Article:

Bright, J.-A., Taylor, D., Curran, J. M., & Buckleton, J. S. (2013). **Developing allelic and stutter peak height models for a continuous method of DNA interpretation**. *Forensic Science International: Genetics*, *7*(2), 296–304.

INSTITUTE OF ENVIRONMENTAL SCIENCE AND RESEARCH LIMITED
Kenepuru Science Centre: 34 Kenepuru Drive, Kenepuru, Porirua 5022 | PO Box 50348, Porirua 5240, New Zealand
T: +64 4 914 0700 **F:** +64 4 914 0770
www.esr.cri.nz

Developing allelic and stutter peak height models for a continuous method of DNA interpretation

Jo-Anne Bright[1,2*], Duncan Talyor[3], James Curran[2], John S. Buckleton[1]

[1] *ESR, Private Bag 92021 Auckland 1025, New Zealand*

[2] *Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand*

[3] *Forensic Science South Australia, 21 Divett Place, SA 5000, Australia*

* Corresponding author at: Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland, 1142, New Zealand. Email address: Jo.bright@esr.cri.nz .

Traditional forensic DNA interpretation methods are restricted as they are unable to deal completely with complex low level or mixed DNA profiles. This type of data has become more prevalent as DNA typing technologies become more sensitive. In addition they do not make full use of the information available in peak heights. Existing methods of interpretation are often described as binary which describes the fact that the probability of the evidence is assigned as 0 or 1 (hence binary) (see for example [1] at 7.3.3). These methods are being replaced by more advanced interpretation methods such as continuous models. In this paper we describe a series of models that can be used to calculate expected values for allele and stutter peak heights, and their ratio SR. This model could inform methods which implement a continuous method for the interpretation of DNA profiling data.

Keywords: DNA interpretation, mixture interpretation, continuous models, stutter

Introduction

The forensic examination of biological evidence often produces low level or mixed DNA profiles, which are regarded as complex profiles. Traditional methods of interpretation are often described as binary which describes the fact that the probability of the evidence is assigned as 0 or 1 (hence binary) (see for example [1] at 7.3.3). These methods are being replaced by more advanced interpretation methods such as continuous models [2, 3]. In this paper we describe a series of models that can be used to calculate expected values for

allele and stutter peak heights, and their ratio, SR [2, 3]. This is motivated by the difficulties traditional methods have with the interpretation of complex profiles [4, 5]. Complicating interpretation of any DNA profile is the occurrence of stutter, an artefact of the PCR amplification of STR loci.

The earliest forms of the binary model considered alleles to be present or absent. Methods were subsequently developed that used heterozygous balance (Hb) to determine whether combinations of genotypes were supported or not. The binary model assigns a value of zero or one to the probability of the profile given the proposed allelic combination (hence the term binary) depending on whether the alleles could pair given Hb . The application of this model makes a number of assumptions including that peak area/height (hereafter height) is proportional to the quantity of template DNA and that the height of 'shared' peaks between individuals is the sum of the peaks from the contributing individuals. This is actually a rephrasing of the assumption that the height of a peak is linearly related to the quantity of DNA. Known shortcomings of the binary model [6, 7] have led to the development of new and improved models that factor in the probability of drop-out [8-11]. Subsequently, fully continuous interpretation models have been developed [3, 12]. These models take the quantitative information from the electropherogram (for example peak heights) and use them to calculate the probability of the peak heights given all the possible genotype combinations for the individual contributors. This approach removes some of the criticism regarding subjectivity [13, 14] in profile analysis and attempt to ensure consistency in DNA interpretation and reporting across different laboratories. Well described probabilistic systems give a detailed accounting of their respective methods. What transpires inside a human expert's mind can be far more opaque than equations provided in peer-reviewed journals.

Continuous methods make assumptions about the underlying behaviour of peak height, or of the variability in the ratio of the two peaks of a heterozygote (Hb), and the ratio of allelic peak height to stutter peak height (SR) to evaluate the probability of a set of peak heights. These models may be developed from empirical data external to the profile under interpretation, by a combination of external data and the profile under consideration, or simply by the profile under consideration. We would tend to favour the combination approach.

In this paper we investigate the underlying behaviour of Hb and SR . We also investigate the relationship between the heights of two alleles of a heterozygote, and the allele and its stutter product. The aim is to build models to inform a continuous interpretation system. Previous work has investigated the variability in Hb in Applied Biosystems' Identifiler™ [15] and MiniFiler™ [16] multiplexes. The continuous model may work by means of modelling the variability in Hb directly but more often works with variability in peak heights themselves [2, 3]. In single source profiles, the variability in Hb reduces as the average peak height (APH) at a locus increases.

The distribution of peak heights varies with the quantity of DNA and is difficult to investigate directly. The investigation could be undertaken by making consistent extractions and amplifications of entirely equivalent templates. In this case we would expect the height of each peak to vary about the same mean. The distribution could be determined directly. However the consistent replication of extraction and amplification template presents some experimental challenges. We are incapable of standardising the template to absolute precision. It is likely that the replicate peak heights would vary about a mean that was also varying. This is because template would vary and then the PCR process would add further variance. Since the two alleles of a heterozygote are as close as we can envisage to replicate extractions and amplifications of the same template, the variation in

(the logarithm of) Hb should be twice that of (the logarithm of) peak height. This suggests that one practical route into modelling the distribution of peak height is through the distribution in Hb.

The variability in SR is routinely estimated by individual laboratories as part of an internal validation of a new multiplex or an analysis platform. Previous work has investigated the longest uninterrupted sequence (LUS) as a predictor of stutter [17, 18]. It has been shown that alleles with large LUS values stutter more than alleles with small LUS values and plausibly amplify less. For any given LUS there will still be stutter peaks above or below expectation. A larger than expected stutter is likely to be caused by stutter events early in the PCR process. This would be expected to lead to a smaller allelic peak. This allows us to define the following hypothesis: If, for any given allele, the stutter peak is above expectation given its LUS value, then we expect the peak height for that allele to be below expectation. If this hypothesis were true, then this would have implications for any continuous model that sought to model stutter as well as allelic peaks independently.

Many laboratories are moving to the European Standard Set of Loci (ESSoL). One of the multiplexes which include these loci is Applied Biosystems' NGM SElect™. We report here an investigation into the variability of Hb and SR in this multiplex. We acknowledge that the concepts are universal across many different STR multiplexes. We have developed a biological model that can easily be grasped by a forensic biologist that is intended for use within any software implementing a continuous interpretation method.

Method

289 single source DNA profiles were analysed using Applied Biosystems' NGM SElect™ (Life Technologies, Carlsbad, CA) multiplex. The samples were saliva stains on FTA® Elute card (Whatman, Maidstone, England) and DNA was recovered off the card using a simple elute method. Prior to amplification all samples were quantified using Applied Biosystems' Quantifiler™ (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. A target of 1 ng of DNA was amplified using NGM SElect™ following the manufacturer's instructions in a 9700 silver block thermal cycler. Amplified products were separated on an Applied Biosystems' 3130xl Genetic Analyser (Life Technologies, Carlsbad, CA) and data was analysed using Applied Biosystems' GeneMapper™ *ID* version 3.2.1 (Life Technologies, Carlsbad, CA) using a 25 RFU limit of detection threshold.

Loci where the alleles were separated by one repeat were discarded because stutter is likely to interfere with the allele height of the low molecular weight allele in an additive manner. These have previously been referred to as *stutter affected heterozygotes*. In total, 2,323 heterozygous loci were identified as being suitable for analysis.

Stutter ratio was defined as

$$SR = \frac{O_{a-1}}{O_a}$$

where O_{a-1} refers to the observed height of the stutter peak, and O_a the parent peak.

LUS was defined as the longest stretch of basic repeat motifs within the allele. The longest uninterrupted sequence (LUS) for each allele was determined using the method of Brookes et al. [17]. LUS values were obtained by looking up the allele designation in the

short tandem repeat DNA internet database (STRBase) [19, 20]. Where multiple values for LUS were available the average LUS value across the reported variants observed was taken. δ_{LUS} was defined as the difference in LUS values for the two alleles of a heterozygote. Heterozygote balance (Hb) was calculated as

$$Hb = \frac{O_H}{O_L}$$

where O_H refers to the height of the high molecular weight allele, and O_L the height of the low molecular weight allele. Statistical analysis was undertaken using R [21] and MS EXCEL™.

Linear modelling was used to test the effect of various explanatory variables on the expected values of SR and Hb. Having chosen a model for the expected value we investigate models to predict the variance about this expectation.

Results

Stutter

The following linear model was proposed to describe the relationship between SR and the explanatory variables LUS and locus, l :

$$SR_i = \beta_{0,l} + \beta_{1,l} LUS_i \quad (1)$$

This was termed the stutter model. Linear modelling of stutter has been reported previously [22, 23]. The model described in this paper was selected after exploratory analysis suggested a nil or small effect of other potential explanatory variables. The plots of SR versus LUS for individual NGM SElect™ loci are given in Appendix 1. The interaction term allows a different slope of the SR vs. LUS line for each locus. The R^2 is 0.83 for the stutter model. The improvement in fit of LUS over simple allele number is demonstrated for the TH01 locus in Figures 1 and 2 where Figure 1 gives SR versus repeat number ($R^2 = 0.02$) and Figure 2 SR versus LUS ($R^2 = 0.58$). Of interest may be the 9.3 allele in TH01. This has the structure $[AATG]_6ATG[AATG]_3$ and hence has a LUS of 6. Inspection of Figures 1 and 2 show that the 9.3 allele sits much better in the trend when placed at an *LUS* of 6.

We would anticipate that $\log(SR)$ would be easier to model because SR is a ratio. Given that the allelic peak height is much bigger than the stutter peak height this effect should be minor.

Figure 1. A plot of stutter ratio versus allele repeat number for the TH01 locus.

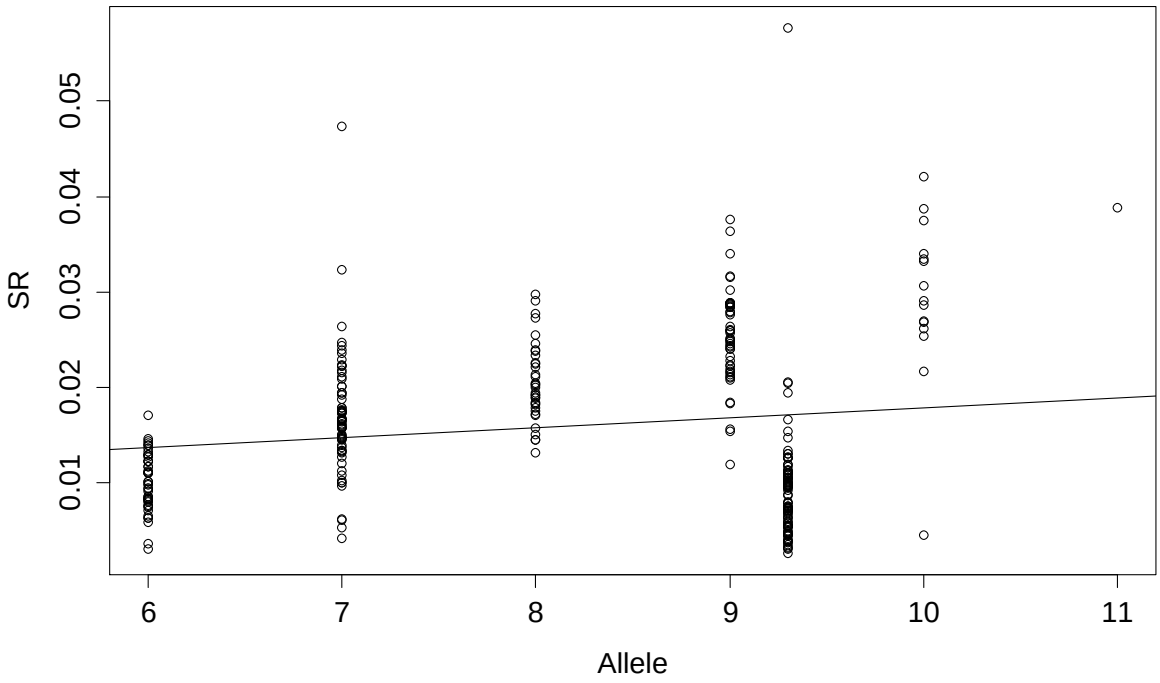
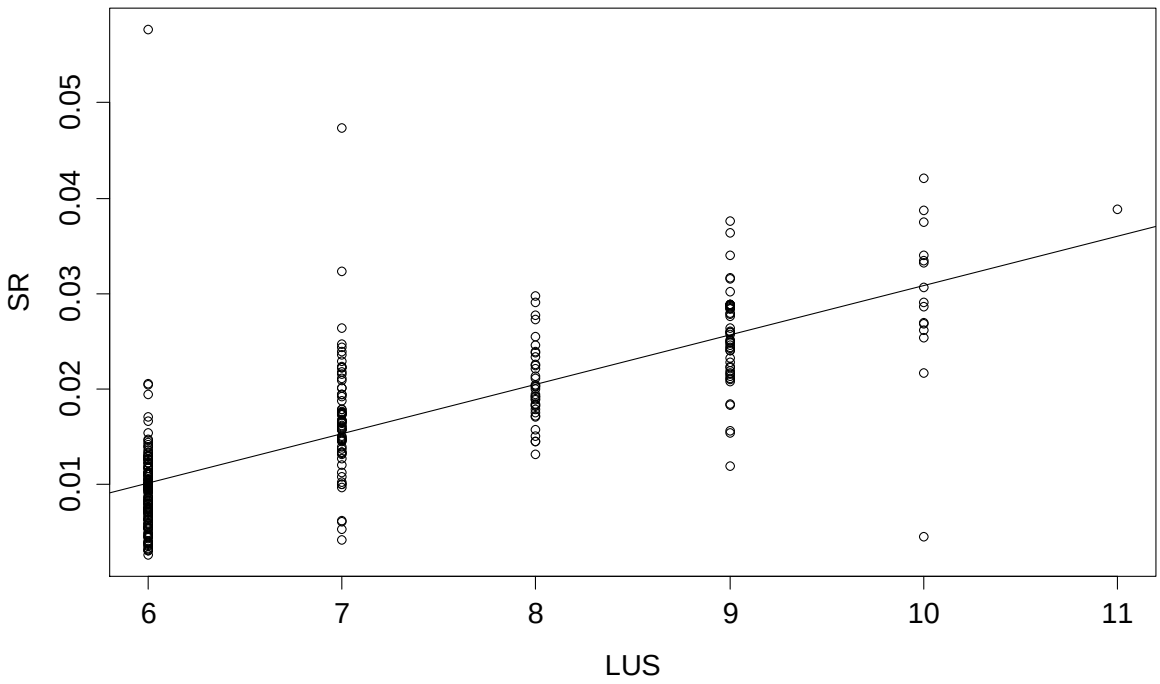


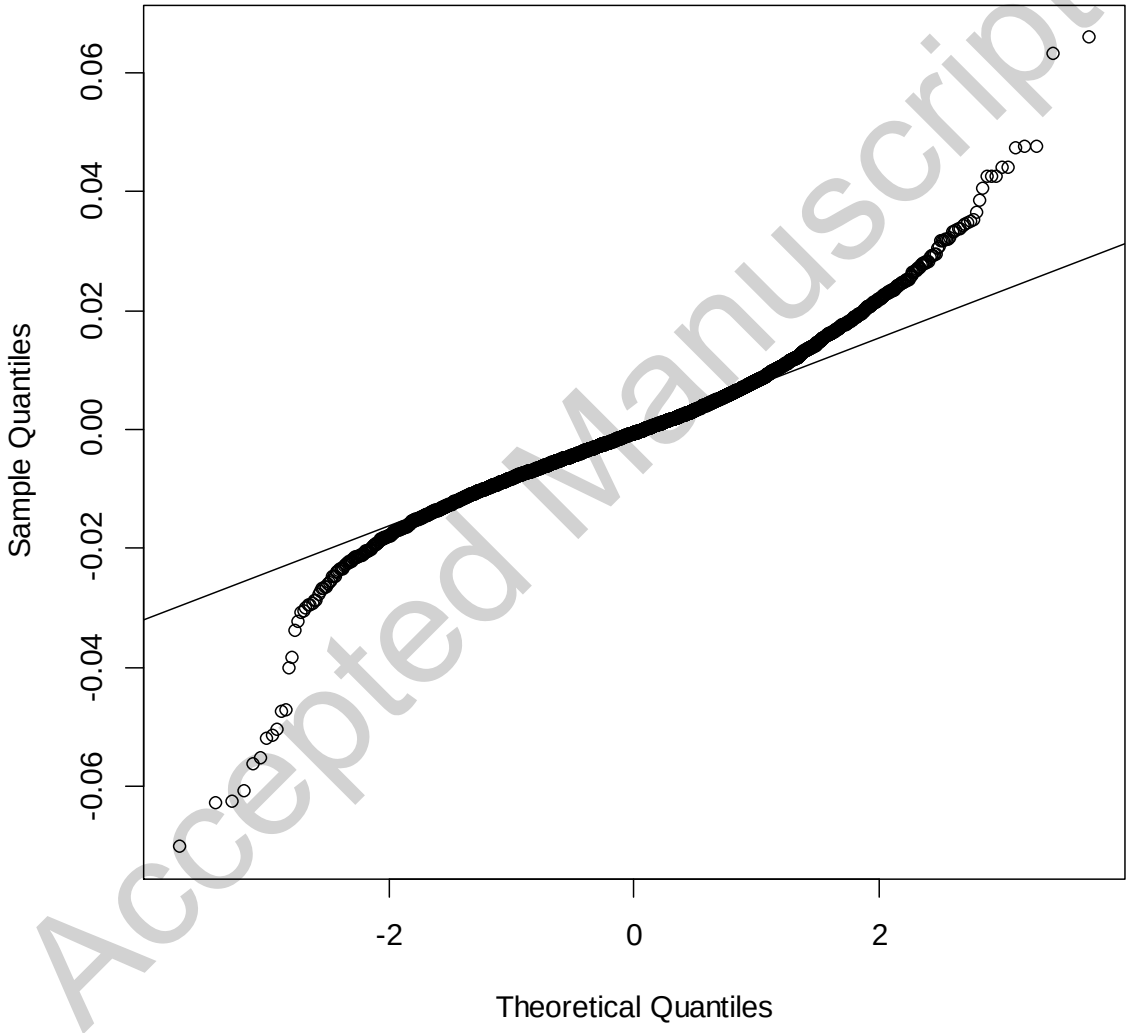
Figure 2. A plot of stutter ratio versus LUS for the TH01 locus.



There is good support for using a linear relationship to model the behaviour of SR with respect to LUS (see Appendix 1). In addition, SR is a standard concept for forensic biologists and so avoiding the introduction of logarithmic scales will improve model acceptance. A summary of the intercepts and slopes, using this model, for every locus in the NGM SElect™ multiplex kit is given in Appendix 2. D2S441 is very poorly described by this model.

A normal quantile-quantile (Q-Q) plot of the residuals from the model versus theoretical quantiles from a normal distribution is presented in Figure 3.

Figure 3. A plot of the Q-Q plot from the full stutter model



The Q-Q plot suggests that the data is symmetric but with heavier tails than the normal distribution. An assumption of approximate normality is plausibly acceptable noting that there are a great many data points in the central region.

The squared residuals were regressed against allele height in order to investigate the factors affecting the variability of SR. There is a significant effect of allele height on the variance of SR ($p = 3.9 \times 10^{-14}$) however, as the coefficient was small (-7.5×10^{-7}), it will have little effect on the predicted variability of SR.

Some alleles show markedly larger variation in SR compared with the expectation. For example, at locus D2S441 the SR for several values of LUS are not well described by the model (refer Appendix 1). Closer inspection suggests that, in many cases, this was caused by an allele that has a complex repeat structure comprising of variant regions with differing LUS values. In another example, for D21S11 30 the sequence has been variously typed as:

$[TCTA]_6\ [TCTG]_5\ [TCTA]_3\ TA\ [TCTA]_3\ TCA\ [TCTA]_2\ TCCA\ TA\ [TCTA]_{11}$

$[TCTA]_5\ [TCTG]_6\ [TCTA]_3\ TA\ [TCTA]_3\ TCA\ [TCTA]_2\ TCCA\ TA\ [TCTA]_{11}$

$[TCTA]_4\ [TCTG]_6\ [TCTA]_3\ TA\ [TCTA]_3\ TCA\ [TCTA]_2\ TCCA\ TA\ [TCTA]_{12}$

$[TCTA]_6\ [TCTG]_6\ [TCTA]_3\ TA\ [TCTA]_3\ TCA\ [TCTA]_2\ TCCA\ TA\ [TCTA]_{10}$

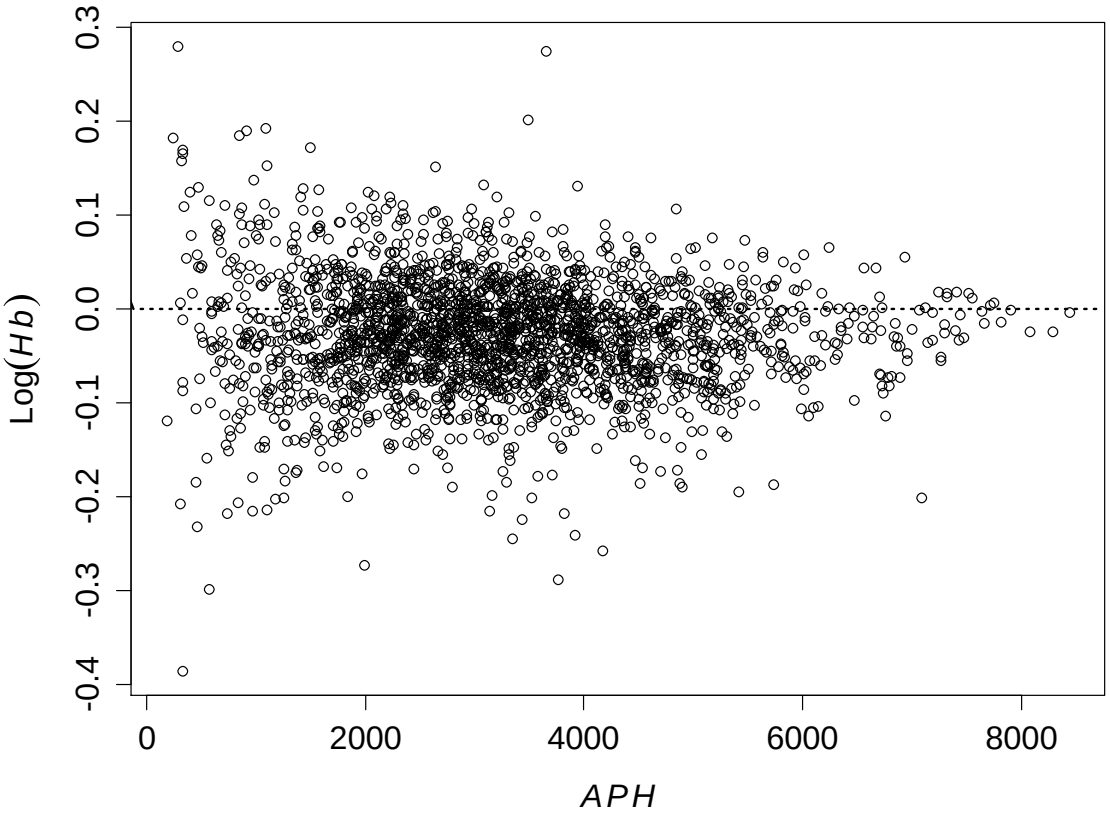
for different variants [19]. Since we will only know the molecular weight and not the sequence when using typical casework electropherograms we have used an average *LUS*. In this case we have used the average of 10,11,11 and 12. Since we have used an average and the sample plausibly contains some of each sequence we expect to see enhanced spread.

If there is indeed an effect of LUS, as observed here and previously [17, 18], then, using the D21S11 example given above we would expect some variants in our set with LUS values of 10,11 and 12. This would lead to distributions in the observed stutter ratio that are centred around a higher value (for $LUS = 12$) and a lower value (for $LUS = 10$) but all are plotted at $LUS = 11$. Hence a wider spread. Such widening is a likely explanation for the heavy tails observed in Figure 3.

Heterozygote balance variability

The relationship between Hb and average peak height (APH) was demonstrated for NGM SElect™ data in Figure 4. The variation in Hb decreases as APH increases. This funnel shape has been observed in other multiplexes [15, 16]. Direct comparison of the distributions shows that there is less variation in Hb with NGM SElect™ compared with that seen in the Identifiler™ and MiniFiler™ multiplexes [15, 24].

Figure 4. $\log(Hb)$ versus APH , 2,323 heterozygote NGM SElect™ loci



It is known [17, 18], and reinforced above, that alleles with large LUS values stutter more. One would expect alleles with large LUS values to have smaller allelic peaks [7] for a given template level. Under this hypothesis, stuttering is one of the determinants of any systematic effect on Hb, and it is the difference in LUS, δ_{LUS}, that should be the explanatory variable for Hb. Since Hb is a ratio we expect $\log(Hb)$ to be more amenable to modelling. This is supported by previous work [15, 16]. It is helpful to consider the concept of the sum of the allelic and stutter peaks, termed total allelic product (T) [7]. This is calculated, for the a^{th} allele, as

$$T_a = O_{a-1} + O_a \tag{2}$$

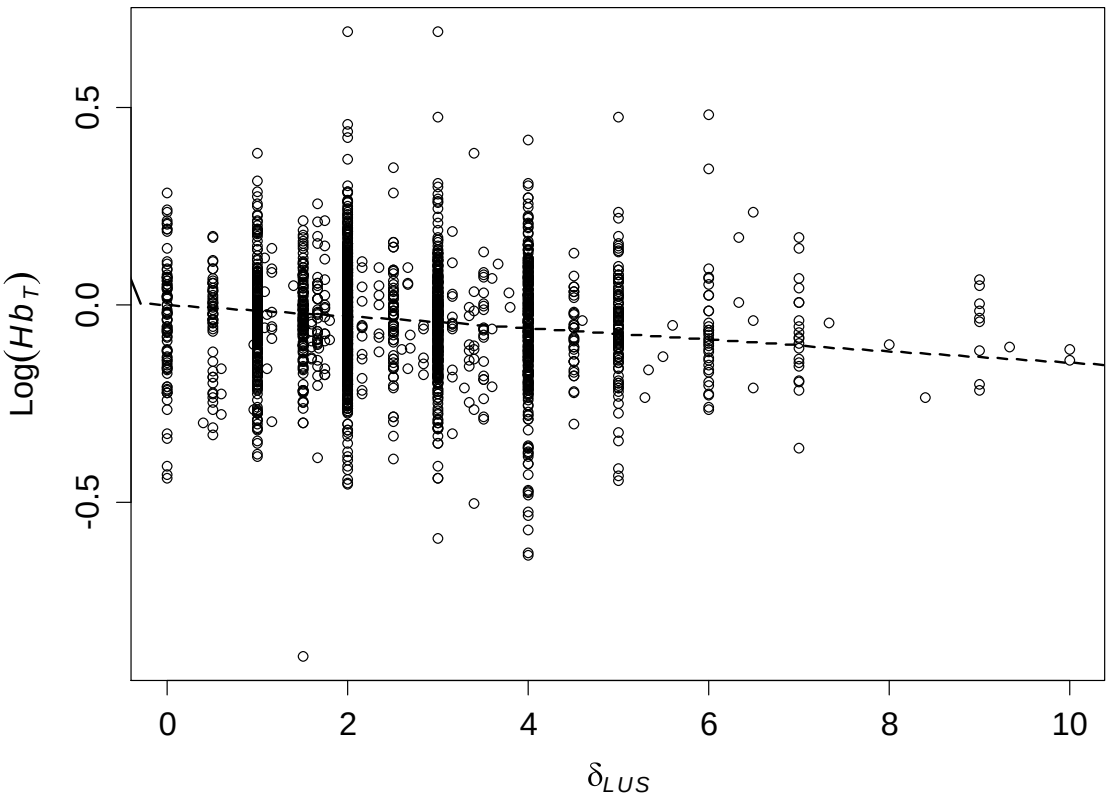
We can now define Hb_T in terms of total allelic product:

$$Hb_T = \frac{T_H}{T_L} \tag{3}$$

where T_H and T_L are the total allelic product values for the high and low molecular weight alleles respectively. If stutter is the only cause of variation in allelic peak height within the two peaks of a heterozygote, then we expect the mean of $\log(Hb_T)$ to be zero, and to have no relationship with δ_{LUS} or any other variable. In Figure 5 we give the plot of $\log(Hb_T)$ vs. δ_{LUS}. The regression line was forced through the origin. There was a small but

significant negative slope to the regression line in Figure 5 ($slope = -0.0047$). A plot of $\log(Hb_T)$ vs. the difference in allele repeats (δ_{AR}) also has a small but significant downwards slope ($slope = -0.0053$, data not shown). We conclude from this that there is something other than just stutter affecting allelic peak height for a given template level. This is likely to be simply due to the reduced amplification efficiency of the larger allele at a heterozygote locus. Of course template level is the primary determinant of peak height but should have no effect on expected Hb_T . After template the next largest effect appears to be stutter ratio and this affects both Hb and peak heights, but should not affect Hb_T . Last there is something else which we, and others, postulate is simply amplification efficiency. This affects peak heights, Hb and Hb_T . We are unable to determine from this analysis whether the behaviour of this last effect, postulated as relative amplification efficiency is better predicted by δ_{AR} or δ_{LUS} however both exhibit a small but significant effect on Hb_T

Figure 5. A plot of $\log(Hb_T)$ versus δ_{LUS}



Modelling peak heights

In this section we model peak heights as opposed to the ratios Hb , Hb_T and SR . In order to develop a model for expected peak height we need first to model the expected value for true mass at each allelic position at a locus T_{an}^{l} . We coin the term mass to subsume considerations of template, degradation and locus amplification effects. The 'true' mass of template DNA is not known. We model mass based on our observations of the data and

understanding of the behaviour of DNA profiles. During modelling of peak heights versus molecular weight for various multiplexes we have observed that some are adequately explained with a linear model whereas some require an exponential model. NGM SElect™ appears to be adequately modelled using the simpler linear model.

For L loci, N contributors and R replicates the height of an allele, a, at locus l, for replicate *r*, from contributor n is modelled as:

$$T^l_{anr} = A^l_{.r}\left(t_n + d_n \times m^l_a\right) X^l_{an} \tag{4}$$

Where:

m^l_a is the molecular weight of allele a at locus l

$A^l_{.r}$ ($l = 1\ldots L$, $r = 1\ldots..R$) is the locus offset at locus l, replicate r

$t_n (n = 1\ldots N)$ is the intercept of the line for mass vs. molecular weight for contributor n

$d_n (n = 1\ldots N)$ is the slope of the line for mass vs. molecular weight for contributor n

X^l_{an} is the count of allele a at locus l in contributor n. $X^l_{an} = 1$ for a heterozygote with allele a and $X^l_{an} = 2$ for a homozygote a.

We refer to the variables A, t, and d collectively as the mass variables **M**. Note that when considering one amplification of a sample we can drop the '*r*' subscripts, which we subsequently do so for simplicity. The locus offset, A^l, allows different amplification efficiencies for each locus. One A^l value may be set arbitrarily, termed 'fixed' and the others allowed to vary, termed 'free'. If A^l is allowed to be completely free it will tend to the midpoint of a heterozygote for single source profiles and to a related position for mixtures. This is unacceptable and would impose a large negative correlation between the peak height residuals. Accordingly we set the probability of each of the $L-1$ free locus specific amplification efficiency parameters A^l for each of the $L-1$ loci as $N(\mu_A, \sigma_A)$ where μ_A is the simple arithmetic average of the A^l values and σ_A is a preset hypervariable. This allows a limited freedom to the A^l variables but penalises any single value that departs significantly from the average. We set a uniform prior on μ_A.

Application of the model for mass and stutter

Mass at an allelic position at a locus can be apportioned to stutter and allele using the following equations where SR is determined from the model.

$$E^l_{(a-1)n} = \frac{SR^l_a\left(T^l_{an}\right)}{1 + SR^l_a} \tag{5}$$

$$E^l_{an} = \frac{T^l_{an}}{1 + SR^l_a} \tag{6}$$

where:

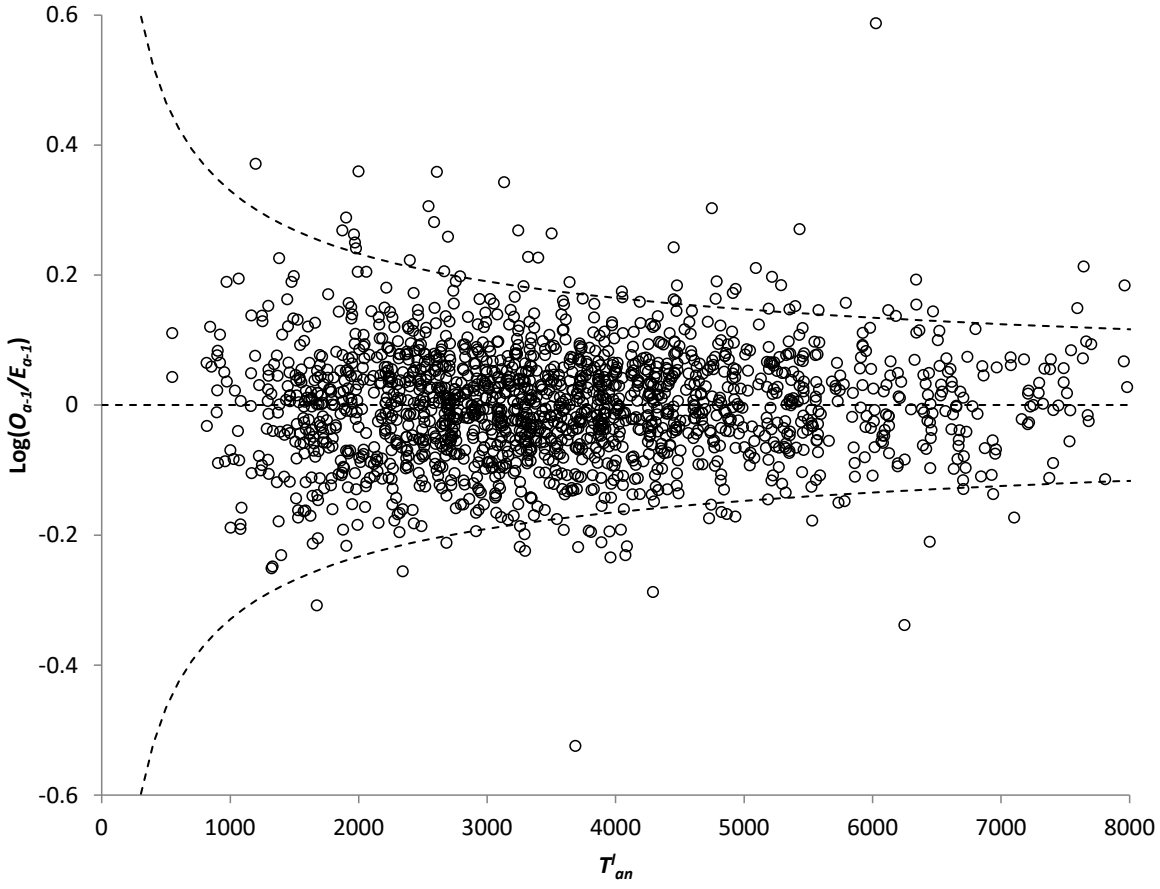
$E^l_{(a-1)n}$ is the expected stutter peak height of the a^{th} allele for the n^{th} contributor at locus l

E^l_{an} is the expected allelic peak height of the a^{th} allele for the n^{th} contributor at locus l.

Mass was assigned for each allele for a subset of 100 samples from the NGM SElect™ dataset. The subset included both heterozygote and homozygote loci but all stutter affected heterozygotes were removed. Mass variables A^l, t_n and s_n were determined by a maximum likelihood method.

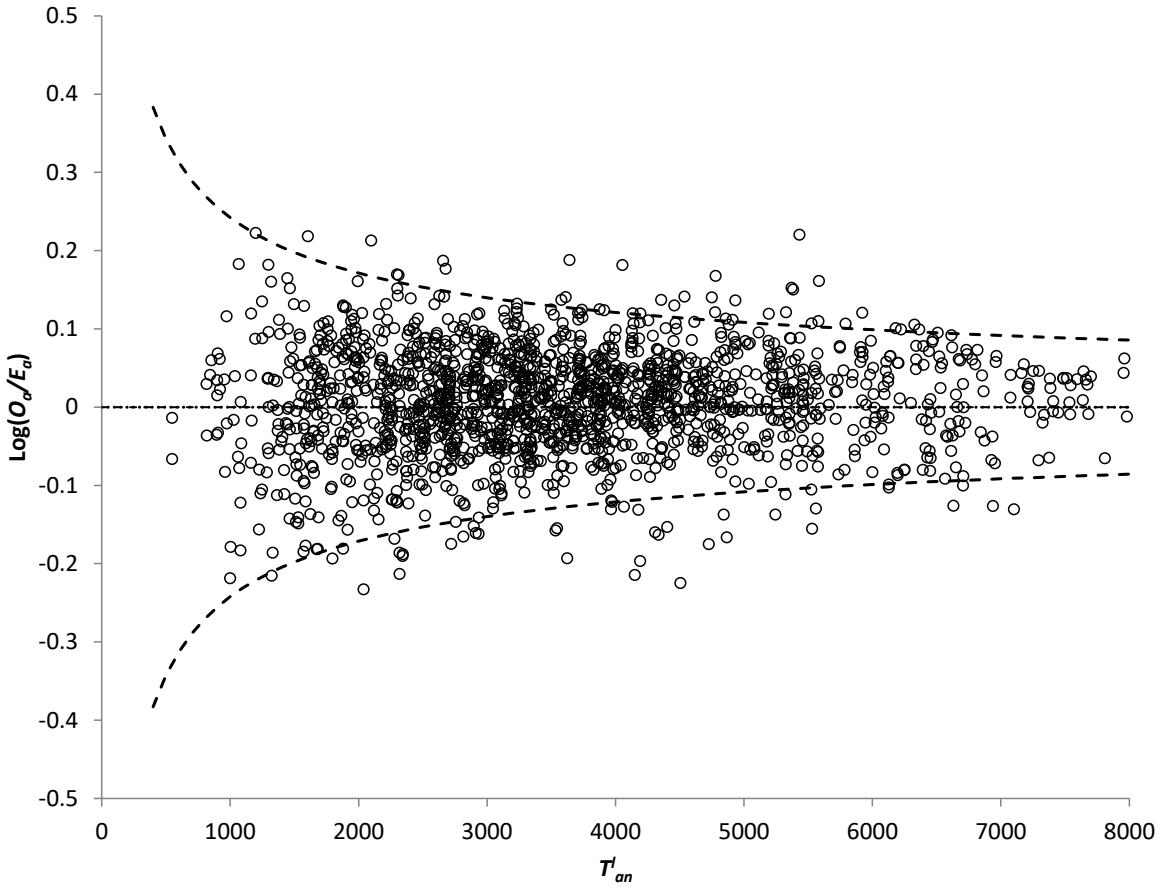
The stutter model (eq 1) was used to calculate the expected stutter ratio for each allele. Weusten and Herbergs [25] suggest that the relative standard deviation on the numbers of chains should be inversely proportional to the square root of the expected number of DNA strands entering the amplification. This suggests that the 95% standard error intervals on stutter ratio should have the shape $k/\sqrt{T^l_{an}}$. Figure 6 is a plot of the logarithm of the ratio of observed and expected heights are plotted against T^l_{an}

Figure 6. A plot of $\log \frac{O_{a-1}}{E_{a-1}}$ vs T^l_{an} for the stutter peaks. The dotted lines approximate +/-2 standard error intervals.



Subsequently the expected heights of the allele peaks were calculated for each sample. The variance of the allele model is examined in Figure 7 where the logarithm of the ratio of observed and expected heights are plotted against T^l_{an}. The x-axis has been truncated in Figure 7 at 8000 RFU to avoid saturation effects. At allele heights above approximately 8000 RFU, the data points tend to rise above the trend. These data points are likely to be affected by saturation of the 3130 camera, where the relationship between amount of DNA and allele height is no longer linear.

Figure 7. A plot of $\log \frac{O_a}{E_a}$ vs T_{an}^l for the allelic peaks. The dotted lines approximate +/-2 standard error intervals.



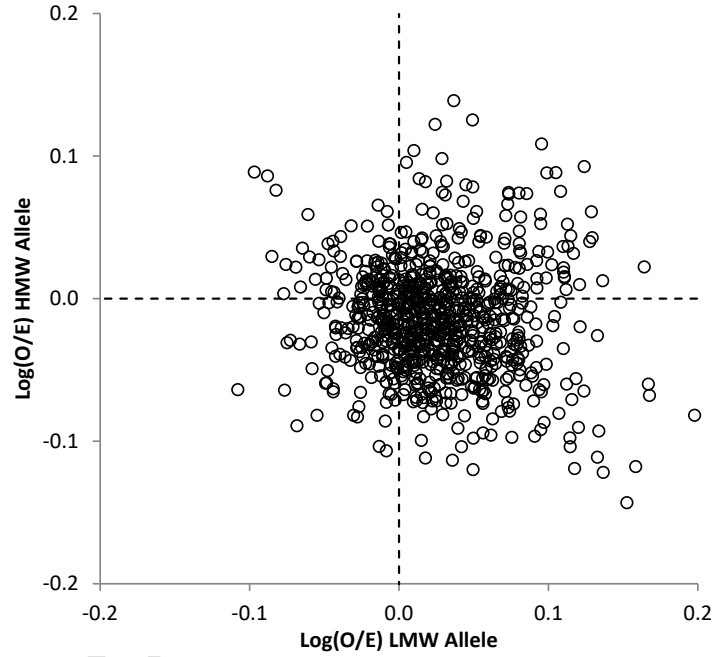
It has been suggested that the variance of the allele model (Figure 7) is inversely proportional to the expected peak height, c_a^2 / E_{an}^l [26]. The dotted lines are $\pm 1.96 \frac{c_a}{\sqrt{E_{an}^l}}$ .where $c = 3.95$ fitted by MLE. These approximate +/-2 standard error intervals are aimed at emphasising the shape of the model fitted to the data. Inspection of these plots indicate that the models are a reasonable description of the data, with few data points observed outside the intervals. The variance is symmetric around mean = 0.

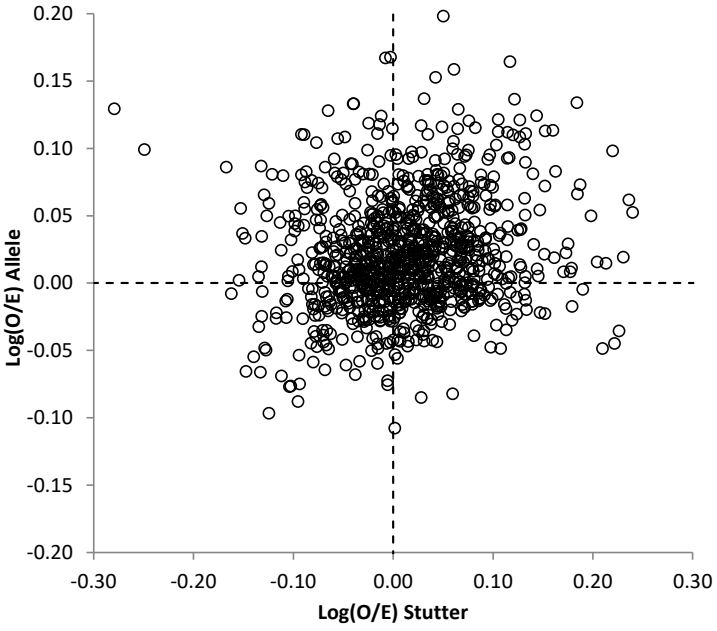
Recall that expected height is developed from the mass variables, **M**. If the predicted T_{an}^l for each of the two alleles of a heterozygote using **M** is correct, then these two variables are conditionally independent given **M**. We could reasonably expect, then, that given **M** the $\log \frac{O_a}{E_a}$ value for each allelic peak of a heterozygote is uncorrelated. However we would still anticipate a negative correlation between the $\log \frac{O_a}{E_a}$ values for the allele and $\log \frac{O_{a-1}}{E_{a-1}}$ for the associated stutter peak.

The correlation between the observed and expected peak heights at each heterozygote locus and between the observed and expected peak heights of allele and stutter was

investigated, graphically (see Figures 8a and 8b, respectively). The Pearson product-moment correlation coefficient was calculated as -0.0795 for $\log \frac{O_H}{E_H}$ for the HMW allele versus $\log \frac{O_L}{E_L}$ for the LMW allele and 0.1157 for $\log \frac{O_a}{E_a}$ allele versus $\log \frac{O_{a-1}}{E_{a-1}}$ stutter.

Figure 8a $\log \frac{O_H}{E_H}$ for the high molecular weight allele vs $\log \frac{O_L}{E_L}$ for the low molecular weight allele for each heterozygote locus and 8b $\log \frac{O_a}{E_a}$ for the allelic peak vs $\log \frac{O_{a-1}}{E_{a-1}}$ for stutter peak.





Unexpectedly the scatter plots in Figures 8a and b indicate that there is no detectable correlation between stutter and allele in this biological model.

Assuming an approximate normal distribution, with a mean of zero, a constant variance for the stutter model, and variance = $\frac{c_a^2}{E_{an}^l}$ for the allele model and variance = $\frac{c_s^2}{E_{an}^l}$ for the stutter model then:

$$\frac{\log O_{(a-1)n}^l}{\log E_{(a-1)n}^l} \sim N\left(0, c_s^2 / E_{an}^l\right)$$

$$\frac{\log O_{an}^l}{\log E_{an}^l} \sim N\left(0, c_a^2 / E_{an}^l\right)$$

Plots to check for normality for the allele and stutter models indicate that the assumption of normality is sustainable (data not shown). Both tails of the distribution appear heavy. Additional exploratory modelling of the data (data not shown) including fitting a gamma distruibution does not improve the fit.

Discussion

Previous publications have suggested that *LUS* is a better explanatory variable for *SR* than allele designation. This is confirmed for the NGM SElect™ multiplex. However one locus, D2S441 is very poorly described by this model. One plausible explanation is that the sequence data needs re-evaluation.

Weusten and Herbergs [25] have suggested that the 95% standard error intervals on stutter ratio should have the shape $k/\sqrt{T_{an}^l}$. This equation was plotted as dotted lines in Figure 6, supporting the theory.

When considering the mean value of *Hb* we expect no effect of template although template is thought to affect the variance about this mean. Stutter ratio does have an effect on mean *Hb* especially when the alleles differ significantly in *LUS*. SR alone however is not the only factor in predicting mean Hb. This can be observed in Figure 5. Larger alleles amplify less efficiently. This is likely to be due to an amplification effect with the longer lengths of DNA resulting in lower peak heights.

The concept of mass (*T*) was introduced in order to model allele heights and stutter heights. *T* was described by the molecular weight of the allele and the three mass parameters; amplification efficiency, intercept, and slope. In this research, mass parameters were determined using maximum likelihood. More elegant methods such as MCMC exist [27].

T_{an}^l and SR were combined to calculate expected heights for stutter and allele. The approximate linearity of the investigative plots showed an acceptable fit to the log normal distribution. Both tails appear heavy which does not suggest that the gamma models being considered by some commentators are a total solution [2, 28, 29]. The correlation graphs, Figures 8a and 8b, show no detectable relationship between the expected heights of alleles and their corresponding stutter, and the HMW and LMW alleles at a heterozygous locus. This suggests that the independence model may be sustainable.

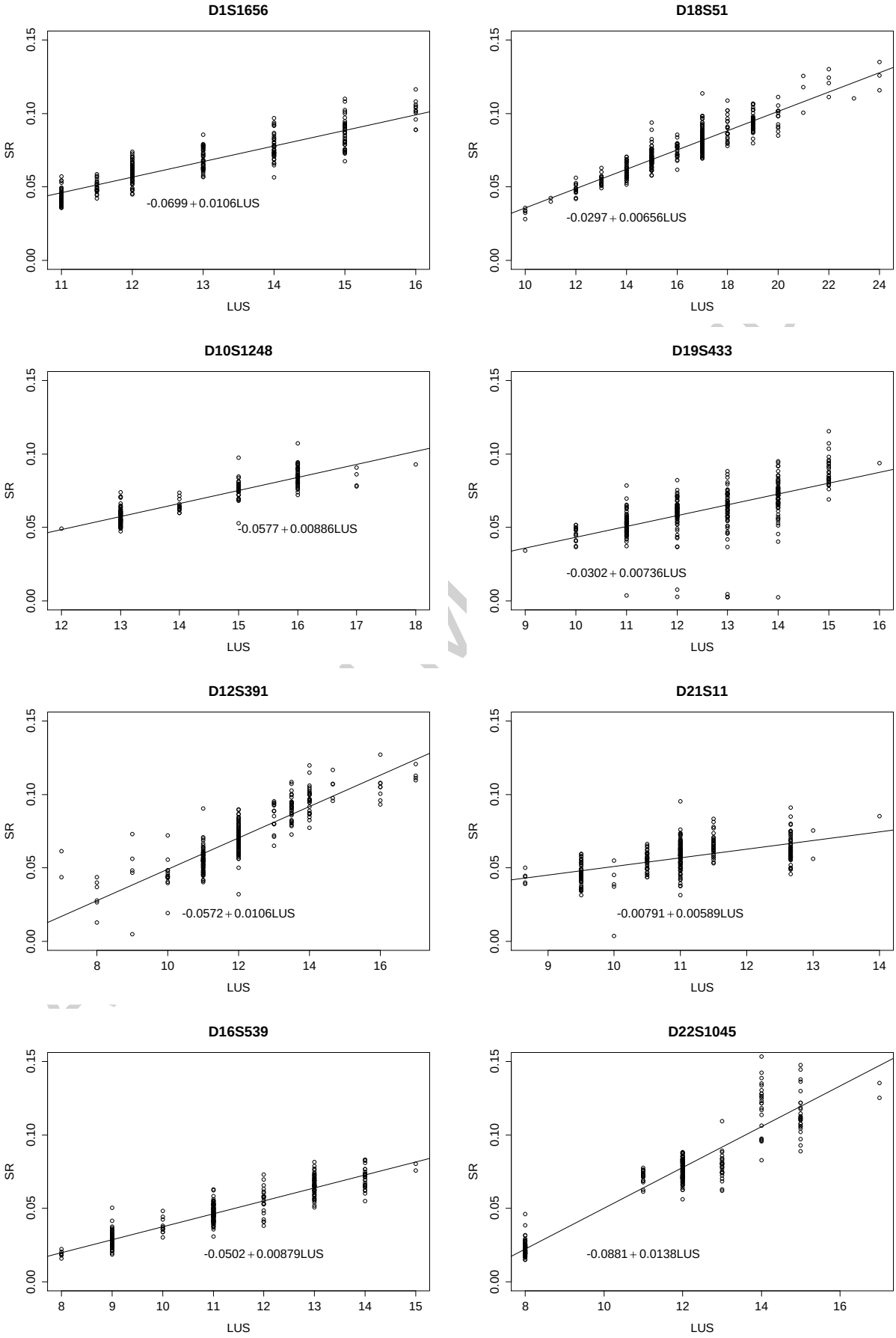
We have described a model that can be used to predict expected values and variances for

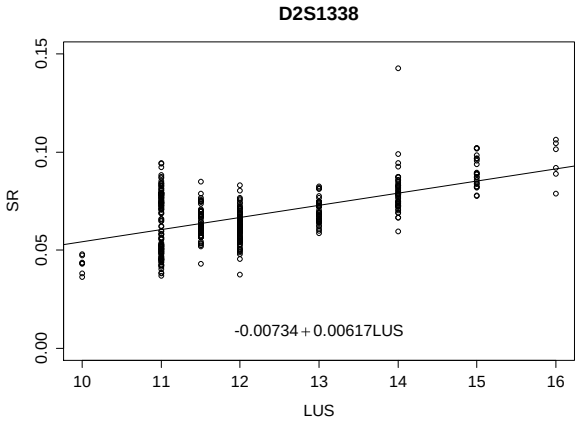
SR but further give models for predicted allele and stutter heights and the variances about these predictions. We did not find a correlation between higher than expected allele peak and lower than expected stutter peaks.

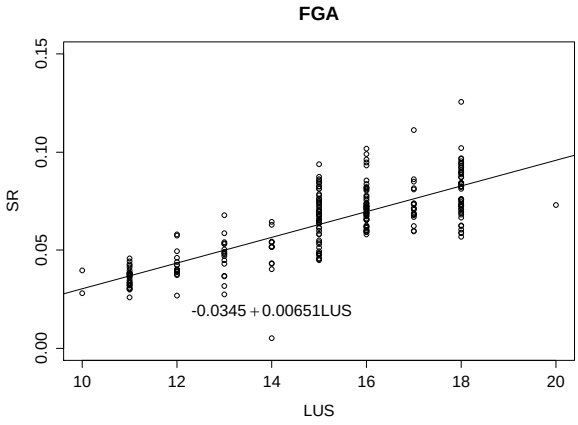
Acknowledgements

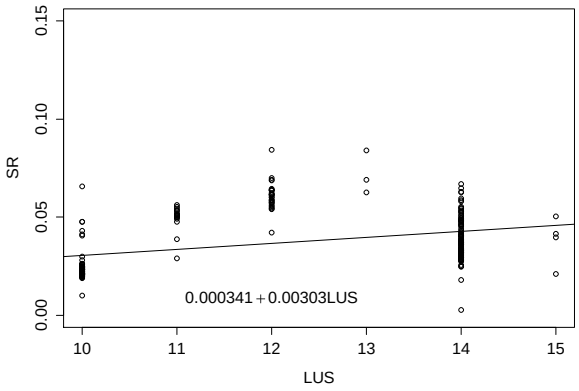
This work was supported in part by grant 2011-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. We gratefully acknowledge the comments of Catherine McGovern, Stuart Cooper and two anonymous reviewers which have greatly improved this paper.

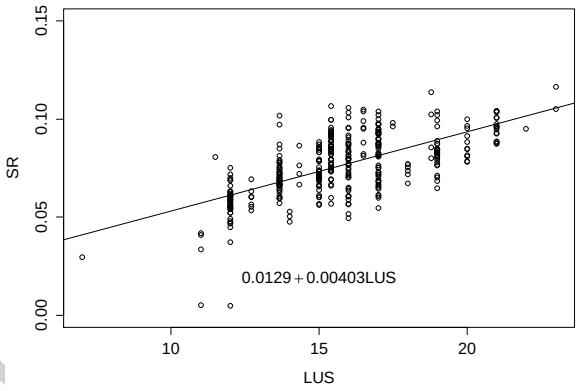
Appendix 1 Stutter ratio vs *LUS* for individual NGM SElect™ loci

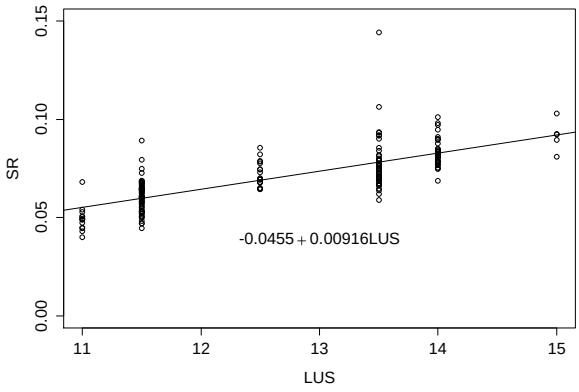


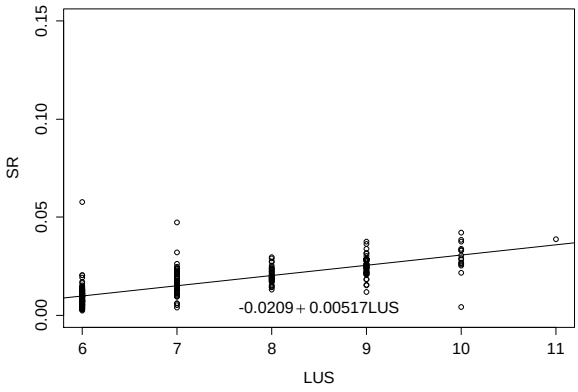

D2S1338
SR
LUS
-0.00734 + 0.00617LUS

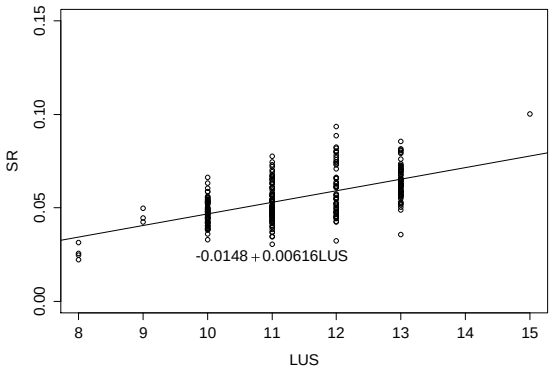

FGA
SR
LUS
-0.0345 + 0.00651LUS

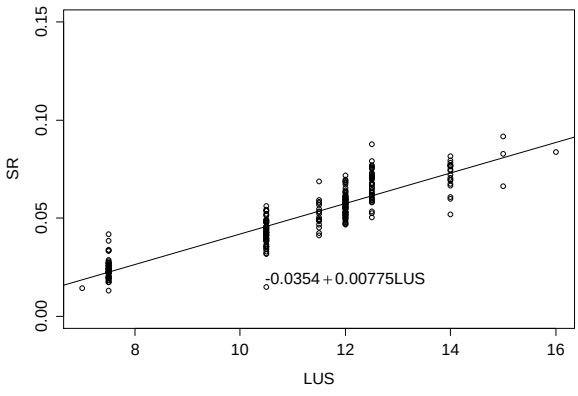

D2S441
SR
LUS
0.000341 + 0.00303LUS


SE33
SR
LUS
0.0129 + 0.00403LUS


D3S1358
SR
LUS
-0.0455 + 0.00916LUS


TH01
SR
LUS
-0.0209 + 0.00517LUS


D8S1179
SR
LUS
-0.0148 + 0.00616LUS


vWA
SR
LUS
-0.0354 + 0.00775LUS

Appendix 2 Summary of the stutter model $SR_i = \beta_{0,l} + \beta_{1,l} LUS_i$

Locus	Intercept	Slope
D10S1248	-0.0576	0.0089
D12S391	-0.0571	0.0107
D16S539	-0.0502	0.0088
D18S51	-0.0297	0.0066
D19S433	-0.0302	0.0074
D1S1656	-0.0699	0.0106
D21S11	-0.0079	0.0059
D22S1045	-0.0881	0.0139
D2S1338	-0.0073	0.0062
D2S441	0.0004	0.0031
D3S1358	-0.0455	0.0092
D8S1179	-0.0148	0.0062
FGA	-0.0344	0.0066
SE33	0.0129	0.0041
TH01	-0.0208	0.0052
vWA	-0.0354	0.0078

References

1. T.M. Clayton and J.S. Buckleton, Mixtures, in Forensic DNA Evidence Interpretation. 2004, CRC Press: Boca Raton, Florida. p. 217-74.
2. R.G. Cowell, S.L. Lauritzen and J. Mortera, Probabilistic modelling for DNA mixture analysis. Forensic Science International: Genetics Supplement Series, 2008. 1(1): p. 640-2.
3. M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, et al., Validating TrueAllele® DNA Mixture Interpretation. Journal of Forensic Sciences, 2011. 56(6): p. 1430-1447.
4. J.-A. Bright, P. Gill and J. Buckleton, Composite profiles in DNA analysis. Forensic Science International: Genetics, 2012. 6(3): p. 317-321.
5. H. Kelly, J.-A. Bright, J. Curran and J. Buckleton, The interpretation of low level DNA mixtures. Forensic Science International: Genetics, 2012. 6(2): p. 191-197.
6. J. Buckleton and C.M. Triggs, Is the 2p rule always conservative? Forensic Science International, 2006. 159: p. 206-209.
7. J.S. Buckleton, C.M. Triggs and S.J. Walsh, Forensic DNA Evidence Interpretation. 2004, Boca Raton, Florida: CRC Press.
8. H. Haned, Forensim: An open-source initiative for the evaluation of statistical methods in forensic genetics. Forensic Science International: Genetics, 2011. 5(4): p. 265-268.
9. H. Haned and P. Gill, Analysis of complex DNA mixtures using the Forensim package. Forensic Science International: Genetics Supplement Series, 2011. 3(1): p. e79-e80.

10. D.J. Balding and J. Buckleton, Interpreting low template DNA profiles. Forensic Science International: Genetics, 2009. 4(1): p. 1-10.
11. K.E. Lohmueller and N. Rudin, Calculating the Weight of Evidence in Low-Template Forensic DNA Casework. Journal of Forensic Sciences, 2012: p. n/a-n/a.
12. I.W. Evett, P.D. Gill and J.A. Lambert, Taking account of peak areas when interpreting mixed DNA profiles. Journal of the Forensic Sciences, 1998. 43(1): p. 62-69.
13. L. Geddes, What are the chances? The New Scientist, 2010. 207(2774): p. 8-10.
14. L. Geddes, C. King and C. Stier, Between Prison and Freedom. The New Scientist, 2010. 207(2773): p. 8-11.
15. J.-A. Bright, J. Turkington and J. Buckleton, Examination of the variability in mixed DNA profile parameters for the Identifiler™ multiplex. Forensic Science International: Genetics, 2009. 4(2): p. 111-114.
16. J.-A. Bright, E. Huizing, L. Melia and J. Buckleton, Determination of the variables affecting mixed MiniFiler™ DNA profiles. Forensic Science International: Genetics, 2011. 5(5): p. 381-385.
17. C. Brookes, J.-A. Bright, S. Harbison and J. Buckleton, Characterising stutter in forensic STR multiplexes. Forensic science international. Genetics, 2012. 6(1): p. 58-63.
18. P.S. Walsh, N.J. Fildes and R. Reynolds, Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Res., 1996. 24: p. 2807-2812.
19. J.M. Butler and D.J. Reeder. Short Tandem Repeat DNA Internet DataBase. Available from: www.cstl.nist.gov/biotech/strbase.
20. C.M. Ruitberg, D.J. Reeder and J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Research, 2001. 29(1): p. 320 - 322.
21. R Development Core team, R: A language and environment for statistical computing. 2004, Vienna, Austria: R Foundation for Statistical Computing.
22. M.W. Perlin, G. Lancia and S.-K. Ng, Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. American Journal of Human Genetics, 1995. 57: p. 1199-1210.
23. S.C. Bakker, R.J. Sinke and P.L. Pearson, Differences in stutter intensities between microsatellites are related to length and sequence of the repeat, in Unravelling the genetics of schizophrenia and ADHD. 2005: University of Utrecht: Utrecht. p. 81-94.
24. J.-A. Bright, E. Huizing, L. Melia and J. Buckleton, Determination of the variables affecting mixed MiniFiler(TM) DNA profiles. Forensic Science International: Genetics, 2011. 5(5): p. 381-385.
25. J. Weusten and J. Herbergs, A stochastic model of the processes in PCR based amplification of STR DNA in forensic applications. Forensic Science International: Genetics, 2012. 6: p. 17-25.
26. H. Kelly, J.-A. Bright, J.M. Curran and J. Buckleton, Modelling heterozygote balance in forensic DNA profiles. Forensic Science International: Genetics, 2012. 6(6): p. 729-734.

27. J.M. Curran, A MCMC method for resolving two person mixtures. Science & Justice, 2008. 48(4): p. 168-177.
28. R.G. Cowell, S.L. Lauritzen and J. Mortera, Probabilistic expert systems for handling artifacts in complex DNA mixtures. Forensic Science International: Genetics, 2011. 5(3): p. 202-209.
29. R.G. Cowell, Validation of an STR peak area model. Forensic Science International: Genetics, 2009. 3(3): p. 193-199.