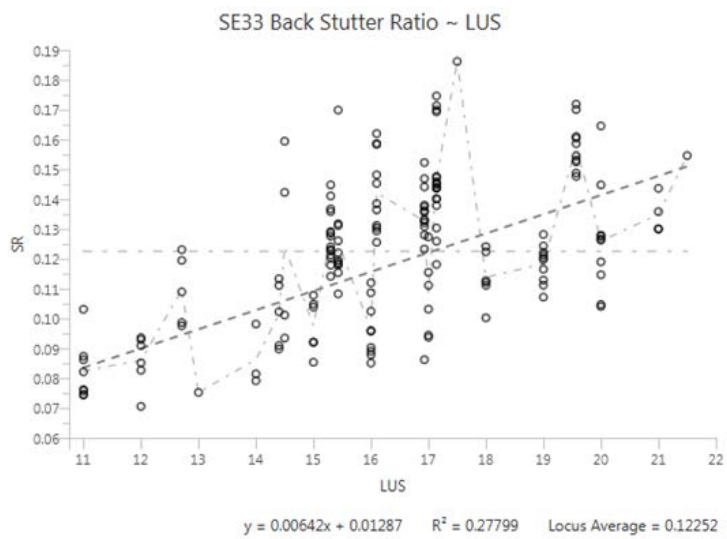

D12S391 Back Stutter Ratio ~ Allele
SR
Allele
y = 0.01063x - 0.10533 R² = 0.85985 Locus Average = 0.10477

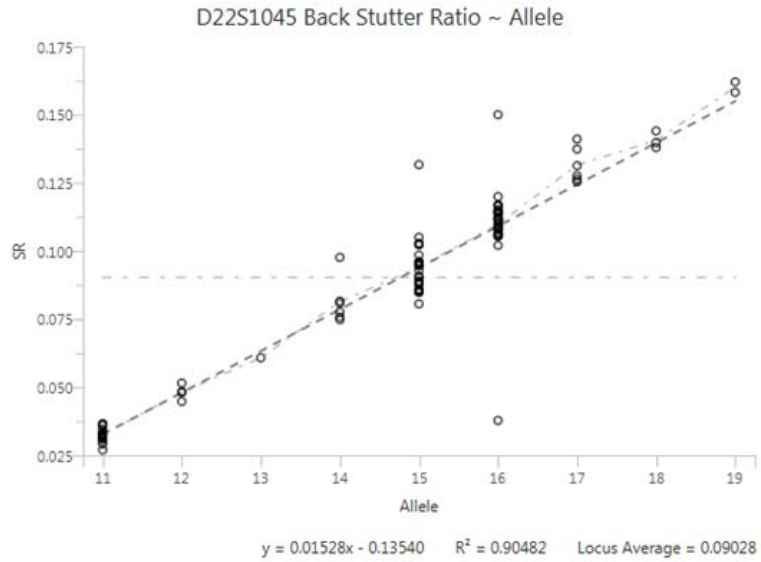

D12S391 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.01439x - 0.06487 R² = 0.79682 Locus Average = 0.10477

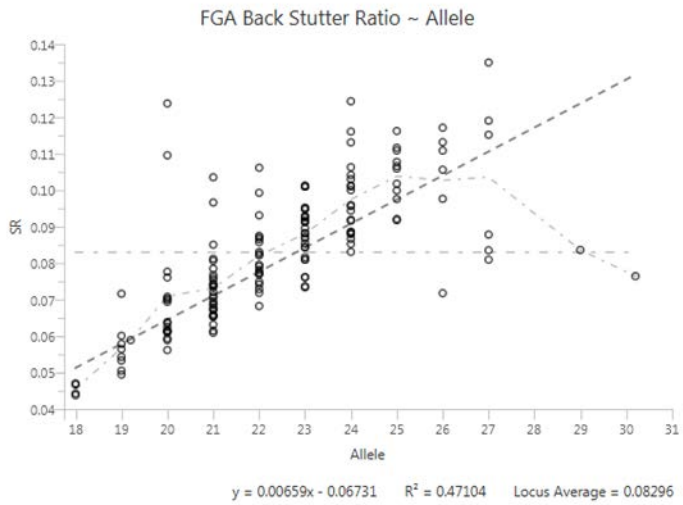

D19S433 Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00981x - 0.06040 R² = 0.60622 Locus Average = 0.08044

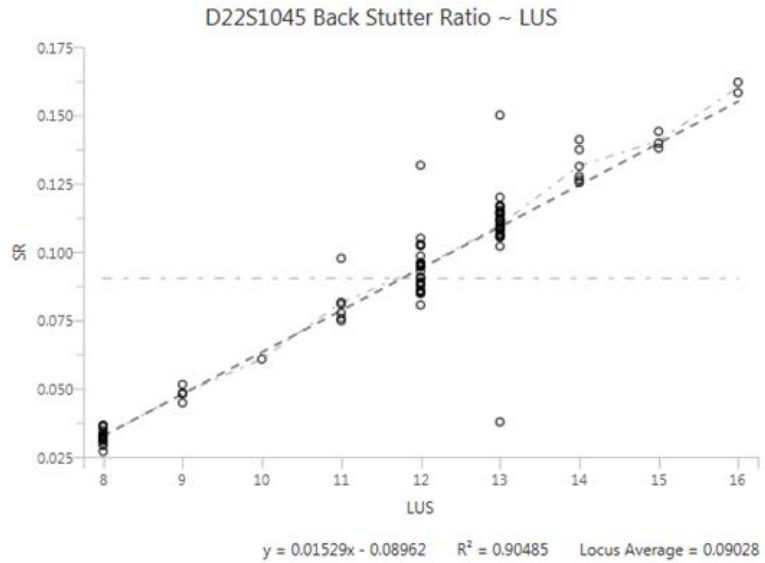

D19S433 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.00896x - 0.03249 R² = 0.61193 Locus Average = 0.08044


SE33 Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00304x + 0.04811 R² = 0.48850 Locus Average = 0.12304

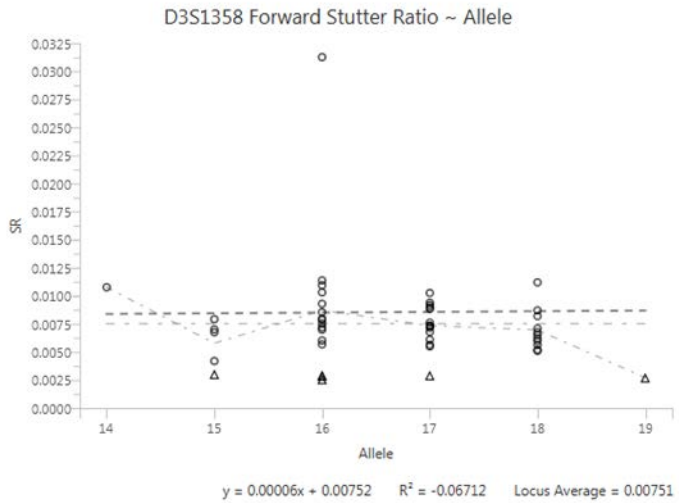

SE33 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.00642x + 0.01287 R² = 0.27799 Locus Average = 0.12252

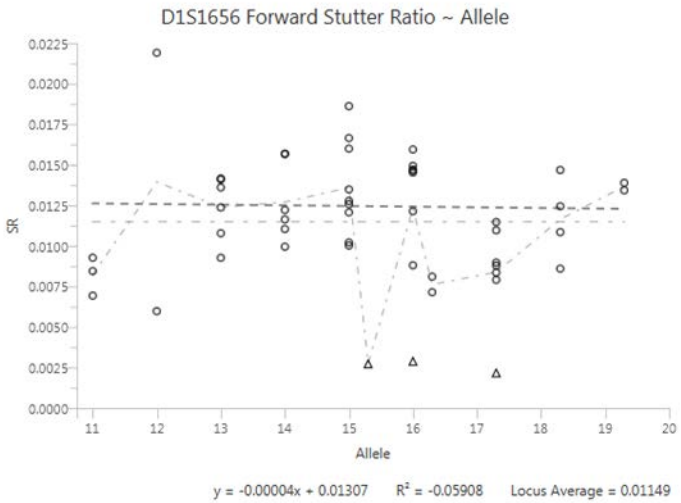

D22S1045 Back Stutter Ratio ~ Allele
SR
Allele
y = 0.01528x - 0.13540 R² = 0.90482 Locus Average = 0.09028

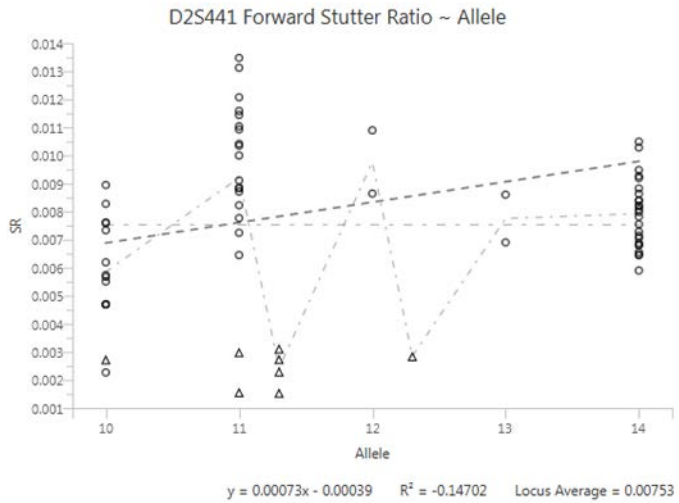

FGA Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00659x - 0.06731 R² = 0.47104 Locus Average = 0.08296

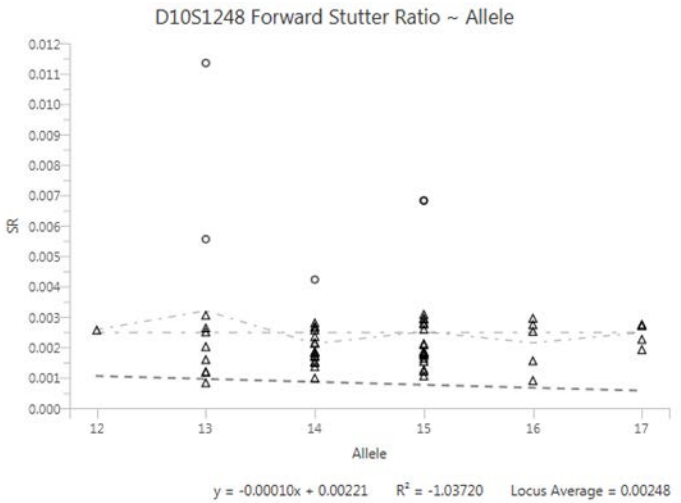

D22S1045 Back Stutter Ratio ~ LUS
SR
LUS
y = 0.01529x - 0.08962 R² = 0.90485 Locus Average = 0.09028

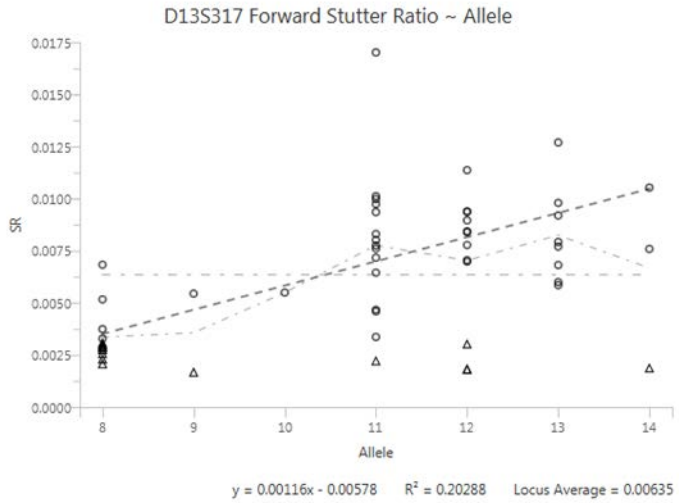
Appendix 2: Forward stutter (1,0) versus allele designation. The dashed line is the regression by allele. The dash-dot line is the average SR per allele. The horizontal dashed line is locus average SR. Please note in most instances the average of the data for the entire locus is used.

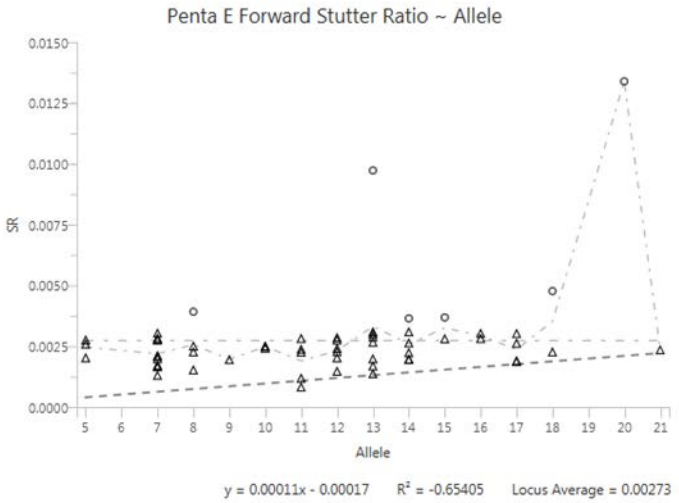


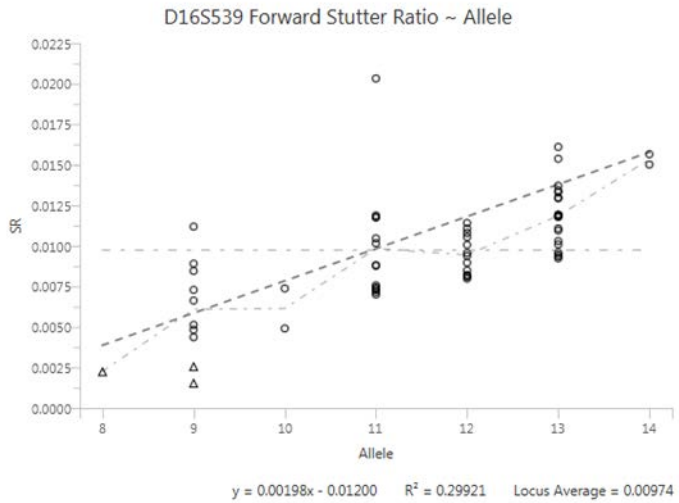


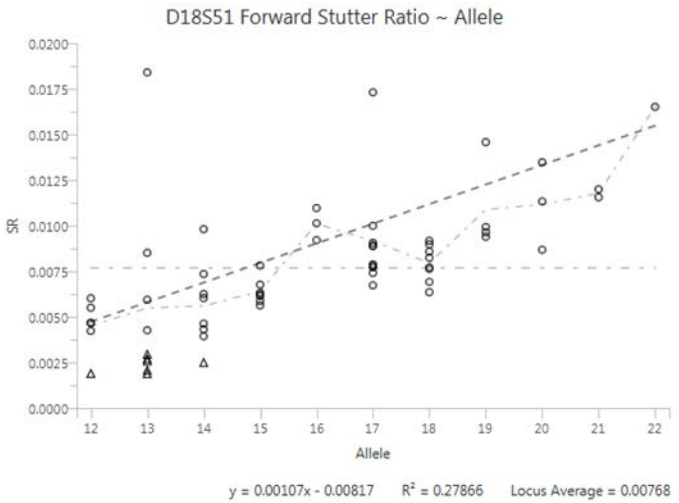


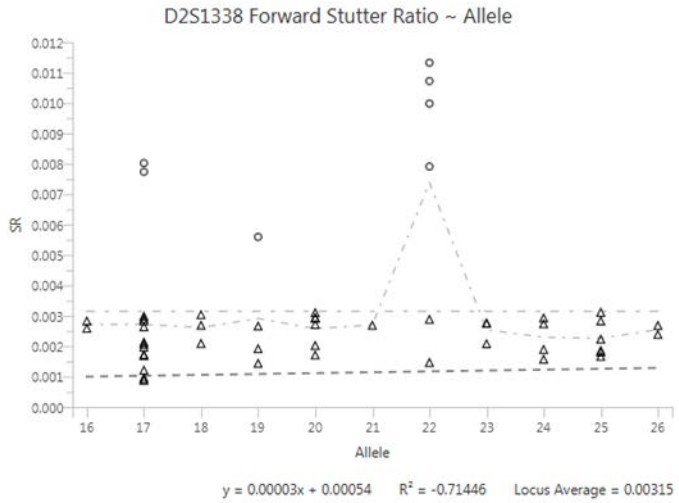


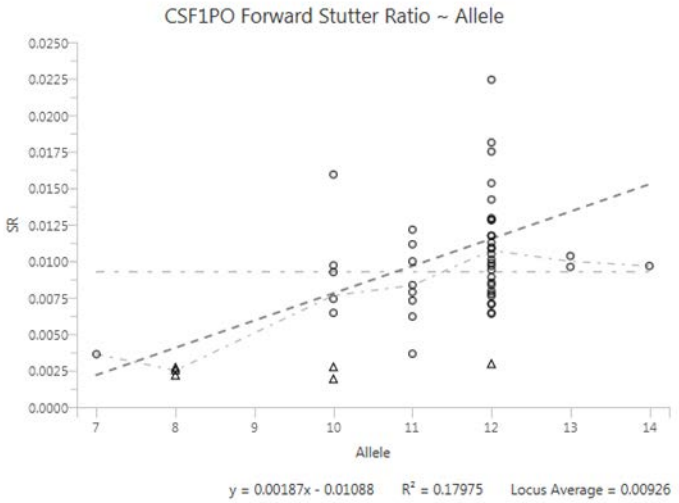

D13S317 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00116x - 0.00578 R² = 0.20288 Locus Average = 0.00635

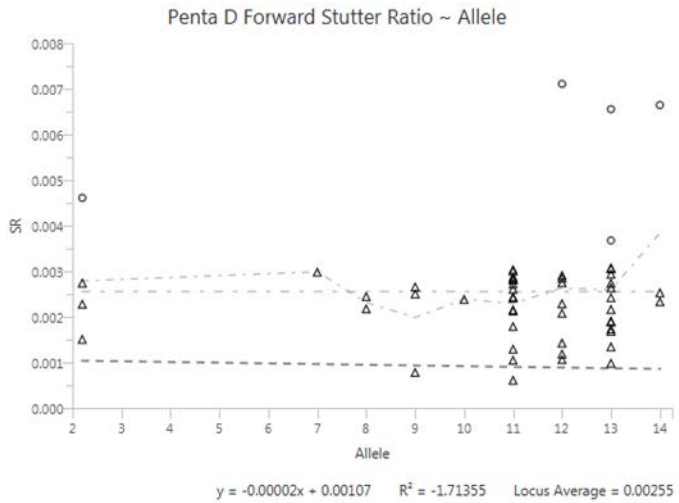

Penta E Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00011x - 0.00017 R² = -0.65405 Locus Average = 0.00273

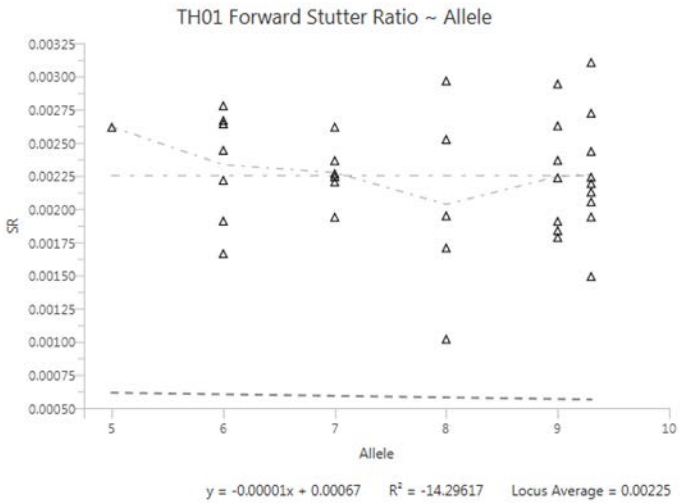

D16S539 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00198x - 0.01200 R² = 0.29921 Locus Average = 0.00974

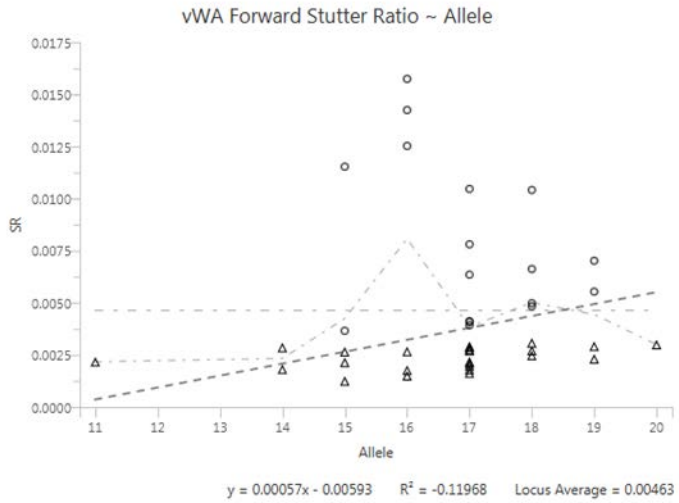

D18S51 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00107x - 0.00817 R² = 0.27866 Locus Average = 0.00768

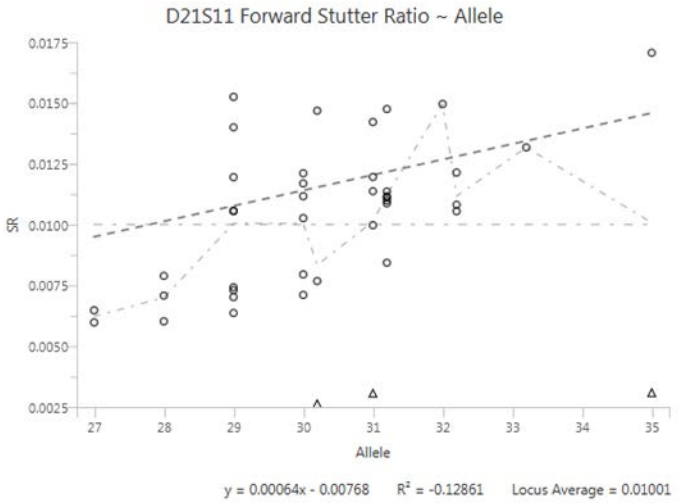

D2S1338 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00003x + 0.00054 R² = -0.71446 Locus Average = 0.00315

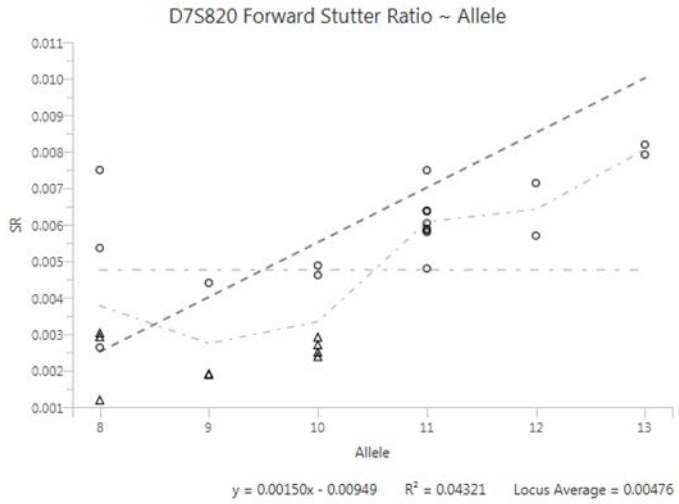

CSF1PO Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00187x - 0.01088 R² = 0.17975 Locus Average = 0.00926

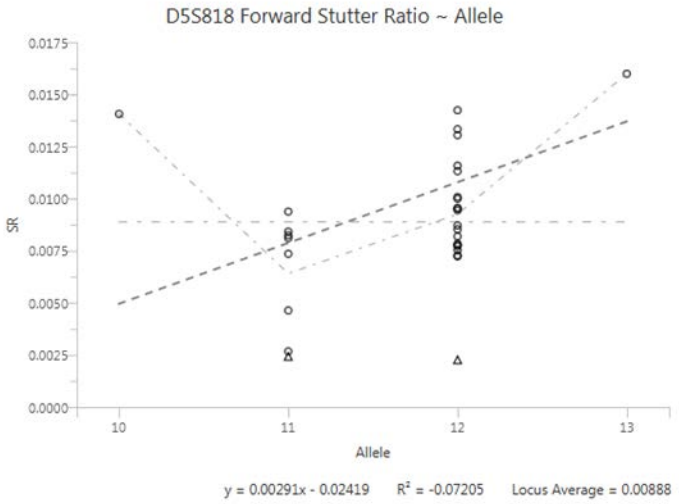

Penta D Forward Stutter Ratio ~ Allele
SR
Allele
y = -0.00002x + 0.00107 R² = -1.71355 Locus Average = 0.00255

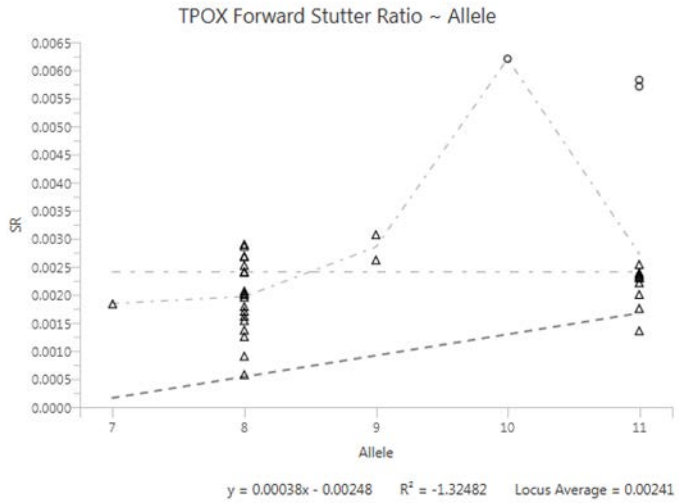

TH01 Forward Stutter Ratio ~ Allele
SR
Allele
y = -0.00001x + 0.00067 R² = -14.29617 Locus Average = 0.00225

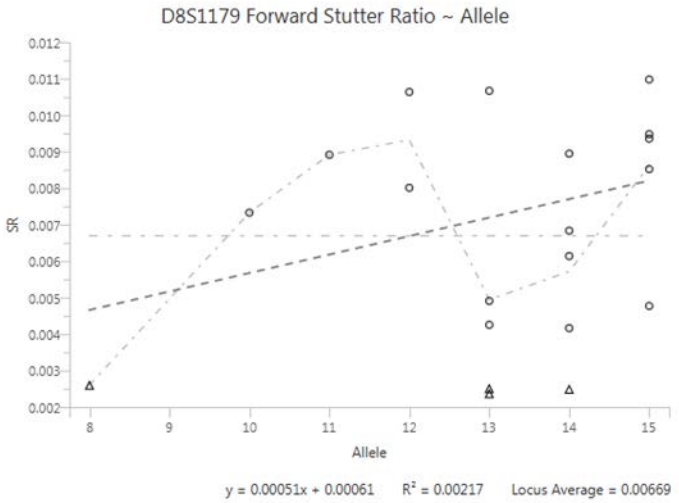

vWA Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00057x - 0.00593 R² = -0.11968 Locus Average = 0.00463

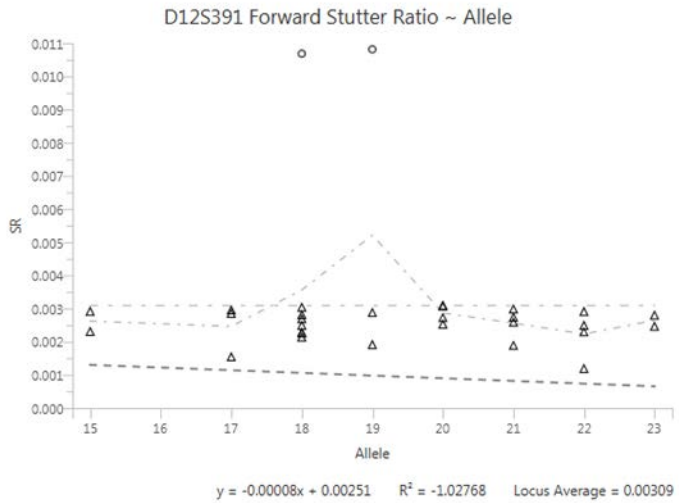

D21S11 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00064x - 0.00768 R² = -0.12861 Locus Average = 0.01001

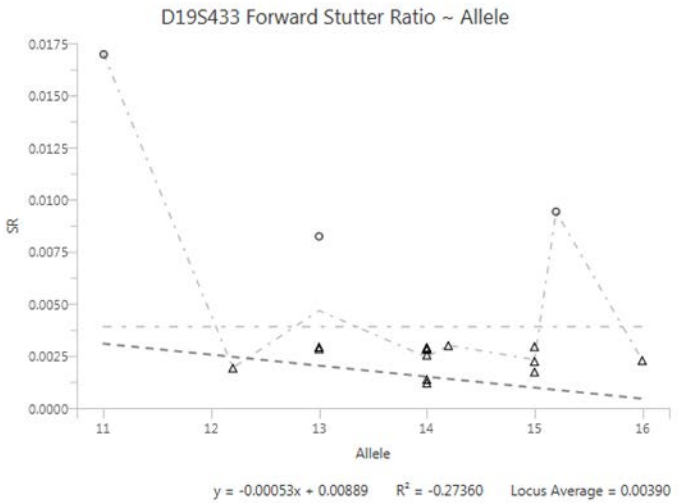

D7S820 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00150x - 0.00949 R² = 0.04321 Locus Average = 0.00476

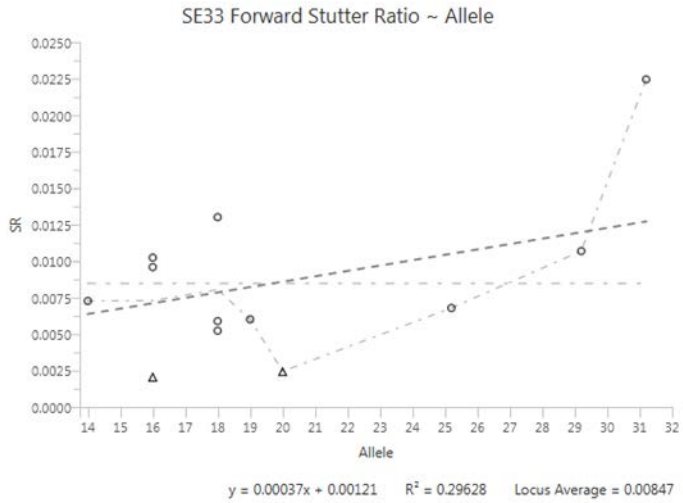

D5S818 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00291x - 0.02419 R² = -0.07205 Locus Average = 0.00888

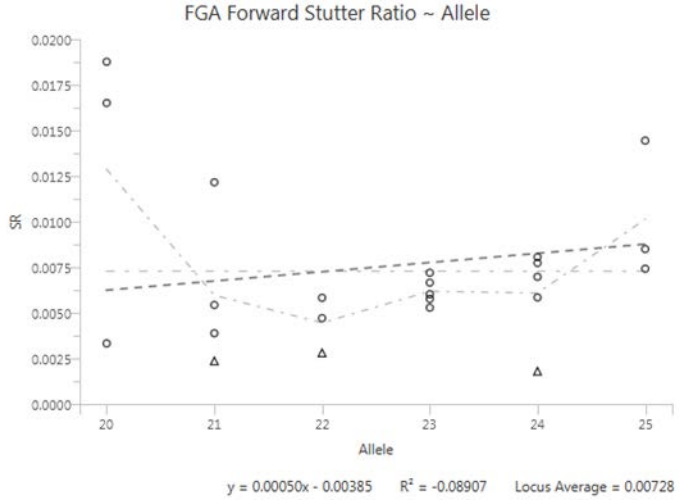

TPOX Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00038x - 0.00248 R² = -1.32482 Locus Average = 0.00241

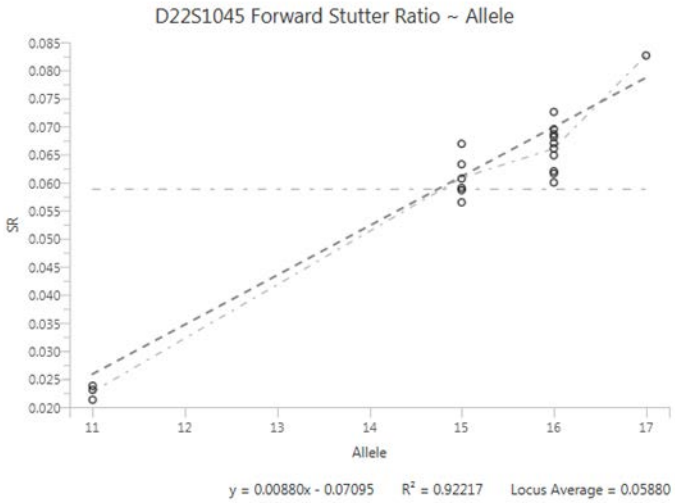

D8S1179 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00051x + 0.00061 R² = 0.00217 Locus Average = 0.00669


D12S391 Forward Stutter Ratio ~ Allele
SR
Allele
y = -0.00008x + 0.00251 R² = -1.02768 Locus Average = 0.00309

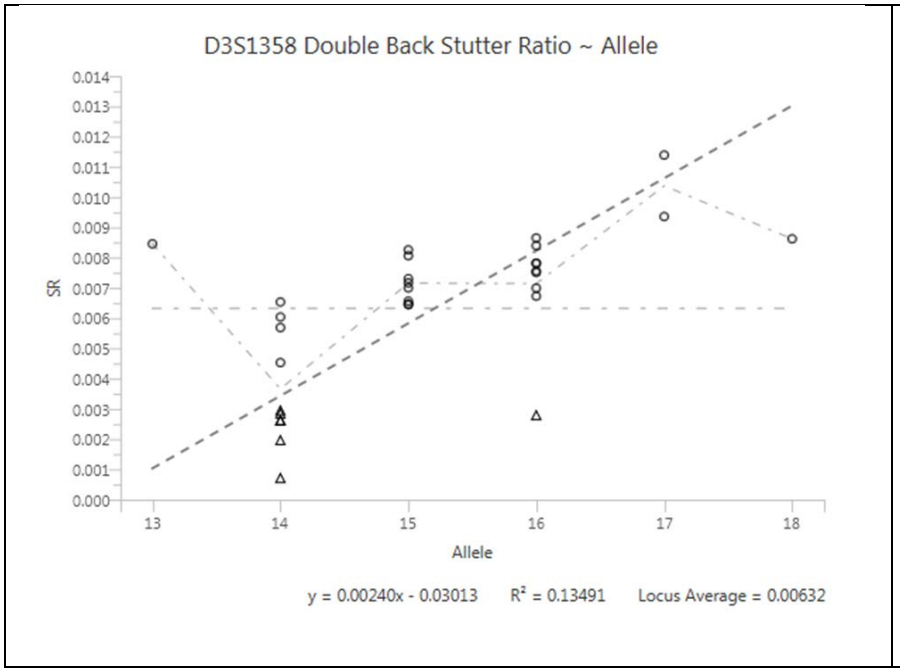

D19S433 Forward Stutter Ratio ~ Allele
SR
Allele
y = -0.00053x + 0.00889 R² = -0.27360 Locus Average = 0.00390

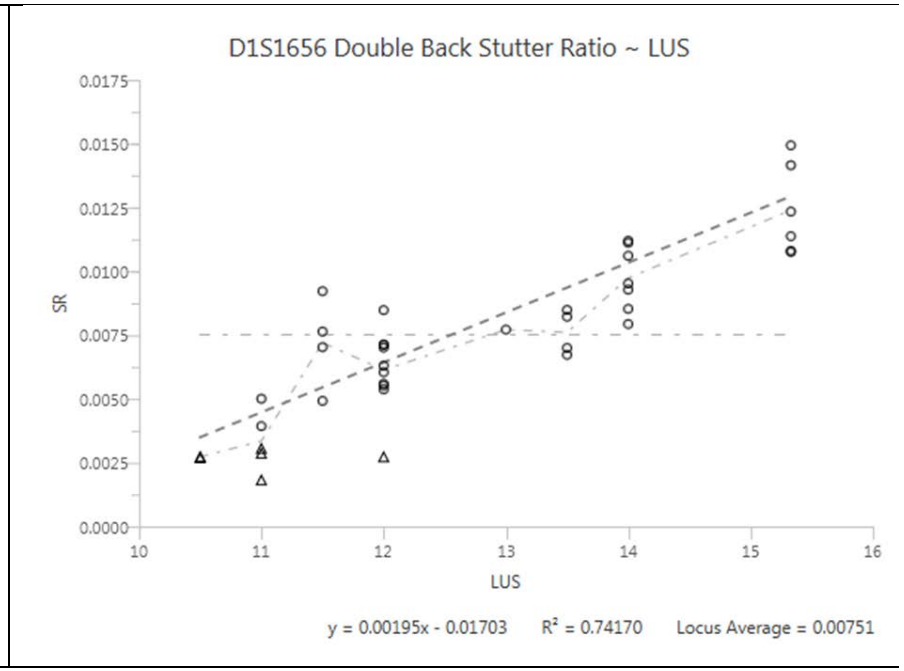

SE33 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00037x + 0.00121 R² = 0.29628 Locus Average = 0.00847

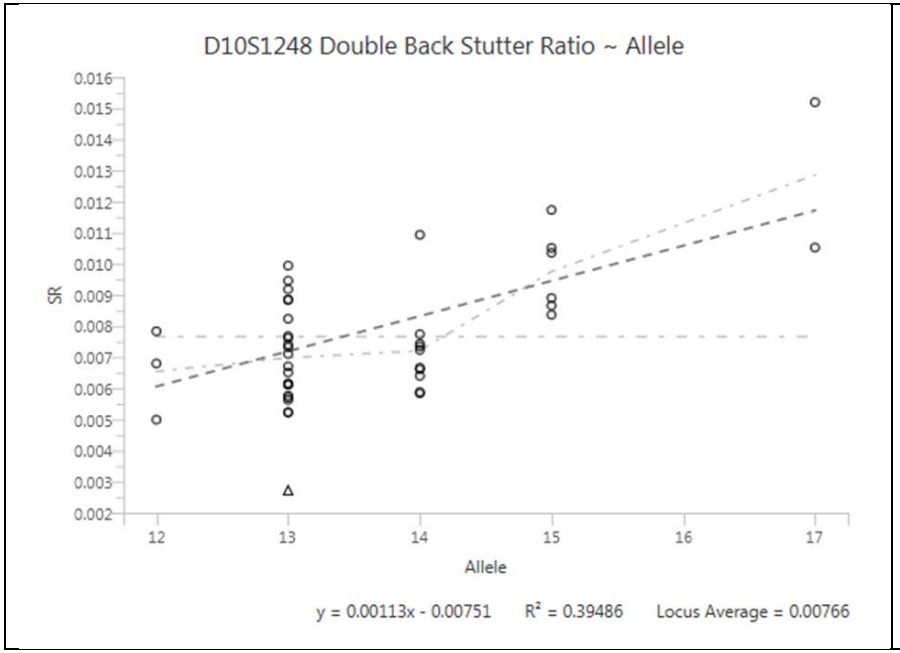

FGA Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00050x - 0.00385 R² = -0.08907 Locus Average = 0.00728

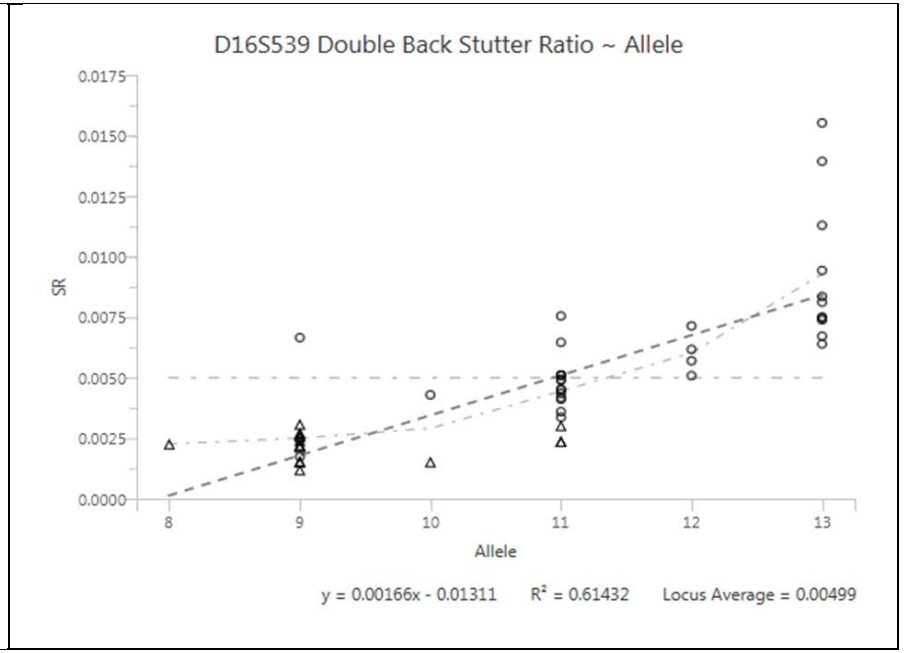

D22S1045 Forward Stutter Ratio ~ Allele
SR
Allele
y = 0.00880x - 0.07095 R² = 0.92217 Locus Average = 0.05880

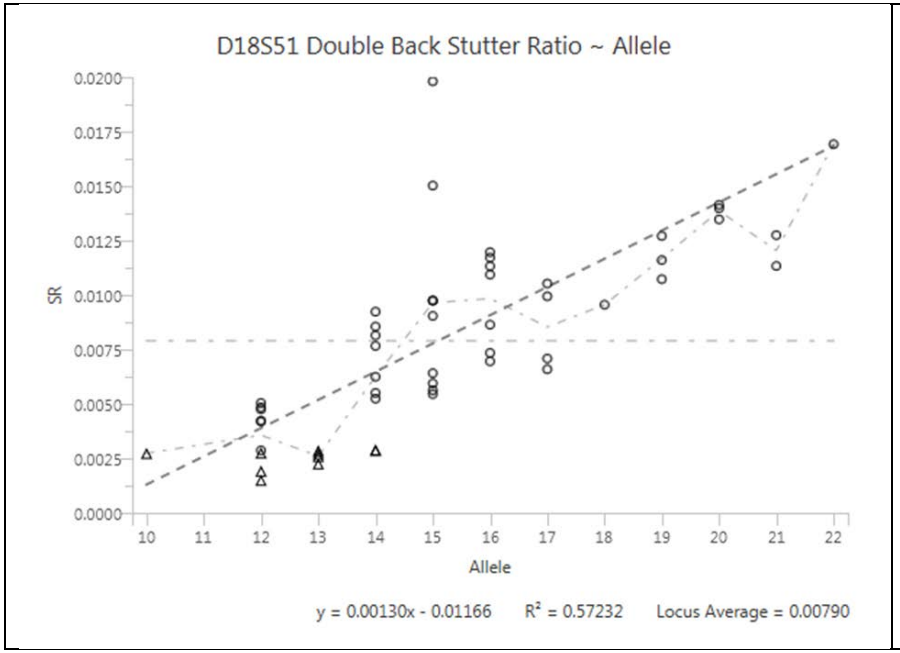
Appendix 3: Double back stutter (-2,0) versus allele designation. The dashed line is the regression by allele. The dash-dot line is the average SR per allele. The horizontal dashed line is locus average SR. Please note for most loci the average of the data for the entire locus is used.

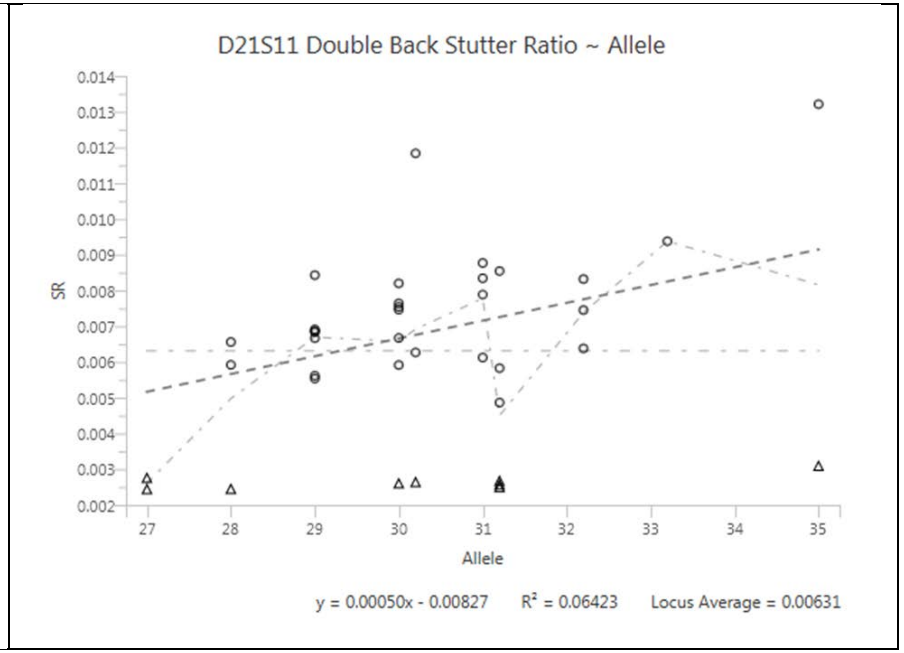


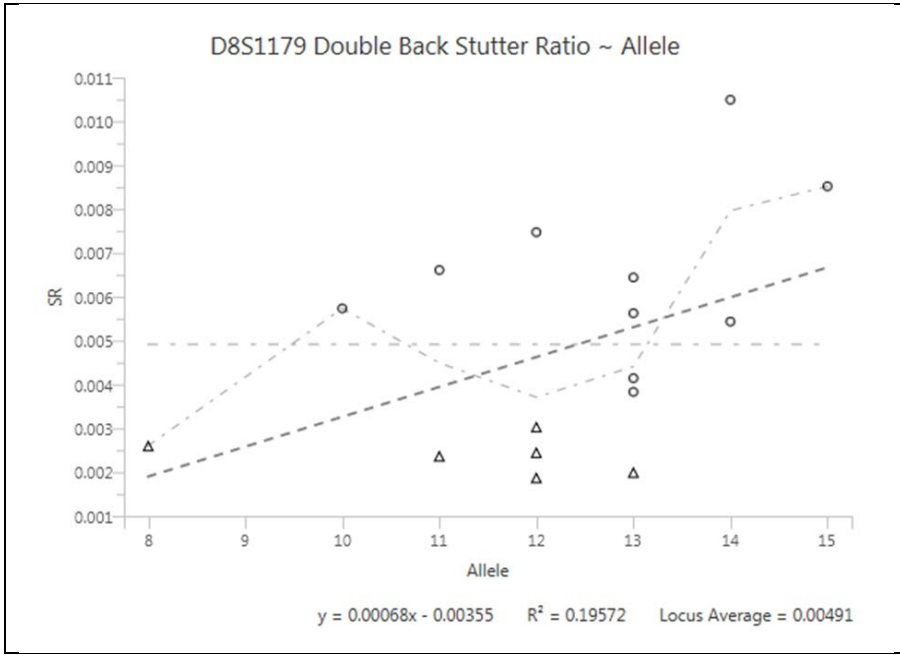


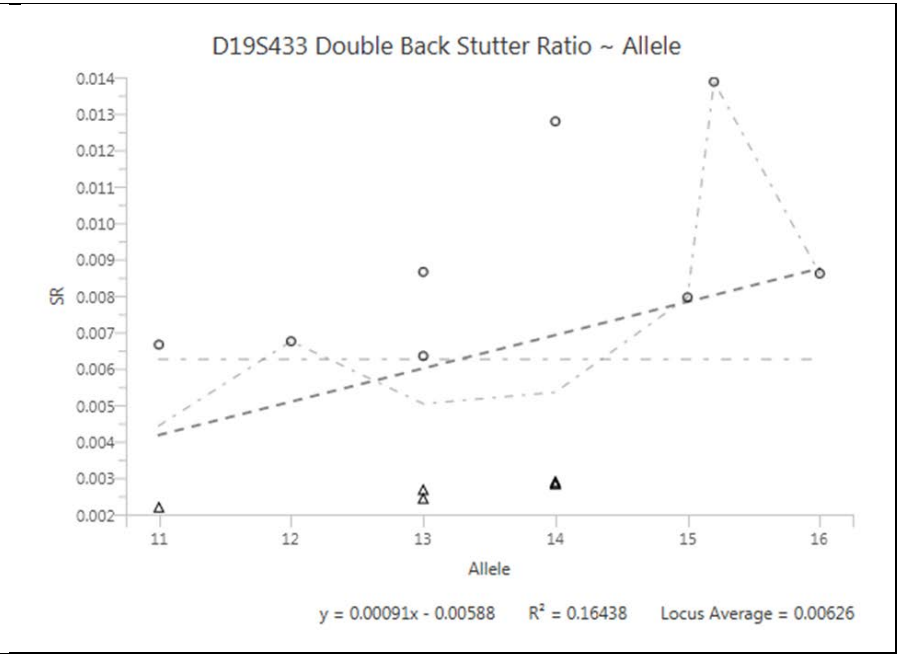

D10S1248 Double Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00113x - 0.00751
R² = 0.39486
Locus Average = 0.00766

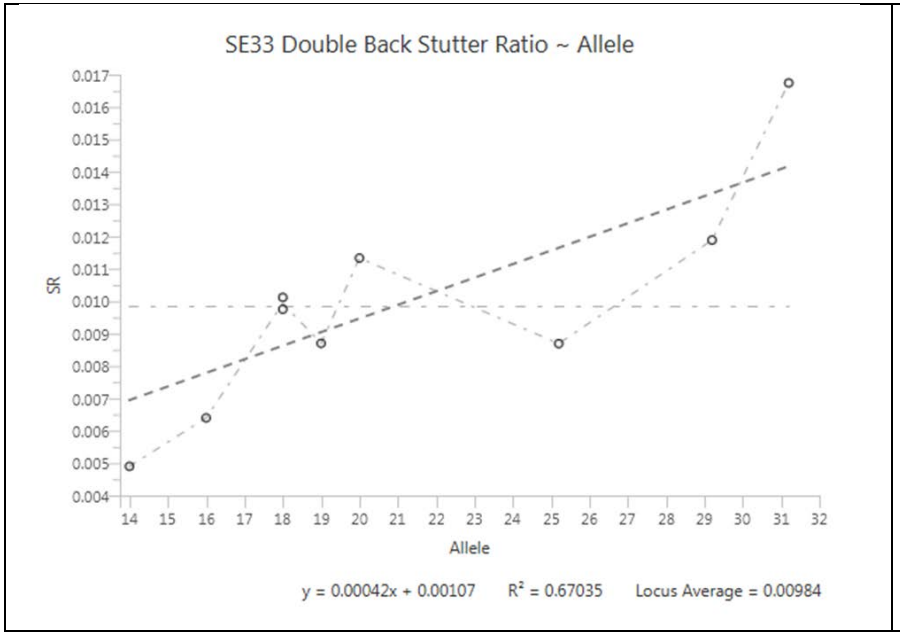

D16S539 Double Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00166x - 0.01311
R² = 0.61432
Locus Average = 0.00499

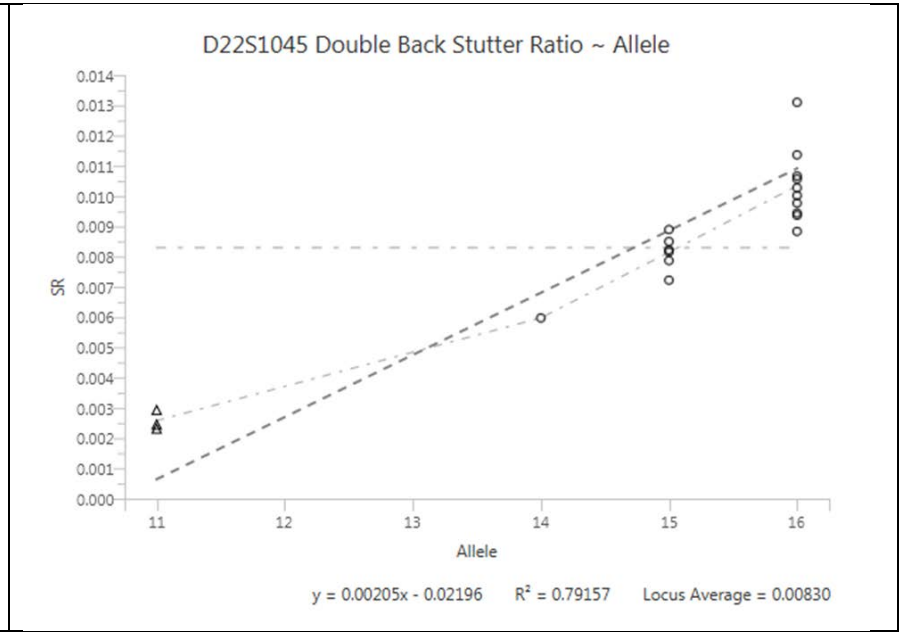

D18S51 Double Back Stutter Ratio ~ Allele
SR
0.0200
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
10 11 12 13 14 15 16 17 18 19 20 21 22
Allele
y = 0.00130x - 0.01166 R² = 0.57232 Locus Average = 0.00790


D21S11 Double Back Stutter Ratio ~ Allele
SR
0.014
0.013
0.012
0.011
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
27 28 29 30 31 32 33 34 35
Allele
y = 0.00050x - 0.00827 R² = 0.06423 Locus Average = 0.00631


D8S1179 Double Back Stutter Ratio ~ Allele
SR
0.011
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
8
9
10
11
12
13
14
15
Allele
y = 0.00068x - 0.00355
R² = 0.19572
Locus Average = 0.00491


D19S433 Double Back Stutter Ratio ~ Allele
SR
0.014
0.013
0.012
0.011
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
11
12
13
14
15
16
Allele
y = 0.00091x - 0.00588
R² = 0.16438
Locus Average = 0.00626


SE33 Double Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00042x + 0.00107
R² = 0.67035
Locus Average = 0.00984


D22S1045 Double Back Stutter Ratio ~ Allele
SR
Allele
y = 0.00205x - 0.02196
R² = 0.79157
Locus Average = 0.00830

Appendix 4: Two base pair/half back stutter (0,-2) versus allele designation. The dashed line is the regression by allele. The dash-dot line is average SR per allele. The horizontal dashed line is locus average SR. Please note for both loci the average of the data for the entire locus is used.