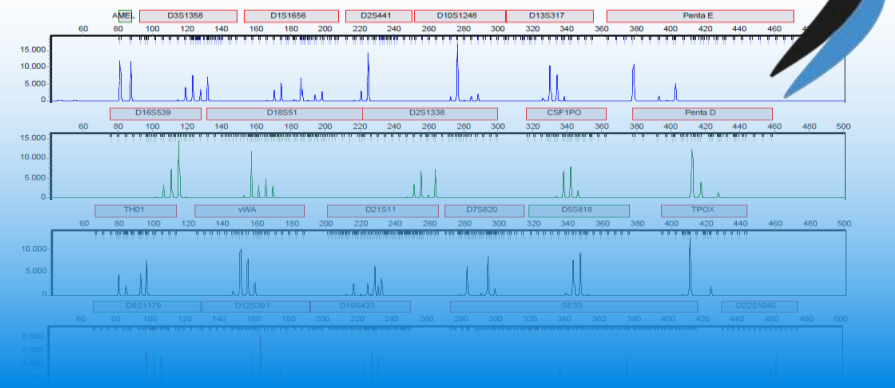
Netherlands Forensic Institute
Ministry of Justice and Security

Interpretation of (mixed) DNA profiles within the NFI



Corina Benschop

13/11/2020

DNA profile interpretation QOL-00711

Step 1:
Define whether the profile is single source or a mixture

Step 2:
Identify the number of contributors to the mixture

Step 3:
Estimate the relative ratio of the individuals contributing to the mixture, and, if possible deconvolve the mixture

Step 4:
Determine whether the profile is suitable for comparison to a reference profile/ suitable for storage or search in the DNA database

Step 5A:
Compare the evidence profile to the reference profile(s)

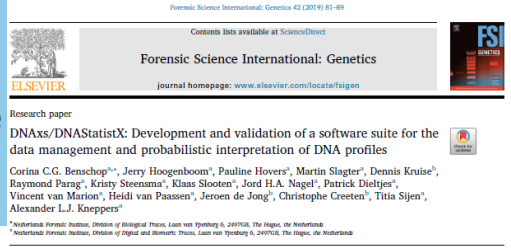
Step 5B:
DDB storage/search

Step 6:
Statistical analyses to define the weight of evidence

Today

DNA xS

Netherlands Forensic Institute
Ministry of Justice and Security


Forensic Science International: Genetics 42 (2019) 81–89

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

DNAxs/DNAStatistX: Development and validation of a software suite for the data management and probabilistic interpretation of DNA profiles

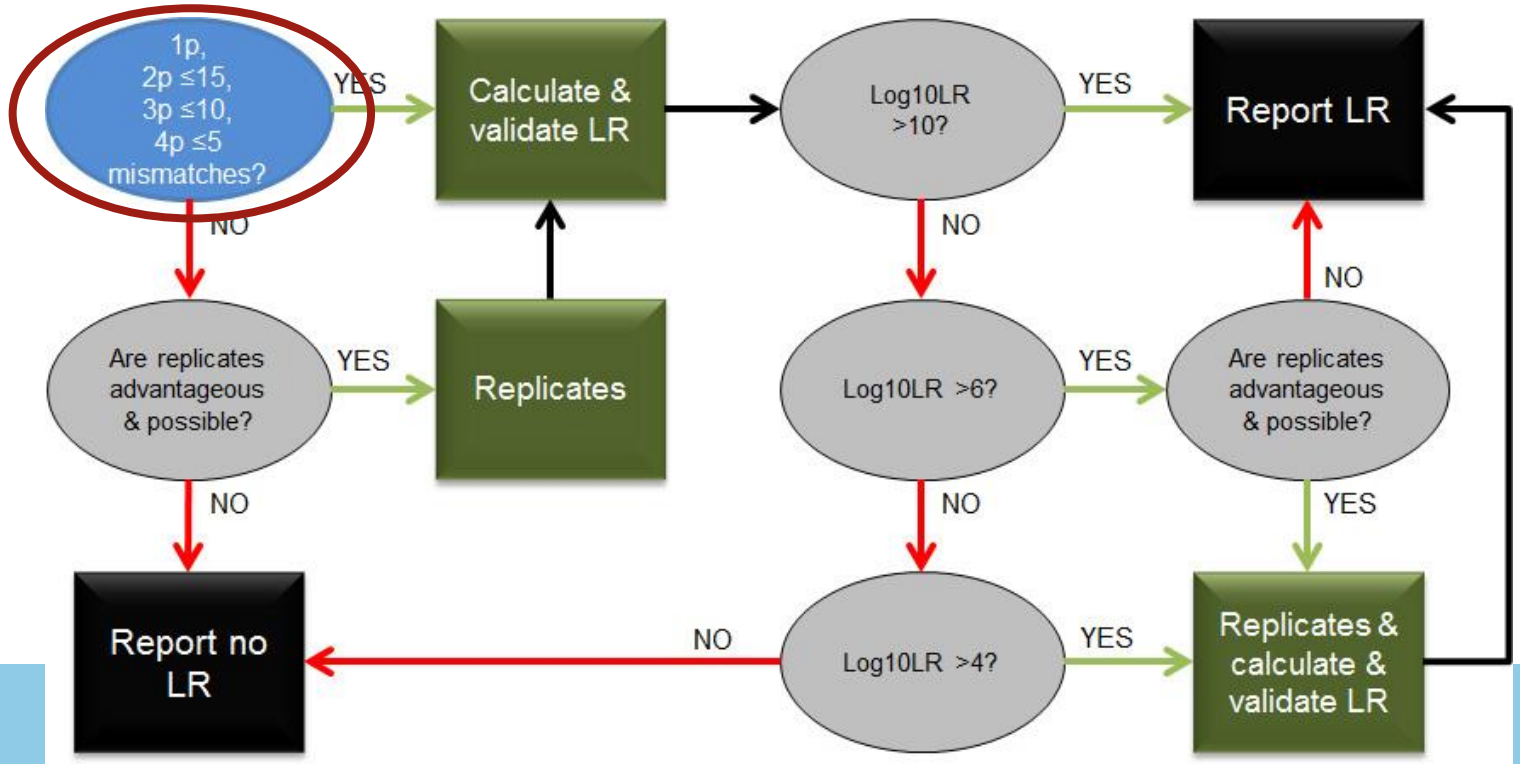
Corina C.G. Benschop[a,*], Jerry Hoogenboom[a], Pauline Hovers[a], Martin Slagter[a], Dennis Kruise[b], Raymond Parag[a], Kristy Steensma[a], Klaas Slooten[b], Jord H.A. Nagel[a], Patrick Dieltjes[a], Vincent van Marion[a], Heidi van Paassen[a], Jeroen de Jong[a], Christophe Creeten[b], Titia Sijen[a], Alexander L.J. Kneppers[a]

[a] Netherlands Forensic Institute, Division of Biological Traces, Laan van Ypenburg 6, 2497GB, The Hague, the Netherlands
[b] Netherlands Forensic Institute, Division of Digital and Biometric Traces, Laan van Ypenburg 6, 2497GB, The Hague, the Netherlands

Guidelines for use in forensic casework

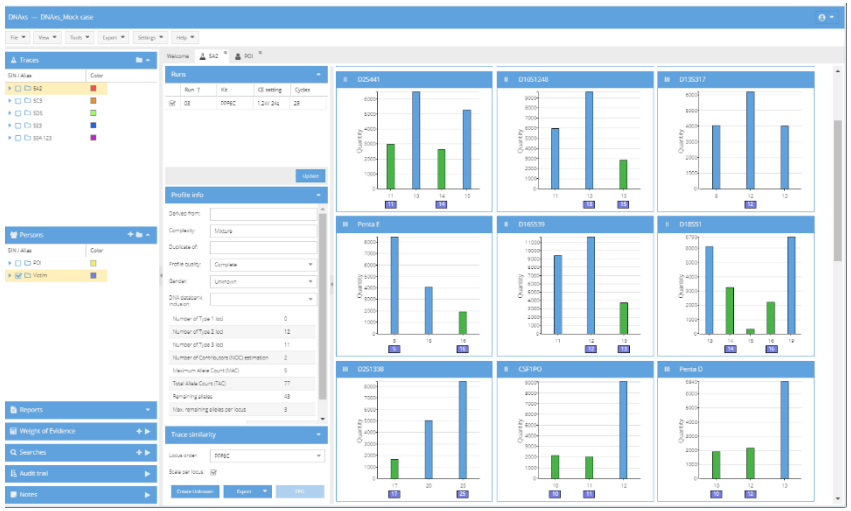
- Based on results from research, casework experience, including efficiency and usefulness of calculations.
- Clearly, the case circumstances, availability of other stains and profiles within the case are factors considered in the process.

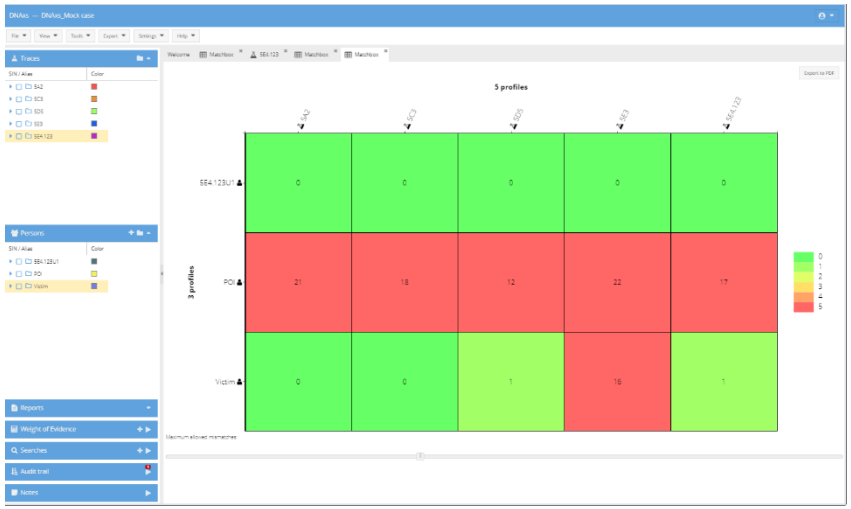
Summarized guidelines to be used as helpful tool in the decision making process are as follows:

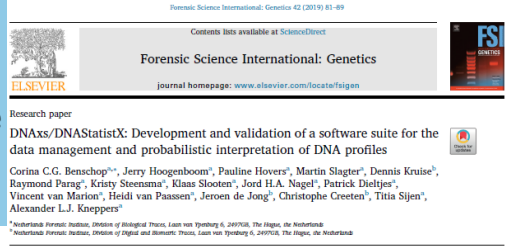


Mismatches?

- Alleles of, for example, a person of interest (POI) that are not observed in the trace profile. I.e. unseen alleles, and in case of true donors, drop-outs.






Forensic Science International: Genetics 42 (2019) 81–89

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

DNAxs/DNAStatistX: Development and validation of a software suite for the data management and probabilistic interpretation of DNA profiles

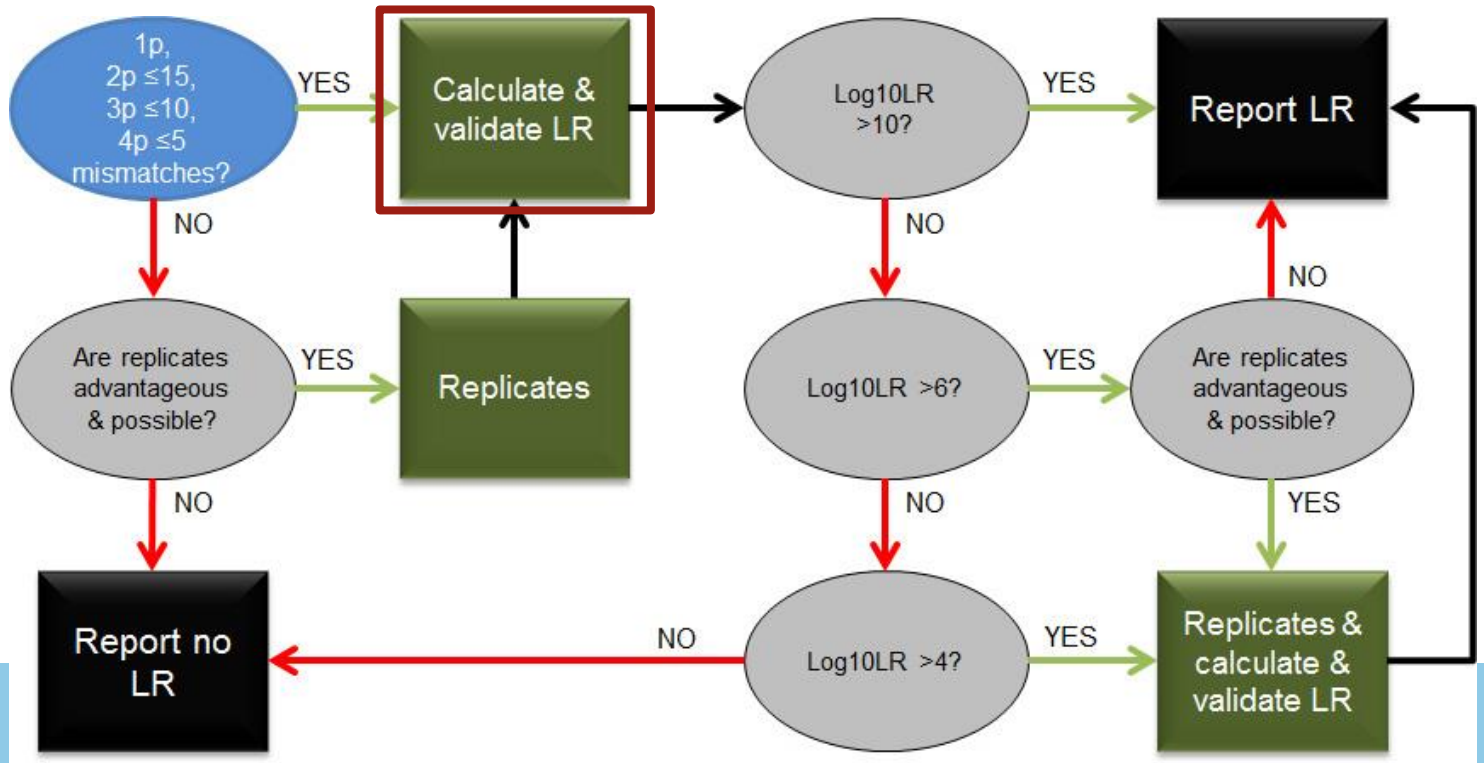
Corina C.G. Benschop[a,*], Jerry Hoogenboom[a], Pauline Hovers[a], Martin Slagter[a], Dennis Kruise[b], Raymond Parag[a], Kristy Steensma[a], Klaas Slooten[b], Jord H.A. Nagel[a], Patrick Dieltjes[a], Vincent van Marion[a], Heidi van Paassen[a], Jeroen de Jong[a], Christophe Creeten[b], Titia Sijen[a], Alexander L.J. Kneppers[a]

[a] Netherlands Forensic Institute, Division of Biological Traces, Laan van Ypenburg 6, 2497GB, The Hague, the Netherlands
[b] Netherlands Forensic Institute, Division of Digital and Biometric Traces, Laan van Ypenburg 6, 2497GB, The Hague, the Netherlands

Guidelines for use in forensic casework

- Based on results from research, casework experience, including efficiency and usefulness of calculations.
- Clearly, the case circumstances, availability of other stains and profiles within the case are factors considered in the process.

Summarized guidelines to be used as helpful tool in the decision making process are as follows:



Weight of Evidence

- The likelihood ratio (LR) is a measure for weight of evidence.
- Using LR calculations we weigh hypotheses H1 (or Hp) and H2 (or Hd), given observed data.
- LR gives a weight of whether the data supports H1 or H2.

LR value	
LR = 1	Neutral evidence, the data support none of the hypotheses
LR > 1	The data supports that H1 is true
LR <1	The data supports that H2 is true

Binary | qualitative | quantitative

- Binary (classical) likelihood ratio calculation
  - Considers genotypic probabilities

- Qualitative/ semi-continuous model:
  - Considers genotypic probabilities
  - Considers drop-out and drop-in

- Quantitative/ continuous model:
  - Considers genotypic probabilities
  - Considers peak height information
  - Drop-out and drop-in based on peak height models

Underlying model

- LR is a function of the genotypic frequencies
  - Hardy Weinberg Equilibrium, assume independent association of the alleles within loci:
    - Genotype 'aa' has frequency p_a^2
    - Genotype 'ab' has frequency $2p_ap_b$
  - Population Sub-structure model
    - Introduces "Fst/Theta" correction
    - Genotype probabilities increase (reduction in LR)
    - Balding-Nichols sampling formula
- Linkage equilibrium
  - Markers bear independent information
  - The product rule: Multiply between loci

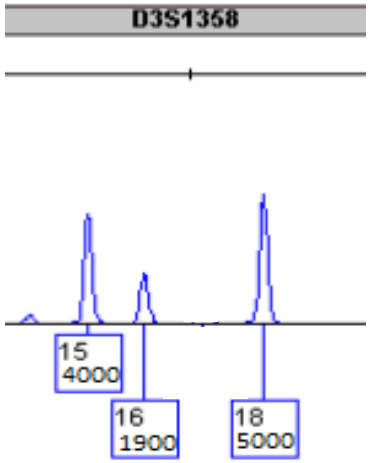
Steps in computing likelihood ratios

User: Define propositions (number of contributors, person of interest etc.)

	Binary
Define possible genotype combinations	☑
Determine genotypic probabilities	☑
Calculate profile likelihoods Sum up probabilities/locus & multiply between loci	☑
Calculate LR: likelihood Hp / likelihood Hd	☑

Example of a binary (classical) likelihood ratio calculation

Example: High-template two-person mixture



	D3S1358
Evidence	15, 16, 18
Victim	15, 18
POI	16, 18

$$LR = \frac{Hp{:}\,Victim + POI\ contributed\ to\ the\ sample}{Hd{:}\,Victim + Unknown\ person\ (unrelated\ to\ the\ POI)\ contributed\ to\ the\ sample}$$

$$LR = \frac{Hp{:}\,V + POI}{Hd{:}\,V + U}$$

Under Hp

$Hp: V + POI$	**D3S1358**
Evidence	15, 16, 18
Victim	15, 18
POI	16, 18

In this case, there are no unknowns

$$Pr(Evidence \mid H_p) = 1$$

Sum up the probabilities for all plausible genotypes

Calculate profile likelihood

Under Hd

$Hd: V + U$

	D3S1358
Evidence	15, 16, 18
Victim	15, 18
Unknown	?

$2p_{15}p_{16}$

$2p_{16}p_{18}$

p_{16}^2

$$Pr(Evidence \mid H_d) = 2p_{15}p_{16} + 2p_{16}p_{18} + p_{16}^2$$

$$(2 \times 0.25 \times 0.04) + (2 \times 0.04 \times 0.15) + (0.04^2) = 0.0336$$

- The victim's profile explains alleles 15 and 18
- The unknown has to have allele **16**: allele **16** is constrained

Under Hp and Hd

$Hp: V + POI$

	D3S1358
Evidence	15, 16, 18
Victim	15, 18
POI	16, 18

Likelihood under Hp

$$Pr(Evidence \mid H_p) = 1$$

$Hd: V + U$

	D3S1358
Evidence	15, 16, 18
Victim	15, 18
Unknown	15, **16**
	16, 18
	16, **16**

Likelihood under Hd

$$Pr(Evidence \mid H_d) = 2p_{15}p_{16} + 2p_{16}p_{18} + p_{16}^2$$

Under Hp and Hd: LR for 1 locus

$Hp: V + POI$

	D3S1358
Evidence	15, 16, 18
Victim	15, 18
POI	16, 18

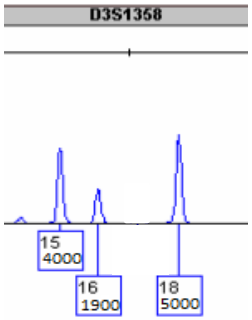
$Hd: V + U$

	D3S1358
Evidence	15, 16, 18
Victim	15, 18
Unknown	15, **16**
	16, 18
	16, **16**

$$LR = \frac{1}{2p_{15}p_{16} + 2p_{16}p_{18} + p_{16}^2}$$

$$LR = \frac{1}{0.0336} = 29.76$$

Example with more loci: Multiply between loci



	D3S1358
Evidence	15, 16, 18
Victim	15, 18
Unknown	15, **16**
	16, 18
	16, **16**

	Locus 2
Evidence	10, 12, 14, 15
Victim	14, 15
Unknown	**10**, **12**

	Locus 3
Evidence	9, 9.3
Victim	9, 9.3
Unknown	9, 9.3
	9, 9
	9.3, 9.3

$2p_{15}p_{16} + 2p_{16}p_{18} + p_{16}^2$ X $2p_{10}p_{12}$ X $2p_9p_{9.3} + p_9^2 + p_{9.3}^2$

Steps in computing likelihood ratios

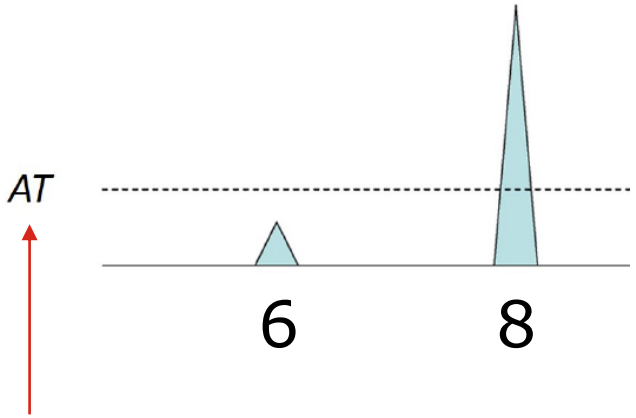
User: Define propositions (number of contributors, person of interest etc.)

	Binary	Qualitative
Define possible genotype combinations	☑	☑
Determine genotypic probabilities	☑	☑
Estimate drop-out probabilities	☒	☑
Calculate profile likelihoods Sum up probabilities/locus & multiply between loci	☑	☑
Calculate LR: likelihood Hp / likelihood Hd	☑	☑

The drop-out phenomena

Evidence

Contributor with genotype 6/8

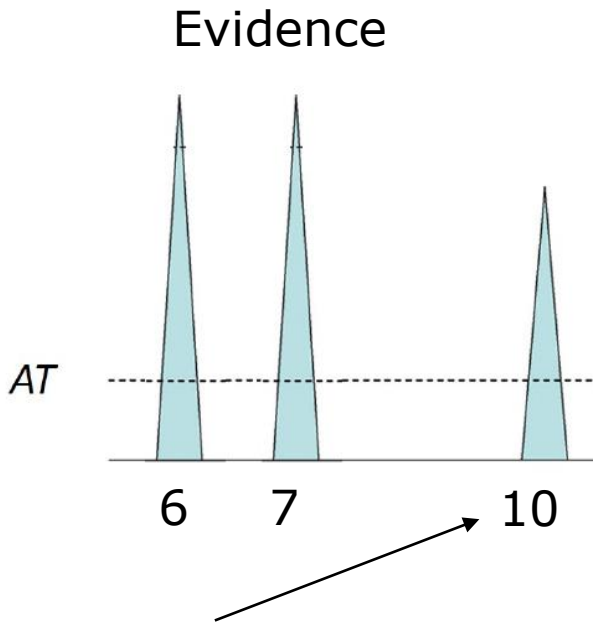


Analytical threshold (AT) used to separate true signal from noise

Small DNA-amount leads to low peak falling below AT
Corresponding allele removed from evidence

The drop-in phenomena

Contributor with genotype 6/7

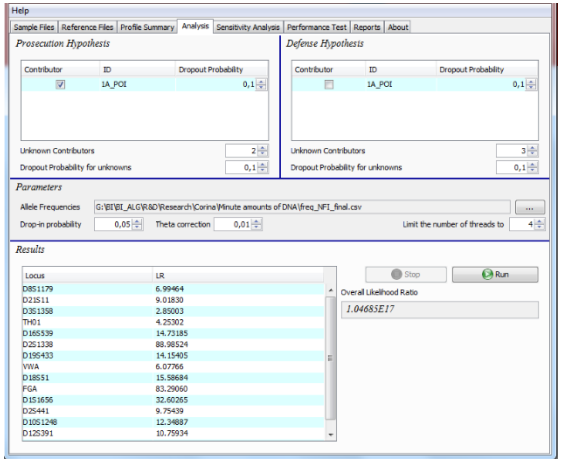


A spurious allele 10 observed, not belonging to the contributor(s)

Typical more likely with smaller peak heights than higher peak heights

LR with drop-out and drop-in *using the qualitative model*

- Main theory described in:
  - Haned at al, FSIG 2012
  - DNA commission ISFG, FSIG 2012
  - Gill et al, FSI 2007
  - **Curran et al, FSI 2005**
- Two key parameters in the model:
  - Drop-out & drop-in



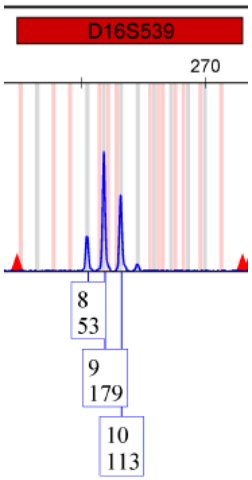
Introducing drop-out parameter *d*

- An allele drops out with a probability of *d*
- An allele does not drop out with a probability of 1 – *d*

- Allele drop-out from a heterozygote genotype: *d*
- Allele drop-out from a homozygote genotype: $d' = d^2$

The Q-allele: a possible non-observed allele *explained as drop-out*

$Hp: POI + U$



Q-allele, can be anything except the observed alleles (here 8, 9, 10)

No drop-in →

	D16S539
Evidence	8, 9, 10
POI	9, 10
Unknown	8, 8 8, 9 8, 10 8, Q

Assuming drop-in →

Additional possibilities	Allele drop-in
9,9	8
9,10	8
9,Q	8
10,10	8
10,Q	8
Q,Q	8

This increases the number of terms under Hd

Steps in computing likelihood ratios

User: Define propositions (number of contributors, person of interest etc.)

	Binary	Qualitative	Quantitative
Estimate parameters (by optimizer) Mixture proportion, peak height (PH) expectation, PH variance, degradation slope, stutter	✗	✗	✓
Define possible genotype combinations	✓	✓	✓
Determine genotypic probabilities	✓	✓	✓
Estimate drop-out probabilities	✗	✓	✗
Compute PH weights	✗	✗	✓
Calculate profile likelihoods Sum up probabilities/locus & multiply between loci	✓	✓	✓
Calculate LR: likelihood Hp / likelihood Hd	✓	✓	✓

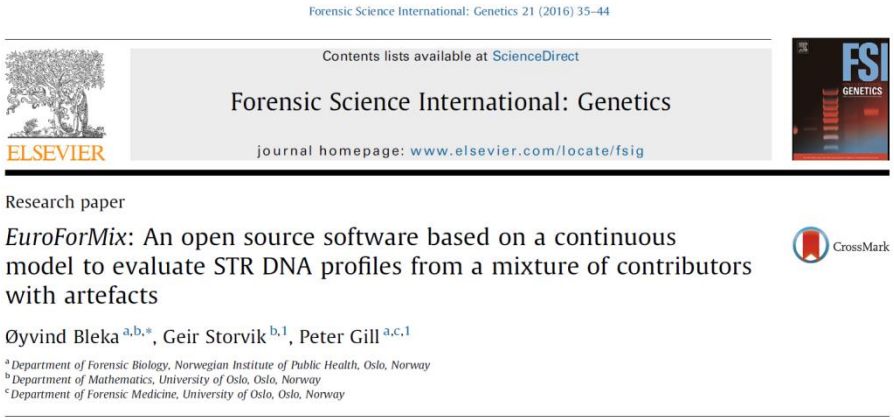
Various continuous models available

- EuroForMix — Open source, freely available
  - Bulletproof
- DNAStatistX
- Kongoh
- LikeLTD

- CeesIT — Commercial or proprietary
- LiRaHT
- TrueAllele
- STRmix
- DNAmixtures
- DNA View Mixture Solution
- Genoproof Mixture
- maSTR
- …

DNAStatistX is based on EuroForMix

- euroformix.com



Forensic Science International: Genetics 21 (2016) 35–44

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts

Øyvind Bleka [a,b,*], Geir Storvik [b,1], Peter Gill [a,c,1]

[a] Department of Forensic Biology, Norwegian Institute of Public Health, Oslo, Norway
[b] Department of Mathematics, University of Oslo, Oslo, Norway
[c] Department of Forensic Medicine, University of Oslo, Oslo, Norway

Appl. Statist. (2015)
64, *Part* 1, *pp.* 1–32

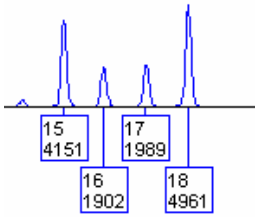
Analysis of forensic DNA mixtures with artefacts

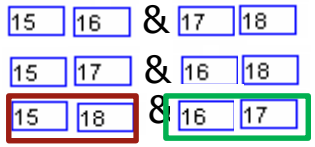
R. G. Cowell,
City University London, UK
T. Graversen and S. L. Lauritzen
University of Oxford, UK
and J. Mortera
Università Roma Tre, Italy

- EuroForMix uses the gamma distribution to model peak heights
  - Alike DNA·VIEW Mixture Solution, LiRaHT, DNA-mixtures, LikeLTD, Kongoh

What to expect from a quantitative model?

- Benefit from using peak heights in case
  - Contributors have distinctive peak heights and LR <highest reporting threshold
  - DNA-profiles show degradation





Trend: Larger weight of evidence for true contributors
Non-contributors that cannot be excluded based on allele matching might be excluded based on peak heights

- No/not much difference compared to semi-continuous model when peak heights are not very informative
  - In case donors contributed equally
  - In case all peaks are within the stochastic range
  - In case one can condition on known contributors

Forensic Science International: Genetics 25 (2016) 85–96

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles

Øyvind Bleka [a,b,*], Corina C.G. Benschop [c], Geir Storvik [b], Peter Gill [a,d]

Example of an LR calculation performed using a quantitative/ continuous model

Software: EuroForMix (EFM) or DNAStatistX

Estimate parameters

The continuous model uses parameters:

- Mixture proportions
- Peak height expectation
- Peak height variance
- Degradation slope
- Backward stutter (n-1) proportion
- Forward stutter (n+1) proportion

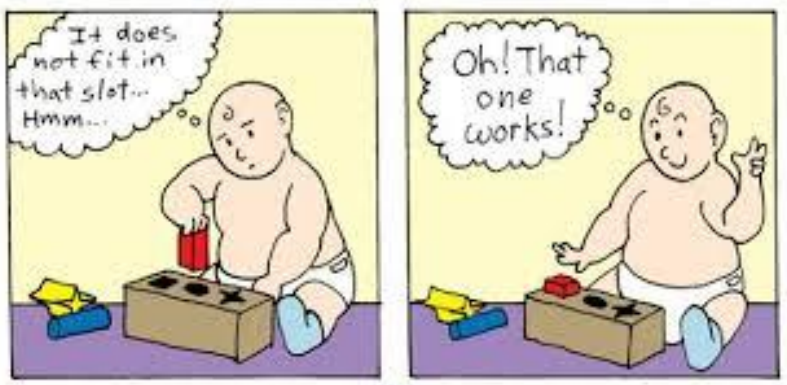
$\boldsymbol{\beta}$

To be **estimated** under $H = H_p$ and $H = H_d$ separately

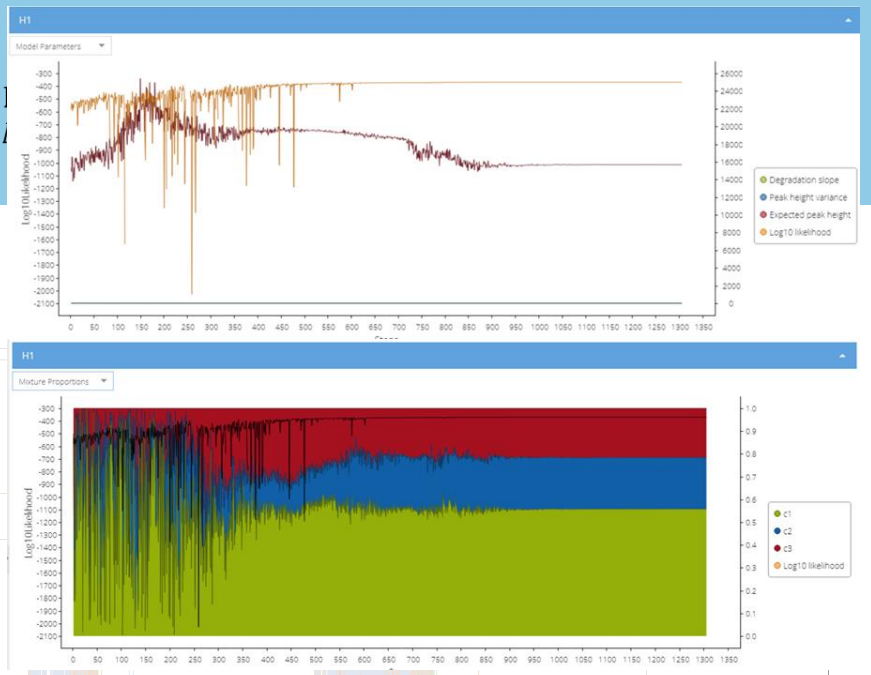
Using evidence profile data!

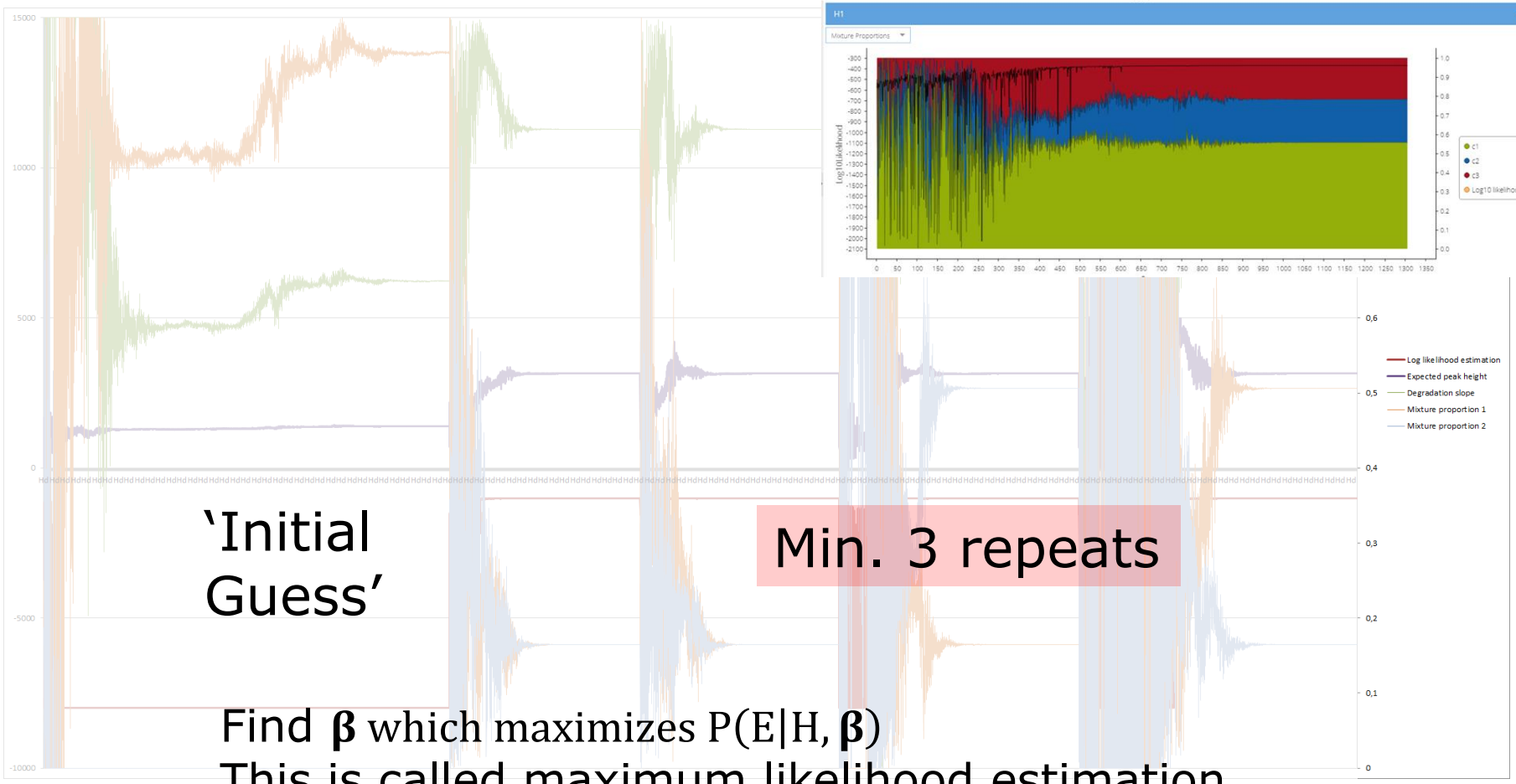
However, these are unknown, thus…

Trial and error using optimizer



Optimizer



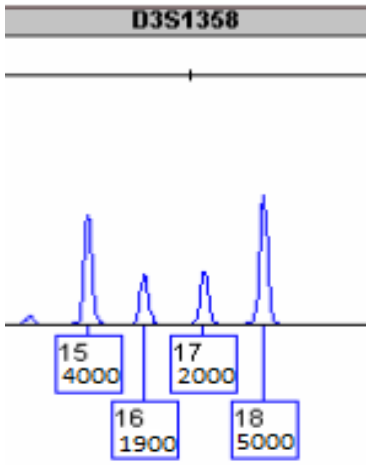


'Initial Guess'

Min. 3 repeats

Find $\boldsymbol{\beta}$ which maximizes $P(E|H, \boldsymbol{\beta})$
This is called maximum likelihood estimation

Example: Two-person mixture



	D16S539
Evidence	15, 16, 17, 18
POI	16, 17

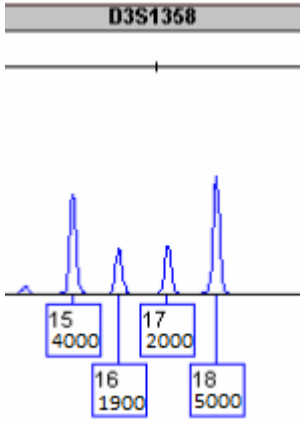
Assumption: two-person mixture

$$LR = \frac{Hp: POI + Unknown\ contributed\ to\ the\ sample}{Hd: Two\ Unknown\ persons\ (unrelated\ to\ the\ POI)\ contributed\ to\ the\ sample}$$

$$LR = \frac{Hp: POI + U}{\boldsymbol{Hd: U + U}}$$

$$LR = \frac{Hp\colon\ POI + U}{\boldsymbol{Hd}\colon\ \boldsymbol{U} + \boldsymbol{U}}$$

Define possible genotype combinations



Possible genotype combinations when regarding drop-out and drop-in:

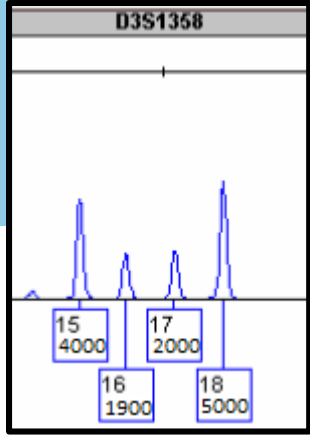
Donor A	Donor B	Drop-in
15/16	17/18	-
15/17	16/18	-
15/18	16/17	-
17/18	15/16	-
16/18	15/17	-
16/17	15/18	-
But also…		
15/15	16/17	18
15/18	16/d.o.	17
d.o.	d.o.	15/16/17/18
Etc…		

All possible genotypes per contributor (n=15):

Donor A allel 1	Donor A allel 2
15	15
15	16
15	17
15	18
16	16
16	17
16	18
17	17
17	18
18	18
Q	Q
15	Q
16	Q
17	Q
18	Q

Drop out

With two unknowns: $15 \times 15 = 225$ combinations

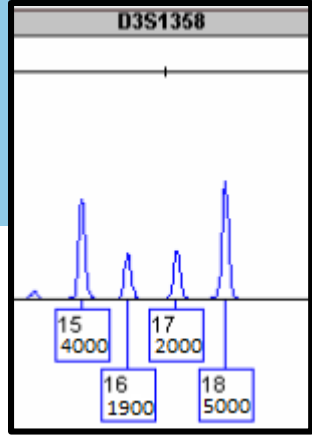


Determine genotype probabilities

Using allele frequencies

Genotype probability heterozygote	Genotype probability homozygote	Frequency of a dropped-out allele	Drop-in probability	Probability for no drop-in
$2P_aP_b$	P_a^2	$P_Q=1-P_a-P_b-P_c-P_d$	$PrC*P_a$	$1-PrC$

Donor A	Donor B	Drop-in	Genotype probability	Drop-in probability
15/16	17/18	-		
15/17	16/18	-		
15/18	16/17	-		
17/18	15/16	-		
16/18	15/17	-		
16/17	15/18	-		
But also…				
15/15	16/17	18		
15/18	16/d.o.	17		
d.o.	d.o.	15/16/17/18		
Etc…				

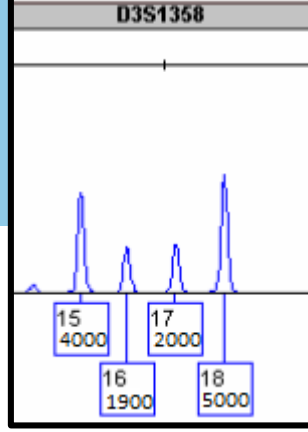


Likelihoods excl. peak height information

- Calculate **likelihood weights** (per allele)
  Likelihoods **excl.** peak height information (qual model)

Donor A	Donor B	Drop-in	Genotype probability	Drop-in probability	Likelihood (product genotype & drop-in probabilities)
15/16	17/18	-	$2P_{15}P_{16}*2P_{17}P_{18}$	1-PrC	0.007929
15/17	16/18	-	$2P_{15}P_{17}*2P_{16}P_{18}$	1-PrC	0.007929
15/18	16/17	-	$2P_{15}P_{18}*2P_{16}P_{17}$	1-PrC	0.007929
17/18	15/16	-	$2P_{17}P_{18}*2P_{15}P_{16}$	1-PrC	0.007929
16/18	15/17	-	$2P_{16}P_{18}*2P_{15}P_{17}$	1-PrC	0.007929
16/17	15/18	-	$2P_{16}P_{17}*2P_{15}P_{18}$	1-PrC	0.007929
But also…					
15/15	16/17	18	$P_{15}{}^2*2P_{16}P_{17}$	$PrC*P_{18}$	4.1837E-06
15/18	16/d.o.	17	$2P_{15}P_{18}*2P_{16}P_Q$	$PrC*P_{17}$	4.7860E-06
d.o.	d.o.	15/16/17/18	$P_Q{}^4$	$PrC^4*P_{15}*P_{16}*P_{17}*P_{18}$	2.5981E-16
Etc…					+
					SUM = 0.047950

Allele likelihoods (weights)

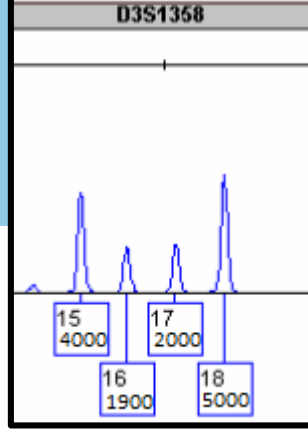


Compute likelihood weights using peak height model

- Find drop-in, drop-out set and for each genotype combination
- Using gamma distribution model for PH (calculate likelihoods)
- Probability of observing a drop-in at a certain peak height: exponential distribution

Allele likelihoods based on PH model

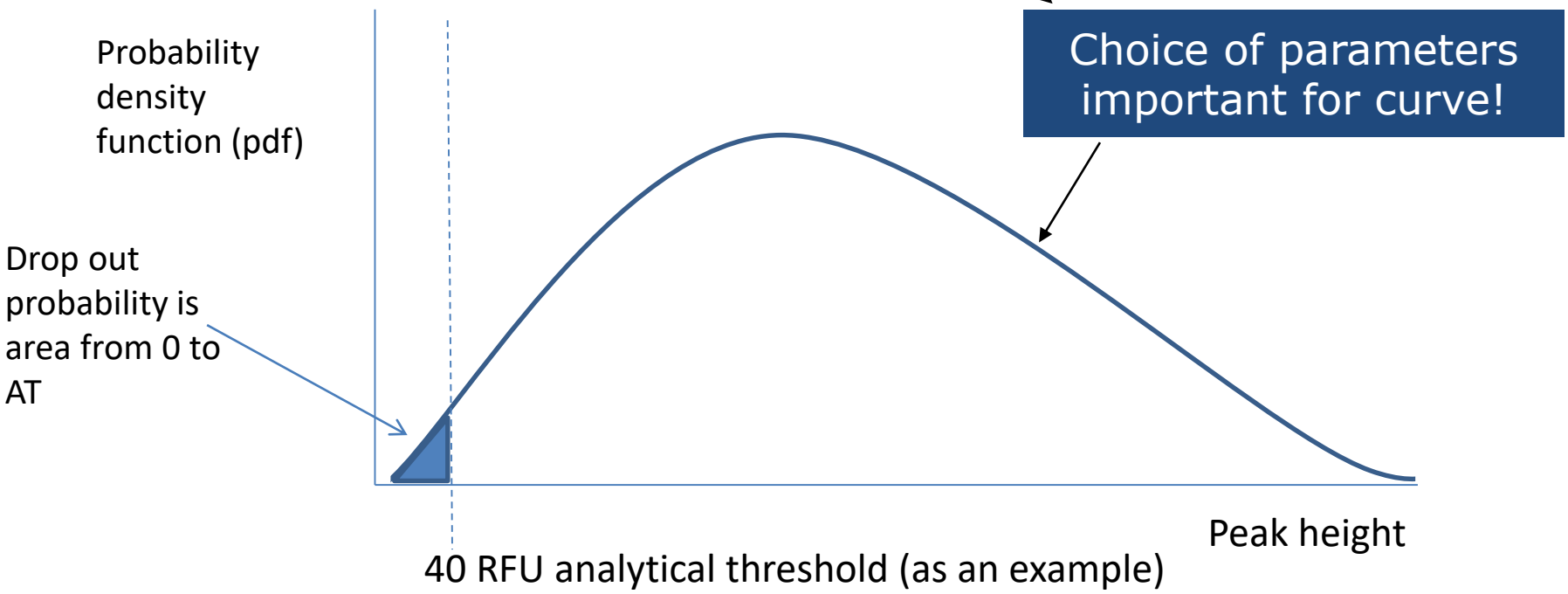
Donor A	Donor B	Drop-in	Genotype probability	Drop-in probability	Likelihood excl PH	PH probability	Likelihood incl PH
15/16	17/18	-	$2P_{15}P_{16}*2P_{17}P_{18}$	1-PrC	0,007929		
15/17	16/18	-	$2P_{15}P_{17}*2P_{16}P_{18}$	1-PrC	0,007929		
15/18	16/17	-	$2P_{15}P_{18}*2P_{16}P_{17}$	1-PrC	0,007929		
17/18	15/16	-	$2P_{17}P_{18}*2P_{15}P_{16}$	1-PrC	0,007929		
16/18	15/17	-	$2P_{16}P_{18}*2P_{15}P_{17}$	1-PrC	0,007929		
16/17	15/18	-	$2P_{16}P_{17}*2P_{15}P_{18}$	1-PrC	0,007929		
But also…							
15/15	16/17	18	$P_{15}^2*2P_{16}P_{17}$	$PrC*P_{18}$	4,1837E-06		
15/18	16/d.o.	17	$2P_{15}P_{18}*2P_{16}P_Q$	$PrC*P_{17}$	4,7860E-06		
d.o.	d.o.	15/16/17/18	P_Q^4	$PrC^4*P_{15}*P_{16}*P_{17}*P_{18}$	2,5981E-16		
	Etc…						



Model

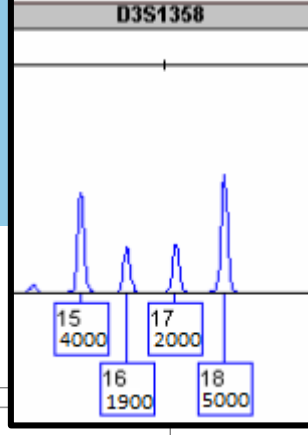
Assumption: Peak heights follow a Gamma distribution

Parameters: Expected peak height, peak height variation, mixture proportions, degradation (and stutter).

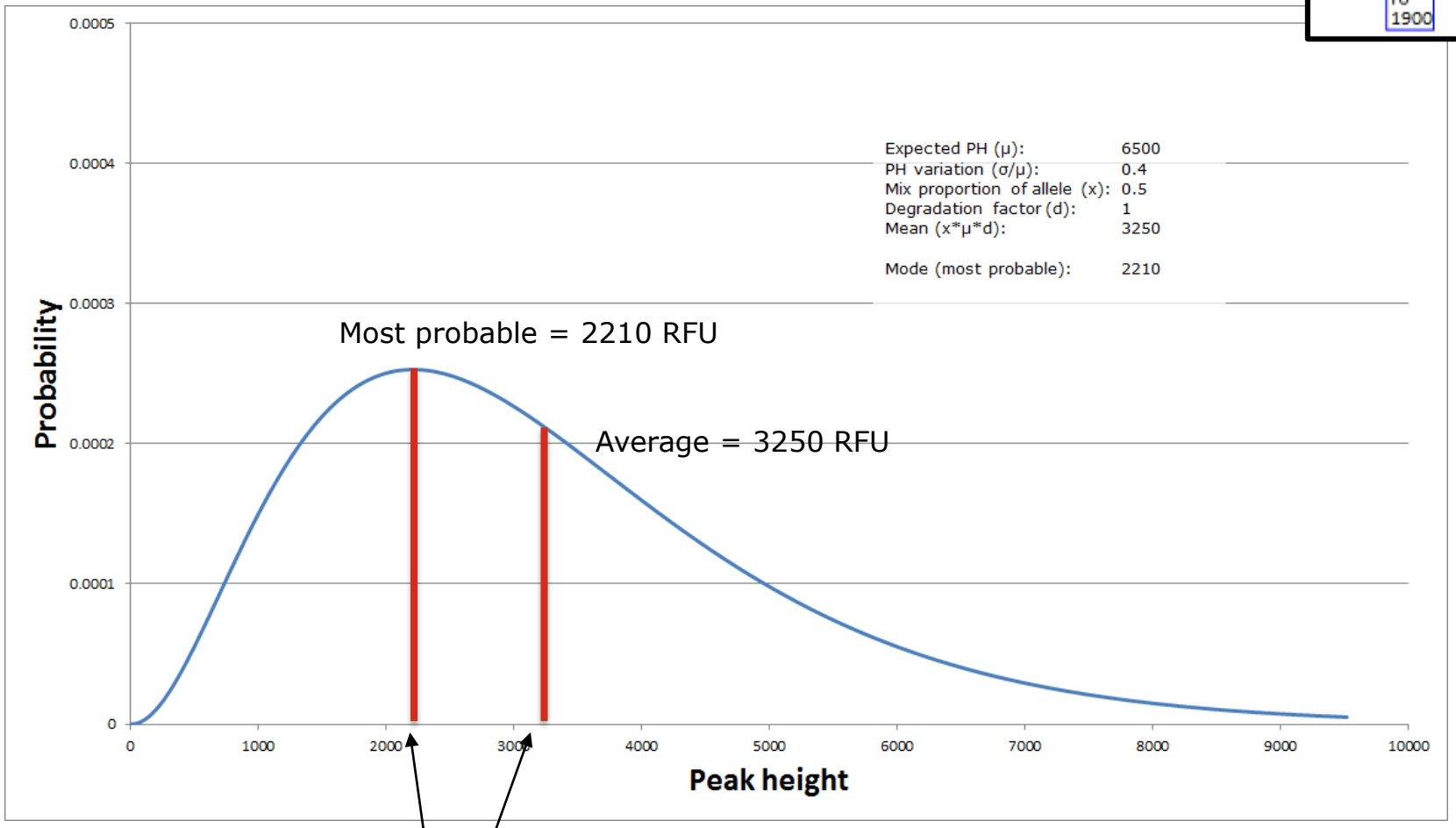


Model

Peak height variation: 0.4
Flat bump → More variation



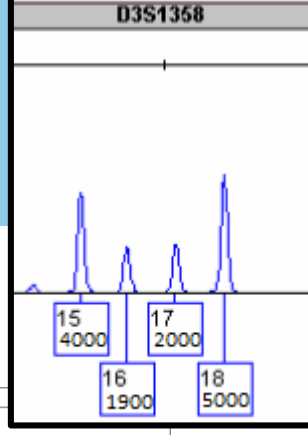
Allele likelihoods (weights)

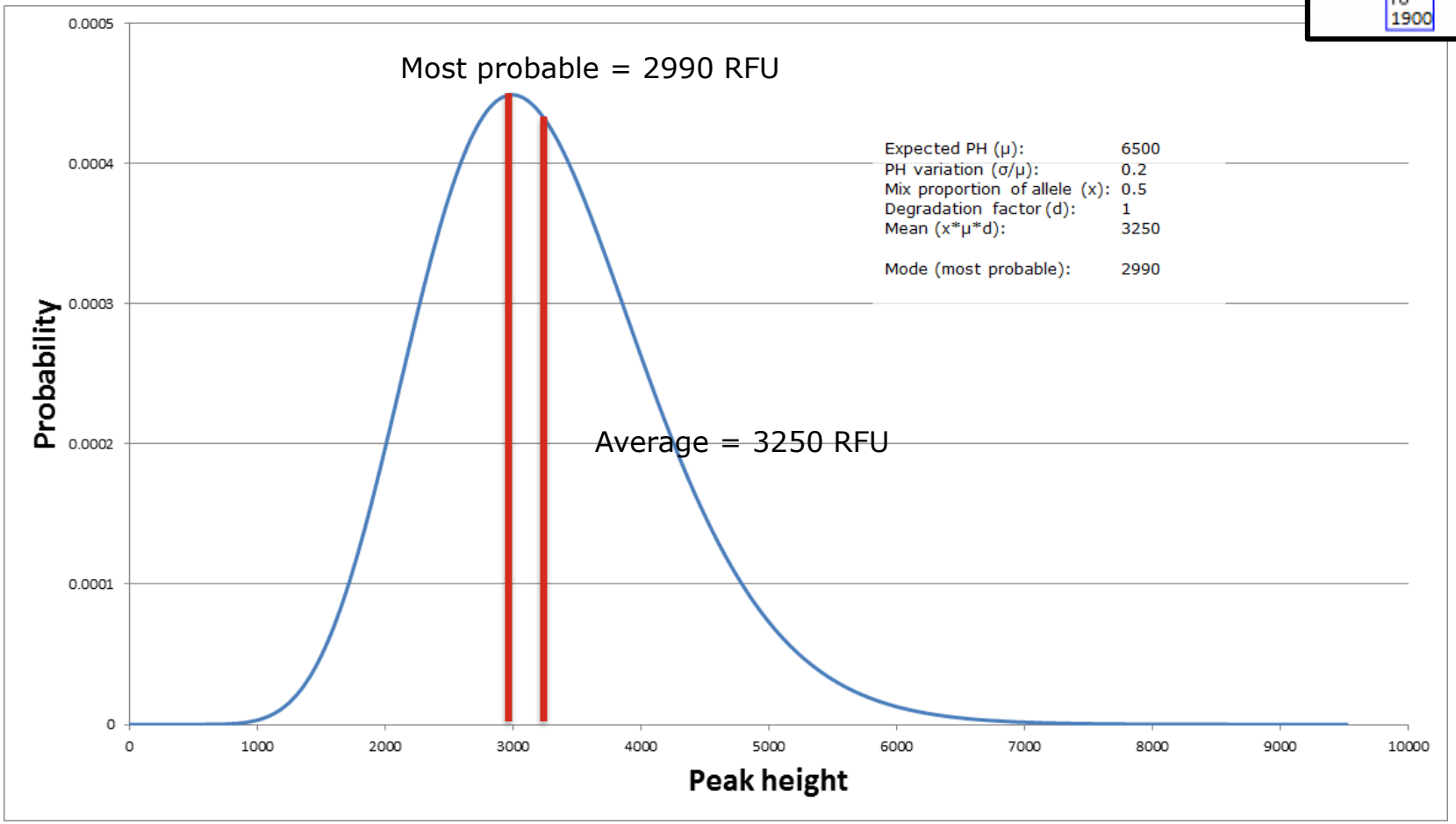


Observed alleles

Model

Peak height variation: 0.2
Steep bump → Less variation

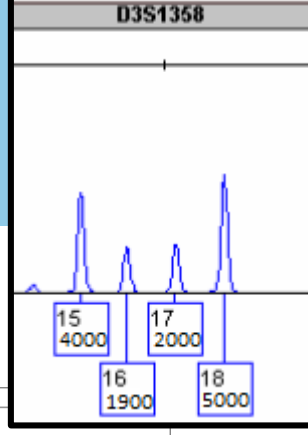


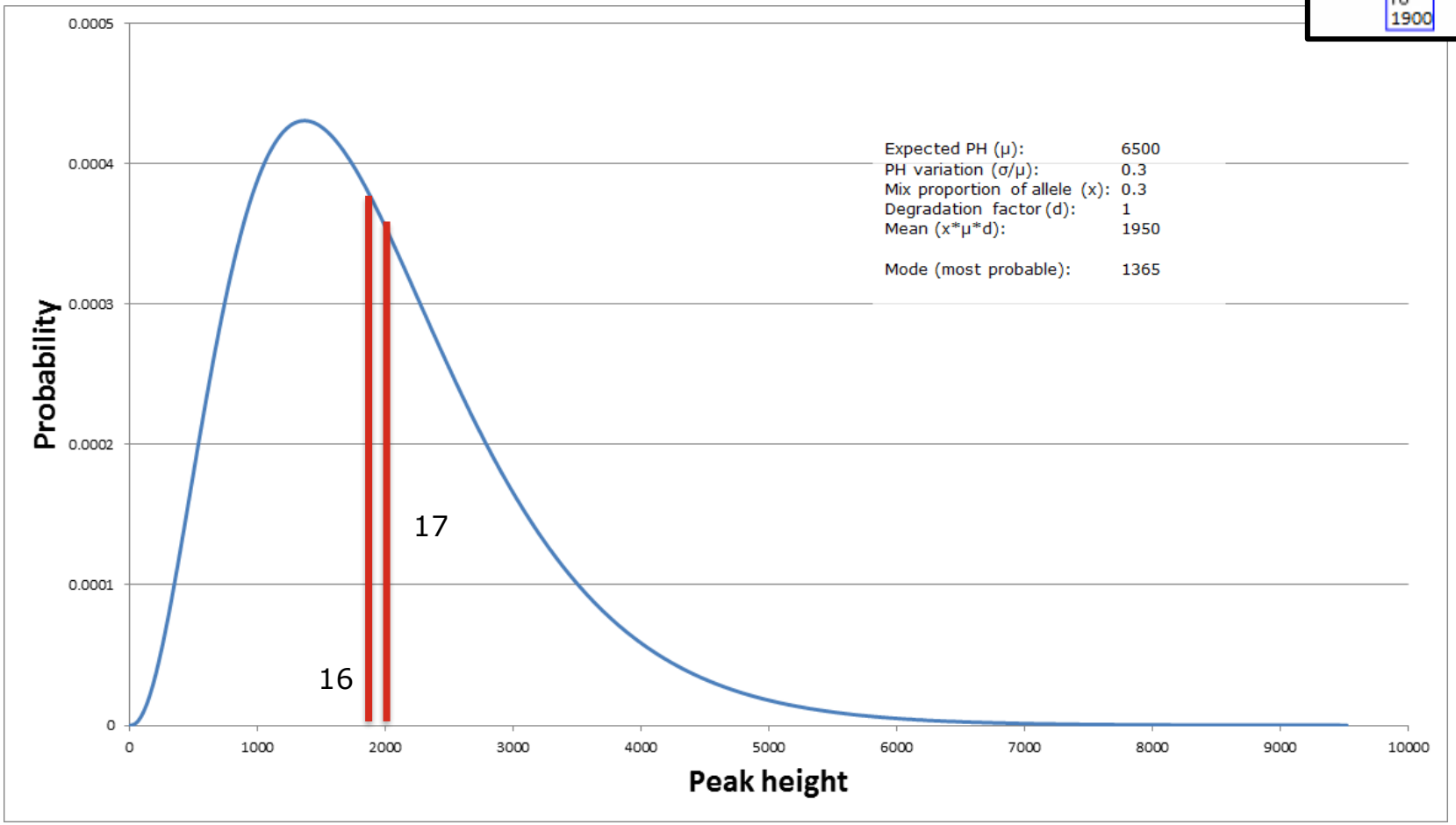


Probable genotype combination (minor)

C1 = 15,18 mixture proportion 0.7

C2 = 16,17 mixture proportion 0.3

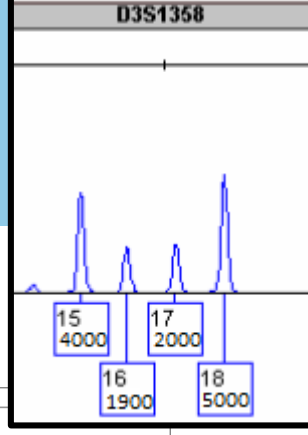


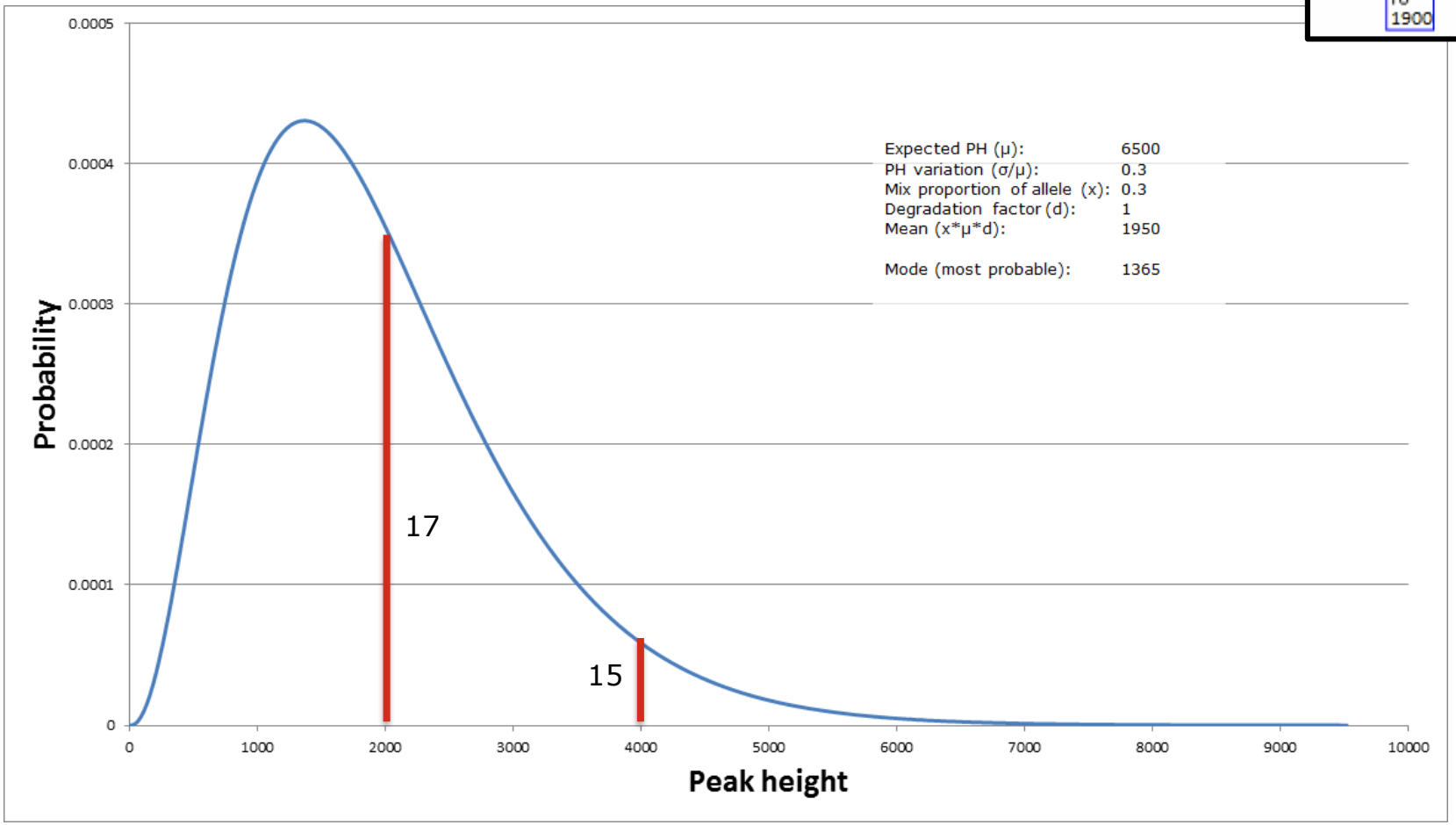


Less probable genotype combination (minor)

C1 = 16,18 mixture proportion 0.7

C2 = 15,17 mixture proportion 0.3

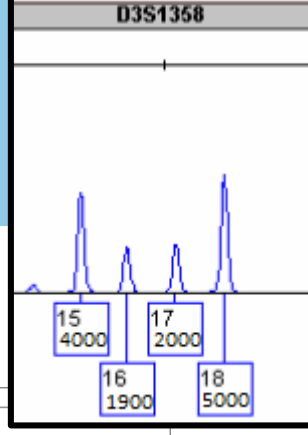


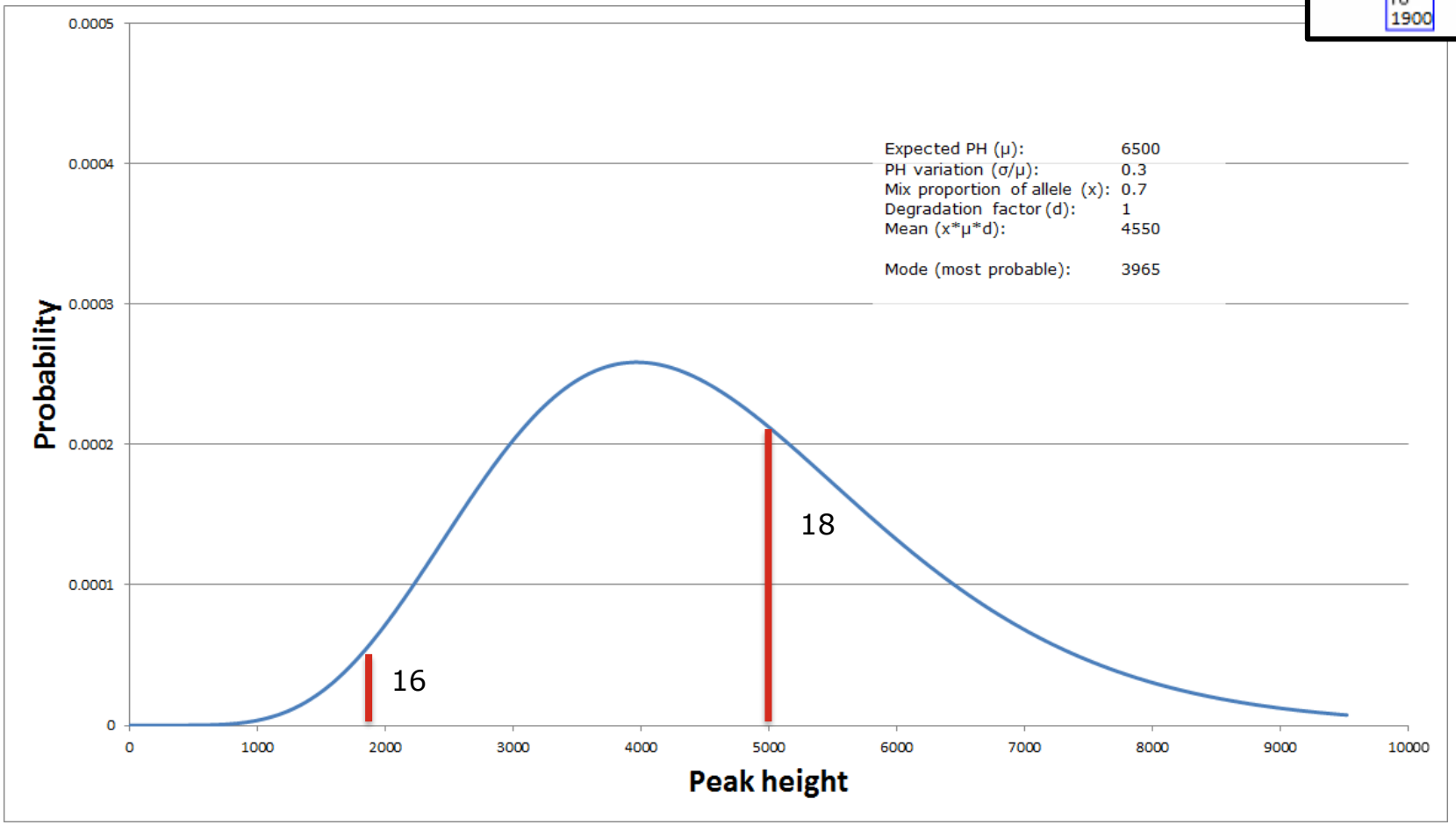


Less probable genotype combination (major)

C1 = 16,18 mixture proportion 0.7

C2 = 15,17 mixture proportion 0.3



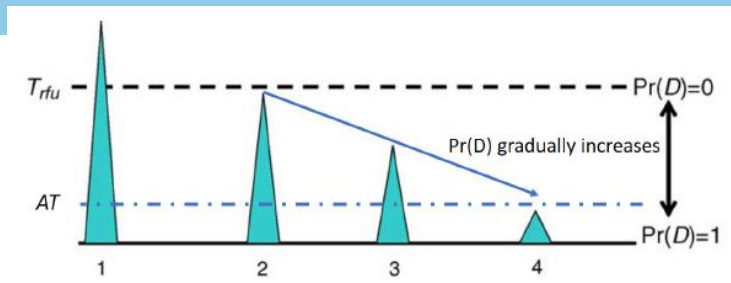


The drop-out model for quantitative models

Netherlands Forensic Institute
Ministry of Justice and Security

The sizes of the peak heights is important

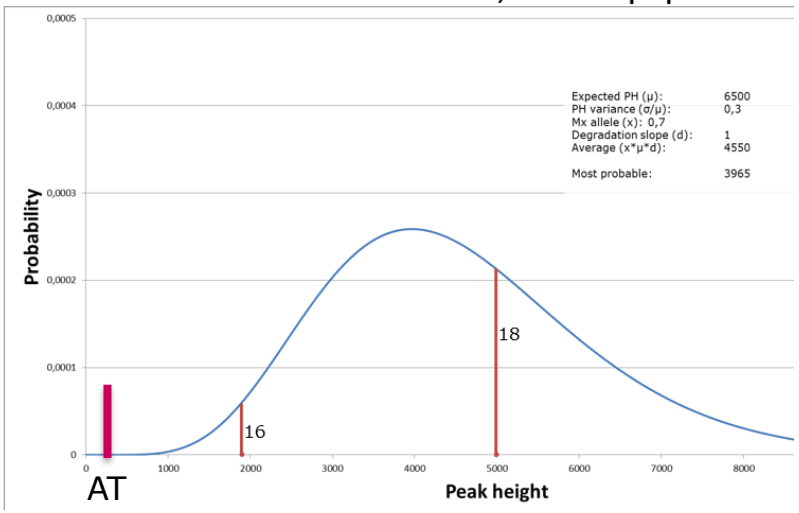
- Large peak heights gives small drop-out probability
- Small peak heights gives high drop-out probability



The likelihood weight for the Q-allele=Pr(dropout)

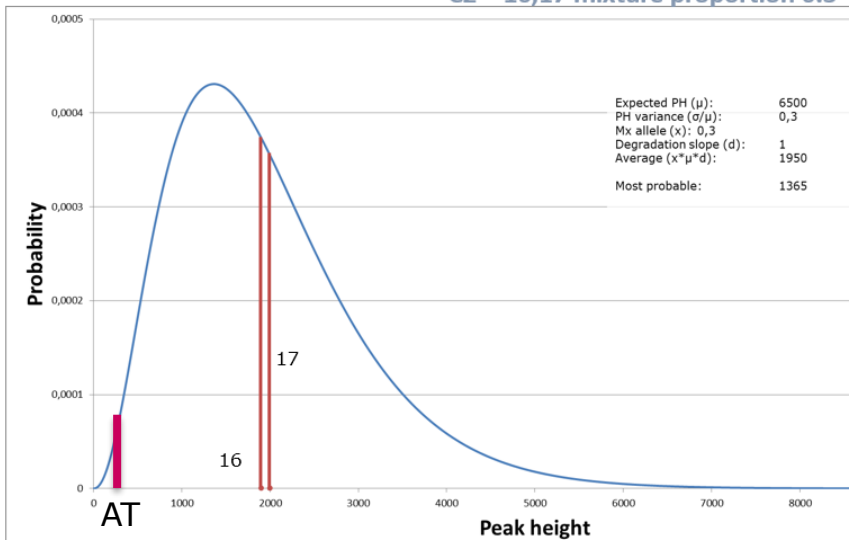
For Major

C1 = 16,18 mixture proportion 0.7
C2 = 15,17 mixture proportion 0.3



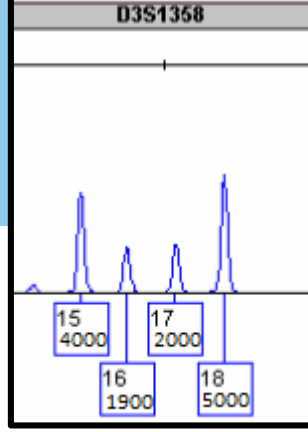
For minor

C1 = 15,18 mixture proportion 0.7
C2 = 16,17 mixture proportion 0.3



Pr(dropout)=0

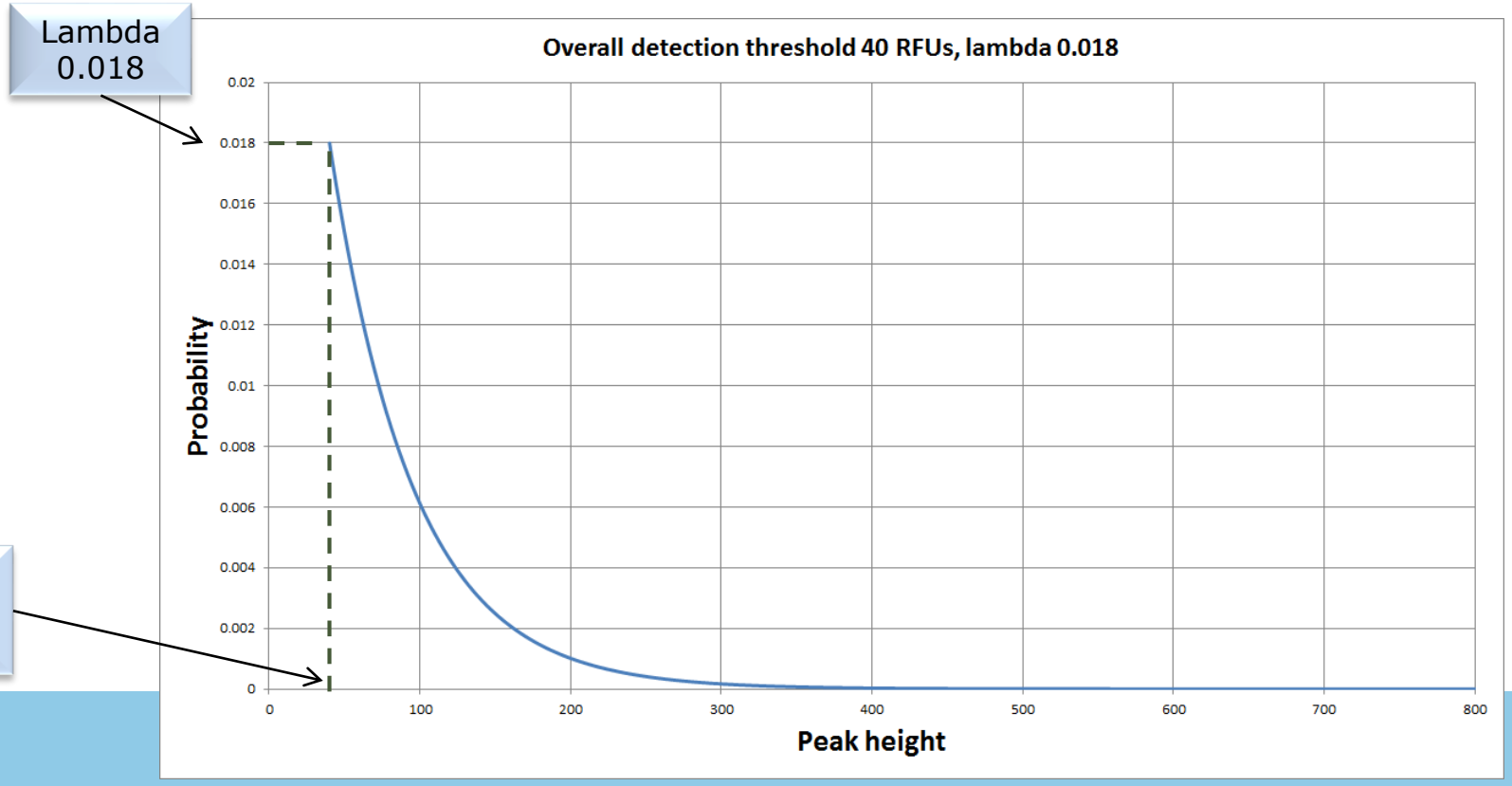
Pr(dropout)>0

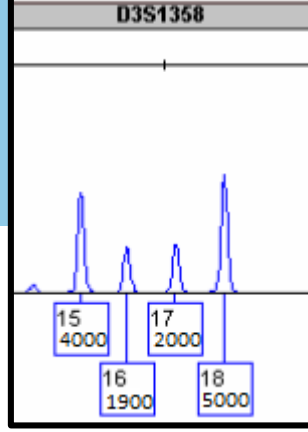


Drop-in

- Model assumption:

  Drop-in peak heights follow an exponential distribution.
- Defined by lambda (λ), which can be empirically determined from casework blanks. By default 0.01. In this example 0.018.





Drop-in

- If allele 17 is a drop-in:
  - $PrC \quad * P_{17} \quad * ExpDistribution(Height_{17} - AT; 0.018)$ Unlikely
  - $0.00426257 * 0.22 \ * ExpDistribution(\quad 2000 - 40; 0.018) = 4.16E^{-19}$

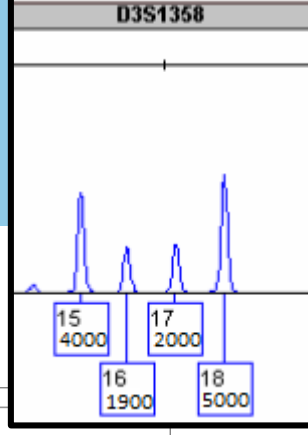
- If allele 17 has a peak height of 80 RFUs instead: More likely
  - $0.00426257 * 0.22 \ * ExpDistribution(\quad 80 - 40; 0{,}018) = 4.56E^{-4}$

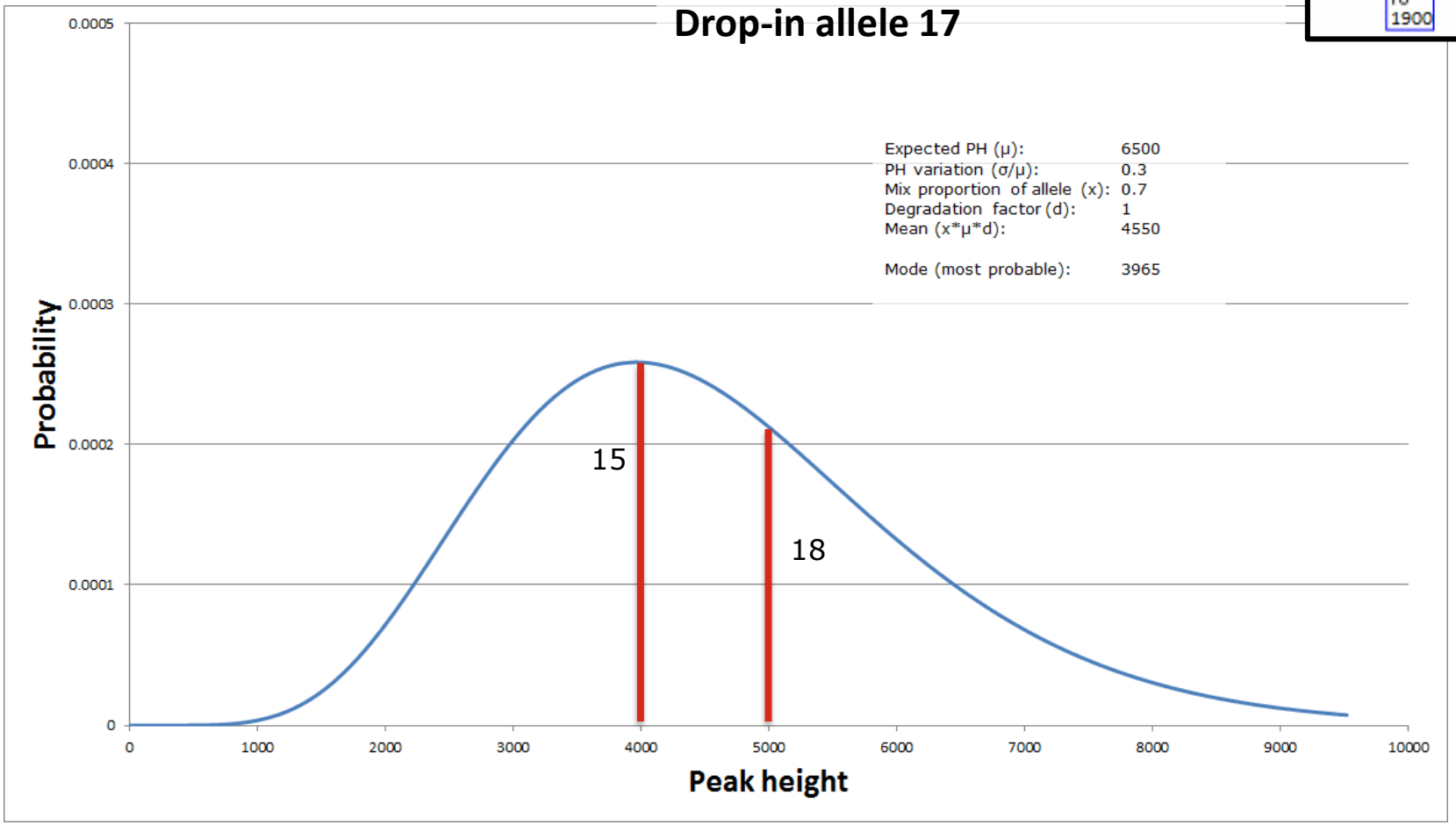
Probable genotype combination (for major)

C1 = 15,18 mixture proportion 0.7

C2 = 16,16 mixture proportion 0.3

Drop-in allele 17



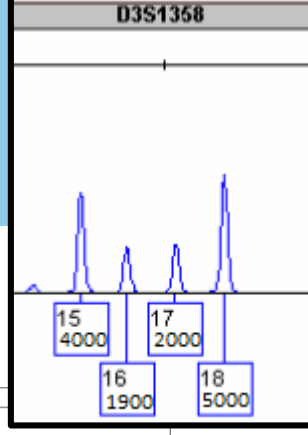


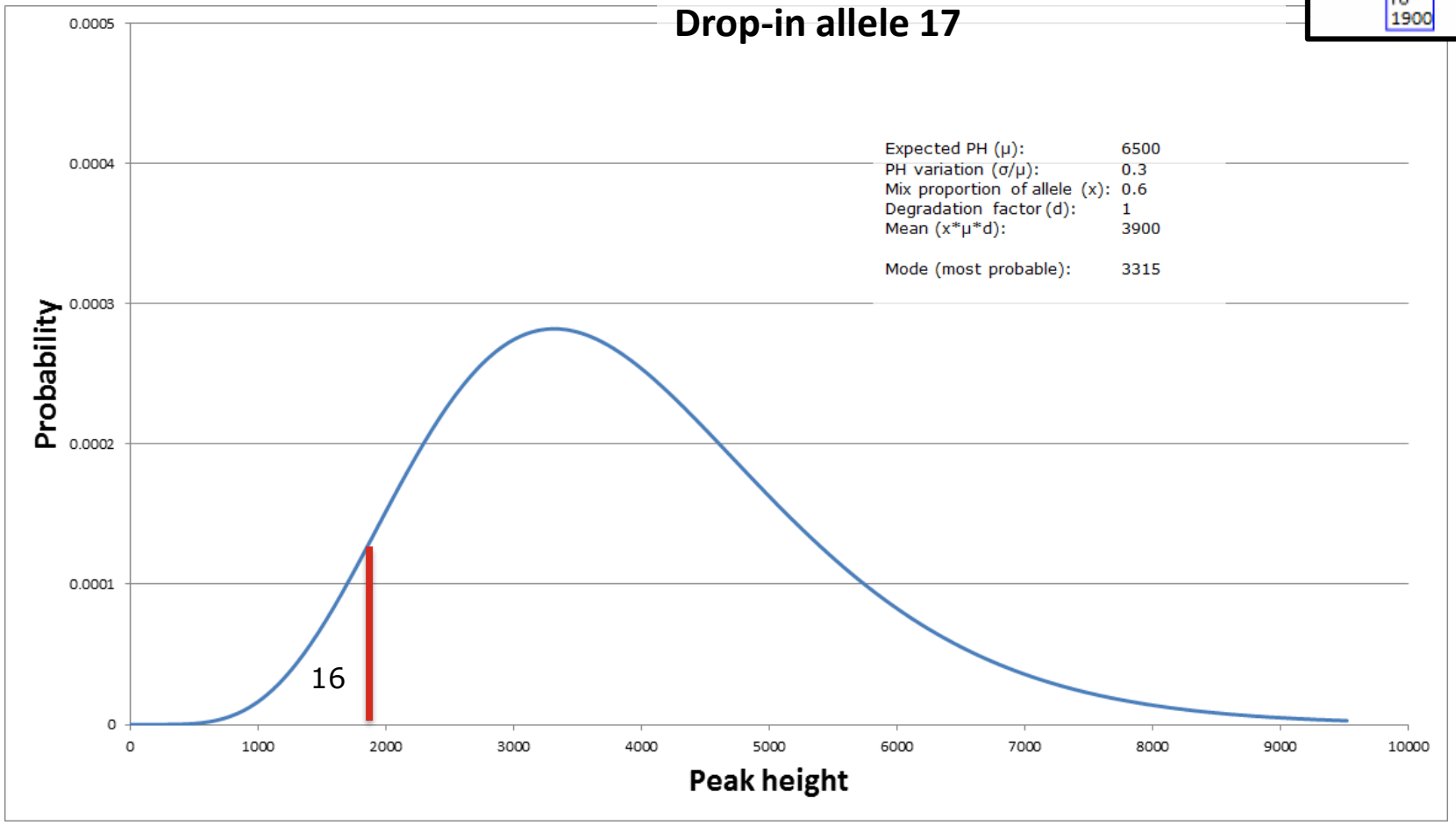
But less probable genotype combination for the minor

C1 = 15,18 mixture proportion 0.7

C2 = 16,16 mixture proportion 0.3

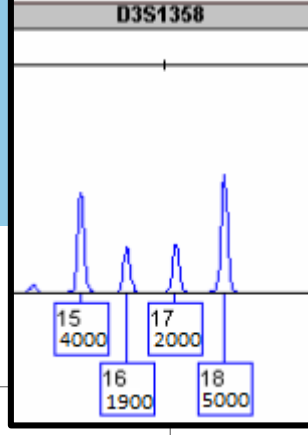
Drop-in allele 17

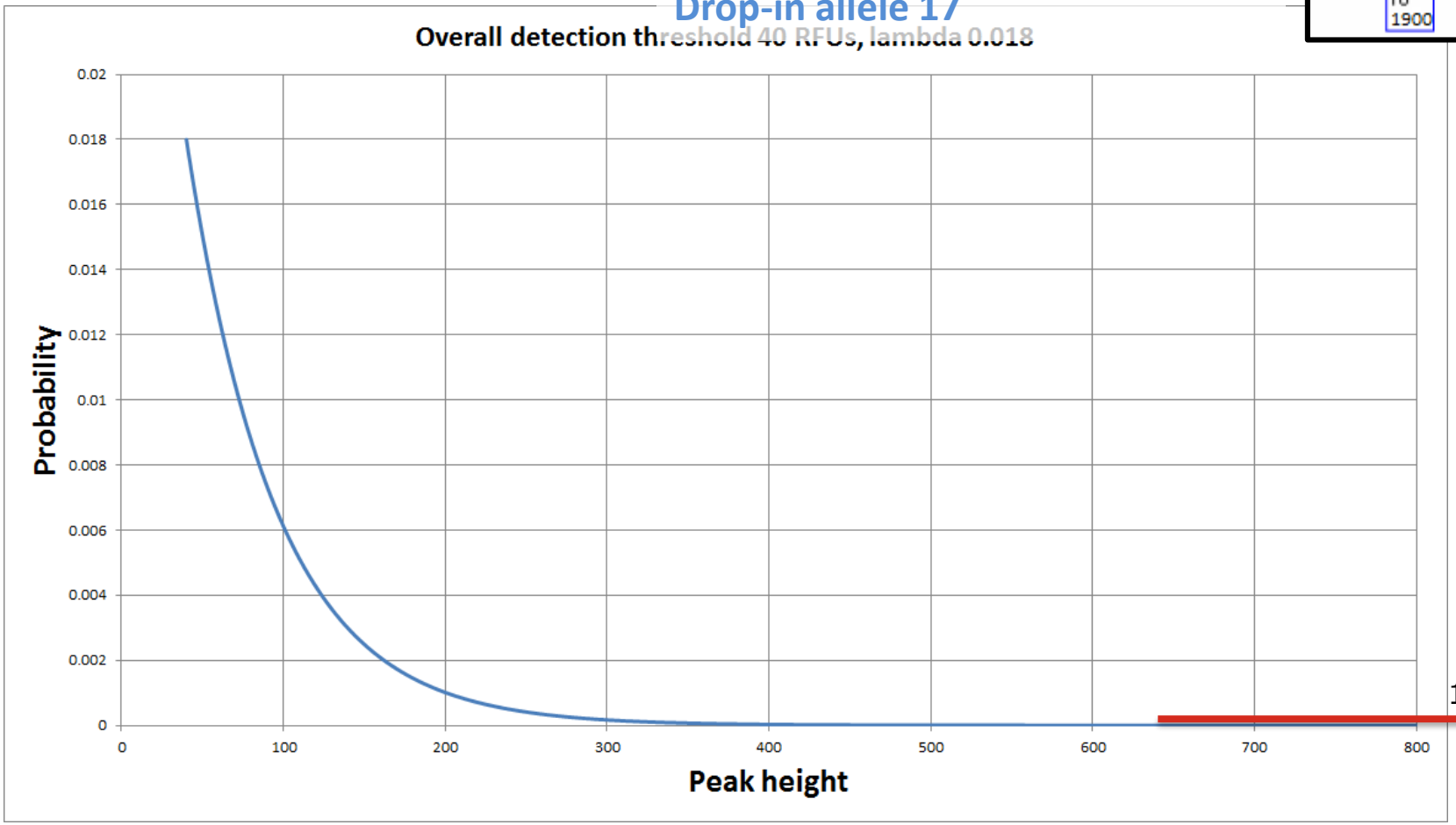




Very unlikely for drop-in

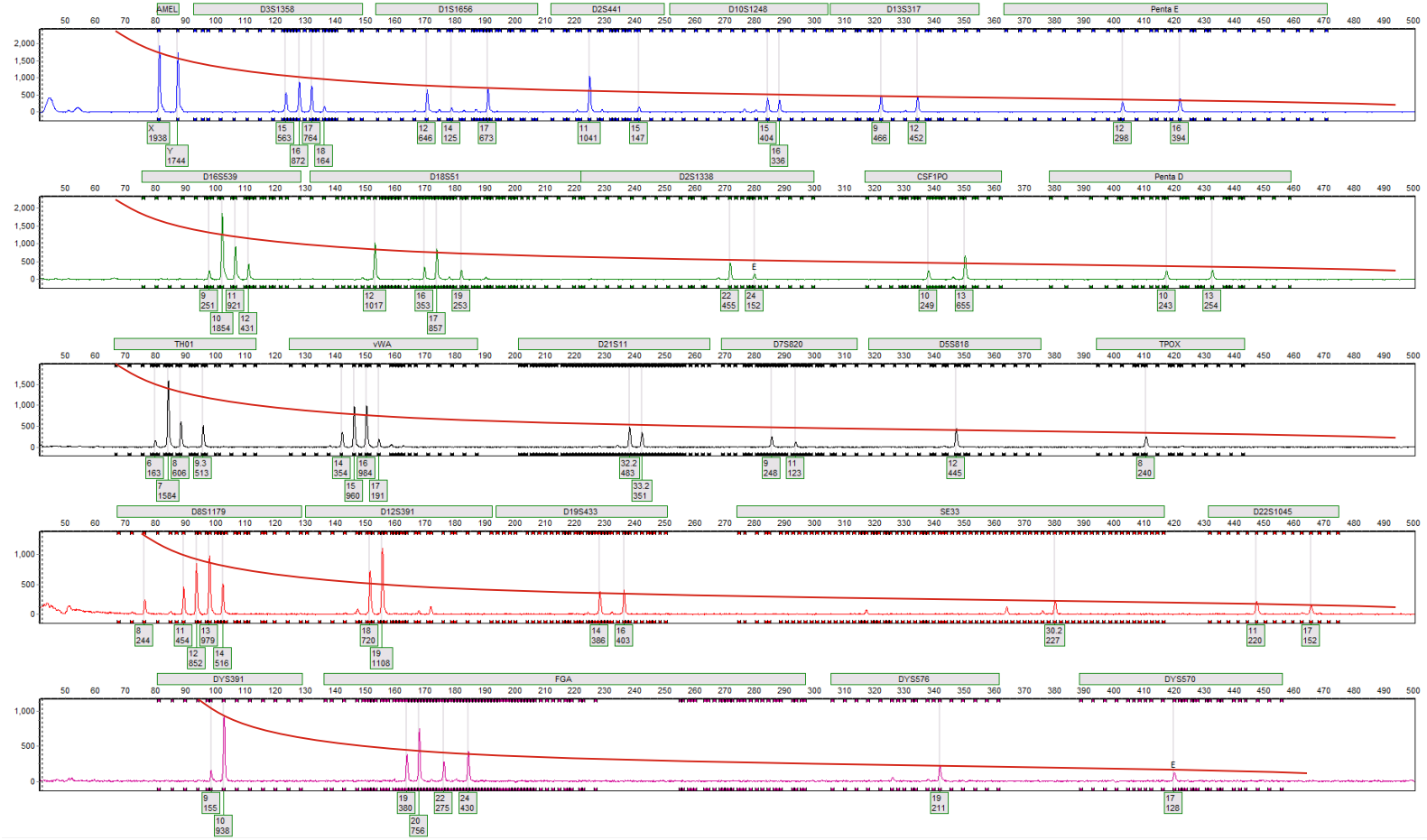
C1 = 15,18 mixture proportion 0.7
C2 = 16,16 mixture proportion 0.3
Drop-in allele 17





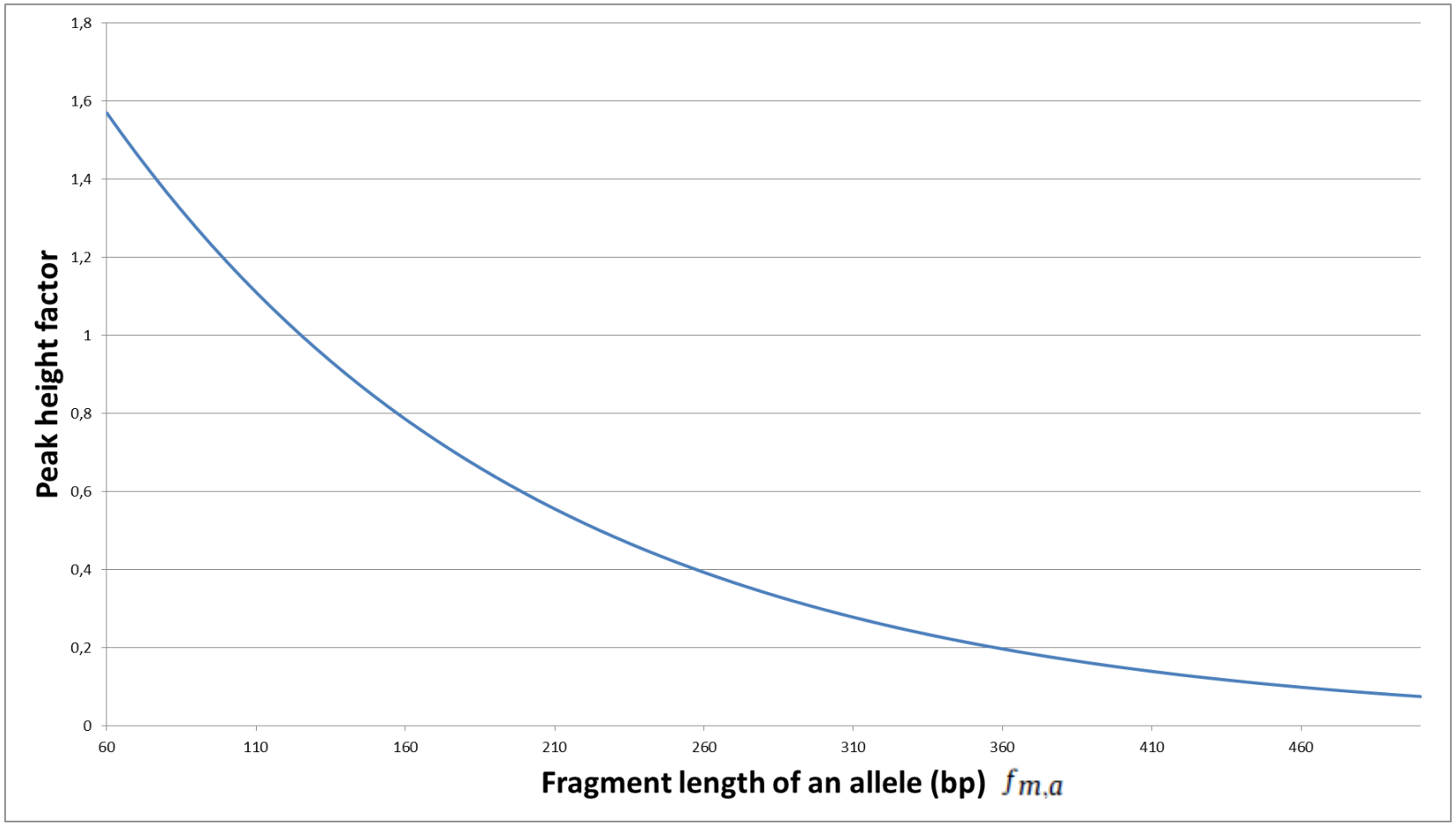
Degradation

$$\beta^{\frac{f_{m,a}-125}{100}}$$



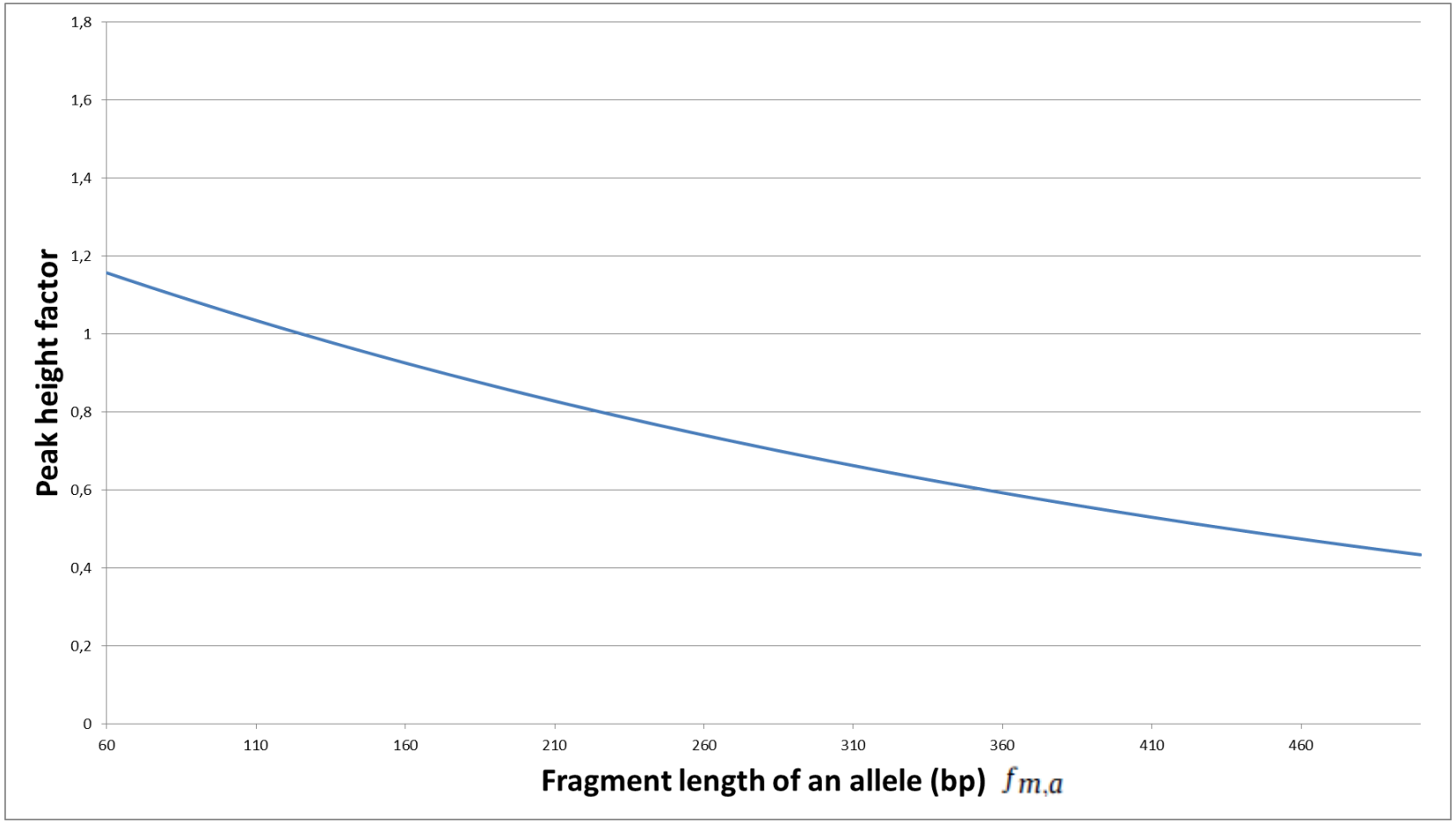
Degradation model β=0.5

$$\beta^{\frac{f_{m,a}-125}{100}}$$



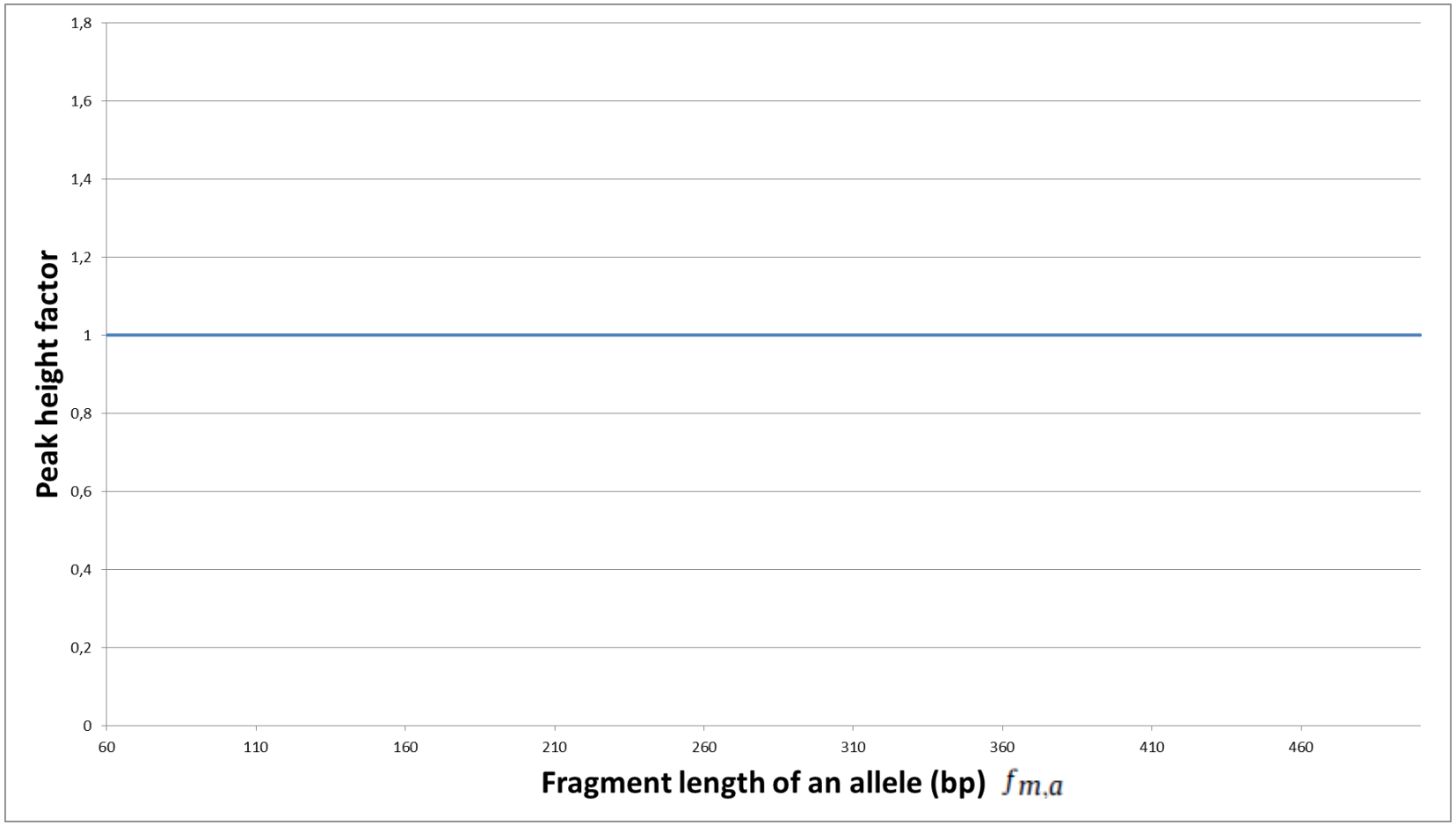
Degradation model β=0.8

$$\beta^{\frac{f_{m,a}-125}{100}}$$

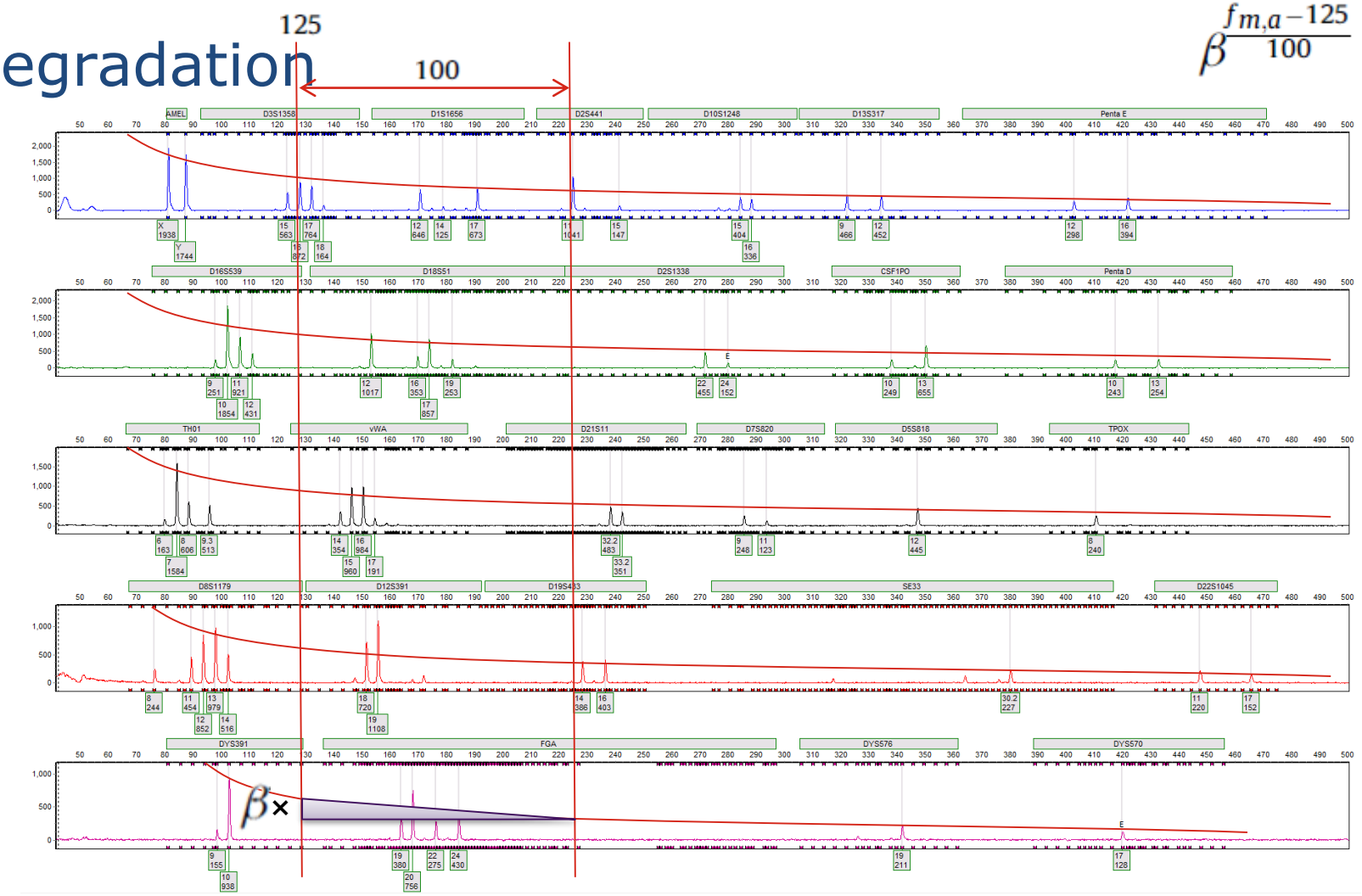


Degradation model $\beta=1$

$$\beta^{\frac{f_{m,a}-125}{100}}$$



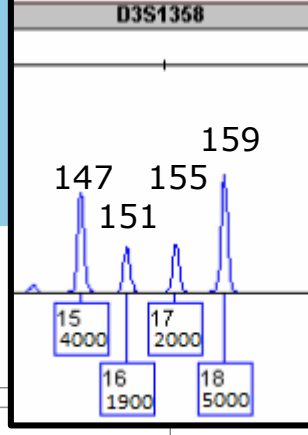
Degradation

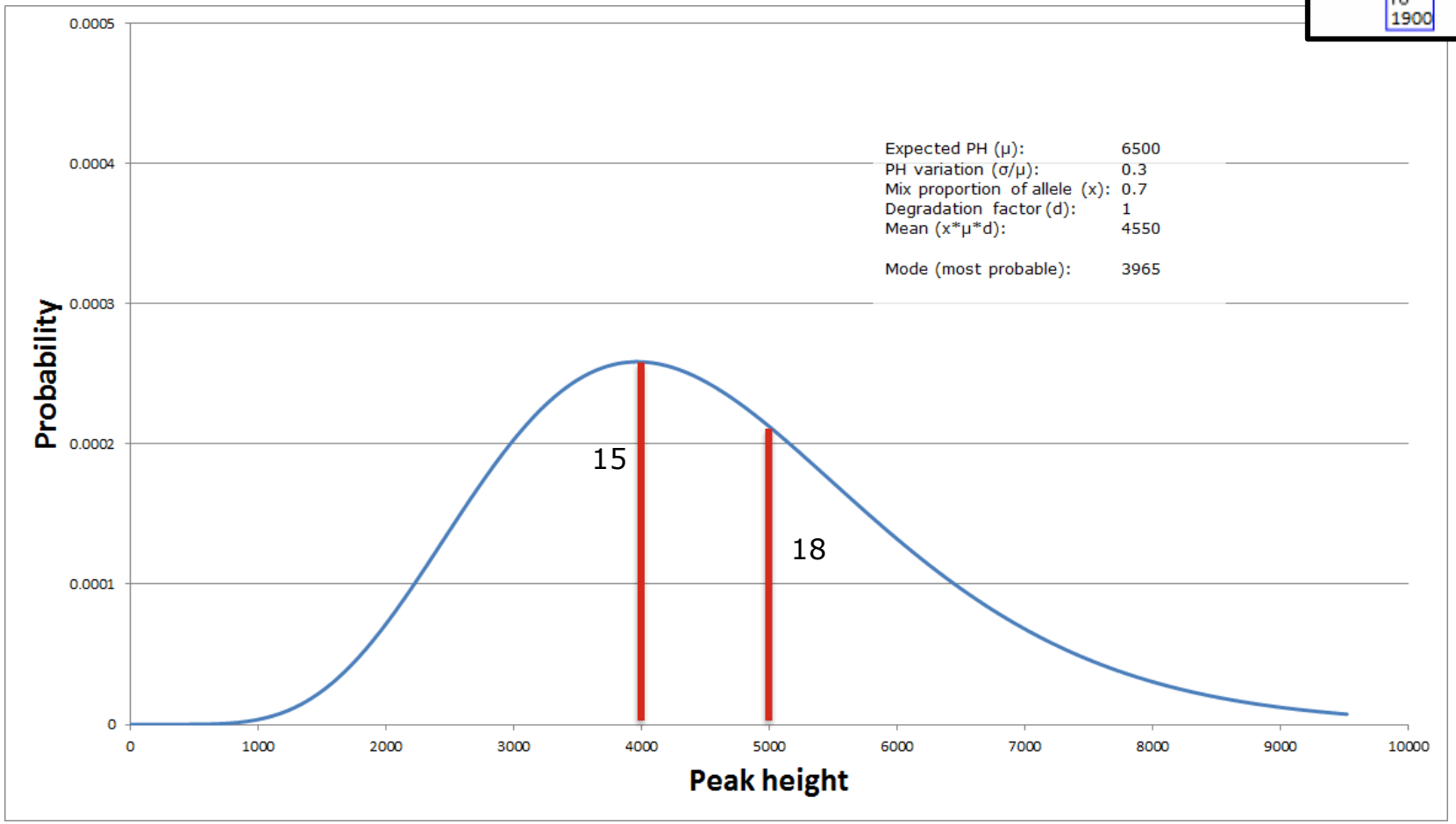


Probable? $\beta=1$

C1 = 15,18 mixture proportion 0.7

C2 = 16,17 mixture proportion 0.3

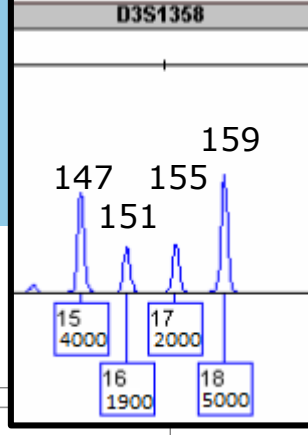


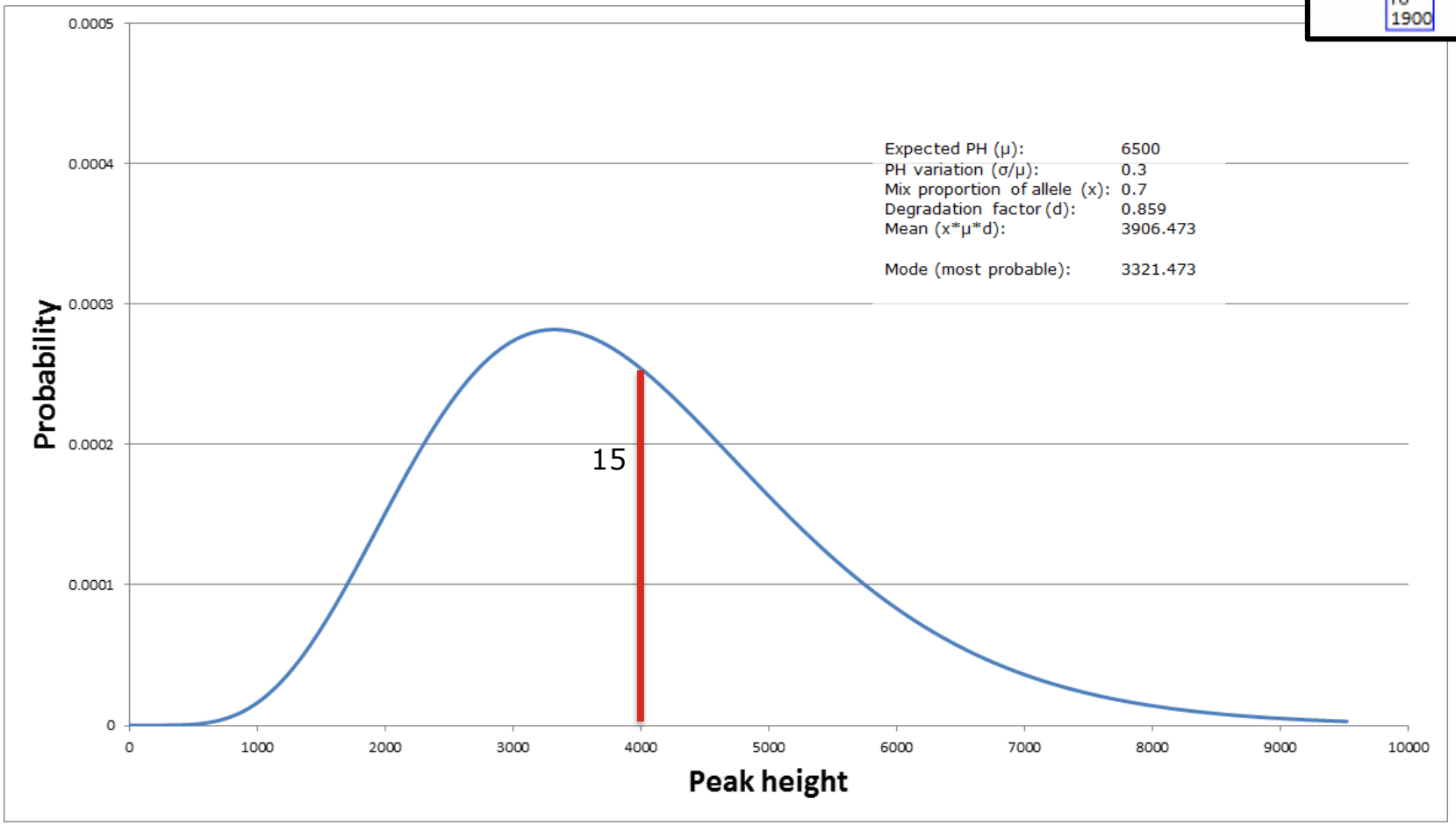


Probable? β=0.5

C1 = 15,18 mixture proportion 0.7

C2 = 16,17 mixture proportion 0.3

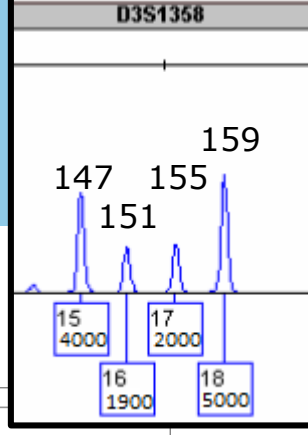


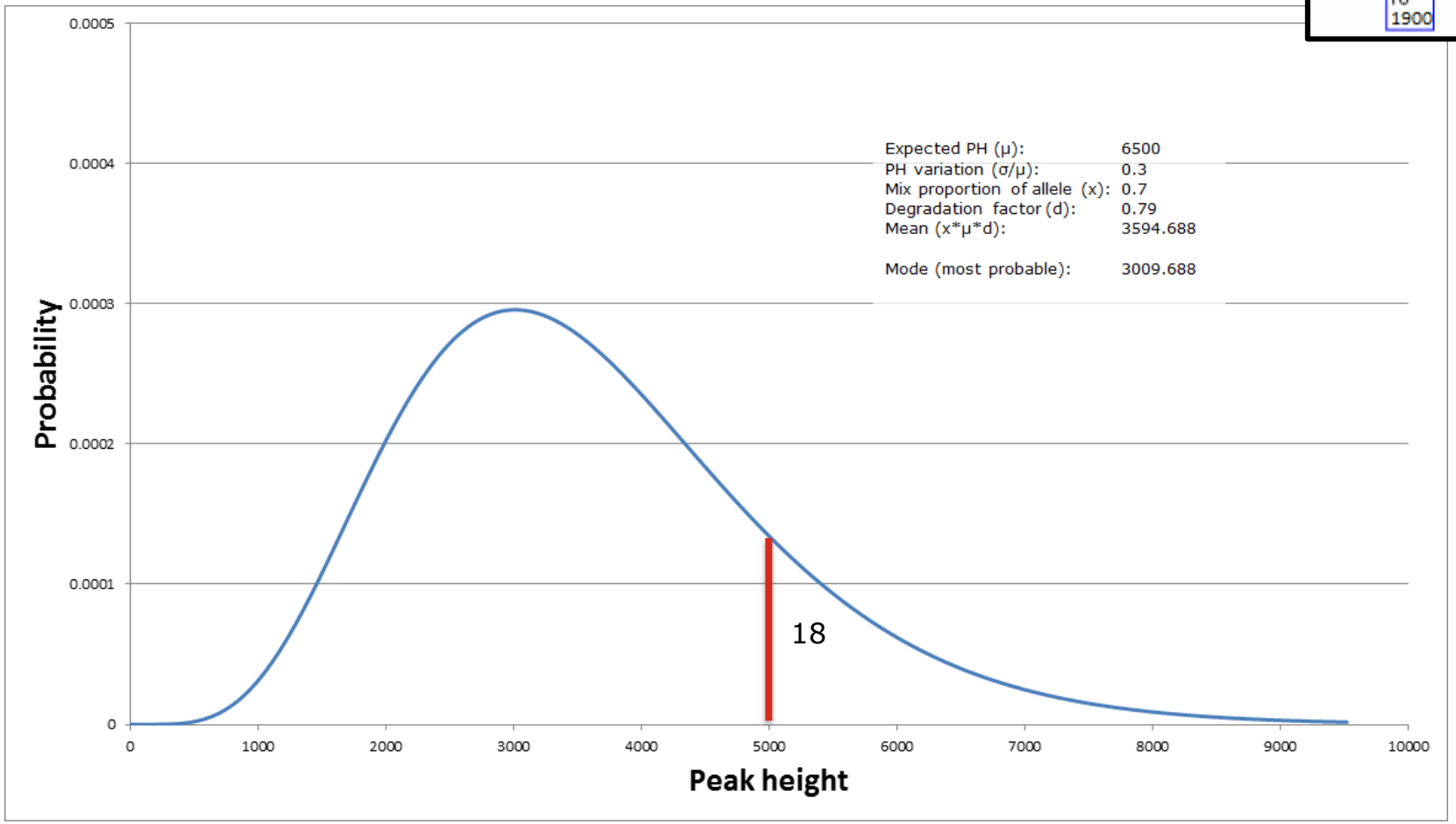


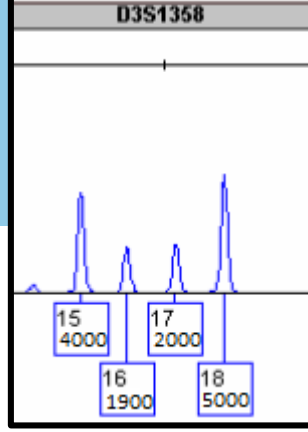
Probable? β=0.5

C1 = 15,18 mixture proportion 0.7

C2 = 16,17 mixture proportion 0.3





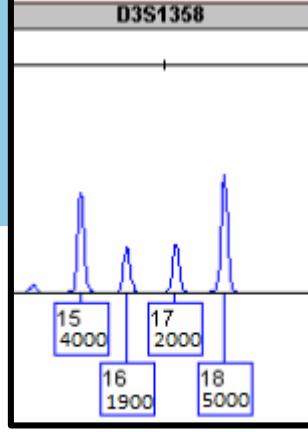


Calculate likelihoods including PH model

- Sum up product within all genotype combinations

Donor A	Donor B	Drop-in	Genotype probability	Drop-in probability	Likelihood incl PH
15/16	17/18	-	$2P_{15}P_{16}*2P_{17}P_{18}$	1-PrC	2.6944E-52
15/17	16/18	-	$2P_{15}P_{17}*2P_{16}P_{18}$	1-PrC	2.5123E-50
15/18	16/17	-	$2P_{15}P_{18}*2P_{16}P_{17}$	1-PrC	3.8438E-15
17/18	15/16	-	$2P_{17}P_{18}*2P_{15}P_{16}$	1-PrC	9.3019E-42
16/18	15/17	-	$2P_{16}P_{18}*2P_{15}P_{17}$	1-PrC	9.9761E-44
16/17	15/18	-	$2P_{16}P_{17}*2P_{15}P_{18}$	1-PrC	6.5203E-79
But also…					
15/15	16/17	18	$P_{15}^2*2P_{16}P_{17}$	$PrC*P_{18}$	1.5107E-89
15/18	16/d.o.	17	$2P_{15}P_{18}*2P_{16}P_Q$	$PrC*P_{17}$	1.4539E-117
d.o.	d.o.	15/16/17/18	P_Q^4	$PrC^4*P_{15}*P_{16}*P_{17}*P_{18}$	0
Etc…					

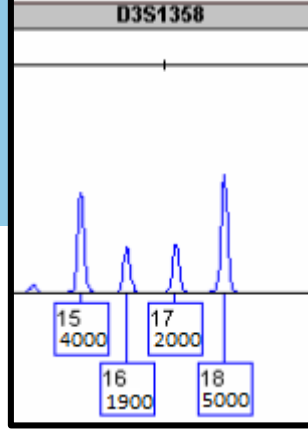
3.0378E-112	1.4539E-117
0	0
	+
	3.8438E-15



Calculate likelihoods multiple replicates

Donor A	Donor B	Drop-in	Genotype probability	Drop-in probability	Likelihood incl PH Replicate 1
15/16	17/18	-	$2P_{15}P_{16}*2P_{17}P_{18}$	1-PrC	2.6944E-52
15/17	16/18	-	$2P_{15}P_{17}*2P_{16}P_{18}$	1-PrC	2.5123E-50
15/18	16/17	-	$2P_{15}P_{18}*2P_{16}P_{17}$	1-PrC	3.8438E-15
17/18	15/16	-	$2P_{17}P_{18}*2P_{15}P_{16}$	1-PrC	9.3019E-42
16/18	15/17	-	$2P_{16}P_{18}*2P_{15}P_{17}$	1-PrC	9.9761E-44
16/17	15/18	-	$2P_{16}P_{17}*2P_{15}P_{18}$	1-PrC	6.5203E-79
But also…					
15/15	16/17	18	$P_{15}^2*2P_{16}P_{17}$	$PrC*P_{18}$	1.5107E-89
15/18	16/d.o.	17	$2P_{15}P_{18}*2P_{16}P_Q$	$PrC*P_{17}$	1.4539E-117
d.o.	d.o.	15/16/17/18	P_Q^4	$PrC^4*P_{15}*P_{16}*P_{17}*P_{18}$	0
Etc…					

3.0378E-112	1.4539E-117
0	0
	+
	3.8438E-15



Calculate profile likelihood

Donor A	Donor B	Drop-in	Genotype probability	Drop-in probability	Likelihood incl PH Replicate 1	Likelihood incl PH Replicate 2	Likelihood
15/16	17/18	-	$2P_{15}P_{16}*2P_{17}P_{18}$	1-PrC	2.6944E-52	2.6475E-45	7.1334E-97
15/17	16/18	-	$2P_{15}P_{17}*2P_{16}P_{18}$	1-PrC	2.5123E-50	2.8226E-46	7.0912E-96
15/18	16/17	-	$2P_{15}P_{18}*2P_{16}P_{17}$	1-PrC	3.8438E-15	1.2048E-14	4.6310E-29
17/18	15/16	-	$2P_{17}P_{18}*2P_{15}P_{16}$	1-PrC	9.3019E-42	2.9156E-41	2.7121E-82
16/18	15/17	-	$2P_{16}P_{18}*2P_{15}P_{17}$	1-PrC	9.9761E-44	2.7347E-40	2.7282E-83
16/17	15/18	-	$2P_{16}P_{17}*2P_{15}P_{18}$	1-PrC	6.5203E-79	6.4067E-72	4.1774E-150
But also…							
15/15	16/17	18	$P_{15}^2*2P_{16}P_{17}$	$PrC*P_{18}$	1.5107E-89	1.4552E-83	2.1984E-172
15/18	16/d.o.	17	$2P_{15}P_{18}*2P_{16}P_Q$	$PrC*P_{17}$	1.4539E-117	2.9109E-115	4.2322E-232
d.o.	d.o.	15/16/17/18	P_Q^4	$PrC^4*P_{15}*P_{16}*P_{17}*P_{18}$	0	0	0
Etc…							+
							4.6311E-29

- Multiply between loci

Locus 1	Locus 2	Locus 3	Locus 4	Locus 5	Locus 6	Locus n	Overall likelihood Hd
4.6311E-29 X	……. X	……. X	…… X	……. X	…… X	……	$P(E\|H,\beta)$

General formula for probabilistic genotyping models

Product rule

Summing over all possible unknown genotypes

$$LR = \frac{P(E|H_p, \beta_p)}{P(E|H_d, \beta_d)} = \frac{\prod_m \sum_g w(E_m, g|\beta_p) P(g|H_p)}{\prod_m \sum_g w(E_m, g|\beta_p) P(g|H_p)}$$

m = marker index
g = genotype combination (for marker m)

β_p, β_d unknown model parameters
Must be estimated

w are the likelihood weights (product over all alleles)

A lot of calculations…

2 person mixture:
(15 + 225 genoty … (23 loci) * ±5.000 optimization steps
= 27.600.000 LE … less than a minute

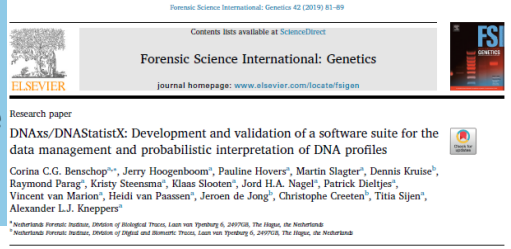
Calculation time is much reduced in the latest versions of EuroForMix and DNAStatistX! ☺

4 person mixture:
(91.125 + 4.100.625 genotypes) … (… …tion steps
= 482.051.250.000 LE calculations → seve… a day)

5 person mixture:
(4.100.625 + 1.252.332.576) * (23 loci) * ± 5.000 optimization steps
= 144.489.818.115.000 LE calculations → takes a long time…

Exploratory approach

- Preferred approach is always to maximize the quality of the sample first
- Statistical tool is not a substitute for sound bio-chemistry
- The LR does not tell you if a proposition is true or not – it can only tell you if one is more likely than the other
- Not a substitute to careful consideration of the circumstances of the case to help formulate the propositions
- Rapid evaluation of multiple sets of propositions is facilitated by software


Forensic Science International: Genetics 42 (2019) 81–89

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

DNAxs/DNAStatistX: Development and validation of a software suite for the data management and probabilistic interpretation of DNA profiles

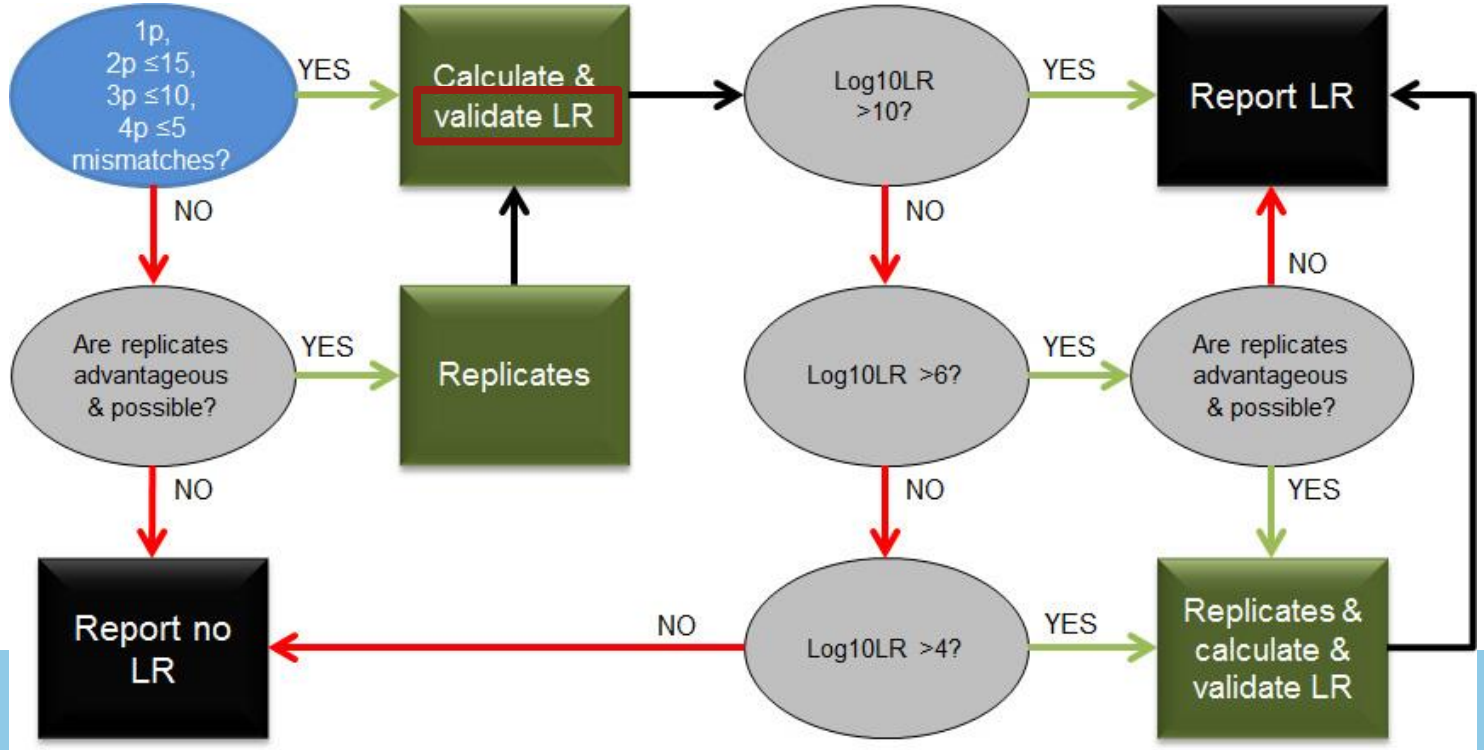
Corina C.G. Benschop[a,*], Jerry Hoogenboom[a], Pauline Hovers[a], Martin Slagter[a], Dennis Kruise[b], Raymond Parag[a], Kristy Steensma[a], Klaas Slooten[a], Jord H.A. Nagel[a], Patrick Dieltjes[a], Vincent van Marion[a], Heidi van Paassen[a], Jeroen de Jong[a], Christophe Creeten[b], Titia Sijen[a], Alexander L.J. Kneppers[a]

[a] Netherlands Forensic Institute, Division of Biological Traces, Laan van Ypenburg 6, 2497GB, The Hague, the Netherlands
[b] Netherlands Forensic Institute, Division of Digital and Biometric Traces, Laan van Ypenburg 6, 2497GB, The Hague, the Netherlands

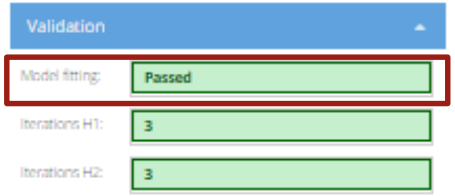
Guidelines for use in forensic casework

- Based on results from research, casework experience, including efficiency and usefulness of calculations.
- Clearly, the case circumstances, availability of other stains and profiles within the case are factors considered in the process.

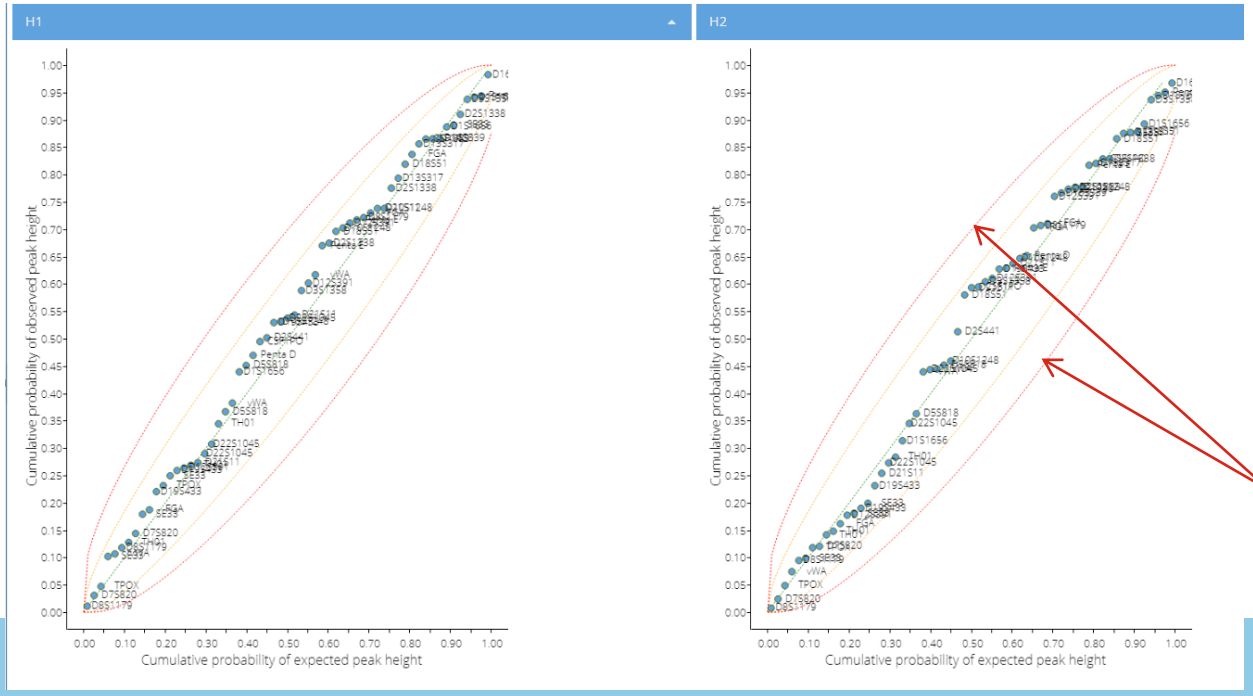
Summarized guidelines to be used as helpful tool in the decision making process are as follows:



Model validation



- The model validation plots are important quality checks.
- In this model validation, the cumulative probabilities for the expected peak height are plotted against the cumulative probabilities for the observed peak heights, resulting in a PP (probability-probability) plot.



The default Bonferroni corrected significance level is set at 0.01.

When at least four values are outside the envelope (0.01-line), the model validation is scored as ‘failed’.

Model validation status

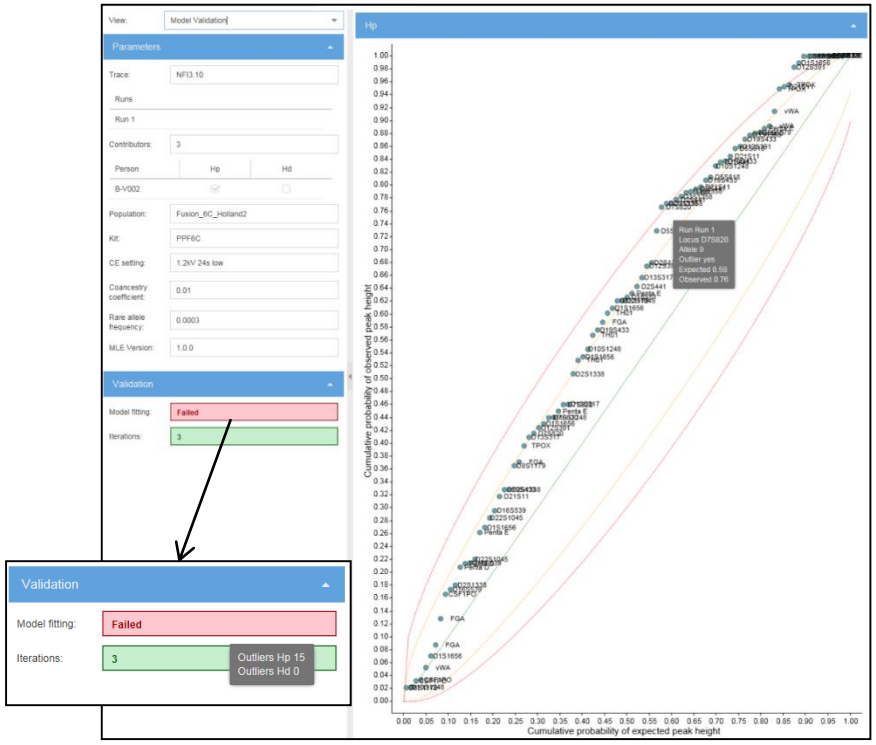
The model validation status can be returned as Passed, Warning, Failed, or Unavailable:

Passed	The model fitting has no or one outlier(s) under H1 and H2 separately.
Warning	The model fitting includes 2 or 3 outliers under H1 and/or H2 separately.
Failed	The model fitting includes more than 3 outliers under H1 and/or H2 separately.

- Model validation may fail due to various reasons

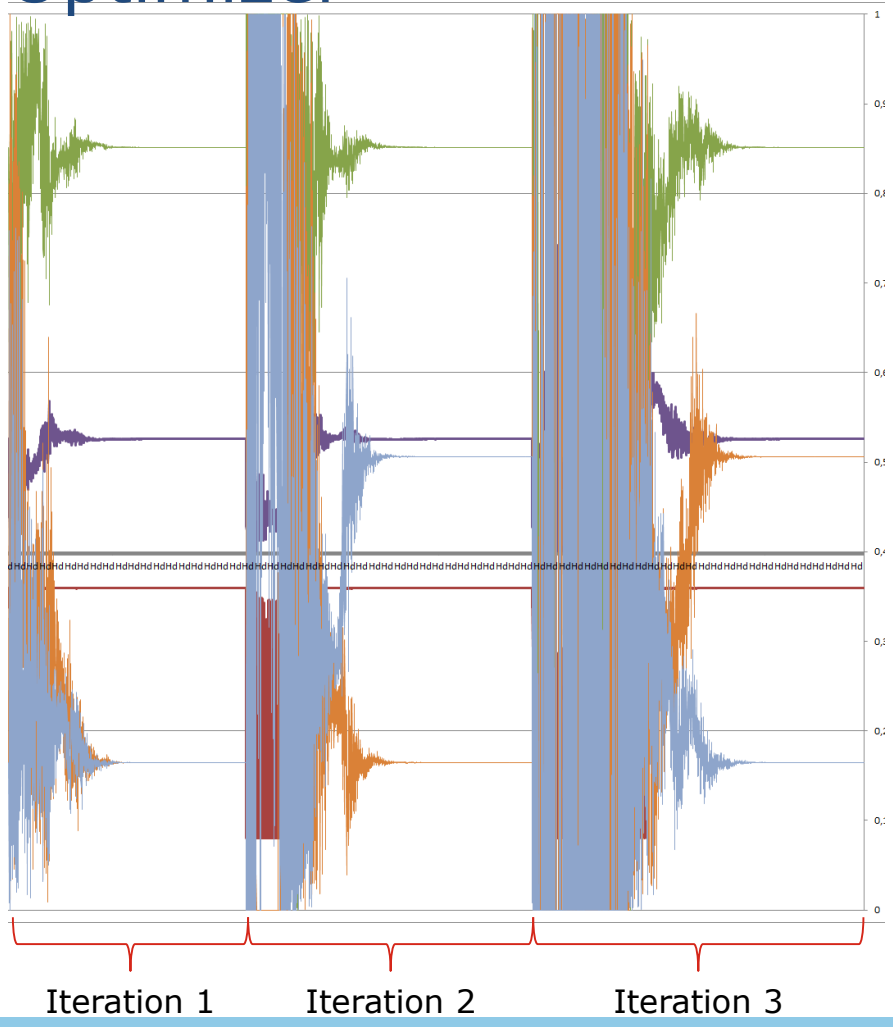
Model fitting (*i.e.* validation) may fail when peak heights cannot be explained according to the propositions and parameters, and can occur as a result of e.g.:

- Incorrect parameter setting such as
  - Lack of degradation model application when data does show degradation
    - Advised action: rerun using degradation model turned on.
  - Underassigned number of contributors (and/or drop-in values) to enable explaining the observed peaks
    - Advised action: rerun using higher number of contributors.
- H1 analyses with a non-contributor as POI (note that model validation often, but not always, fails with a non-contributor)
  - Advised action: none if LR is exclusionary. Results can be reported.
- A peak of improbable height
  - Advised action: check profile.
- Analyses with replicates of extraordinary peak height variation
  - Advised action: Report the LR based on the individual replicates (if model validation passes), but not using replicates analyzed jointly.
- Other than above
  - Advised action: Check results (e.g. EPG vs LR per locus, equality of mixture proportions under H1 and H2, kit settings) and perform a rerun. If model validation still fails, it is advised not to report this LR value.



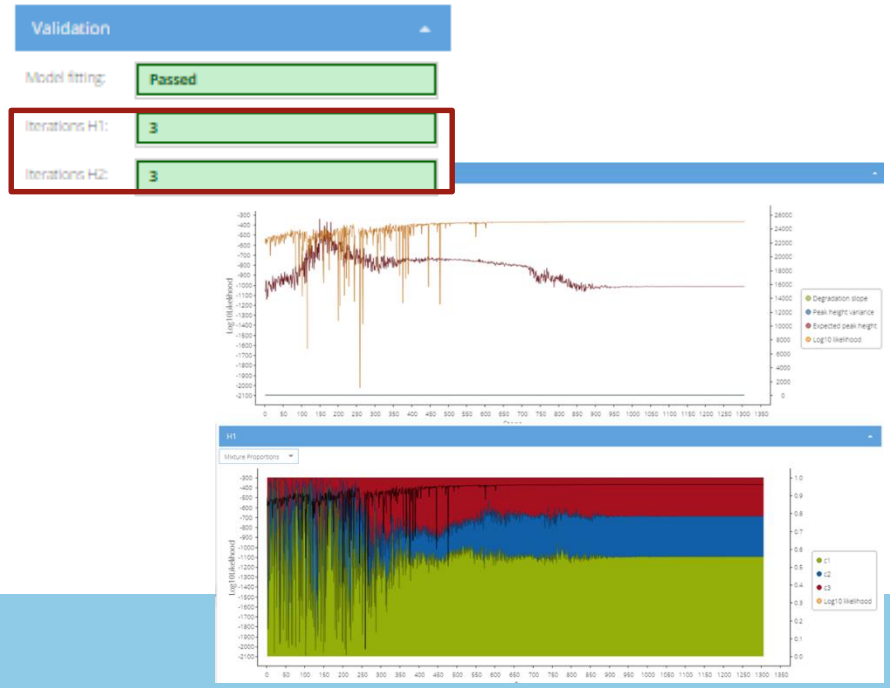
- If failing model validation cannot be explained/ solved, it is advised not to report the LR

Optimizer



Log likelihood estimation
Expected peak height
Degradation slope
Mixture proportion 1
Mixture proportion 2

- Optimizer results are accepted in DNAxs if 3 times sufficiently similar results are obtained.



Optimizer iterations

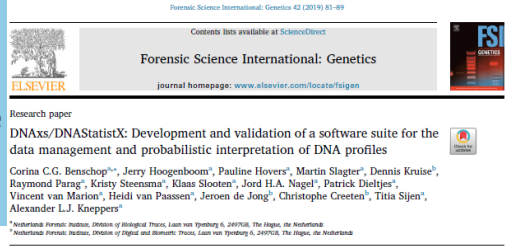
- User manual:

The number of **Iterations** defines the number of repetitions needed to obtain a sufficiently similar optimum. The number of similar results needed is an application parameter and can be set by an administrator (see Settings: Change application settings). The MLE calculation is repeated until the predefined number of similar results (default 3) is reached or when the maximum number or repetitions is met (default 10). The maximum number of repetitions is also an application setting that can be changed by the administrator in the application settings.

- In DNAStatistX, the MLE calculation is repeated until three sufficiently similar results are found with a maximum of ten repetitions.

	The number of similar results is met (the default number of required similar results is 3).
	The predefined number of similar results was not met after the maximum number of iterations.
	No sufficiently similar results were found after the maximum number of iterations.

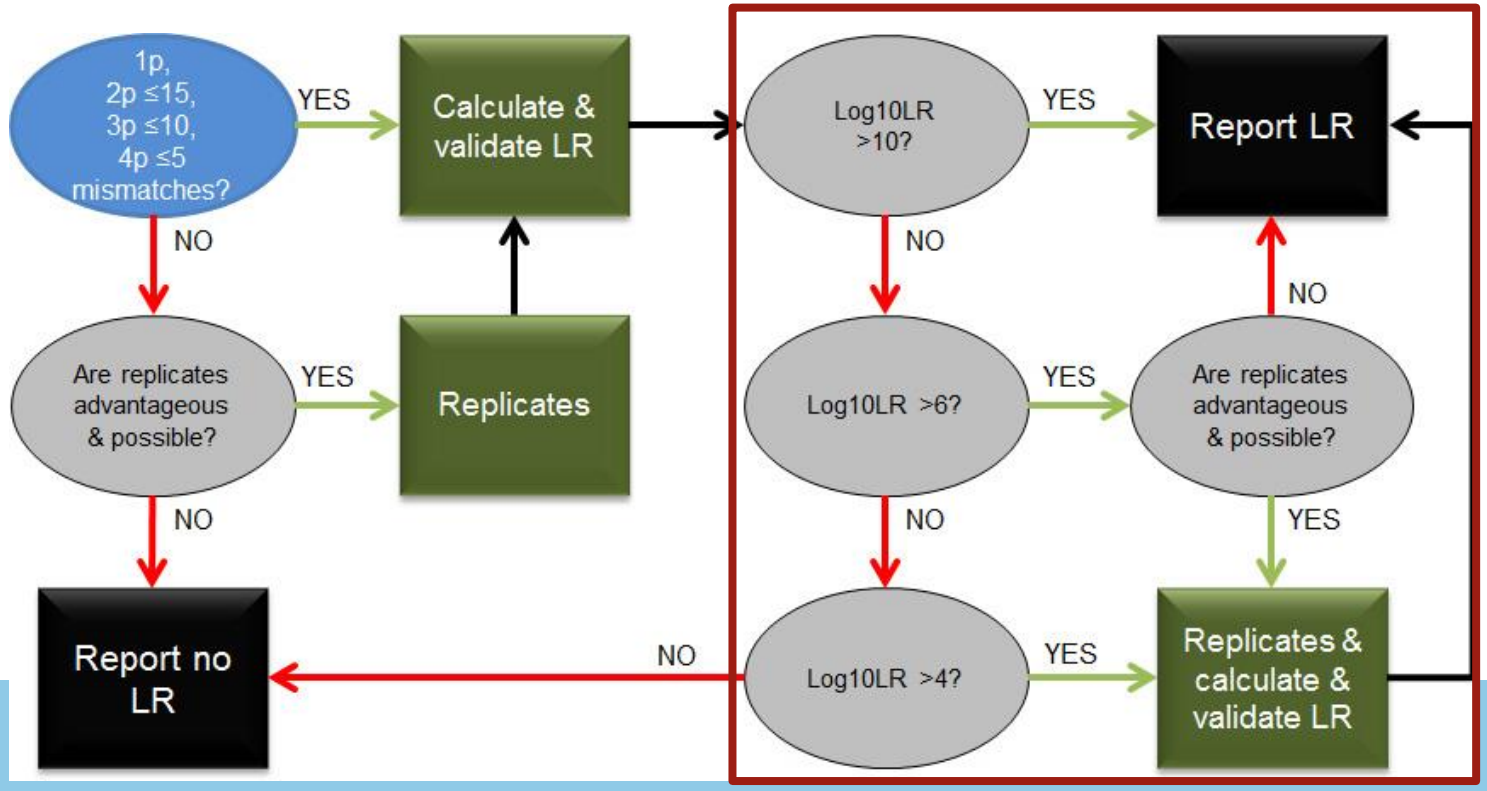
- Iterations may fail under H1 and/or H2 and indicate that no three (default) optimum parameter values were found to be sufficiently similar out of ten iterations.
- With failing iterations it is advised to rerun the calculation.



Guidelines for use in forensic casework

- Based on results from research, casework experience, including efficiency and usefulness of calculations.
- Clearly, the case circumstances, availability of other stains and profiles within the case are factors considered in the process.

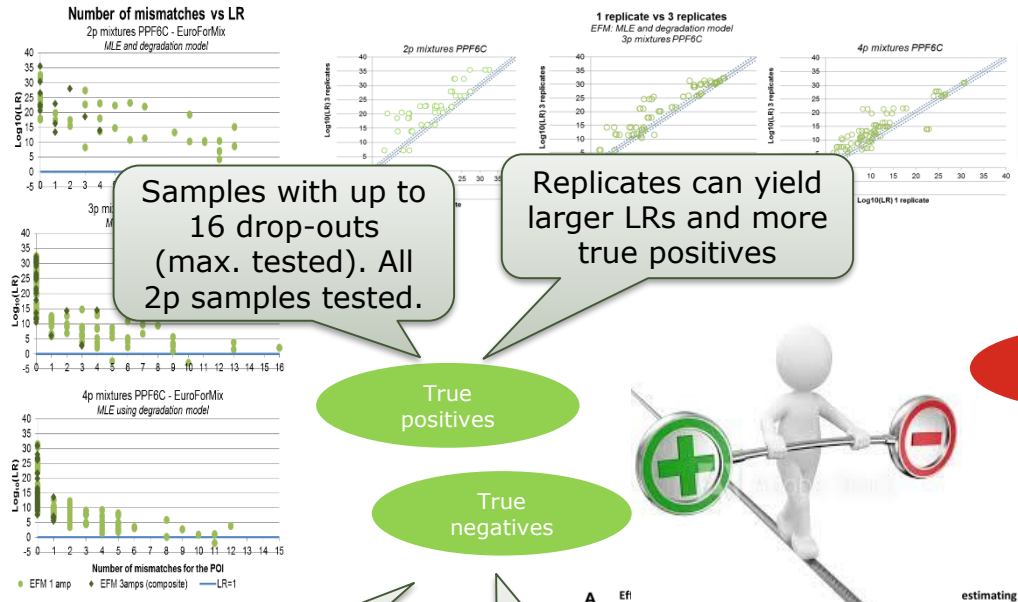
Summarized guidelines to be used as helpful tool in the decision making process are as follows:



Guidelines based on e.g.

Netherlands Forensic Institute
Ministry of Justice and Security

Trends with 2p, 3p and 4p PPF6C profiles



Samples with up to 16 drop-outs (max. tested). All 2p samples tested.

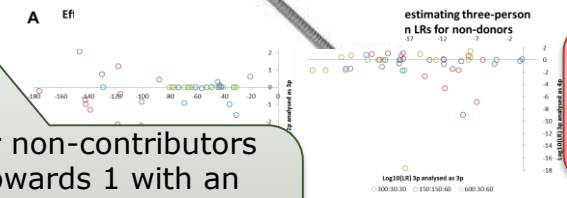
Replicates can yield larger LRs and more true positives

True positives

True negatives

All non-donors (non-relatives) for 2p samples

LRs for non-contributors tend towards 1 with an overestimated number of contributors (not seen when mixtures contain equal contributions).



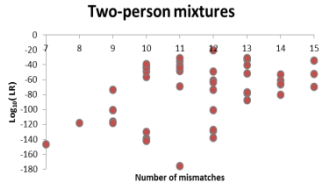
3p mixtures ≥12 mismatches
Max LR=65
4p mixtures ≥5 mismatches
Max LR=123

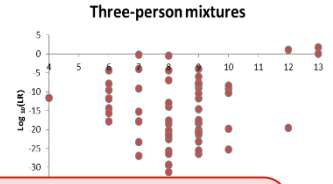
False positives

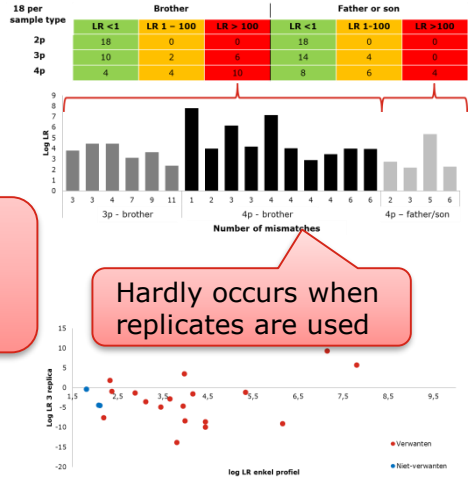
Simulated brothers and fathers/sons that are selected based on a high overlap with the stain profile can yield large LRs

False negatives

Some 3p & 4p samples with ≥5 drop-outs



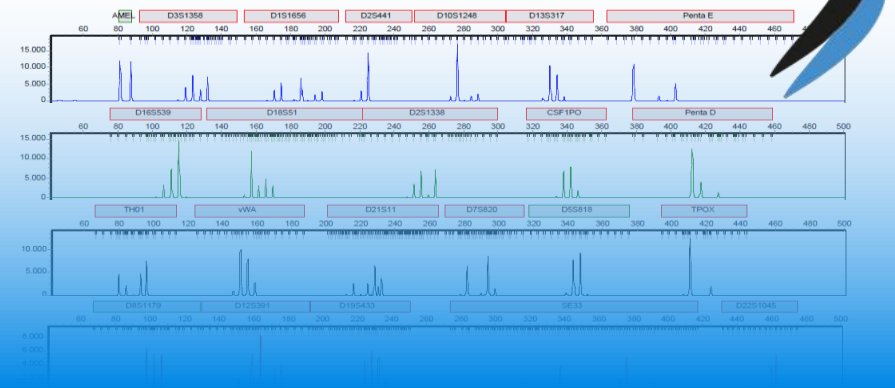




Hardly occurs when replicates are used

Netherlands Forensic Institute
Ministry of Justice and Security

Interpretation of (mixed) DNA profiles within the NFI



Corina Benschop

13/11/2020