

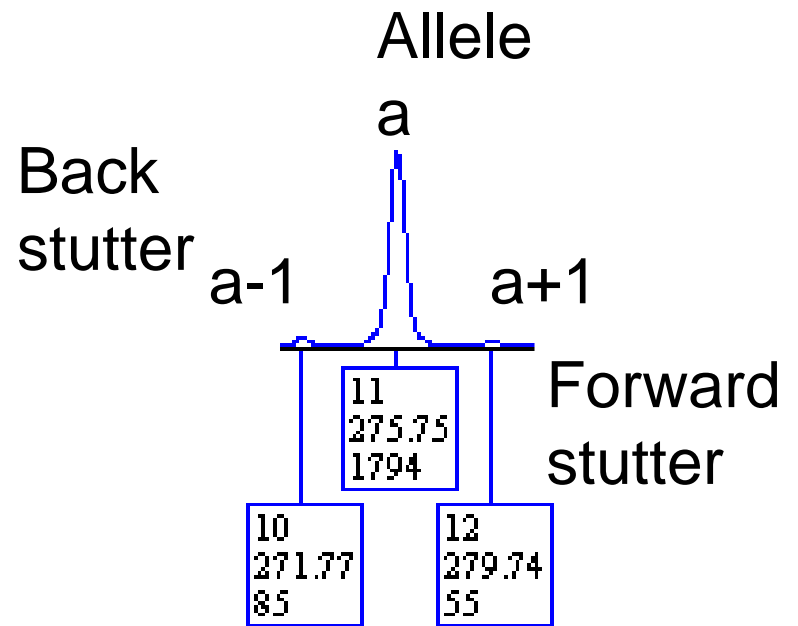
Understanding PCR

- Total allelic product
- Peak height
- Back stutter
- Forward stutter
- Variability in peak height

Total allelic product (TAP)

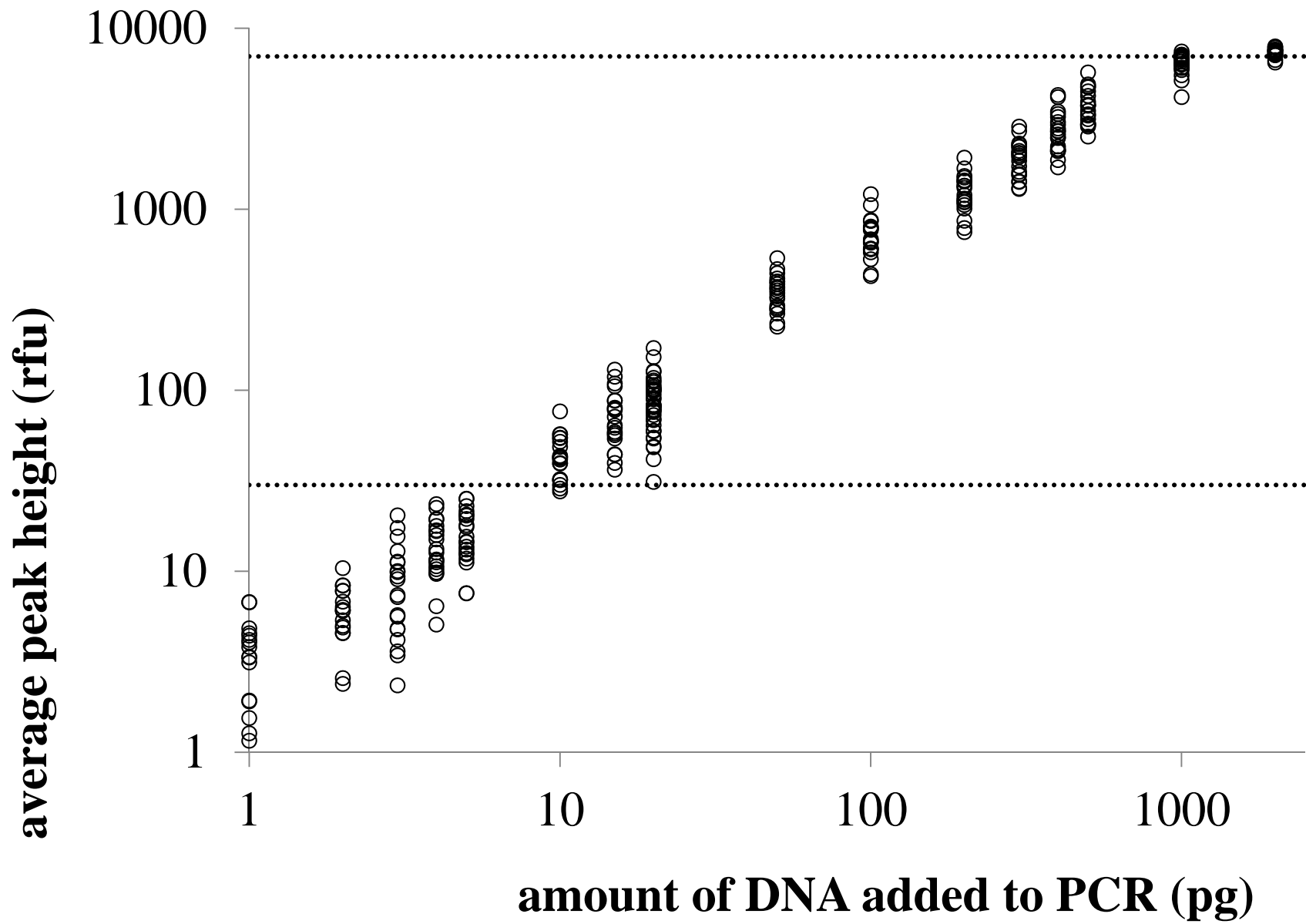
- Total allelic product, $T_a =$ height of allele plus height of stutter peaks

- $T_a = O_{a-1} + O_a + O_{a+1}$

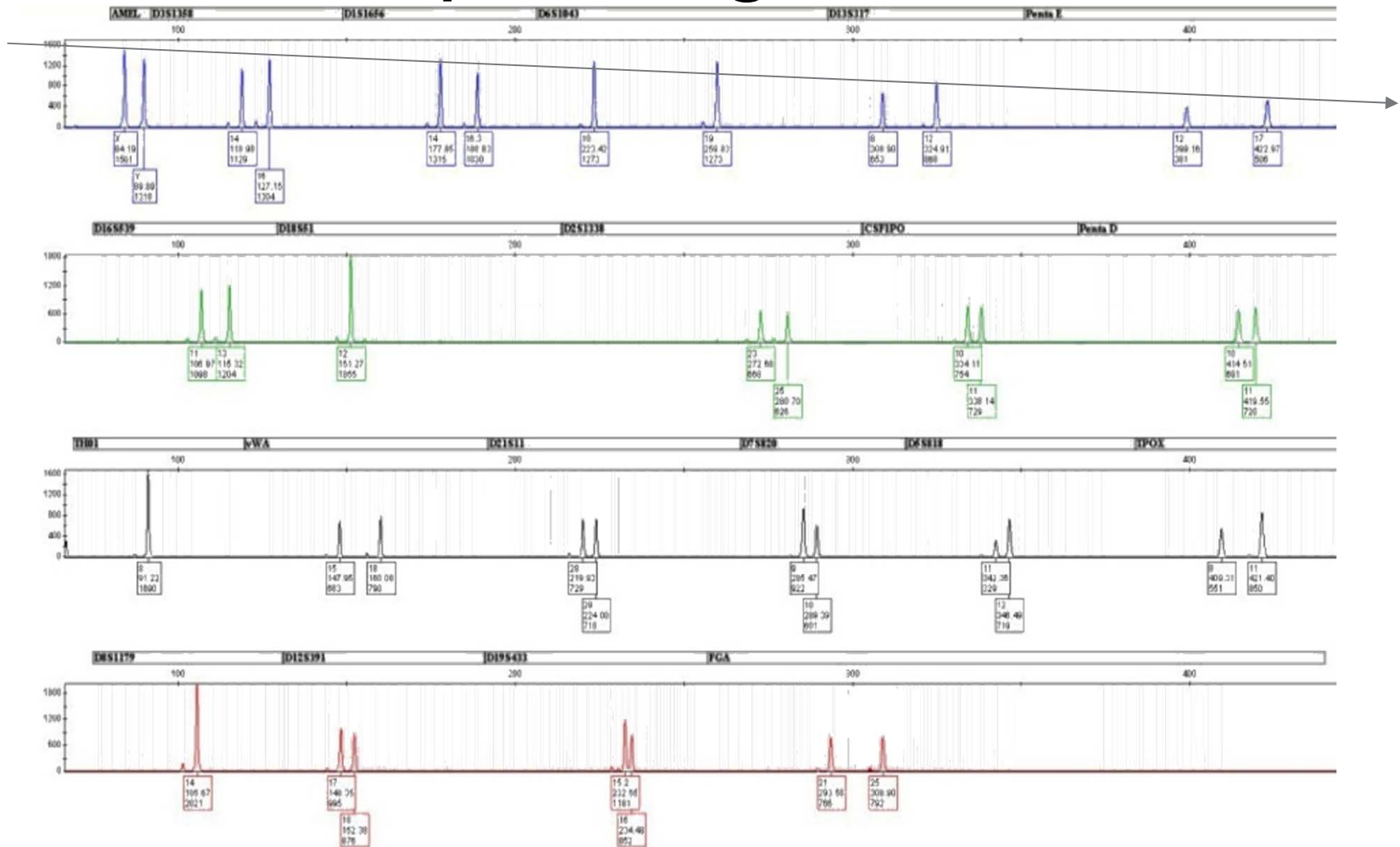


Effects on TAP

- Template
- Degradation
- Locus specific effects
- Additivity

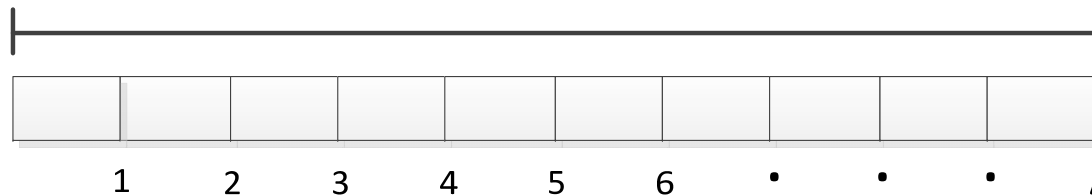


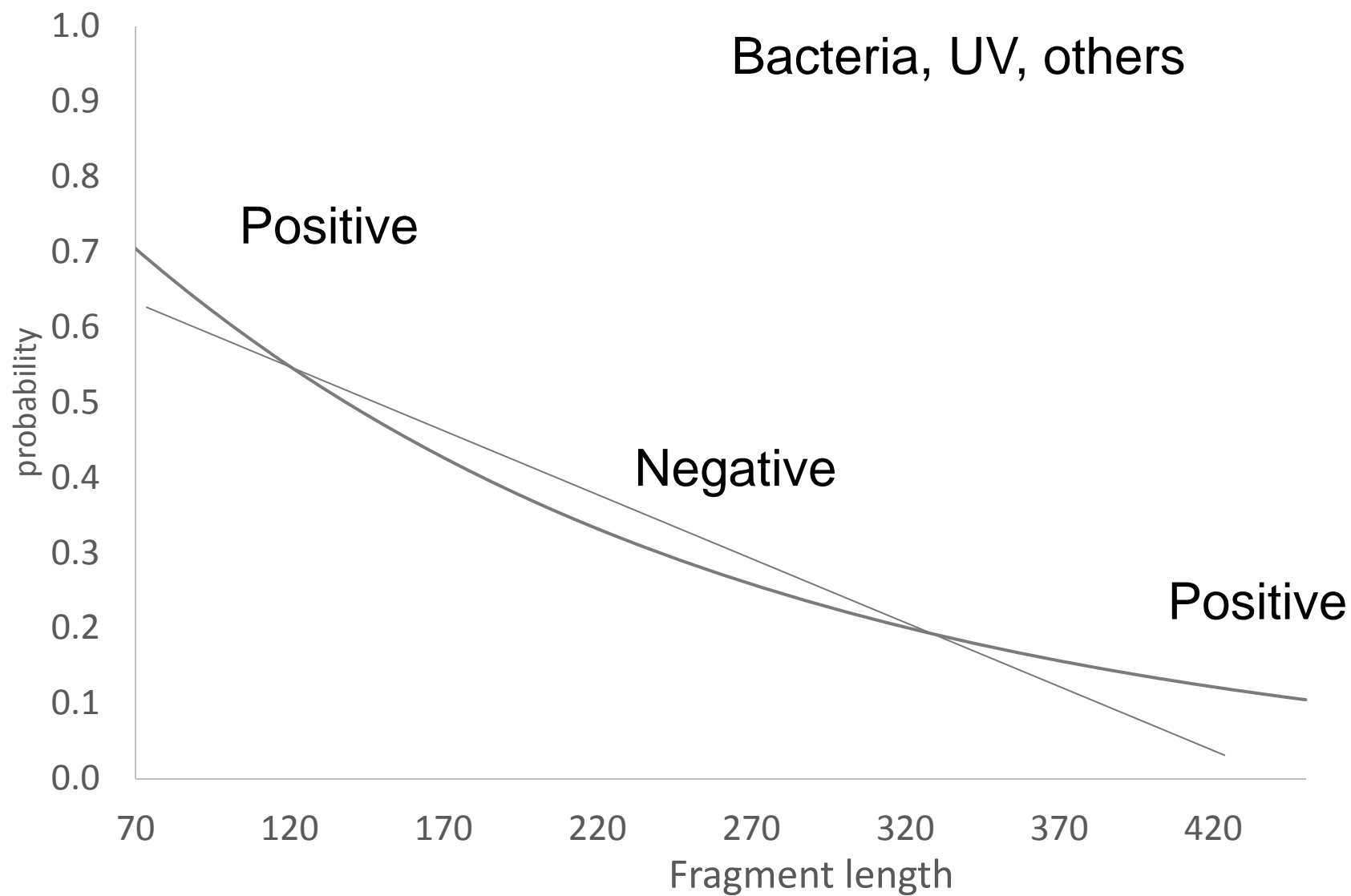
Profile slopes/degradation

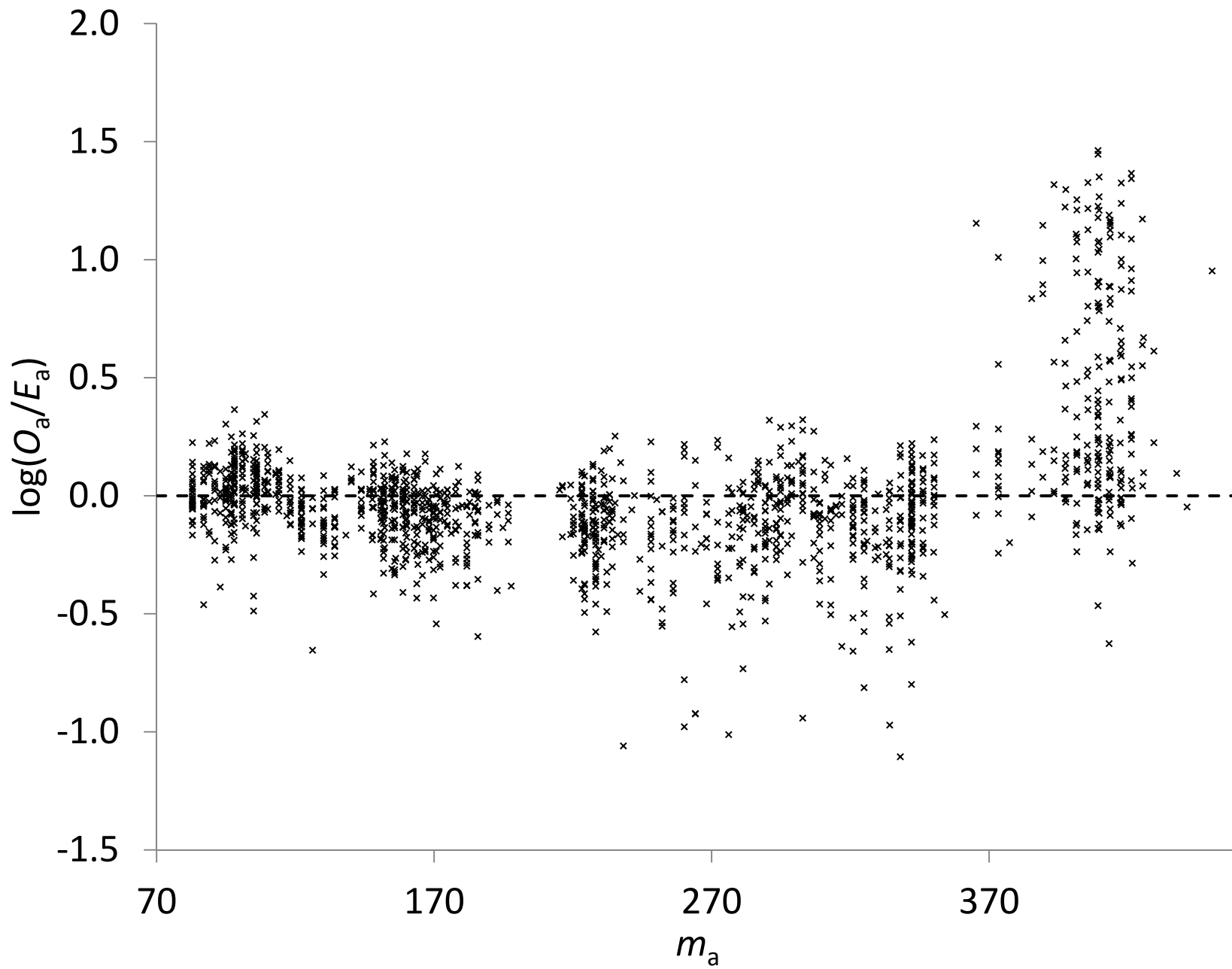


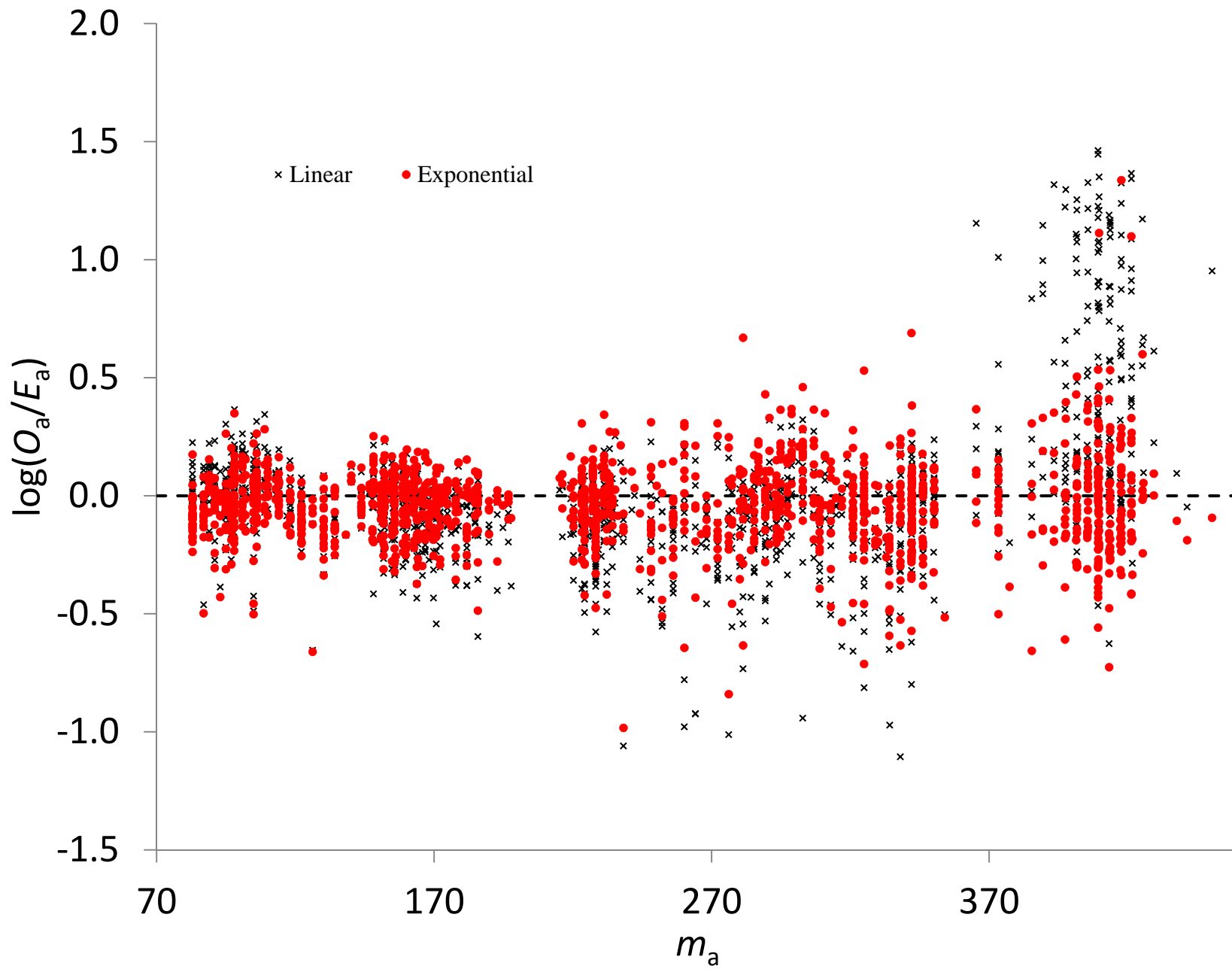
Theory

- Consider that the degradation of the DNA strand was random with respect to location.
- Consider a fragment of length l .
- If the probability of a break is p , at any of the locations $1 \dots l$
- the chance of the full fragment being amplified is $(1-p)^l$.
- Since $1-p$ is less than 1 this equation describes an exponential decline in peak height.

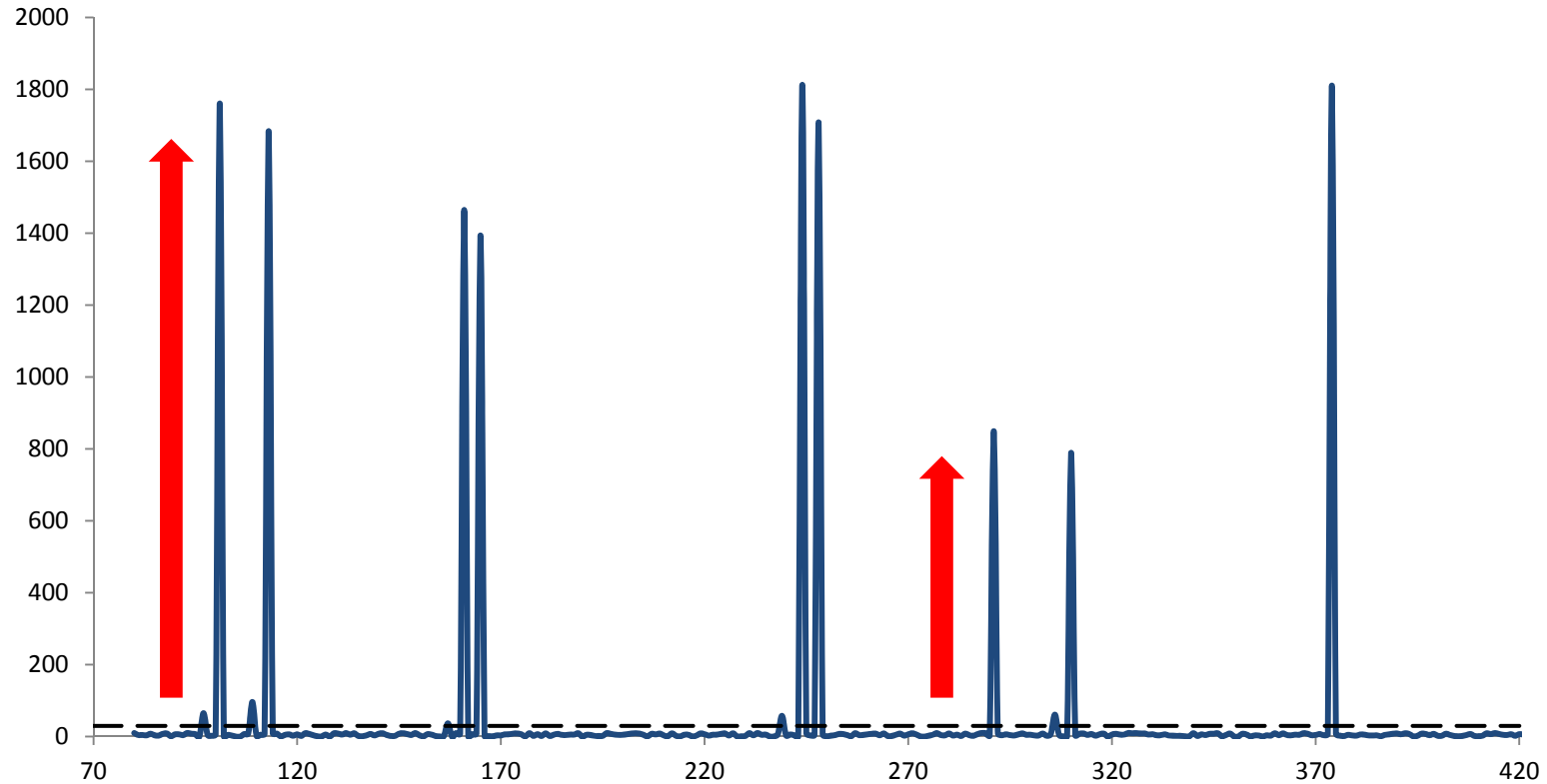








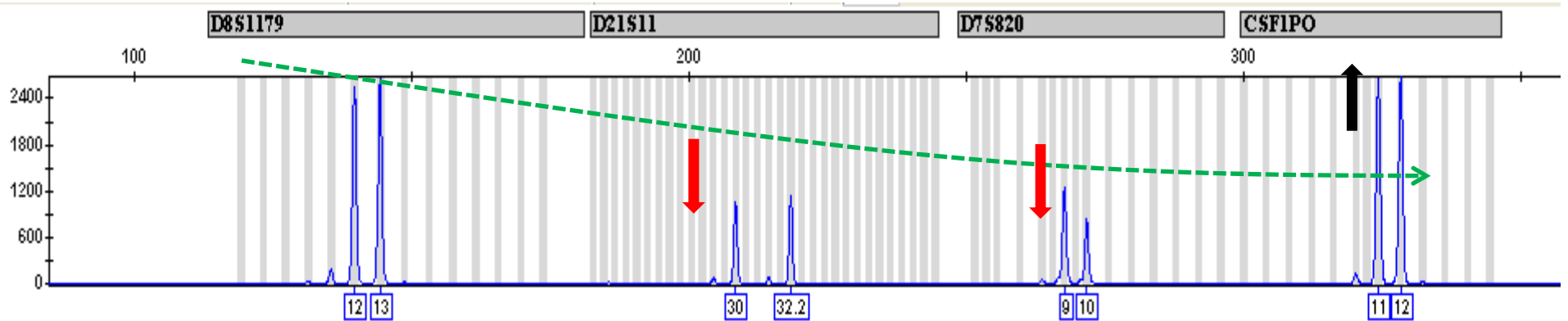
4. Locus specific amplification



4. Locus specific amplification

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

Locus specific amplification example



Assumption of additivity

- For many years we have assumed that the contribution from various sources adds
- Often termed stacking in the US,
- Allele + allele, allele + stutter
- Logical, based on the fact that the enzymes don't know the origin of the template

Recent questioning of additivity

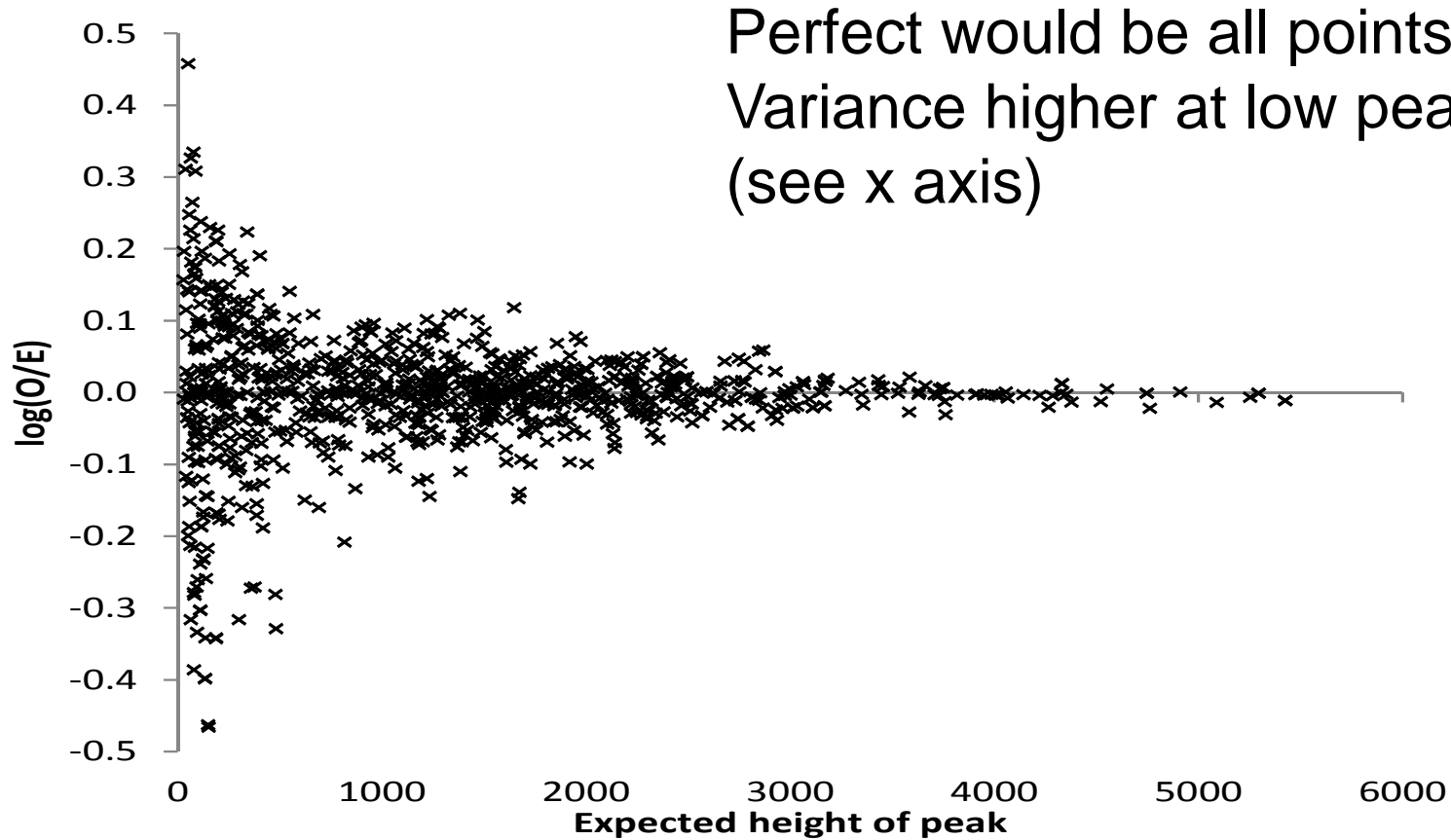
Keith Inman, Norah Rudin, Kirk Lohmueller
California State University East Bay

Presented at California Association of Criminalists
Meeting 2017

..if the relative contributions are not additive, then that calculation is not supportable.

To determine if this practice is scientifically supportable, it would be useful to query a large data set of mixtures created from known profiles, designed specifically to answer this question.

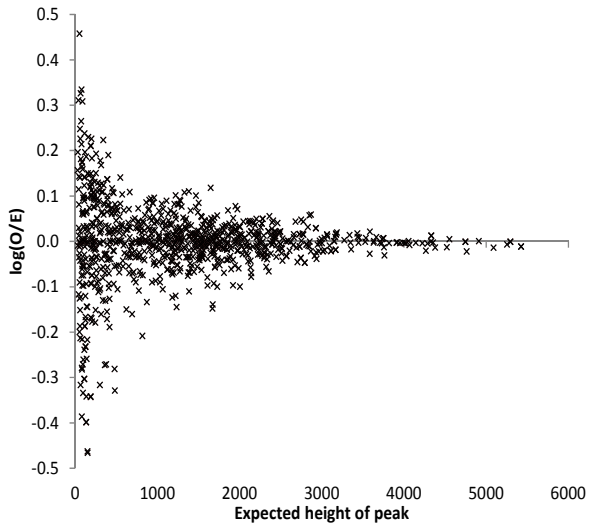
Expected heights developed by fitting a model of template and degradation



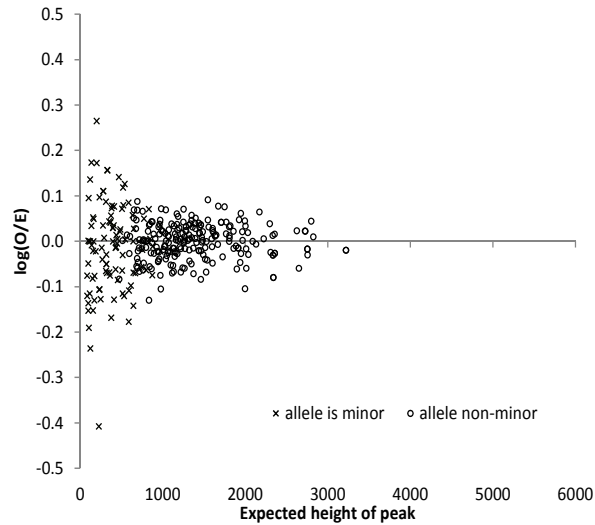
allele

$$\log \frac{O_a^l}{E_a^l} \quad \text{vs} \quad E_a^l$$

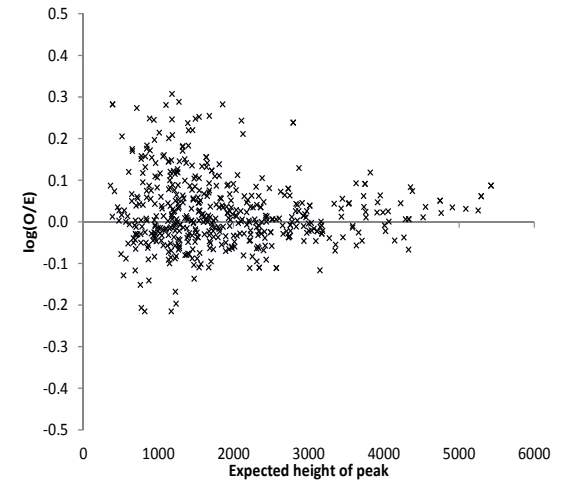
Expected heights developed by fitting a model of template and degradation



allele



composite

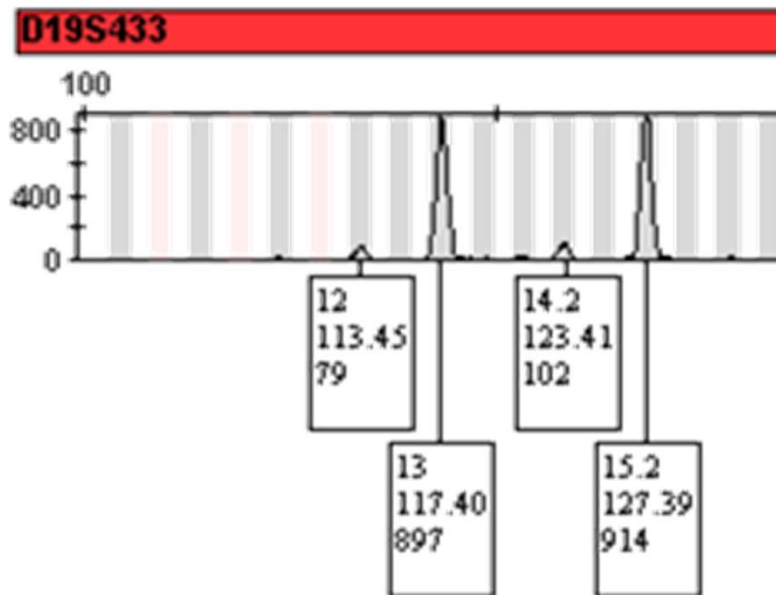


stutter

$$\log \frac{O_a^l}{E_a^l} \quad \text{vs} \quad E_a^l$$

Stutter

- By-product of PCR process
- Generally one repeat unit smaller than target allele



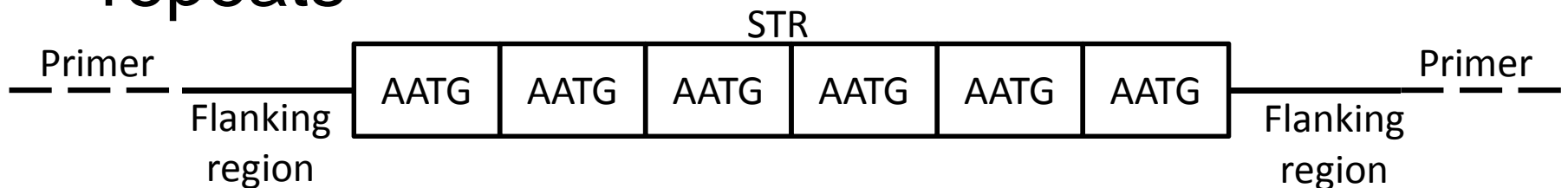
An epg of a heterozygote D19S433 13,15.2 displaying stutter peaks one repeat unit less than the parent alleles

Stutter mechanism

- A proposed mechanism is slipped strand mispairing (SSM).
- During PCR
 - the DNA polymerase enzyme stalls,
 - dissociates from the DNA,
 - the template strand “loops out”,
 - and the new strand is one repeat unit shorter than the template strand

Stutter rates

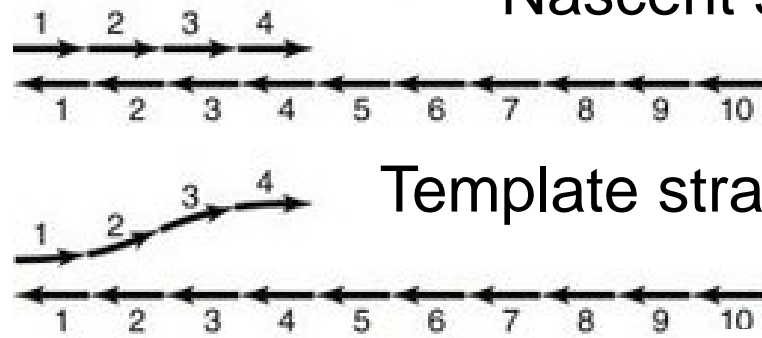
- Degree of stutter formation is related to the type of repeat
- STRs with di- and trinucleotide repeat structures are known to stutter more than tetra- and pentanucleotide repeats
- Most forensic markers are tetranucleotide repeats



Summary (& additional forms of stutter)

Type of stutter	Possible mechanism
Back stutter (minus one full repeat)	Slipped strand mis-pairing (SSM)
Forward stutter (plus one full repeat)	A loop of one repeat units forms in the nascent strand, resulting in insertion of one repeat
Double back stutter (minus 2 full repeats)	A loop of two repeat units forms in template strand resulting in deletion of two repeats (some new evidence suggesting this might be two single stutters)
Minus 2 base pair	SSM but of 2 b repeat within the allele (observed as well as 4 base pair repeat)

(a) Increase in repeat length

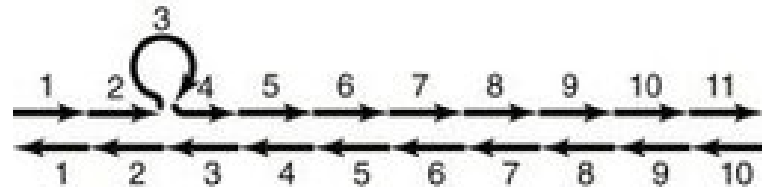


Nascent strand

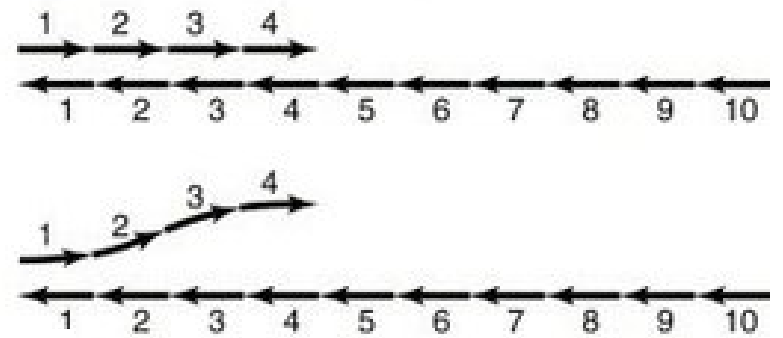
Template strand

Initiation

Dissociation



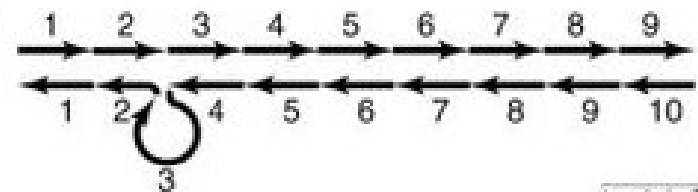
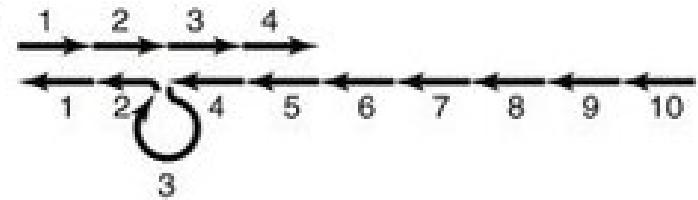
(b) Decrease in repeat length



Or breathing

Rehybridization and misalignment

The new strand is a different length to the template



Stutter ratios

- Back stutter is typically quantified as a stutter ratio (SR):

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

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

Stutter ratios

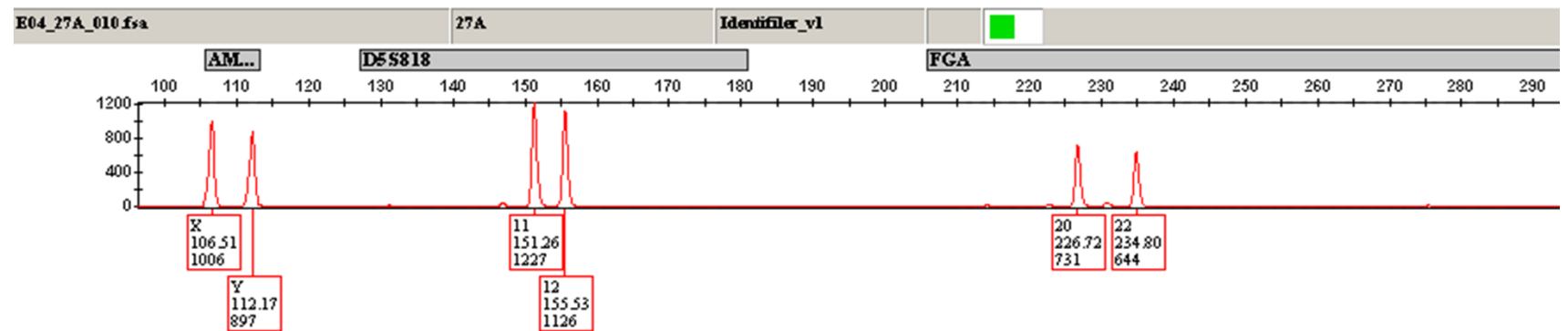
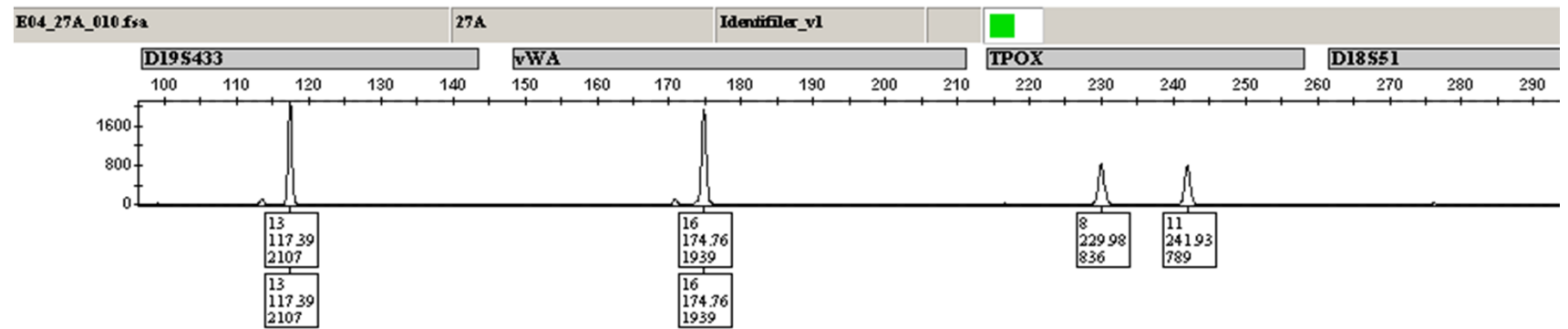
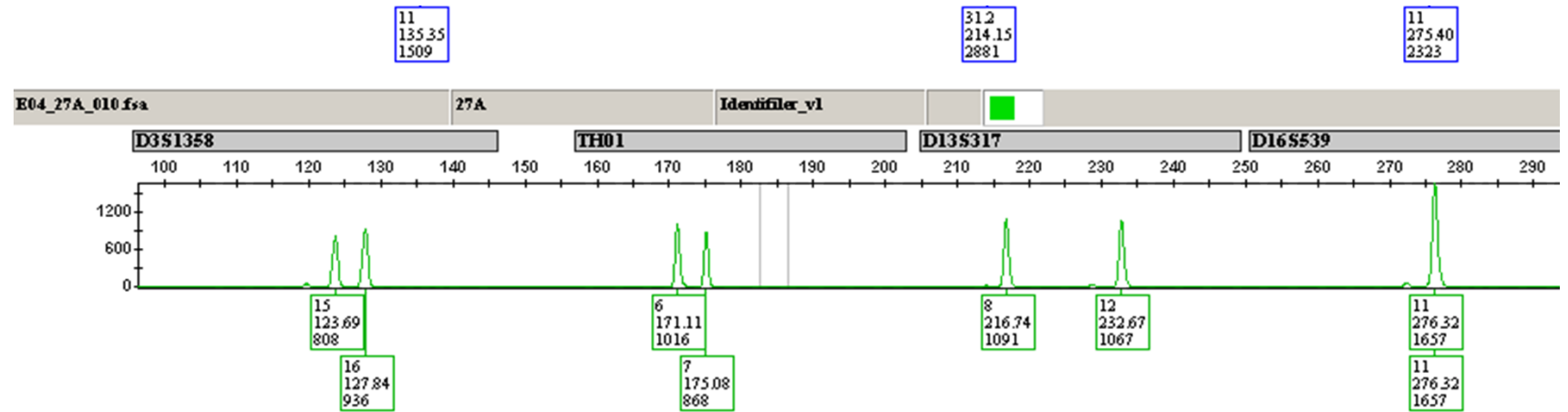
- Forward stutter is typically quantified as a stutter ratio (*FSR*):

$$FSR_a^l = \frac{O_{a+1}}{O_a}$$

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

Stutter ratios

- Traditionally we apply a filter at analysis to remove stutter e.g. 15%
 - Locus specific
 - Profile/Kit specific
- Stutter filters remove the label, but not the peak



Stutter ratios

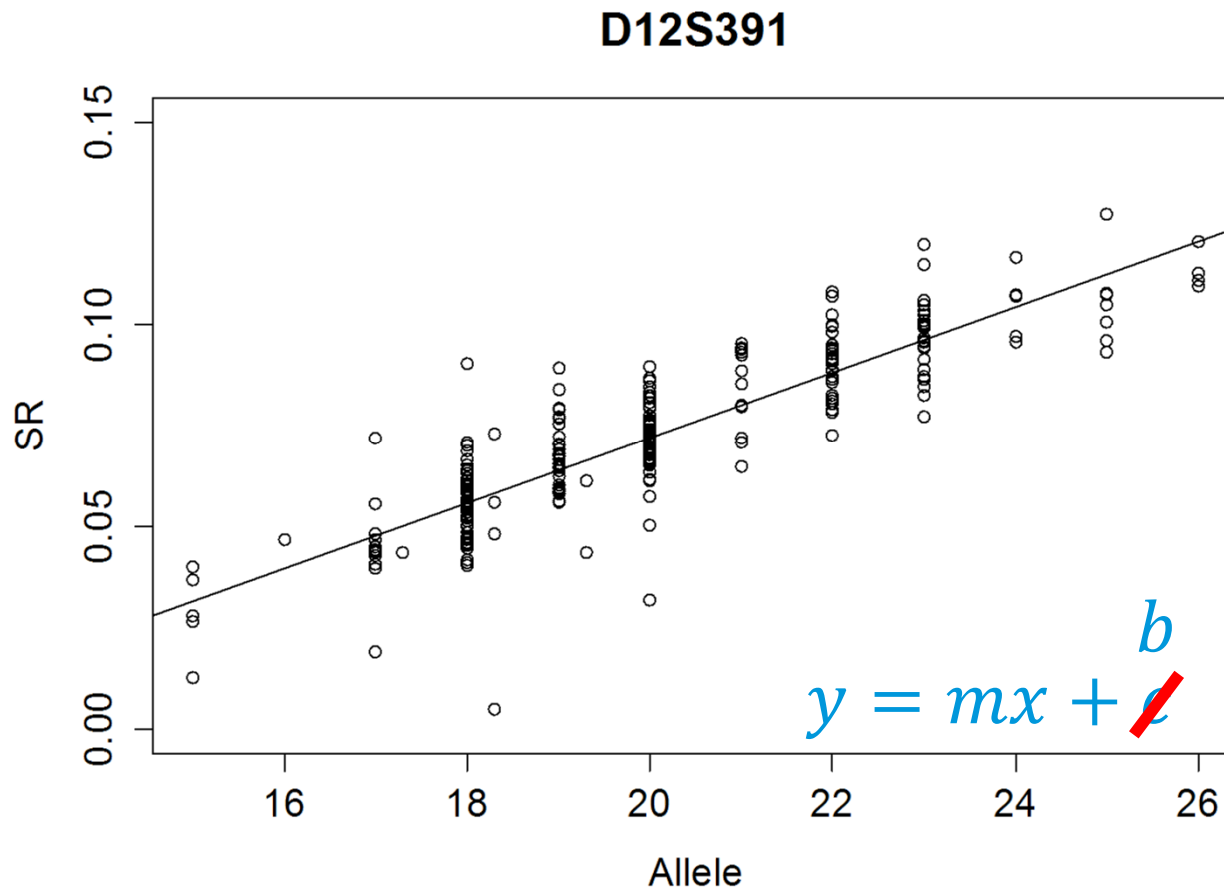
- This is an in/out (binary) decision
- Most important when the minor POI is approximately same height as stutter
- Such a peak may be stutter or stutter/allelic
- Making an in/out decision may be incorrect and have consequences

Stutter products

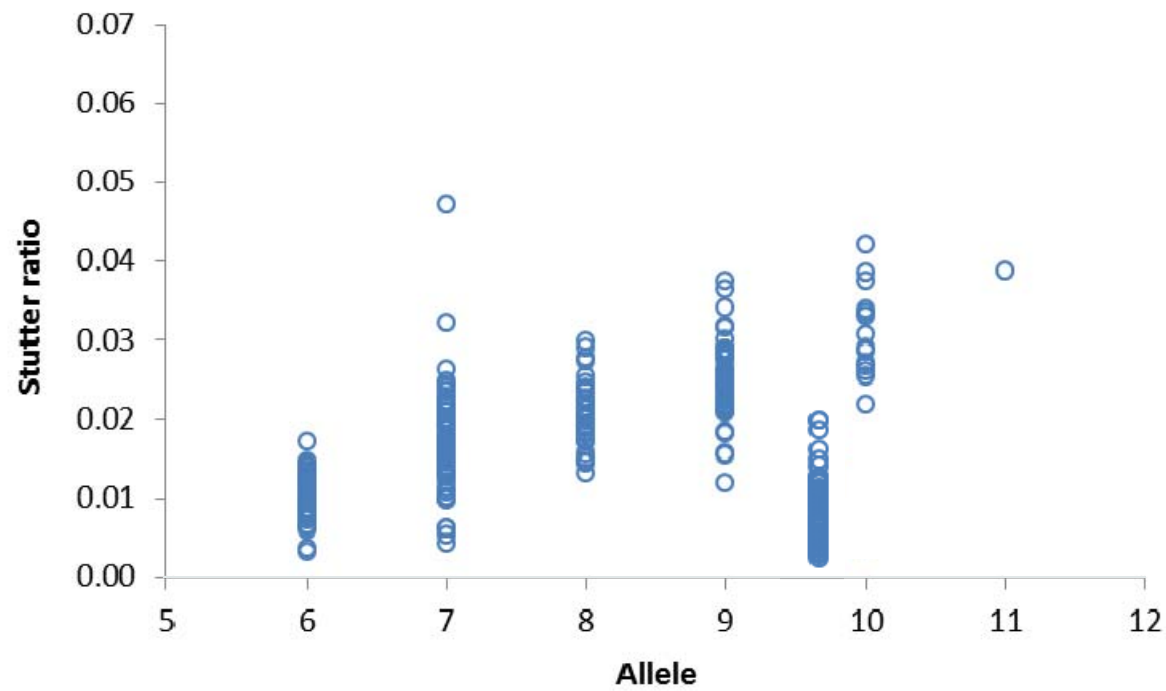
- The products are allelic in all aspects
- Not distinguishable from true allelic products
- Contribute to the complexity of profile interpretation
- Especially when a true contributor's alleles are approximately the same height as the stutter product from another contributor.

Stutter ratios

- Stutter ratios are actually *allele specific*



TH01 stutter



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] ₆ ATG[AATG] ₃	9.3	6

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

Forensic Science International: Genetics

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



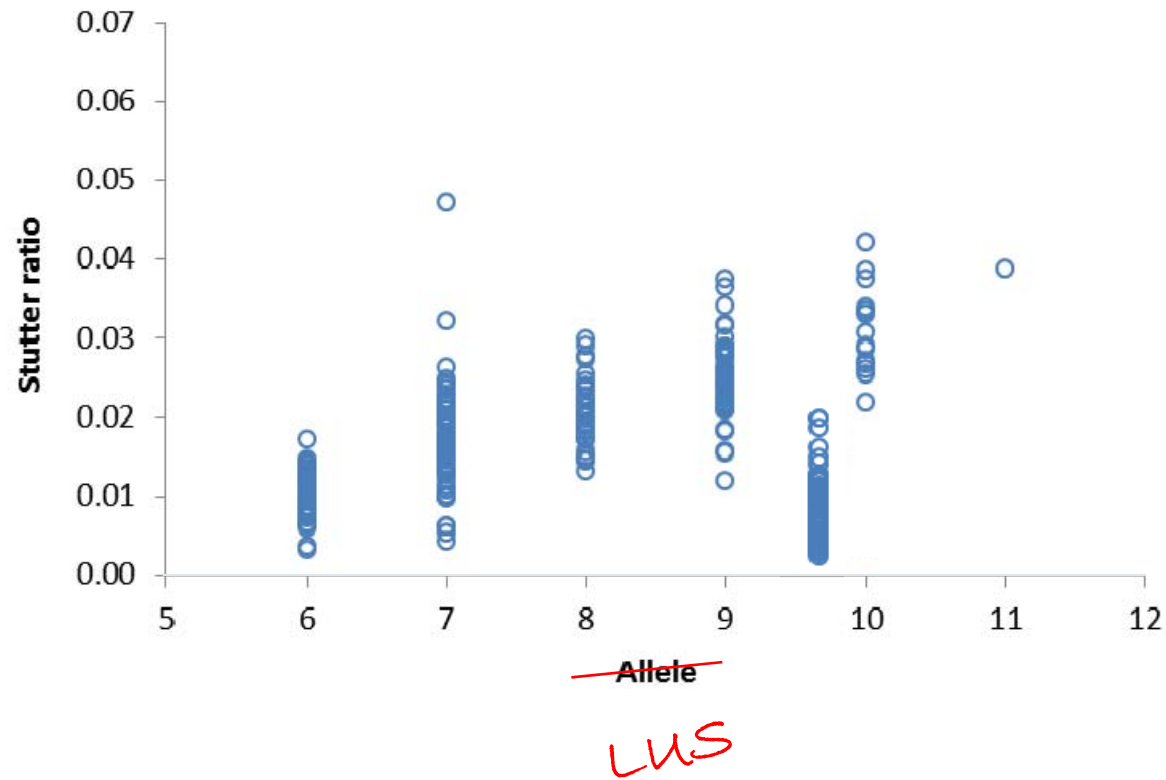
Characterising stutter in forensic STR multiplexes

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

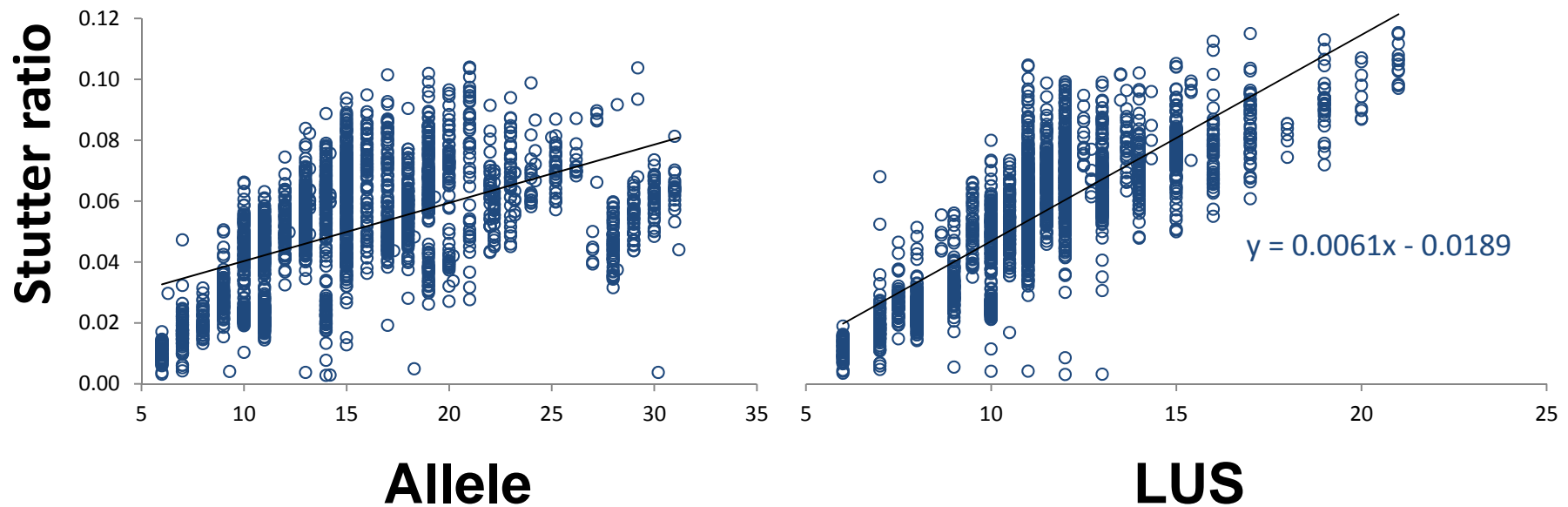
^a Department of Chemistry, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

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

TH01 Stutter ratio versus LUS



Allele versus LUS, NGM Select loci

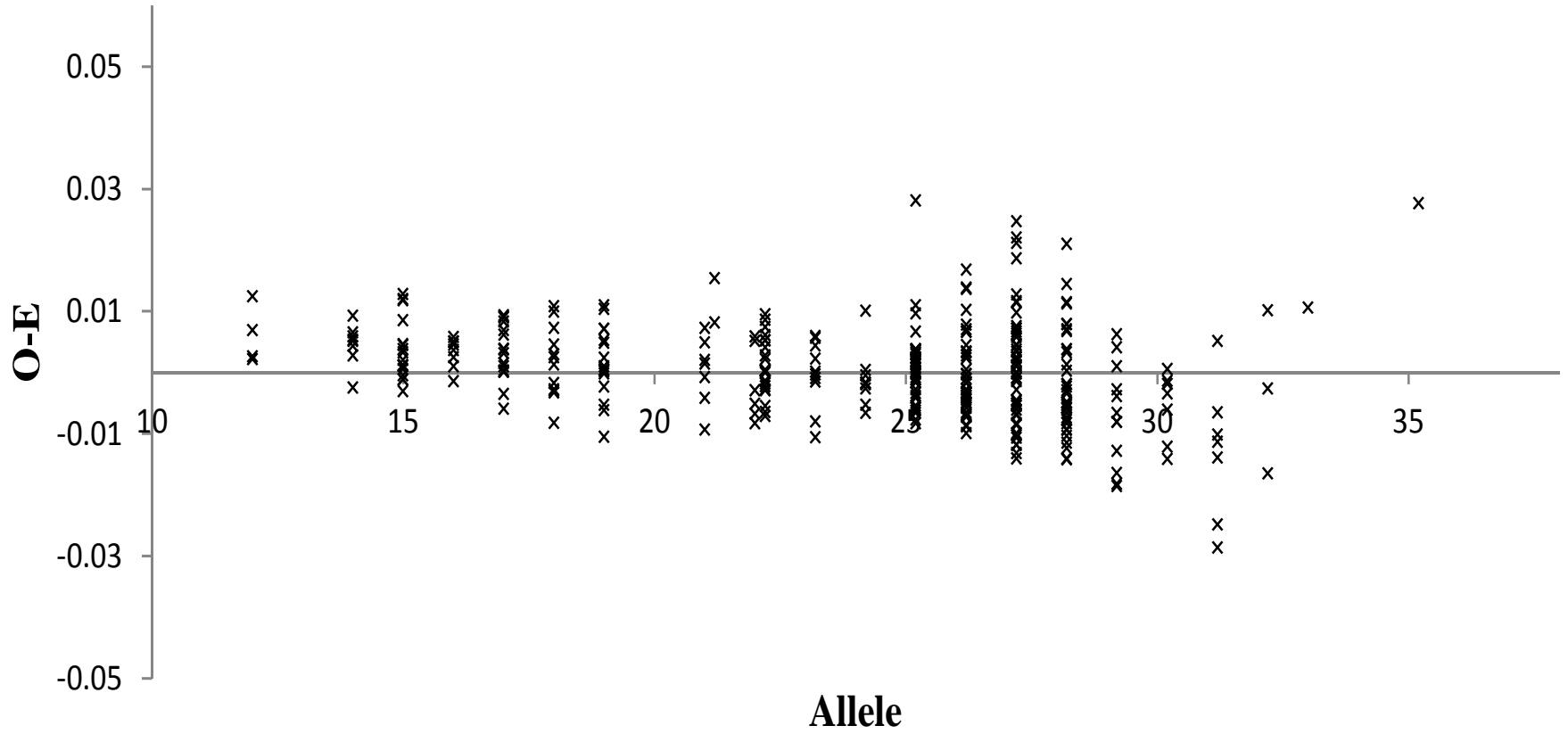


$R^2 = 27\%$

$R^2 = 61\%$

R-squared is a statistical measure of how close the data are to the fitted regression line

21.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₉ AA AAAG [AAAG] ₁₁ G AAGG [AAAG] ₂ AG
21.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₁₁ AA AAAG [AAAG] ₉ G AAGG [AAAG] ₂ AG
22	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₂₂ G [AAAG] ₃ AG
22.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₇ AA AAAG [AAAG] ₁₄ G AAGG [AAAG] ₂ AG
22.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₈ [AG] ₅ [AAAG] ₁₂ G AAGG [AAAG] ₂ AG
22.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₉ AA AAAG [AAAG] ₁₂ G AAGG [AAAG] ₂ AG



$$SR = m \sum_i \max(l_i - 5.83, 0) + c$$

21.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₉ AA AAAG [AAAG] ₁₁ G AAGG [AAAG] ₂ AG
21.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₁₁ AA AAAG [AAAG] ₉ G AAGG [AAAG] ₂ AG
22	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₂₂ G [AAAG] ₃ AG
22.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₇ AA AAAG [AAAG] ₁₄ G AAGG [AAAG] ₂ AG
22.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₈ [AG] ₅ [AAAG] ₁₂ G AAGG [AAAG] ₂ AG
22.2	[AAAG] ₂ AG [AAAG] ₃ AG [AAAG] ₉ AA AAAG [AAAG] ₁₂ G AAGG [AAAG] ₂ AG

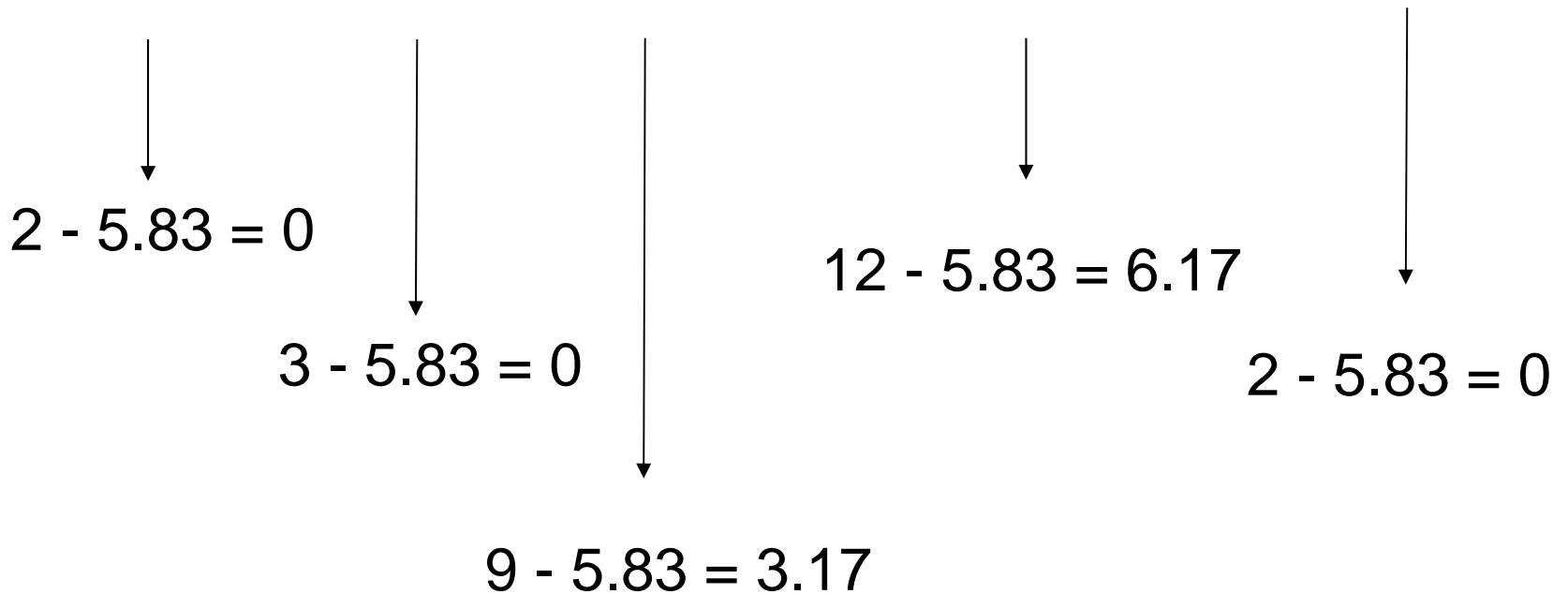
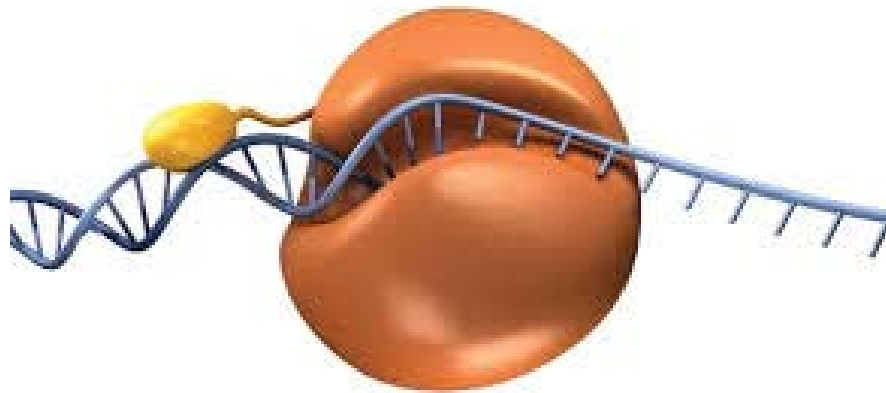
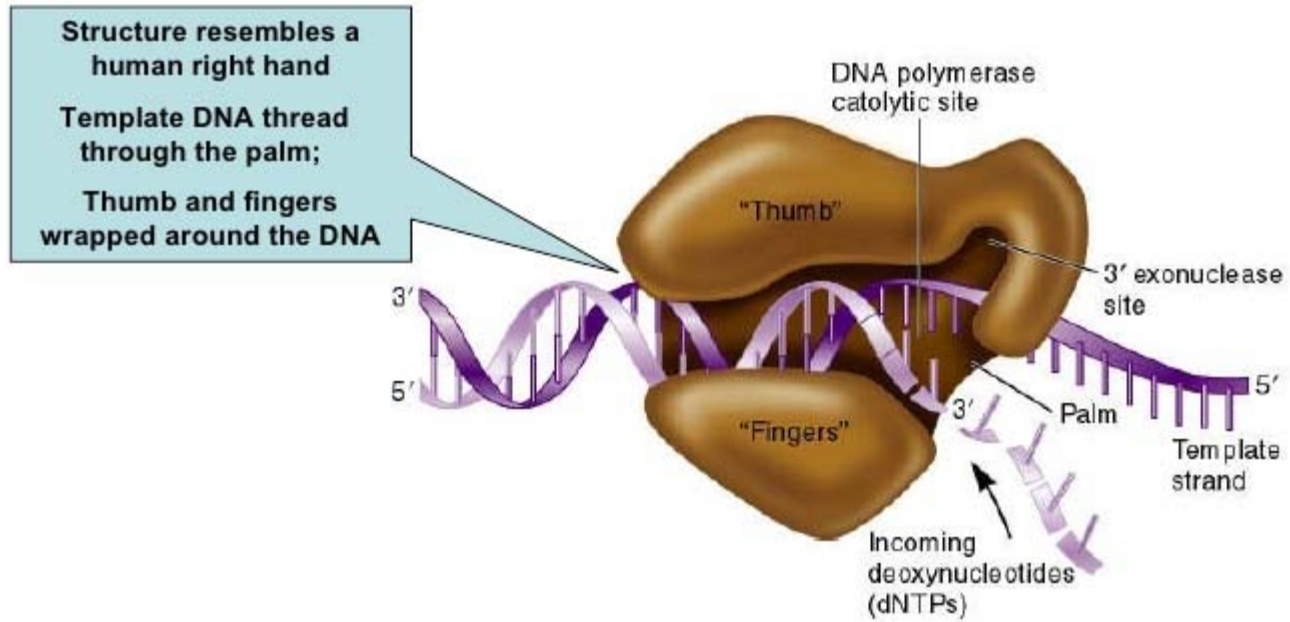
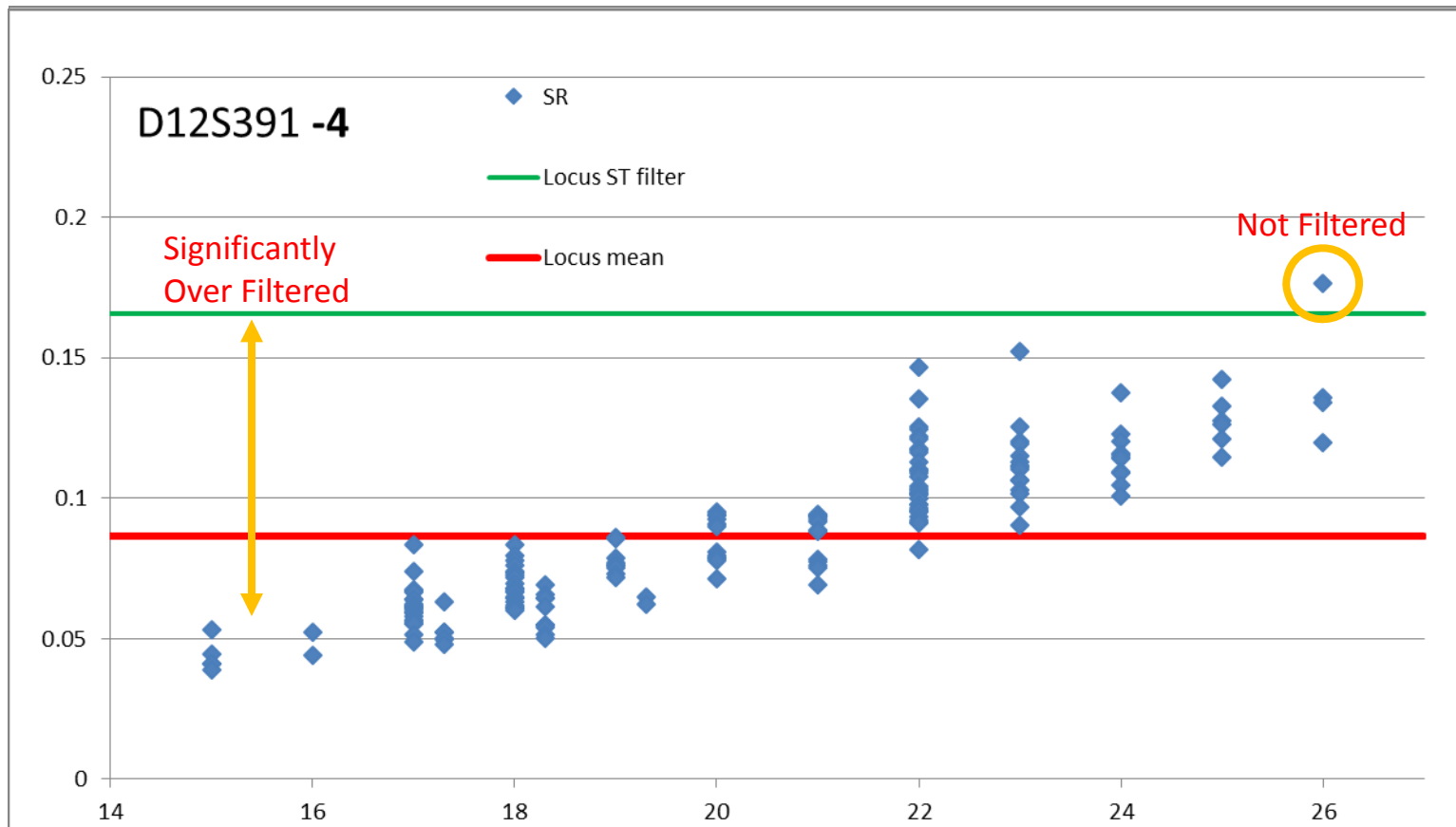


Figure 11.8 Schematic representation of DNA Polymerase III



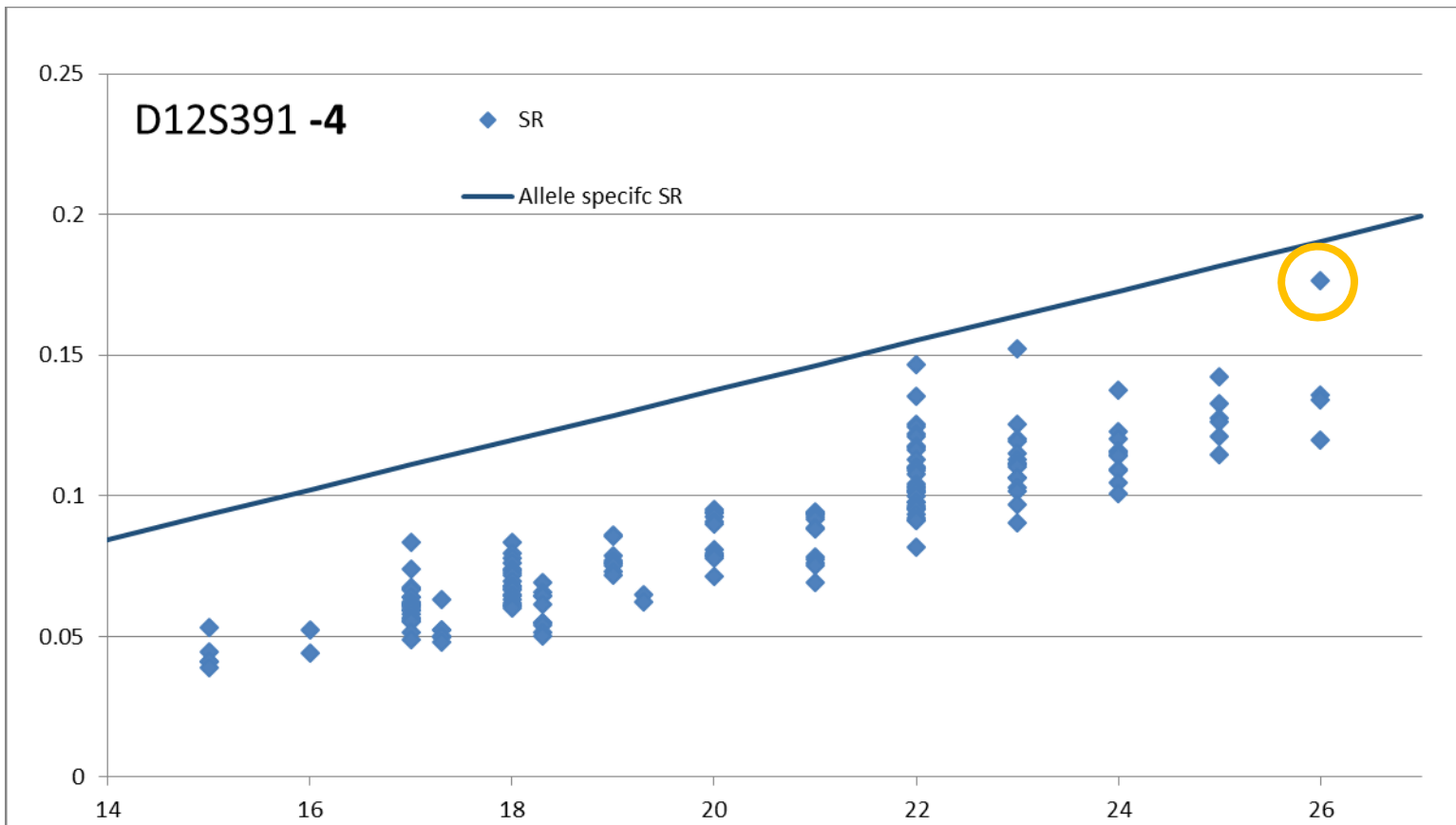


Stutter filters





$$y=0.008x - 0.092 + 2(0.0264)$$

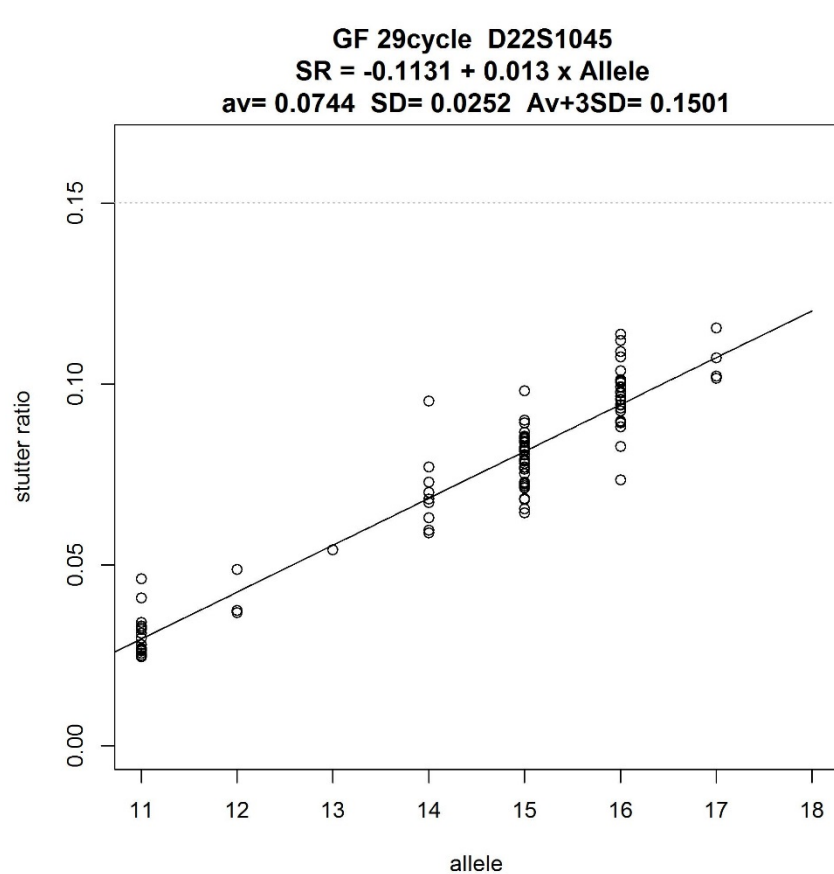


Forward stutter filters

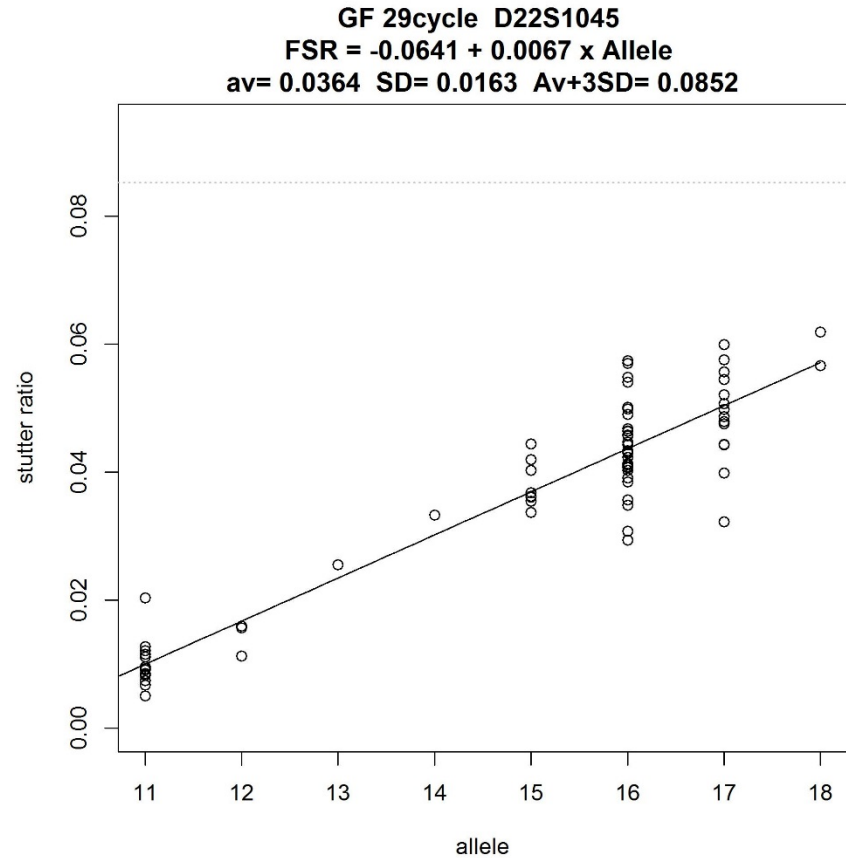
Forward stutter filters usually per locus

Locus	FSR	Locus	FSR
CSF1PO	0.0062	D2S441	0.0055
D10S1248	0.0128	D3S1358	0.0042
D12S391	0.0029	D5S818	0.0077
D13S317	0.0046	D7S820	0.0020
D16S539	0.0059	D8S1179	0.0054
D18S51	0.0045	FGA	0.0030
D19S433	0.0019	SE33	0.0059
D1S1656	0.0052	TH01	0.0006
D21S11	0.0072	TPOX	0.0007
D2S1338	0.0016	vWA	0.0033

Forward stutter filters exception is D22

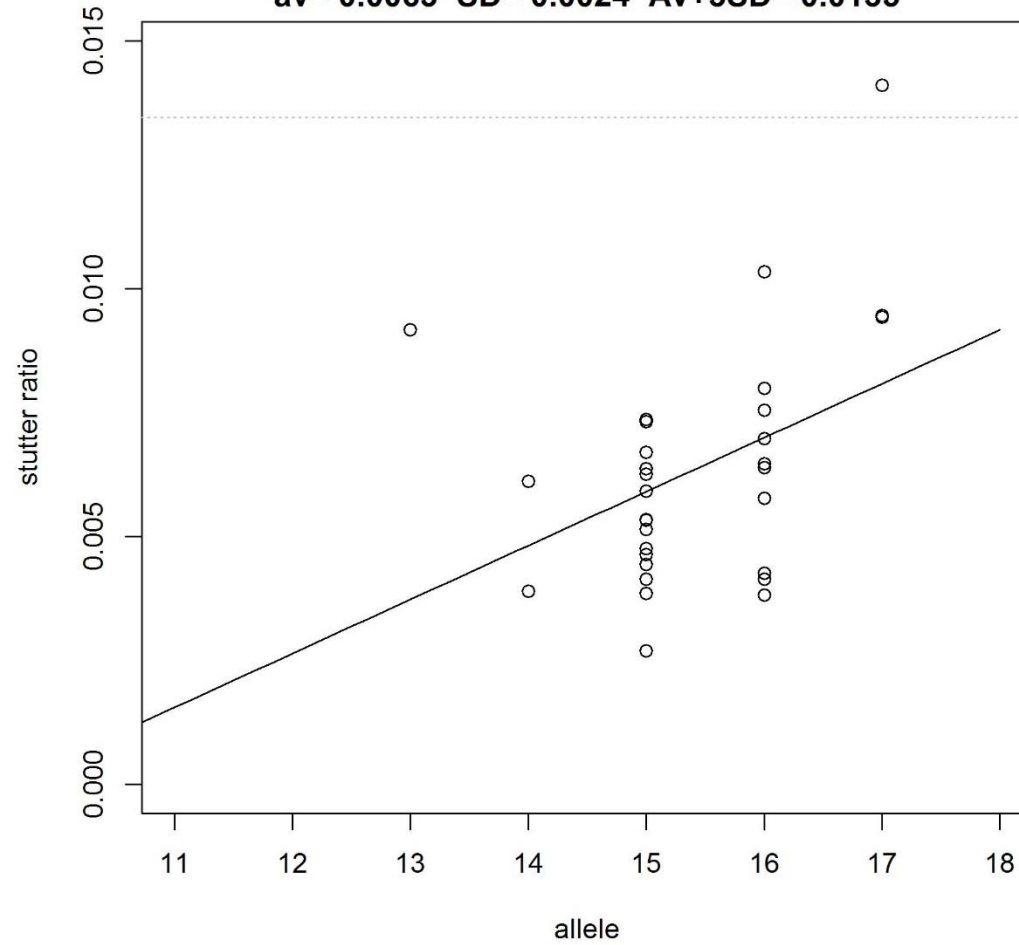


Back stutter

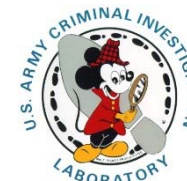


Forward stutter

GF 29cycle D22S1045
DSR = $-0.0104 + 0.0011 \times \text{Allele}$
av= 0.0063 SD= 0.0024 Av+3SD= 0.0135

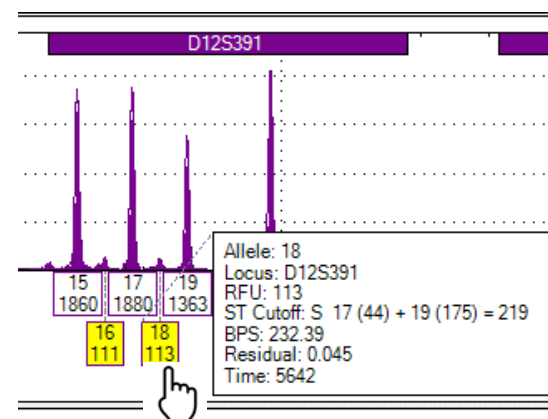
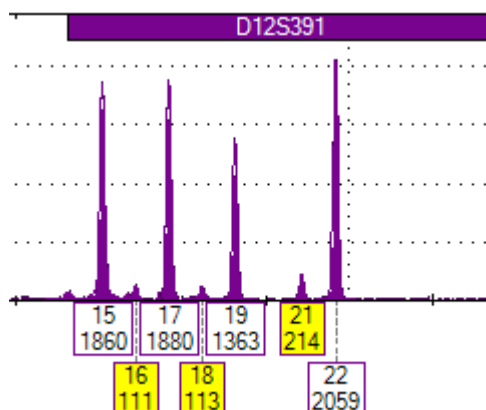


Double back stutter



Combined Stutter

- This means you can no longer consider stutter filters on a simple “%” basis



- Find back stutter contribution
- Find forward stutter contribution
- Add them together



Total Study Results



64 2 person mixtures					
Model	Over	Under	After 3SD	Total	
New	10	5	0	10	
Traditional	61	25	NA	86	
54 3 person mixtures					
Model	Over	Under	After 3SD	Total	
New	7	7	4	11	
Traditional	19	31	NA	50	
54 4 person mixtures					
Model	Over	Under	After 3SD	Total	Over/Under
New	5	5	4	9	0
Traditional	19	27	NA	46	2
Grand total					
"New School"				30	
"Old School"				186	

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

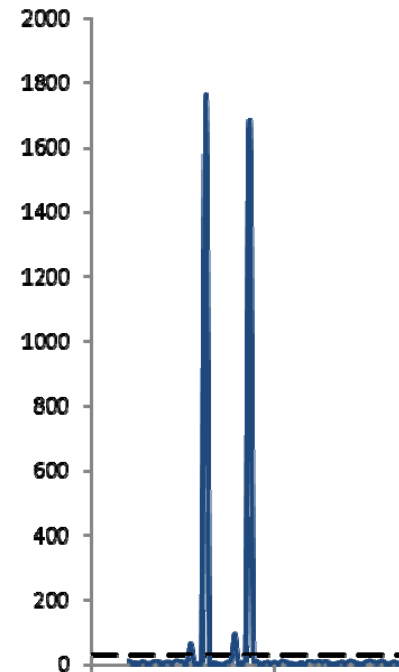
Heterozygote balance

- Also called peak height ratio
- Ratio of two heterozygote peaks at a locus
- Two common definitions

$$Hb_1 = \frac{O_{HMW}}{O_{LMW}}$$

$$Hb_2 = \frac{O_{\text{smaller}}}{O_{\text{larger}}} = PHR$$

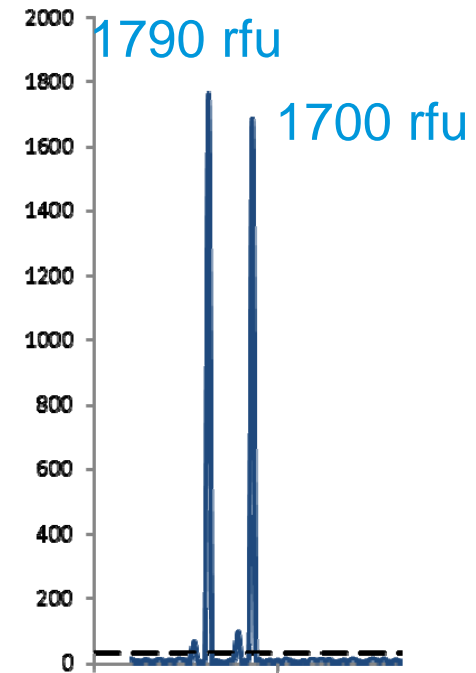
Where O is observed peak height



Heterozygote balance

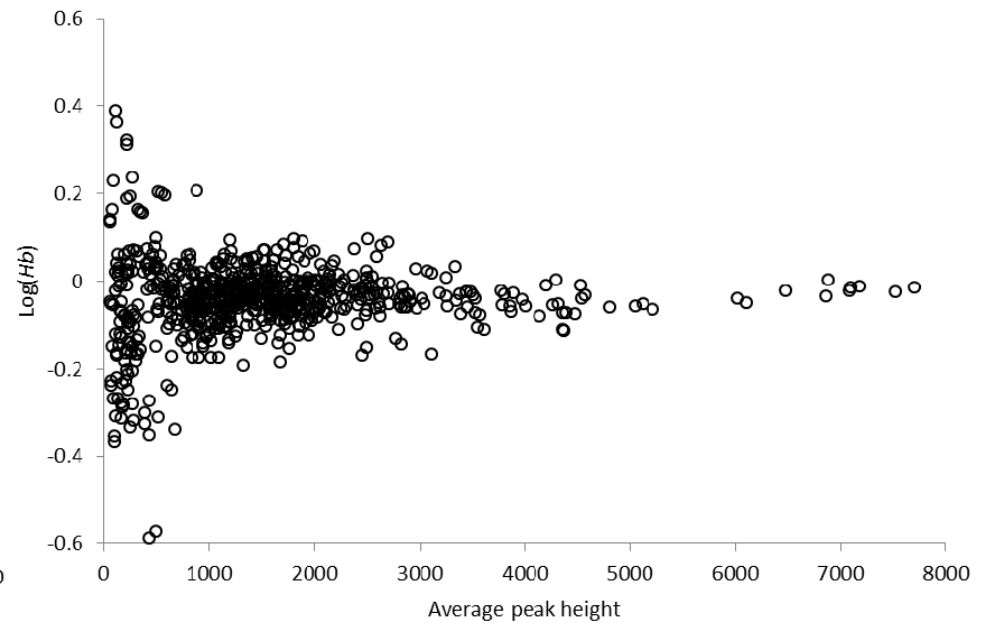
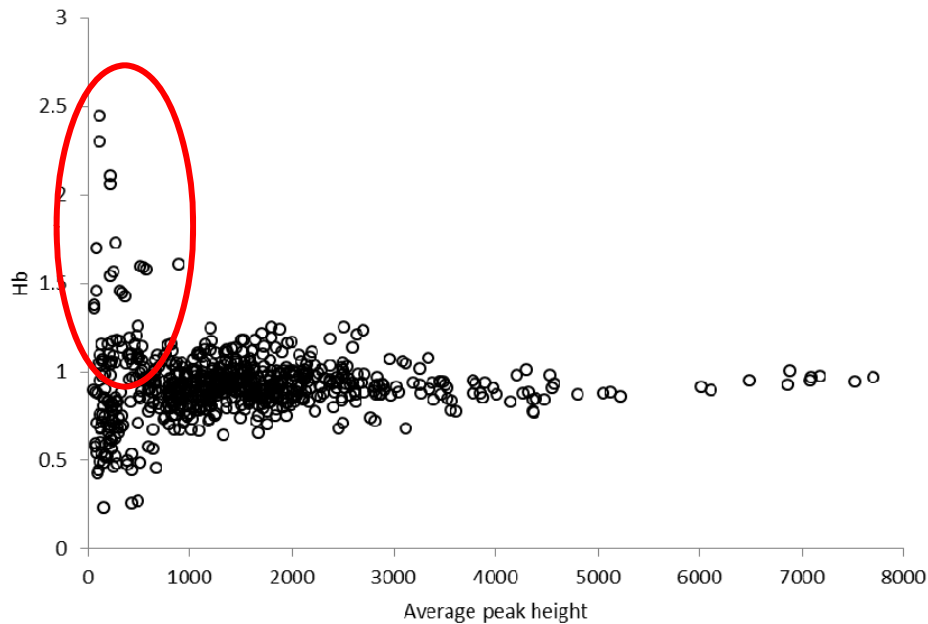
$$Hb_1 = \frac{O_{HMW}}{O_{LMW}} \quad \frac{1700}{1790} = 0.95$$

$$Hb_2 = \frac{O_{\text{smaller}}}{O_{\text{larger}}} = PHR$$



- Hb_1 has the highest information content because it maintains peak order
- Hb_2 may be obtained from Hb_1 but not vice versa

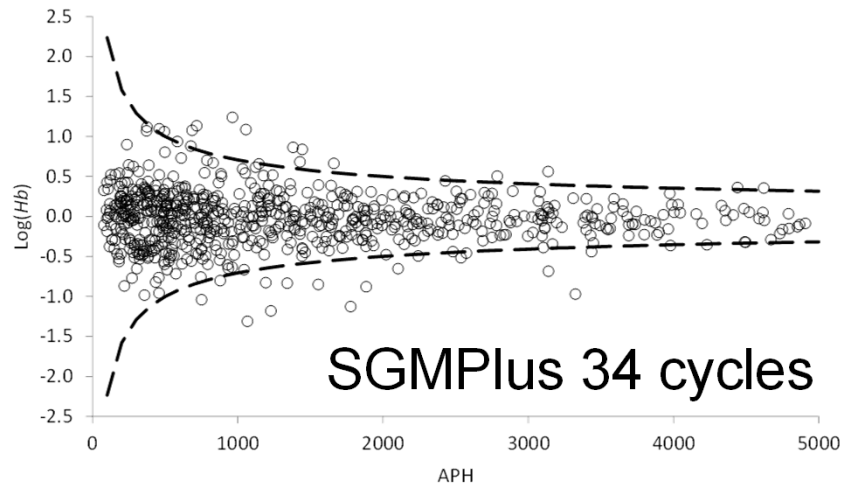
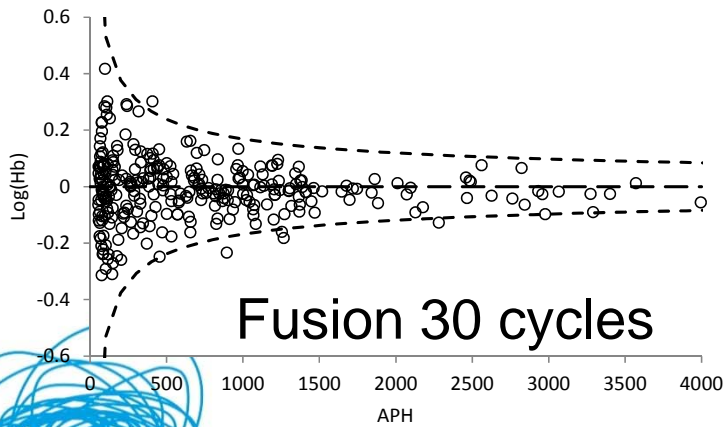
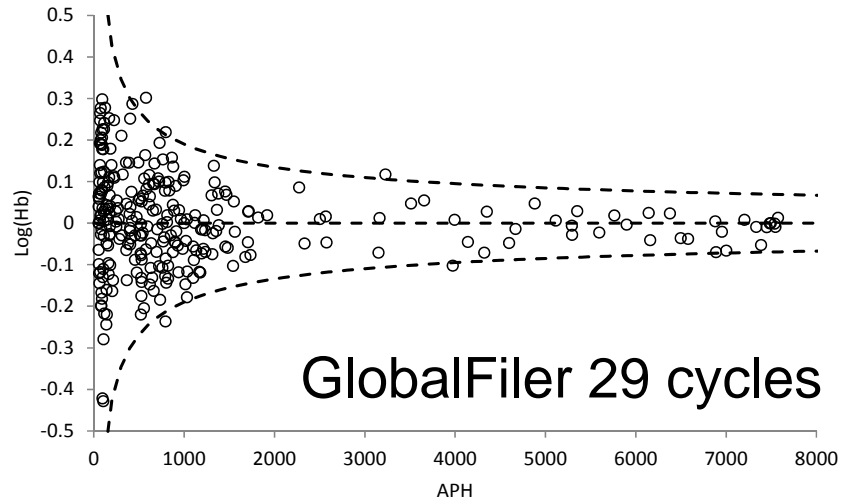
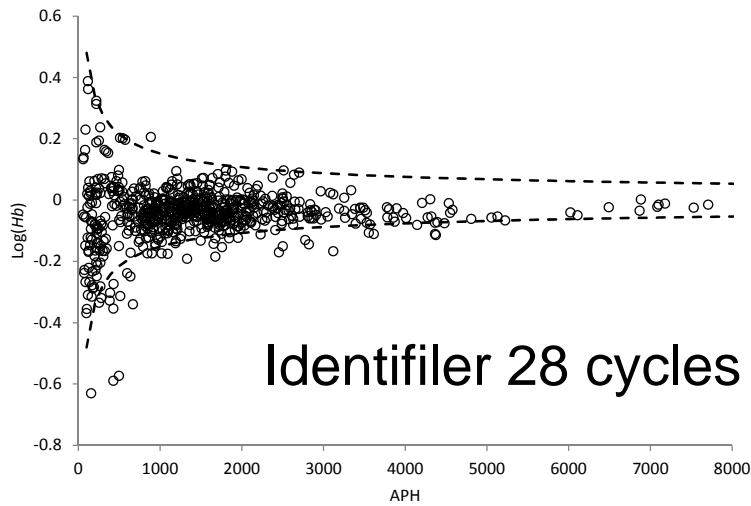
Hb versus average peak height



“Most ratios look better as logs”



Consistent across platforms



Mock samples v Casework

There is a strong propensity towards mock samples in the US

But some criticism that these do not model casework



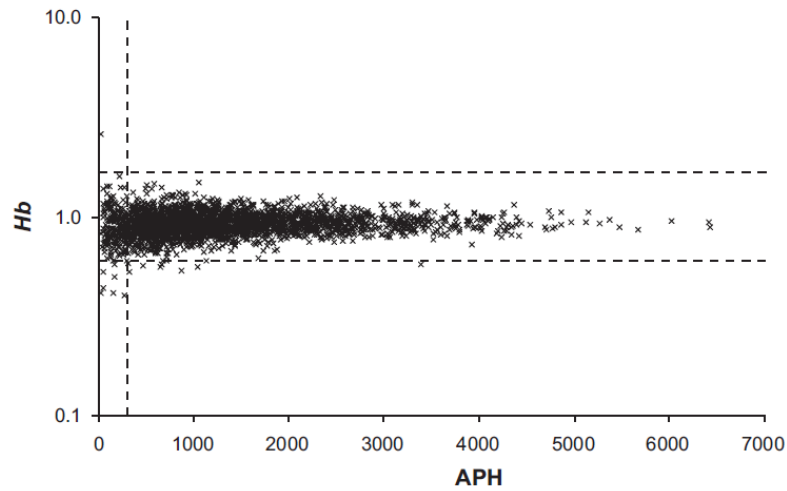


Fig. 1. Heterozygote balance versus APH data from single source pristine ($N = 2889$). The horizontal dotted lines are at $Hb = 0.6$ and 1.66 and the vertical dotted line is at $APH = 300$ RFU.

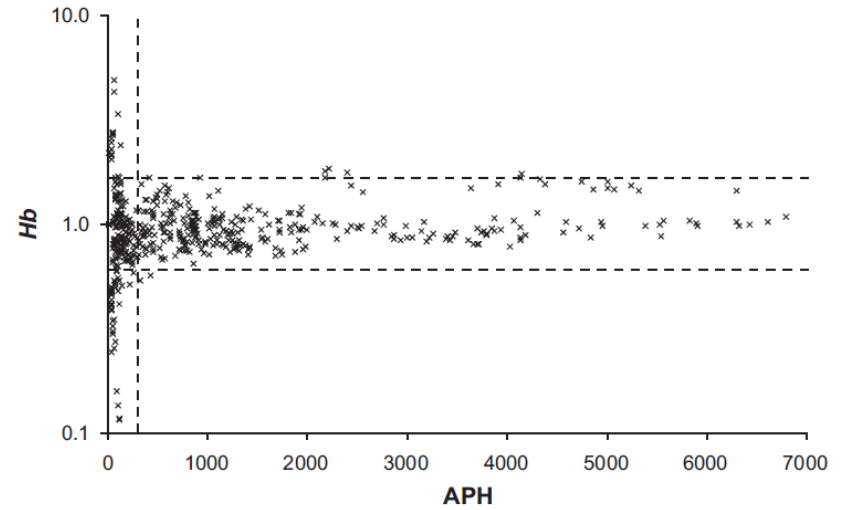


Fig. 3. Heterozygote balance versus APH data from mixed source pristine DNA ($N = 559$).

Mock single source

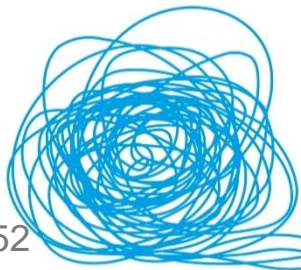
Masked mixed source

Transform

The variance of a sum is the sum of the variances (if independent)

$$\begin{aligned}\log Hb &= \log \frac{H_{HMW}}{H_{LMW}} \\ &= \log H_{HMW} - \log H_{LMW}\end{aligned}$$

$$\text{var } \log Hb = \text{var } \log H_{HMW} + \text{var } \log H_{LMW}$$



Transform

$$\text{var log } Hb = \text{var log } H_{HMW} + \text{var log } H_{LMW}$$

Assume $\text{var log } H_{HMW} = \text{var log } H_{LMW} = \text{var log } H_i$

$$\text{var log } Hb = 2 \text{var log } H_i$$





Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



A stochastic model of the processes in PCR based amplification of STR DNA in forensic applications

Jos Weusten^{a,*}, Jos Herbergs^{b,1}

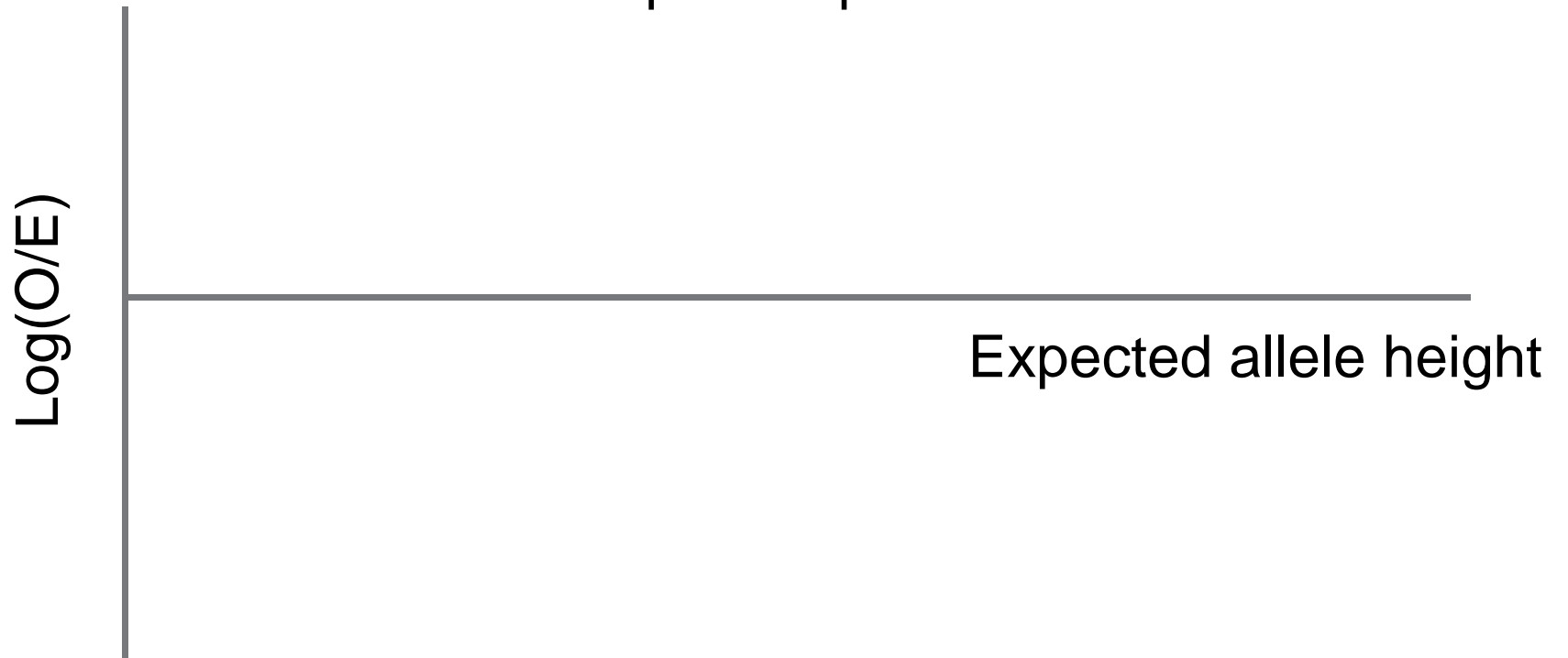
^aDSM Resolve, The Maastricht Forensic Institute, P.O. Box 18, 6160 MD Geleen, Urmonderbaan 22, 6167 RD Geleen, The Netherlands

^bDNalysis Maastricht, The Maastricht Forensic Institute, P.O. Box 1036, 6201 BA Maastricht, Oxfordlaan 70, 6229 EV Maastricht, The Netherlands

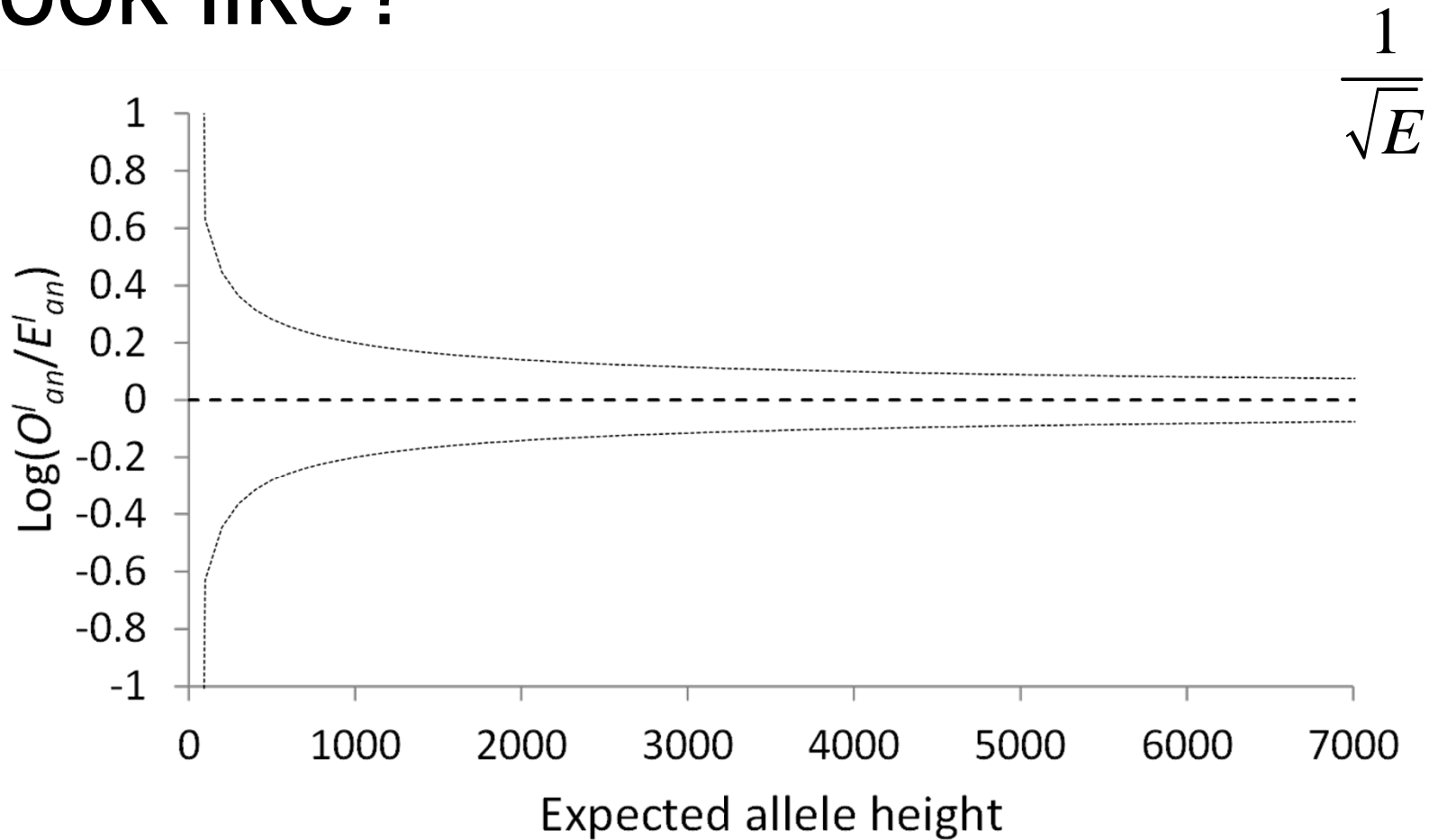
- the relative error $\frac{\sigma}{\mu}$ on the peak areas is proportional to $\frac{1}{\sqrt{N_0}}$ N_0 is the number of starting template

What would a perfect model look like?

If we compare the observed profile to the expected profile...

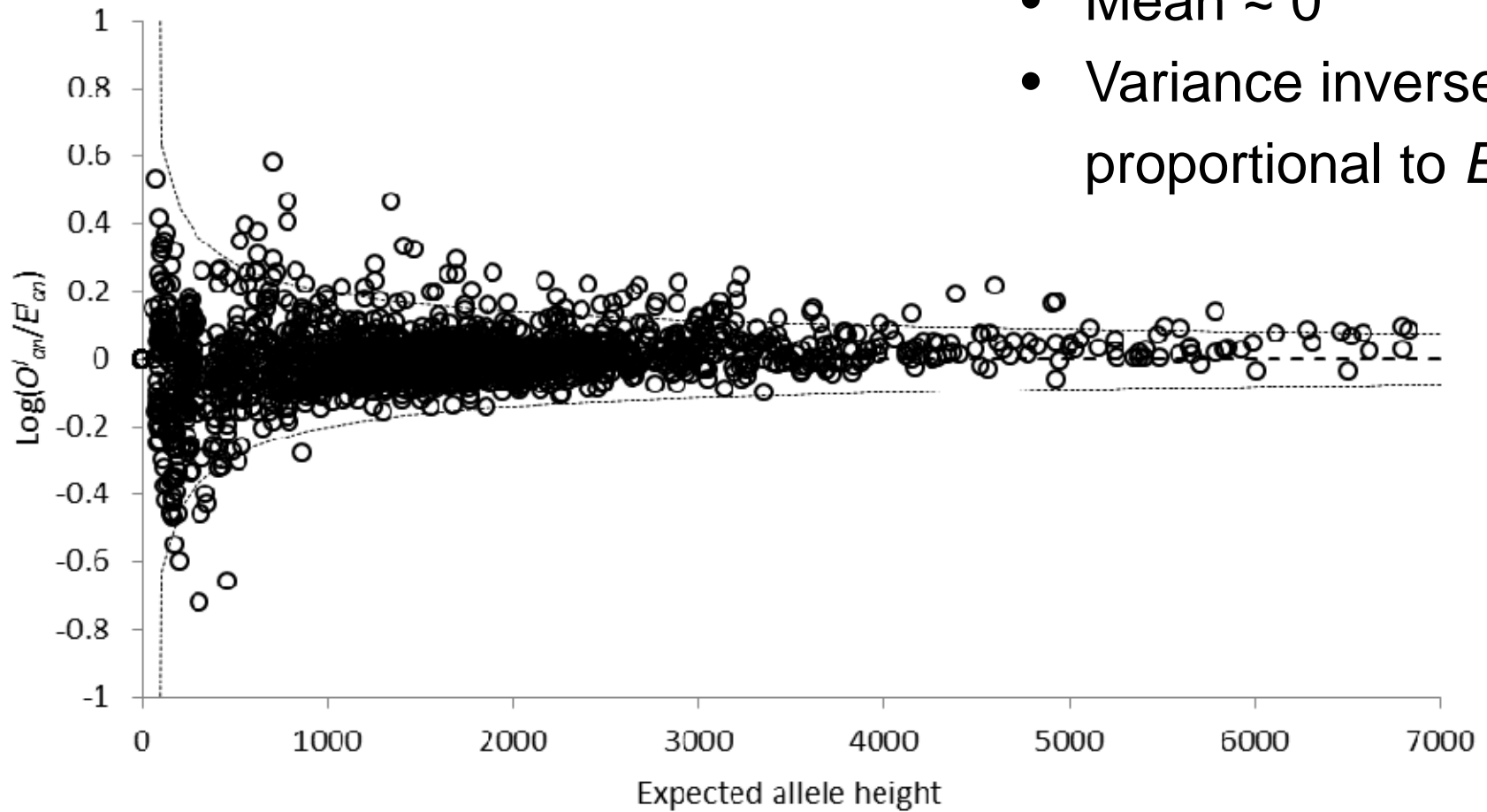


What would a perfect model look like?



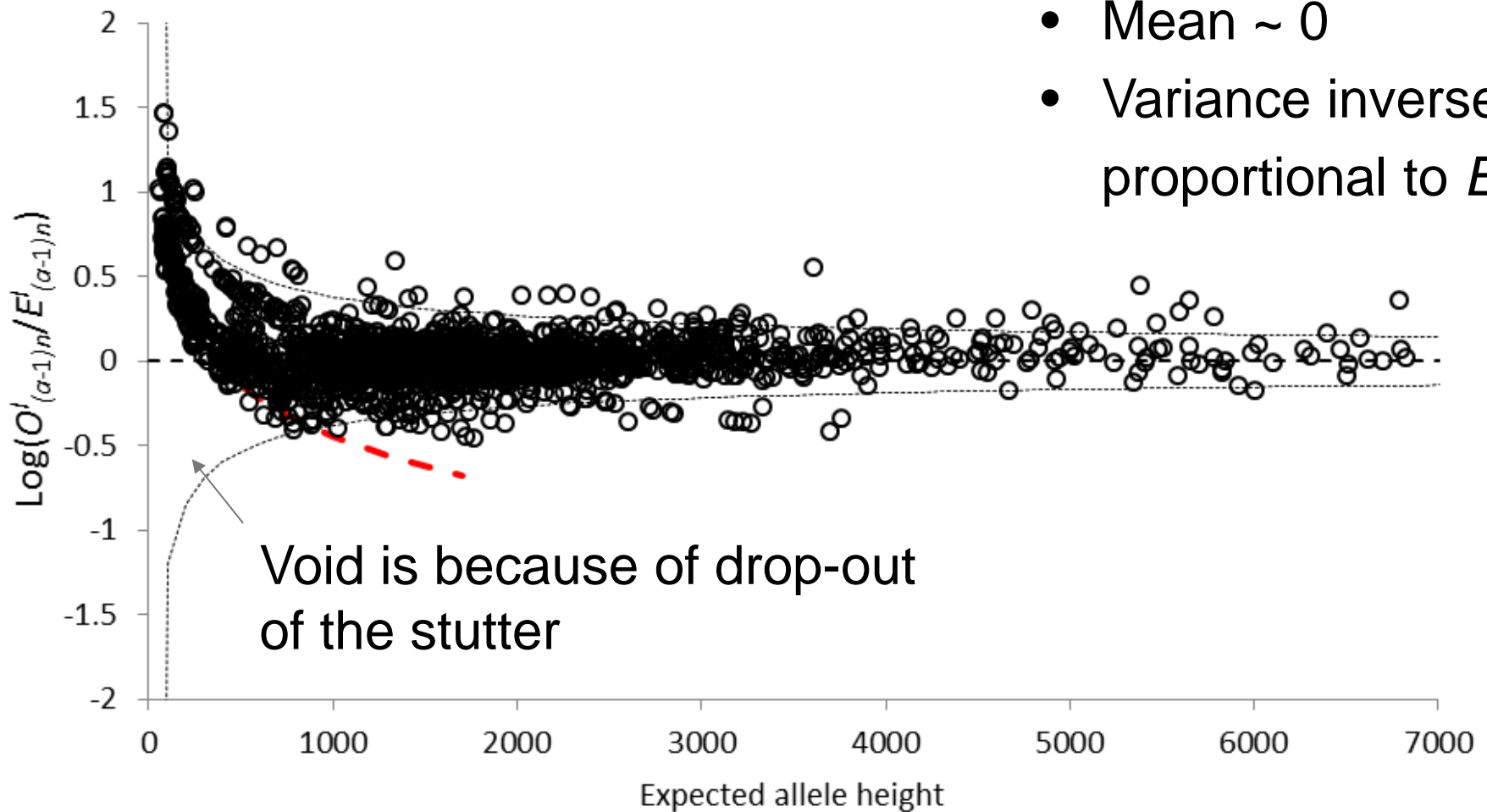
Variance of allele model

- Mean ~ 0
- Variance inversely proportional to E



Variance of stutter model

- Mean ~ 0
- Variance inversely proportional to E



STRMIX™
RESOLVE
MORE DNA
MIXTURES.