Biological Models

ISFG Advanced mixture interpretation workshop

Jo-Anne Bright

Specialist Science Solutions

Manaaki Tangata Taiao Hoki
protecting people and their environment through science

Scope

- What do we know about the biology of DNA profiles?
- How can this inform interpretation models?
- How does knowing expected peak heights help?

1. Heterozygote balance

Forensic Science International: Genetics

Volume 4, Issue 2, February 2010, Pages 111–114

Examination of the variability in mixed DNA profile parameters for the Identifiler™ multiplex

Jo-Anne Bright, Jnana Turkington, John Buckleton

ESR, 120 Mt Albert Road, PB 92021, Auckland, New Zealand

Forensic Science International: Genetics

Volume 6, Issue 6, December 2012, Pages 729–734

Analysis and biostatistical interpretation of complex and low template DNA samples

Modelling heterozygote balance in forensic DNA profiles

Hannah Kelly[a, b], Jo-Anne Bright[a], James M. Curran[b], John Buckleton[a]

[a] ESR, PB 92021, Auckland, New Zealand

[b] Department of Statistics, University of Auckland, PB 92019, Auckland, New Zealand

Heterozygote balance

- Hb is used to:
  - Inform number of contributors to a profile
  - Restrict possible genotype combinations in a mixed DNA profile
- Important to assess bounds on Hb
- Hb rules are based on the expected height variance between a pair of alleles in a heterozygote
- Traditionally, applied across a profile

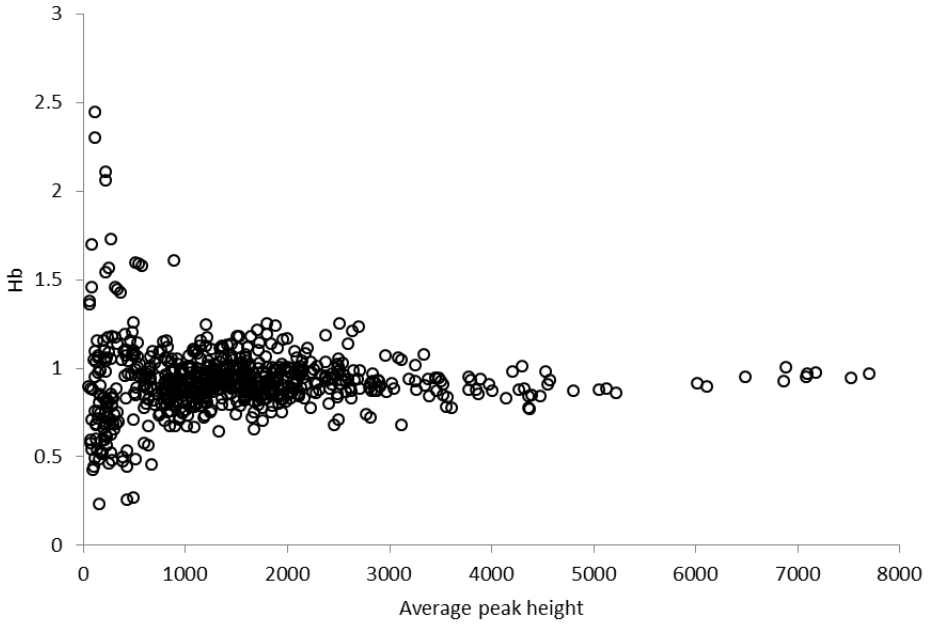
Definition of heterozygote balance

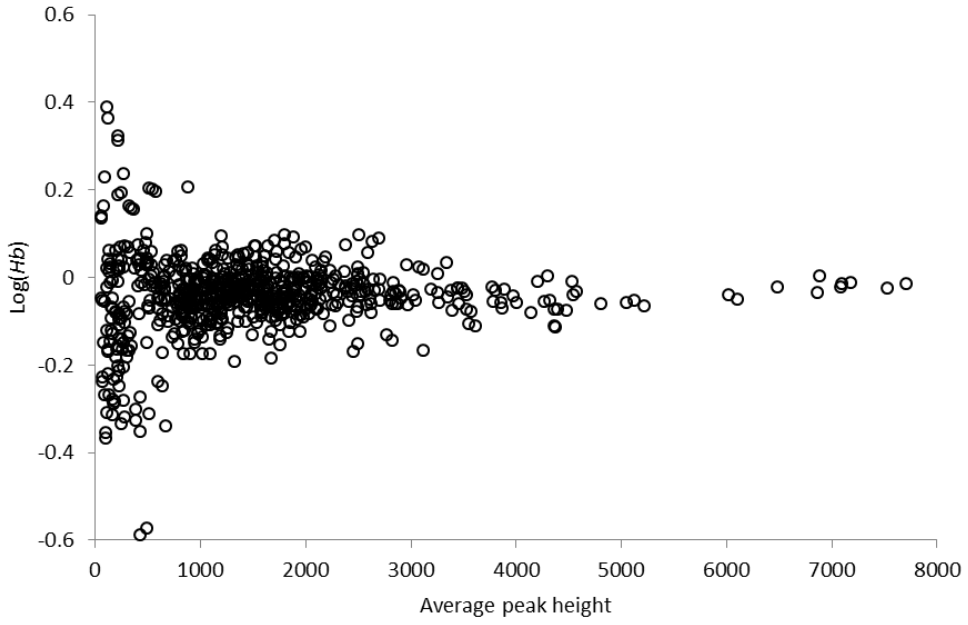
- **Two definitions of heterozygote balance or peak height ratio:**

$$Hb_1 = \frac{O_{HMW}}{O_{LMW}} \qquad Hb_2 = \frac{O_{\text{smaller}}}{O_{\text{larger}}}$$

- **Where *O* is observed peak height**
- **Hb_1 has the highest information content because it maintains peak order**
- **Hb_2 may be obtained from Hb_1 but not vice versa**

Hb versus average peak height

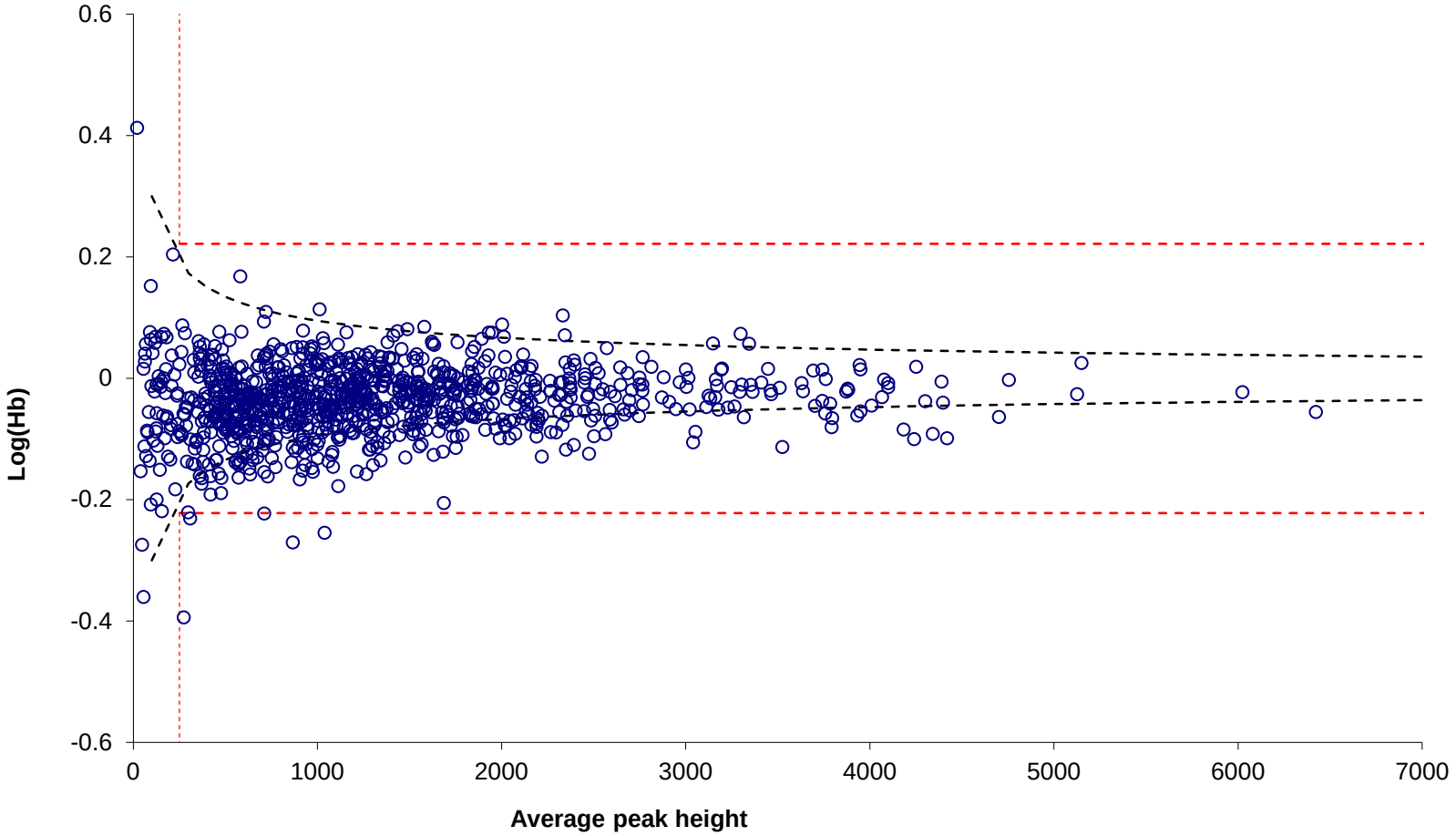




Variability of Hb

Conventional thresholds

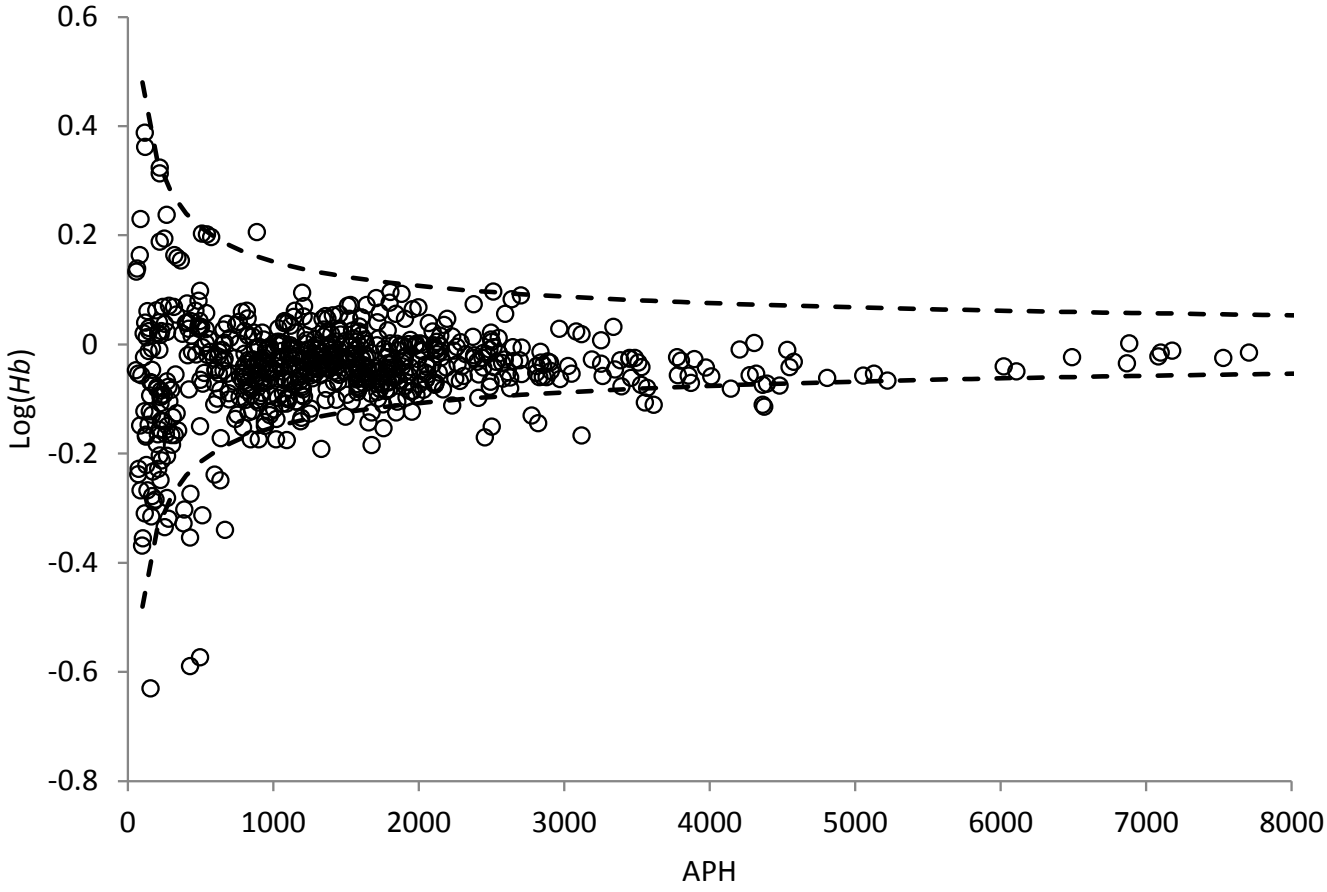
95% intervals



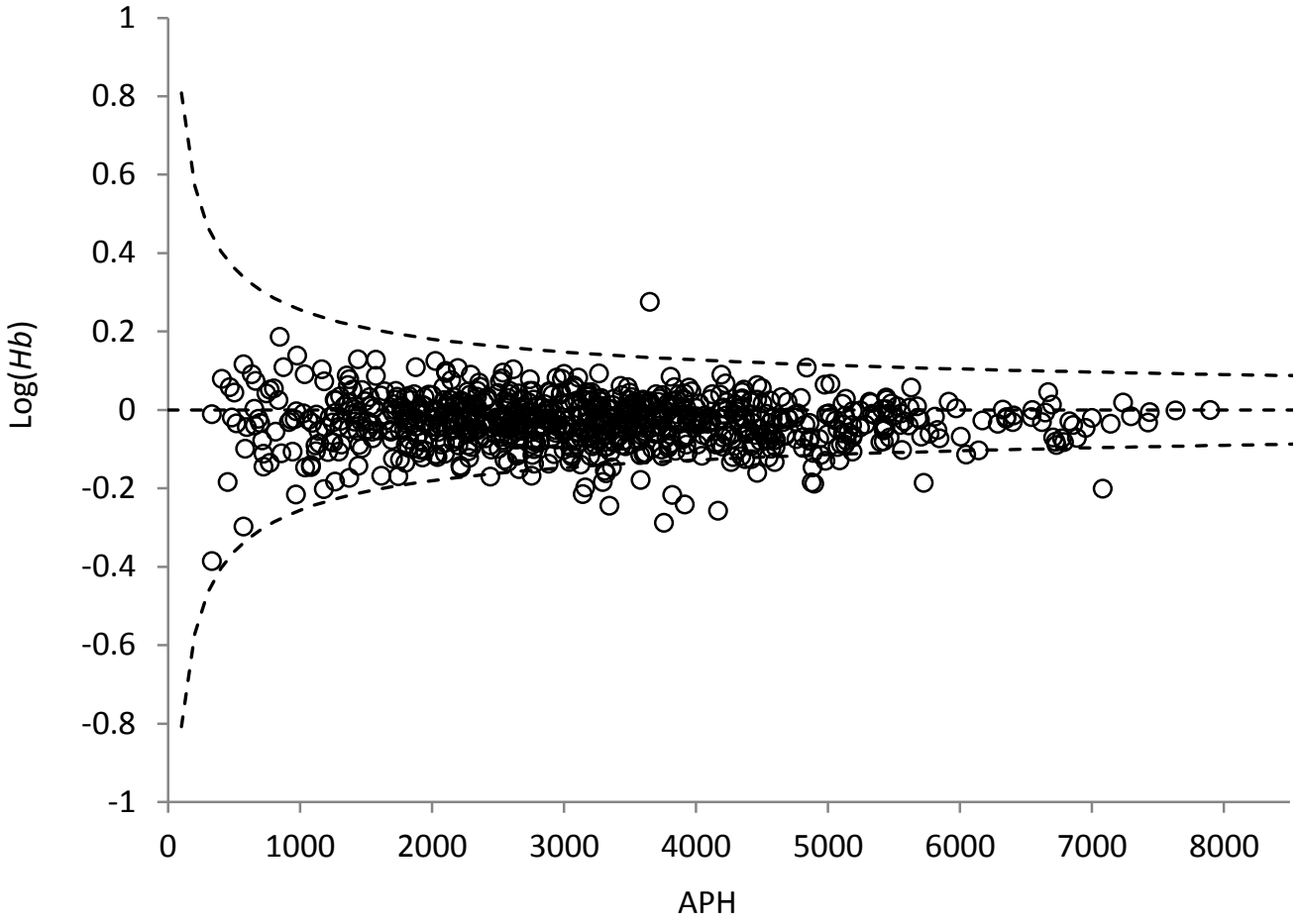
Conclusion

- The mean of heterozygote balance is unaffected by average peak height
- The variance about this mean is much lower at high average peak heights
- This is true over multiple kits and PCR cycle numbers

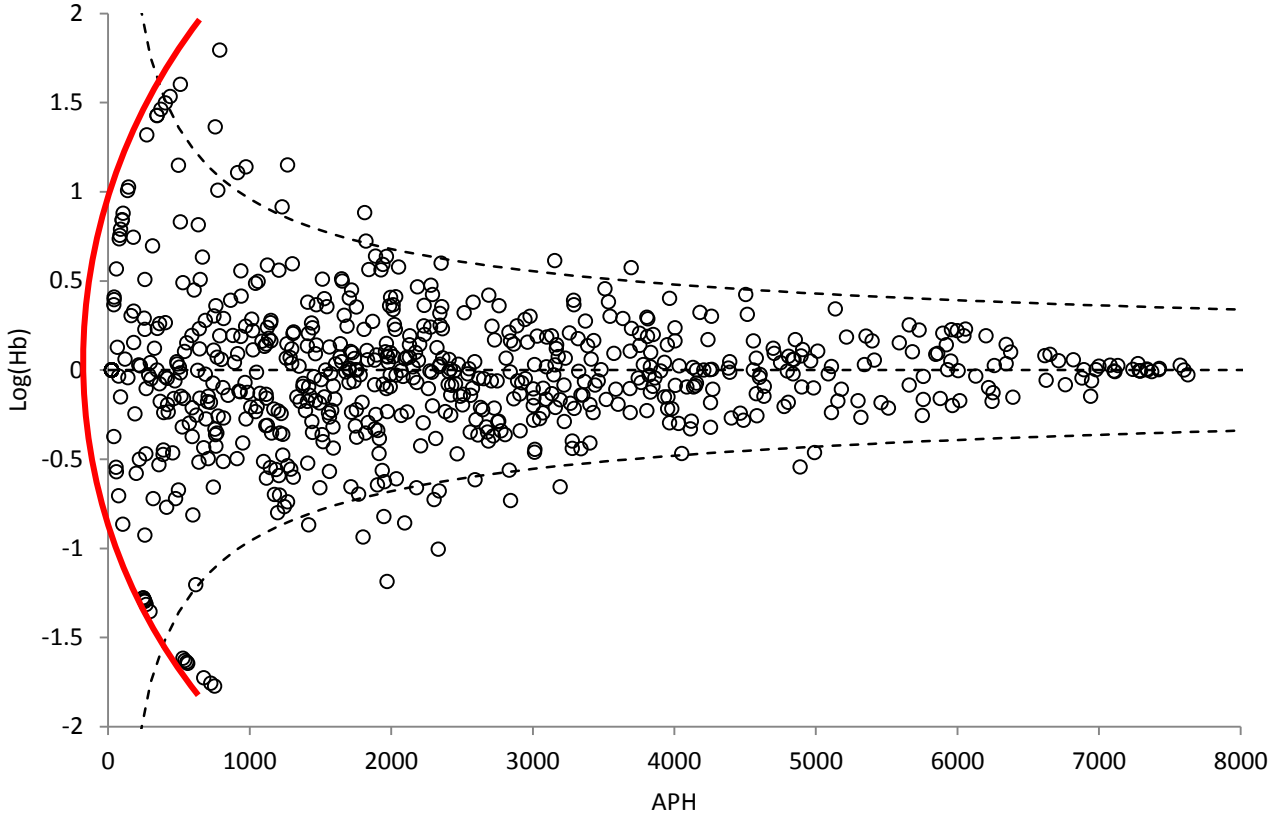
Identifiler 28 cycles



NGM SElect 29 cycles



SGMPlus 34 cycles

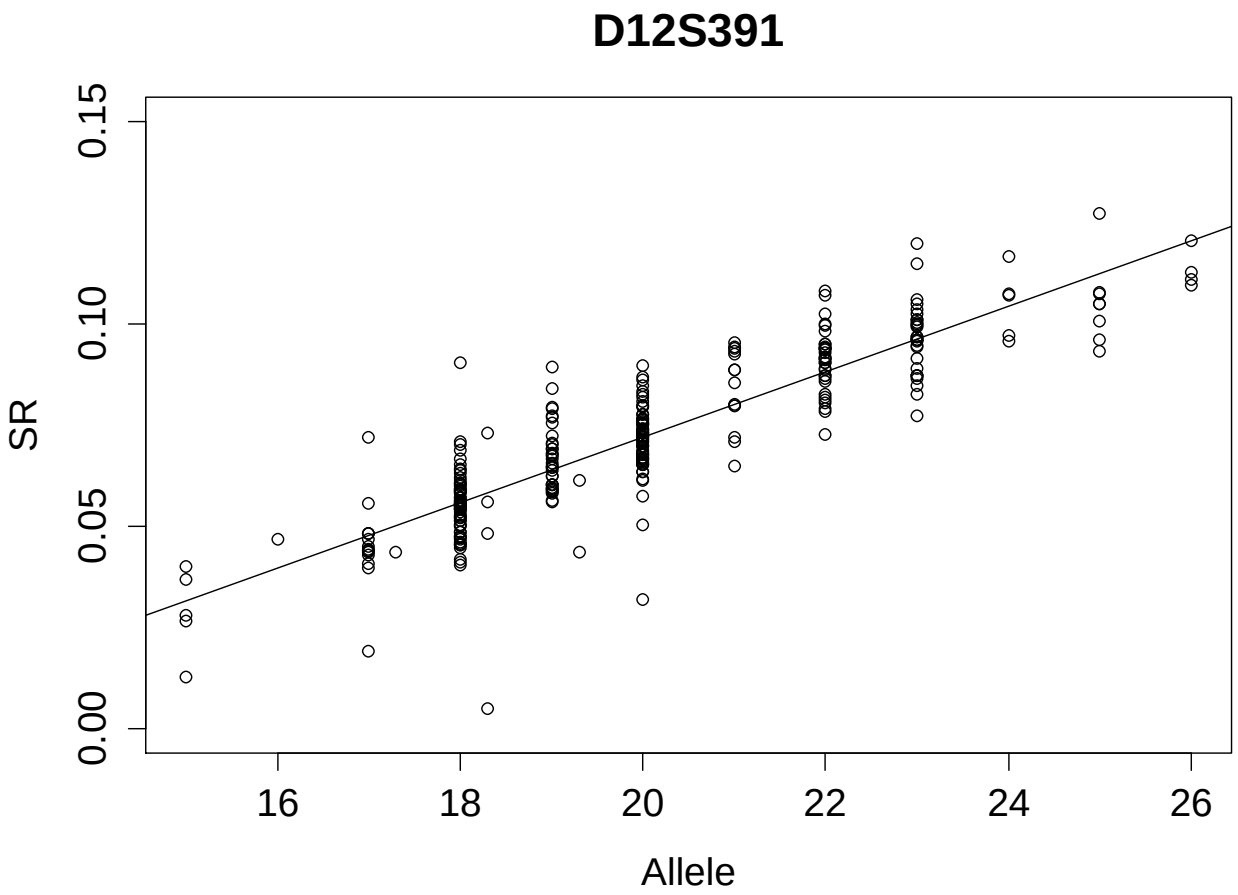


2. Stutter ratios

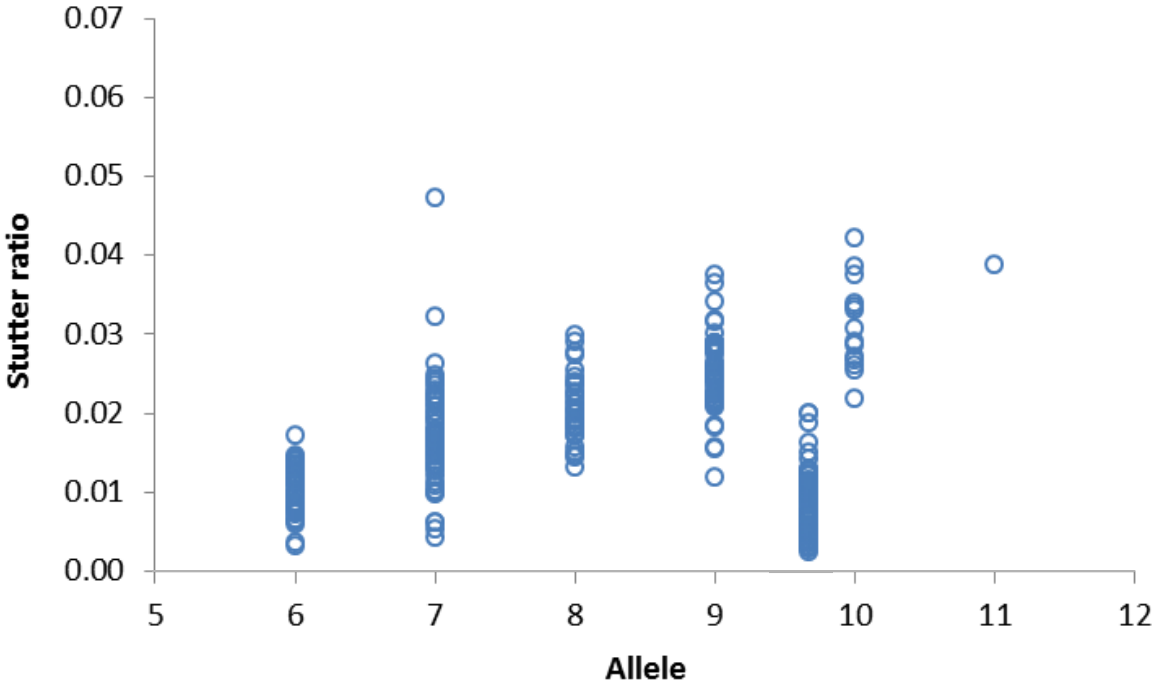
- Traditionally we apply a threshold at analysis to remove stutter
  - Locus specific
  - Kit specific
- What if your minor POI was approximately same RFU as stutter?
- Is removing stutter peaks conservative?
- What if a stutter peak was actually allelic and excluded your POI?

Stutter ratios

- **Stutter ratios are actually allele specific**



TH01 stutter



ELSEVIER

Forensic Science International: Genetics

Volume 6, Issue 1, January 2012, Pages 58–63

Characterising stutter in forensic STR multiplexes

Clare Brookes[a], Jo-Anne Bright[b], SallyAnn Harbison[b], John Buckleton[b,]

[a] Department of Chemistry, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

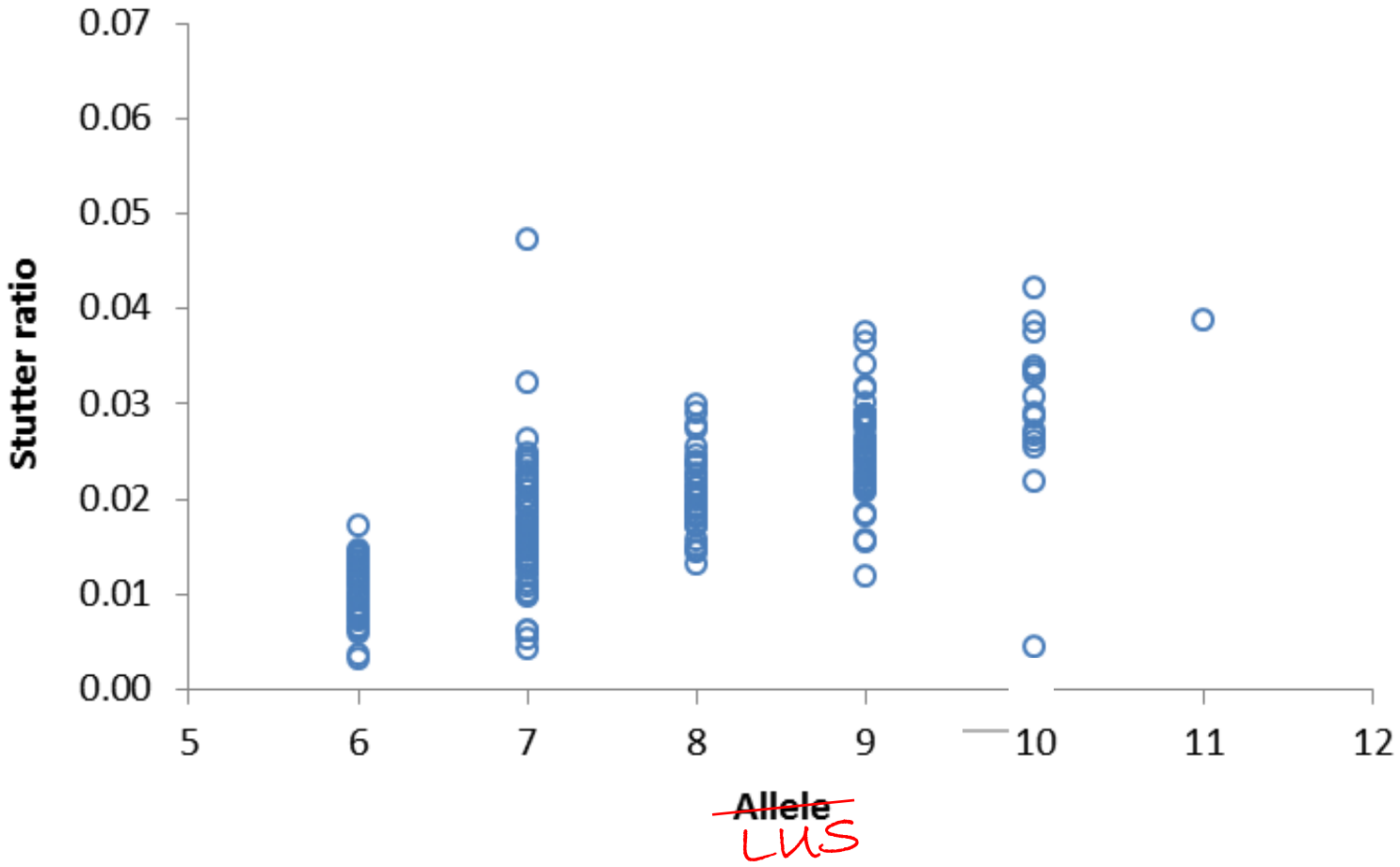
[b] Institute of Environmental Science and Research Ltd, Private Bag 92021, Auckland 1142, New Zealand

TH01 repeat structure

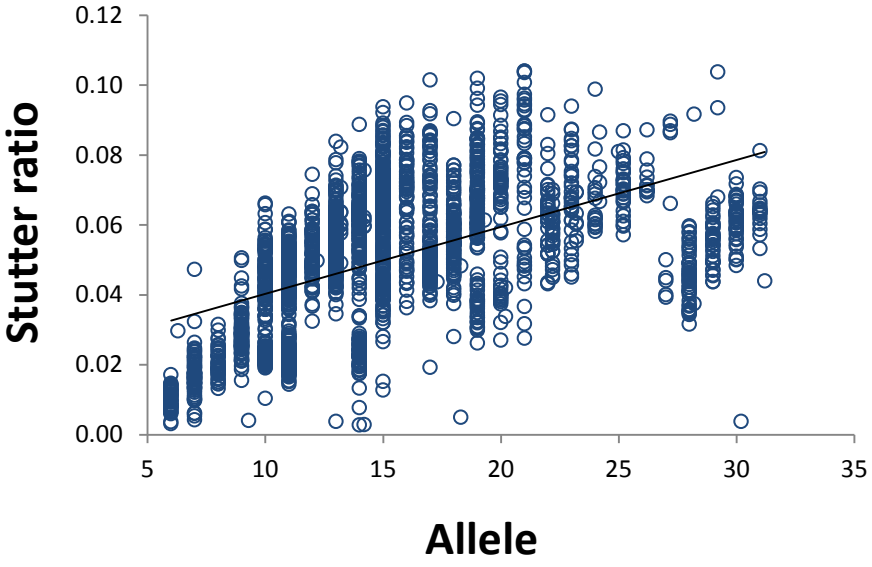
Common TH01 allele sequences		
Repeat structure	Allele	LUS
$[AATG]_6$	6	6
$[AATG]_7$	7	7
$[AATG]_8$	8	8
$[AATG]_9$	9	9
$[AATG]_6ATG[AATG]_3$	9.3	6

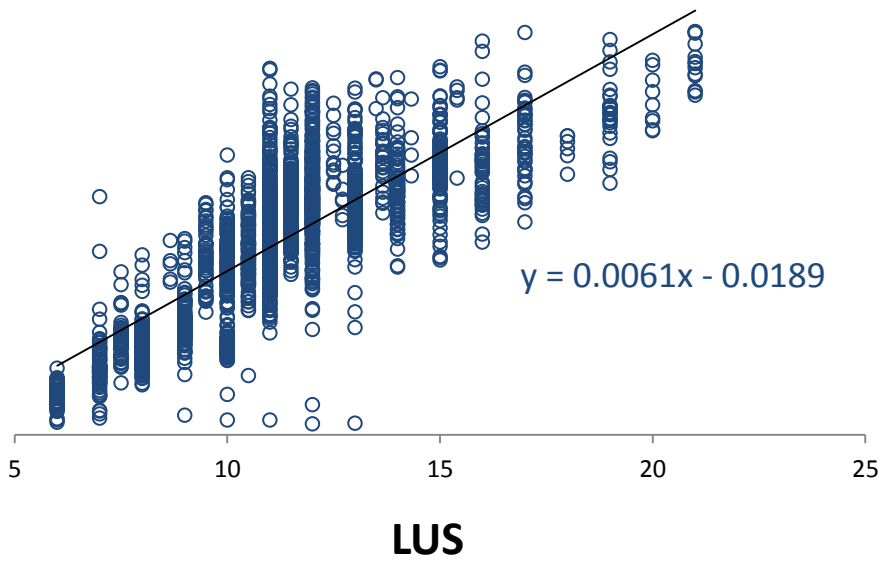
Longest uninterrupted stretch of basic repeat motifs is a good predictor of stutter ratio

TH01 Stutter ratio versus LUS



Allele versus LUS, NGM Select loci





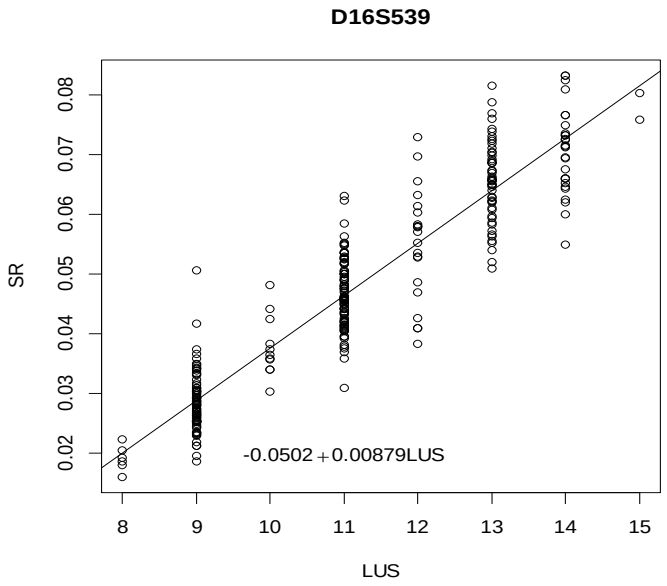
$R^2 = 27\%$

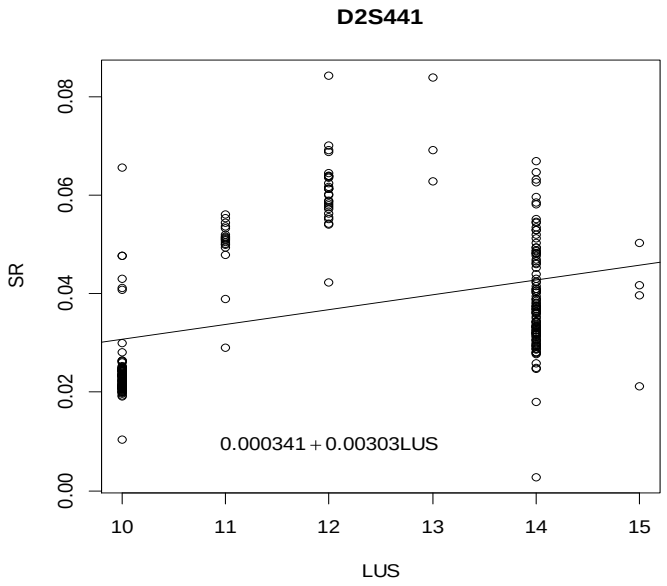
$R^2 = 61\%$

Stutter model

$$SR = m\mathrm{LUS} + c$$

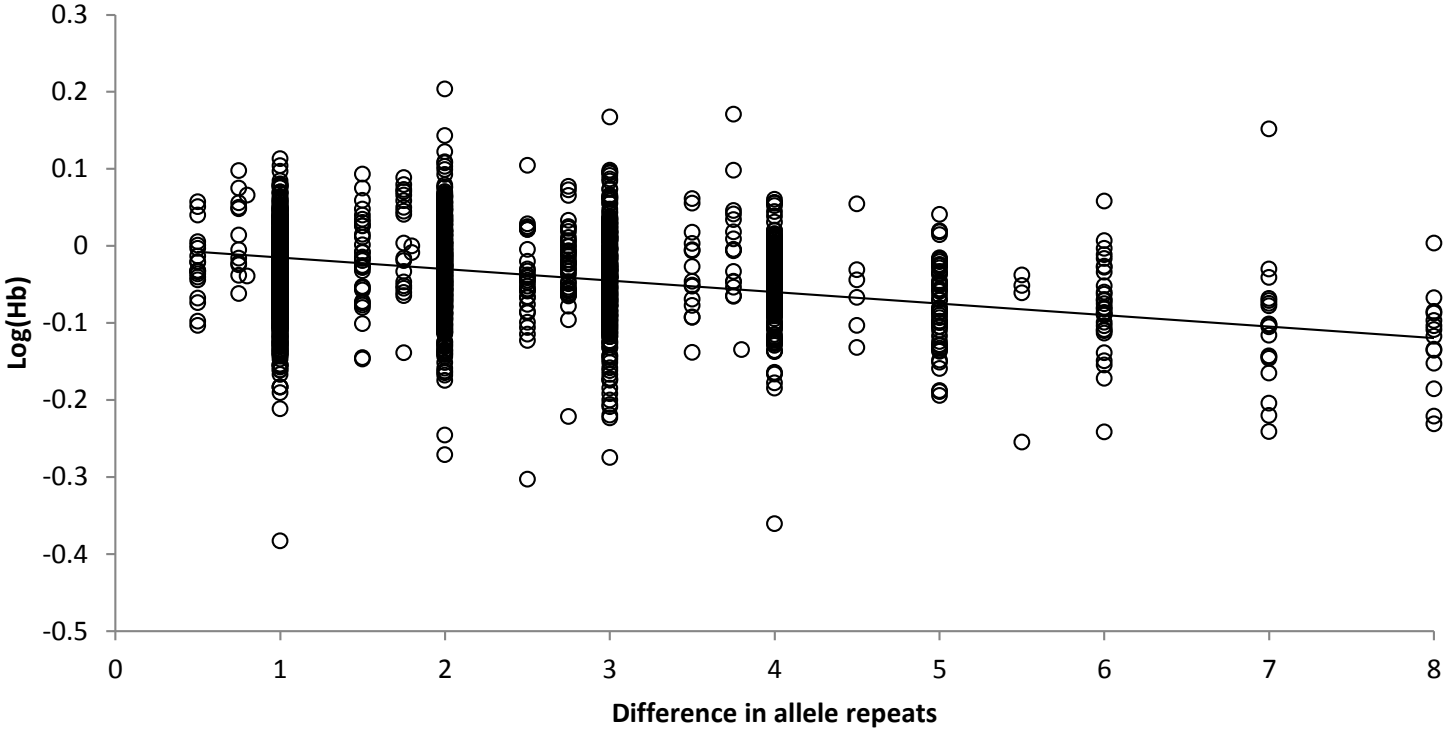
- **Values for slope and intercept can be determined for each marker using regression**





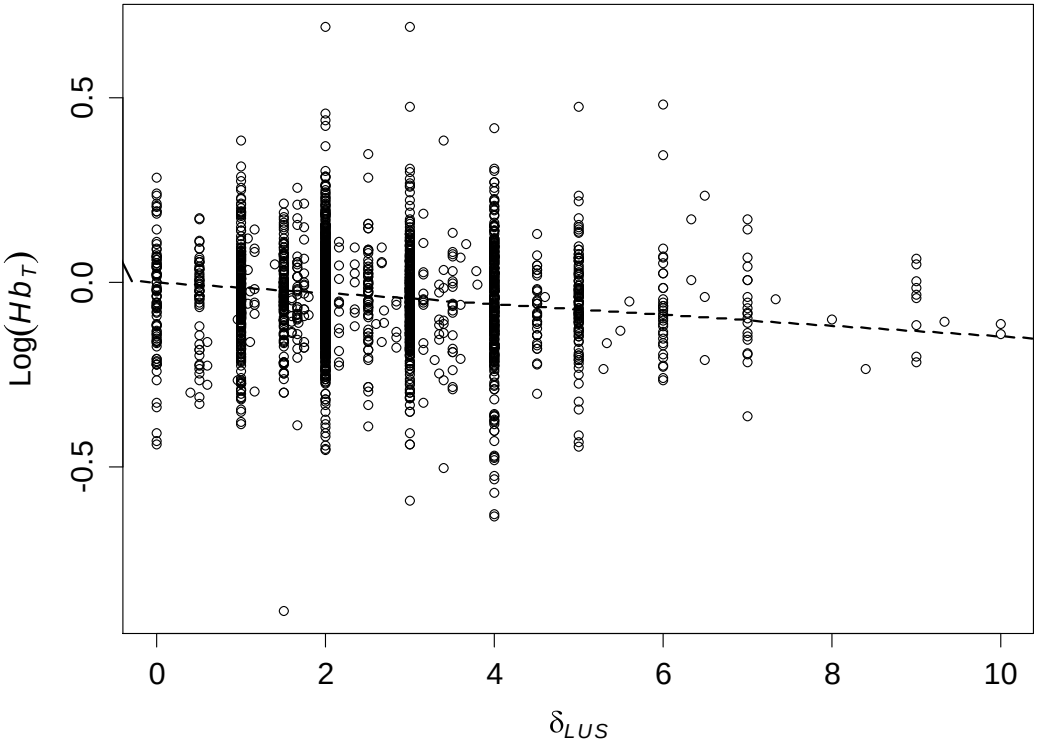
Stutter effect on profile slope

- Longer alleles stutter more.
- Is this the cause of observed general decreases in profile slope?



Stutter effect on profile slope

- Taking into account stutter by calculating total allelic product there's still a small but significant negative slope
- Likely to be simply due to the reduced amplification efficiency of the larger allele at a heterozygote locus



3. Profile slopes

Forensic Science International: Genetics

Volume 6, Issue 1, January 2012, Pages 97–101

Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out

Torben Tvedebrink[a], Poul Svante Eriksen[a, 1], Helle Smidt Mogensen[b, 2], Niels Morling[b, 3]

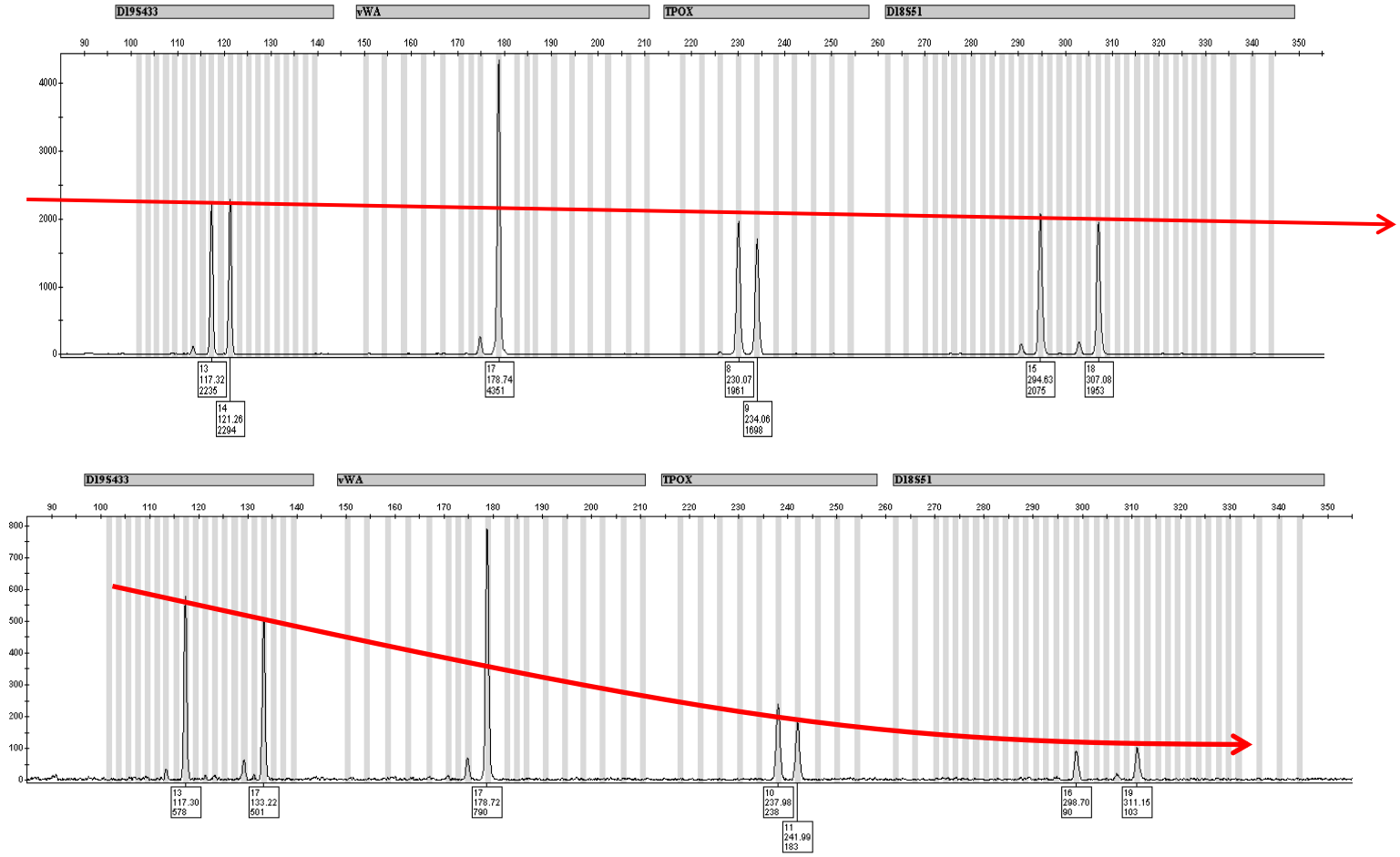
Australian Journal of Forensic Sciences

Degradation of forensic DNA profiles

DOI: 10.1080/00450618.2013.772235

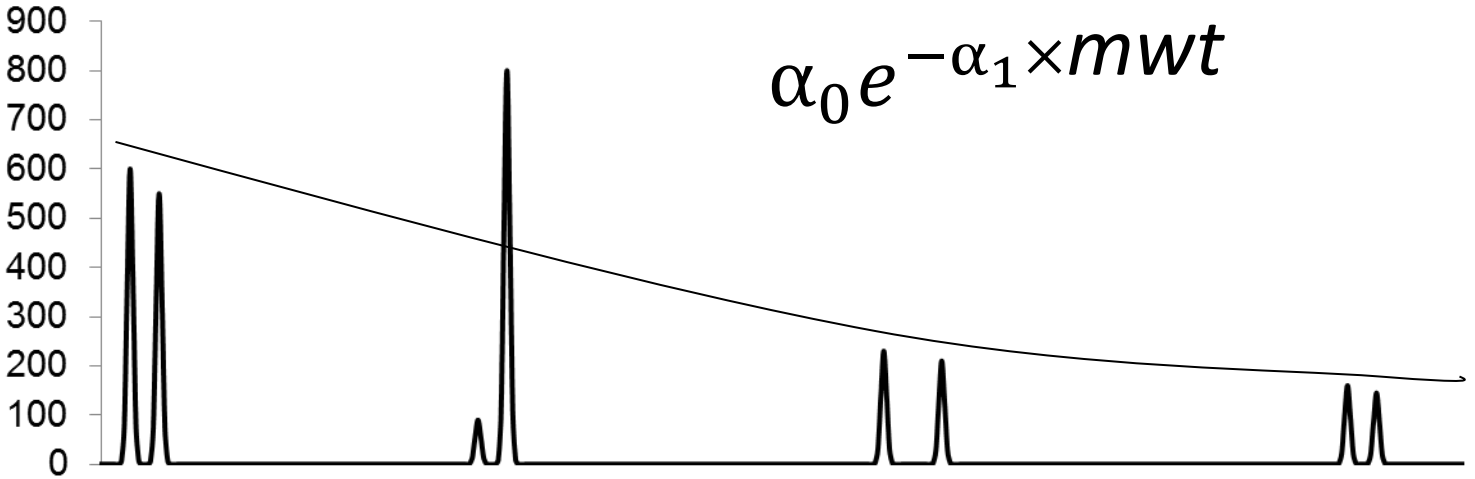
Jo-Anne Bright[ab*], Duncan Taylor[c], James M. Curran[b] & John S. Buckleton[a]

Degradation slopes



Degradation curve

- **Empirical data has shown that for larger multiplexes a DNA slope is best described by an exponential curve**

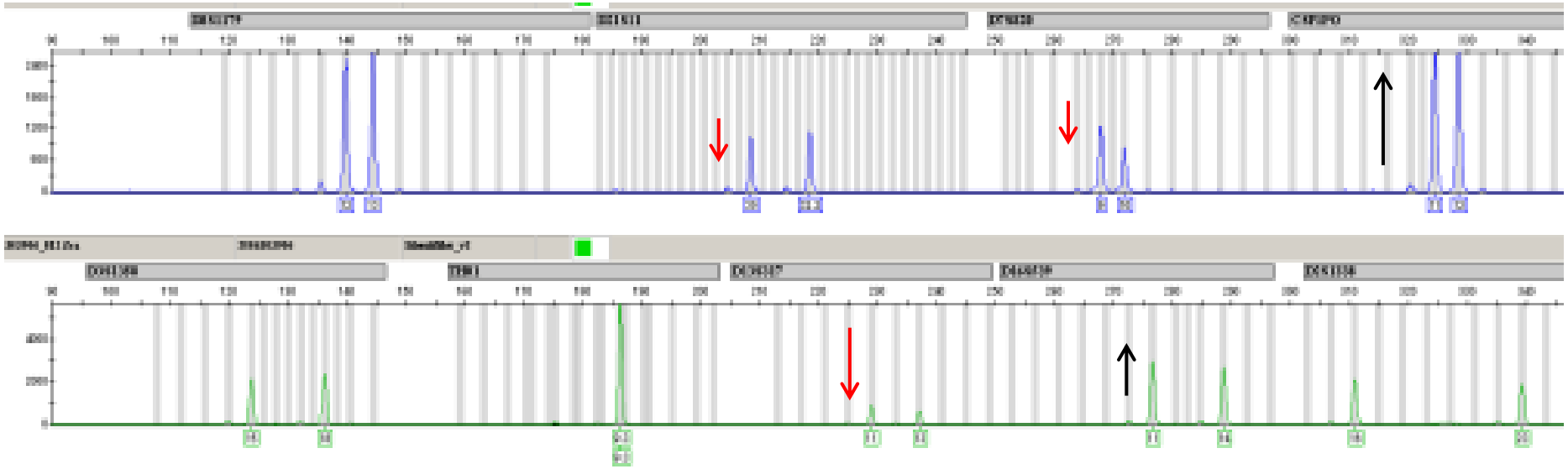


- **Equation describes an exponential curve, intercept α_0, slope α_1 decreasing with molecular weight**

4. Locus specific amplification

- Observation that some loci amplify more efficiently than others
- Results in varying peak heights off the general trend
- Locus offset at each locus allows for this variation

Locus specific amplification example



A biological model – an example

ELSEVIER

Forensic Science International: Genetics

Volume 7, Issue 2, February 2013, Pages 296–304

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

Jo-Anne Bright[a, b], Duncan Taylor[c], James M. Curran[b], John S. Buckleton[a]

[a] ESR Ltd, Private Bag 92021, Auckland, New Zealand

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

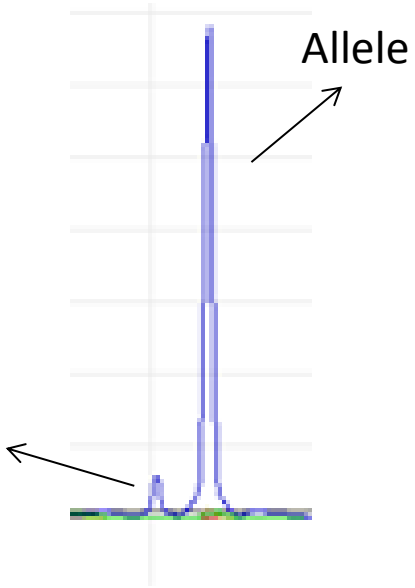
[c] Forensic Science South Australia, 21 Divett Place, SA 5000, Australia

A biological model – an example

- A model that calculates the expected heights of allelic an stutter peaks
- Takes into account:
  - Stutter
  - Degradation
  - Locus effects
- Informed by empirical data
- For use within a continuous method of DNA interpretation

Total allelic product

- 'True' (but unknown) amount of template DNA
- PCR product: allele plus stutter peak heights
- Model template DNA based on our observations:
  - Height of peaks from a single contributor is approximately constant across loci
  - Generally trends downwards with increasing molecular weight
  - Slope may vary between contributors (i.e. degrade at different rates)
  - Individual loci may still be above or below the trend



Modelling total allelic product

- **Mass of an allele at a locus is modelled by the mass parameters:**
  - **Slope d_n (degradation) and intercept t_n (template)**
- **Mass decreases with increasing molecular weight of an allele at a locus (m^l_a)**
- **Locus offset at each locus A^l (locus specific amplification efficiency)**

$$T^{\ell}_{an} = A^{\ell} t_n X^{\ell}_{an} \times e^{-d_n \times m^{\ell}_a}$$

Where X^l_{an} = dose, the count of allele a at locus l for contributor n:

Heterozygote = 1

Homozygote = 2

Peak height estimation

- **The total allelic product from an allele is divided into stutter and allelic peak heights**
- **The height of the stutter and allelic peaks formed from allele *a* contributor *n* are calculated by:**

Allele

$$E_{an}^{\ell} = \frac{T_{an}^{\ell}}{1 + SR_a^{\ell}}$$

Stutter

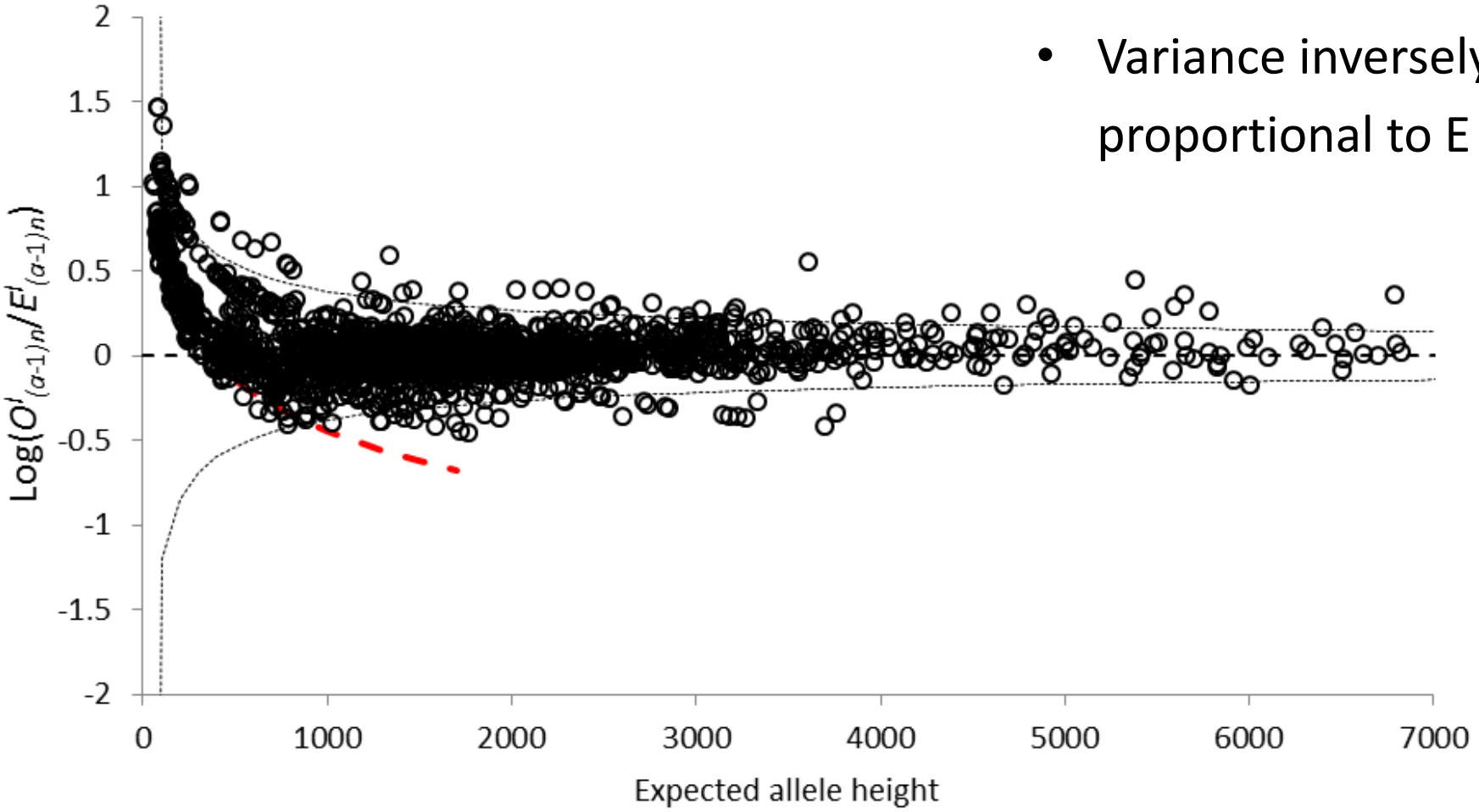
$$E_{(a-1)n}^{\ell} = \frac{SR_a^{\ell}\left(T_{an}^{\ell}\right)}{1 + SR_a^{\ell}}$$

Test of the model

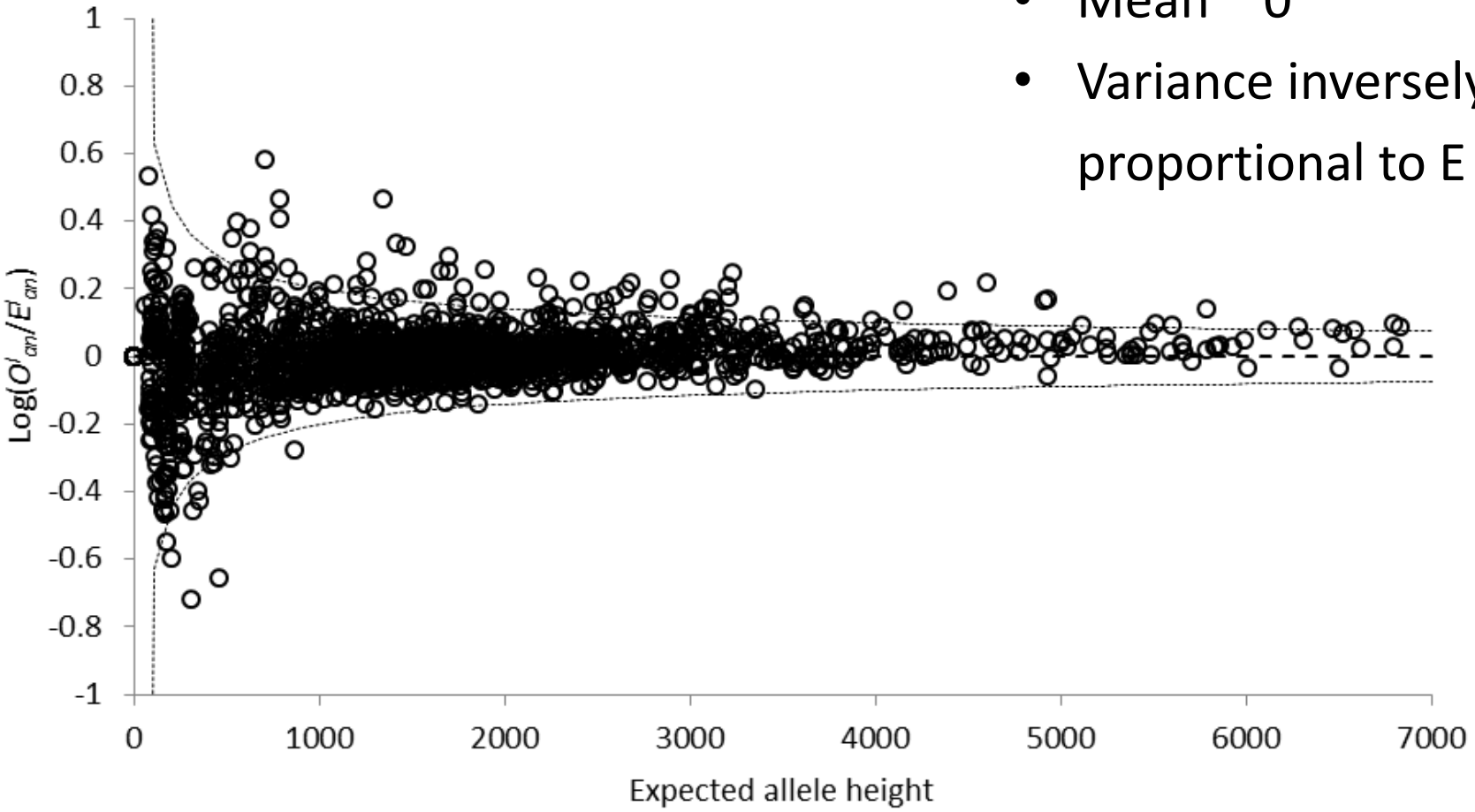
- 99 single source DNA profiles
- Applied Biosystems' Identifiler™ multiplex.
- 50 rfu analysis threshold
- Mass parameters estimated by MLE
- Total allelic product calculated
- Expected height of all allele and stutter peaks calculated
  - Applying the *LUS* model for stutter ratio

Variance of stutter model

- Mean ~ 0
- Variance inversely proportional to E



Variance of allele model



- Mean ~ 0
- Variance inversely proportional to E

Model distribution

Assuming:

- **an approximate normal distribution,**
- **mean of zero,**
- **a variance = $\frac{c^2}{E_{an}^l}$ for the allele model,**
- **and a variance = $\frac{k^2}{E_{an}^l}$ for the stutter model, then:**

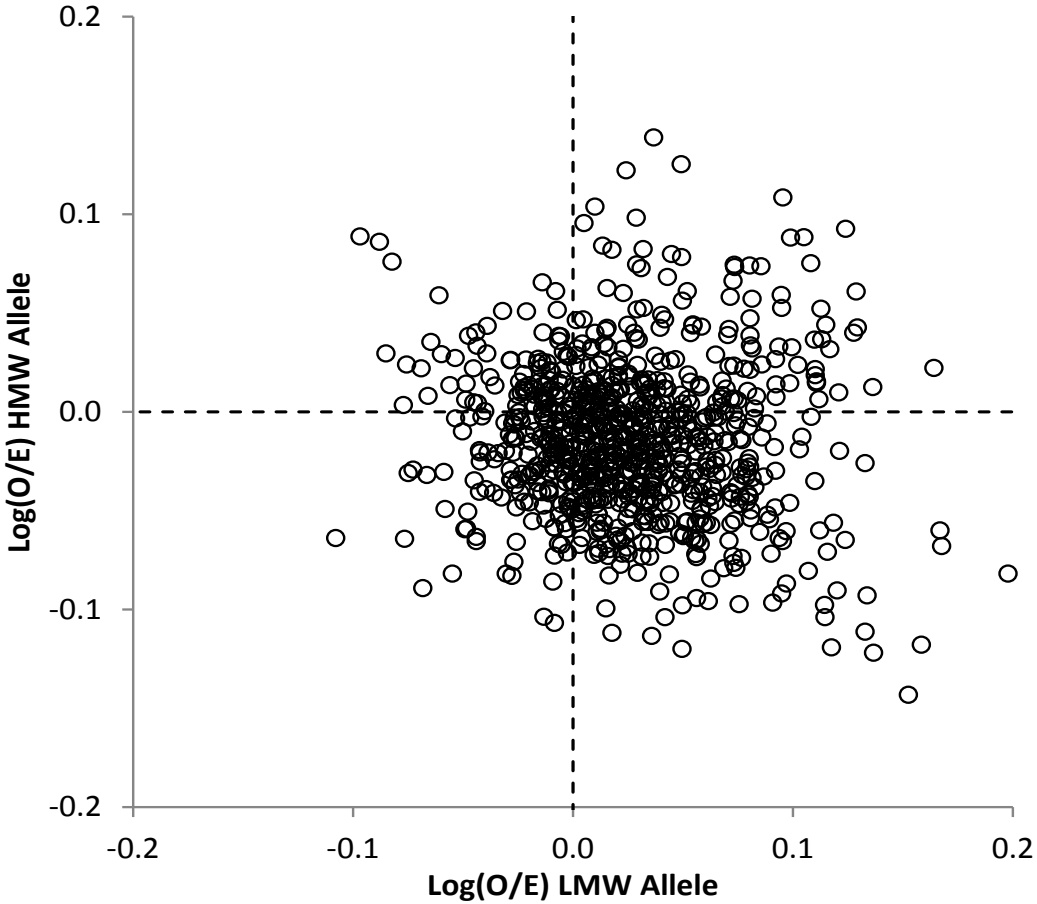
$$\log\left(\frac{O_{(a-1)}}{E_{(a-1)n}^l}\right) \sim N\left(0, \frac{k^2}{E_{an}^l}\right) \text{ for stutter}$$

$$\log\left(\frac{O_a}{E_{an}^l}\right) \sim N\left(0, \frac{c^2}{E_{an}^l}\right) \text{ for alleles}$$

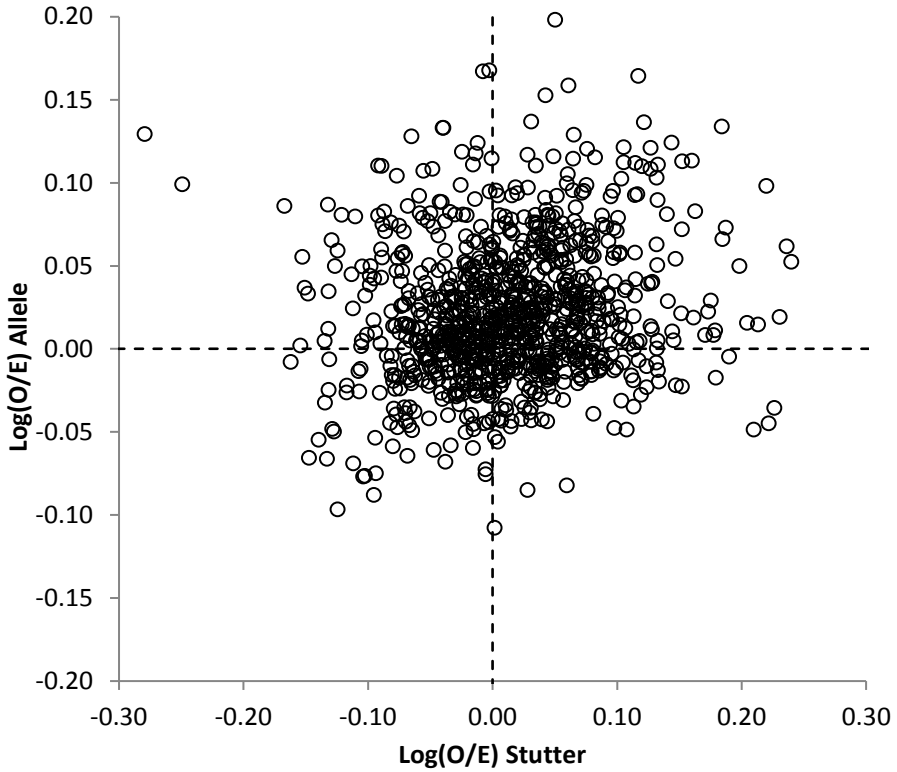
Assumption

- **Assumption of independence across alleles and stutter at a locus**
  - **i.e. peak heights in a profile are not correlated**
- **However, a larger than expected stutter peak is likely to be associated with a smaller than expected allelic peak**
  - **If stutter occurs early in PCR this results in increased stutter height at the detriment to the allele height**
- **For any given allele if the stutter peak is above expectation given the *LUS* we expect the allelic peak to be below expectation**

Log(O/E) HMW vs LMW Allele



Log(O/E) Allele vs Stutter



- **No detectable correlation between stutter and allele in the biological model**

Biological Models

ISFG Advanced mixture interpretation workshop

Jo-Anne Bright

Specialist Science Solutions

Manaaki Tangata Taiao Hoki
protecting people and their environment through science