

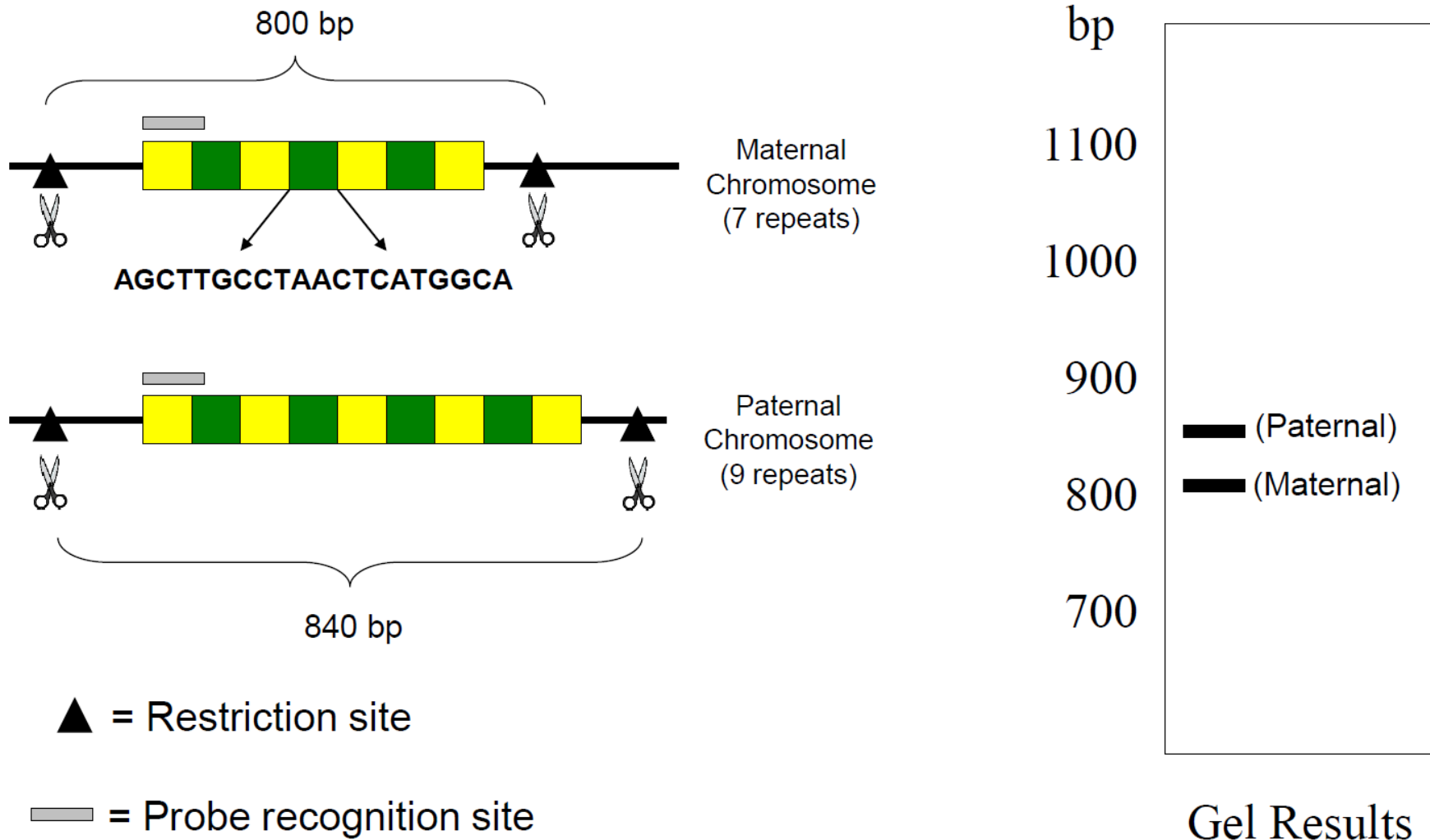
Forensic Genetics

Module 19 – Topic 1

Schedule – Day 1

Intro	Course logistics and introductions	11:30-12:20
Topic 1	STR Typing	12:35-1:25
Topic 2	Probability and Likelihood Ratios	1:40-2:30

Before STRs: RFLP – Repeat Variation (VNTR)



Challenges

- High quality and high quantity stains were required (micrograms of DNA).
- Probes were labeled with radioactive ^{32}P .
- Each of the 5-7 loci took about a week to develop.
- Very laborious and time-consuming

STR Typing

- Forensic DNA interpretation has been centered on the analysis of STRs (*short tandem repeats*), i.e. short DNA sequences that are repeated several times.
- These repeat patterns are located in areas called *loci* and vary among individuals.
- Variants for a given locus are called *alleles* and it is this variation (called polymorphism) that allows us to associate a particular DNA sample with an individual person.

Topic 1 - STR Typing

(Outline)

- Mutations
 - Slipped Strand Mispairing
- PCR-CE Analysis and Anomalies
- Peak Height Variability
 - Template
 - Degradation
 - Locus Specific Amplification Efficiency
 - Stutter
- Mixtures
 - Heterozygous Balance

Mutations

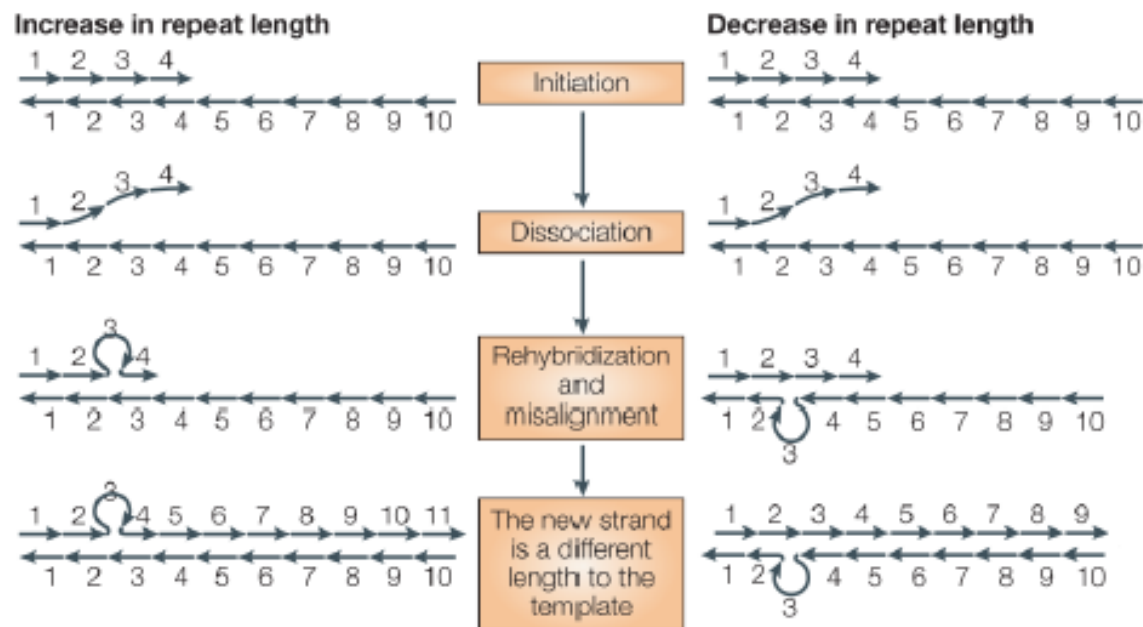
Mutations are the cause of the variation encountered in DNA and one of the reasons that STR loci render highly informative markers in forensic genetics. Most of the mutations are caused by an error during DNA replication (although other mechanisms and external influences can also lead to a change in DNA sequence).

Examples of mutations:

- **Substitutions:** A point mutation where one base is substituted for another, such as a SNP.
- **Indels:** Small insertions/deletions due to the addition of one or more extra nucleotides into the DNA or the loss of a section of DNA.

Slipped Strand Mismatching

STR polymorphisms derive mainly from variability in length. A proposed mechanism for these genetic variations is the *slipped strand mismatching* (SSM) mechanism: the dissociation of replicating DNA strands followed by misaligned re-association.

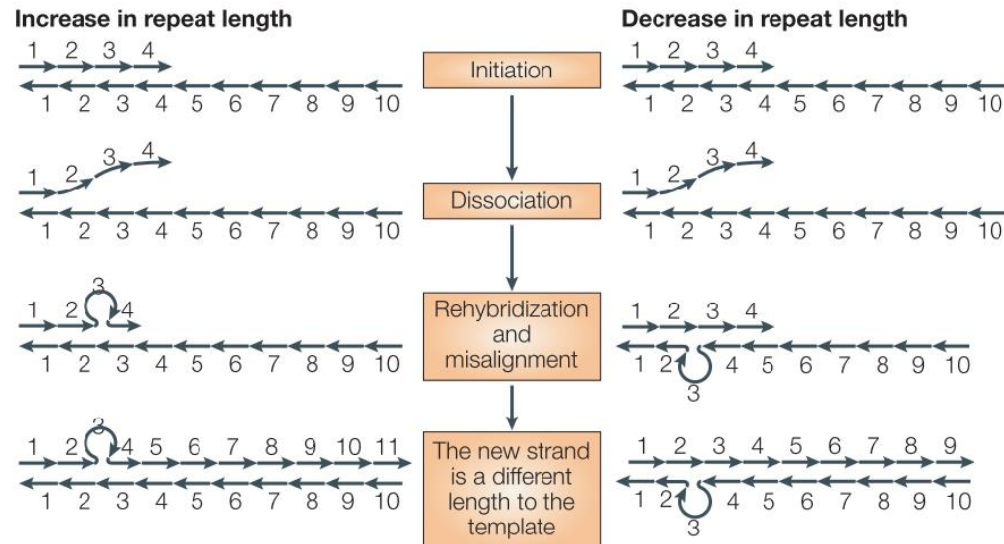


Source: Microsatellites: simple sequences with complex evolution (Ellengren, 2004).

Zoom Poll

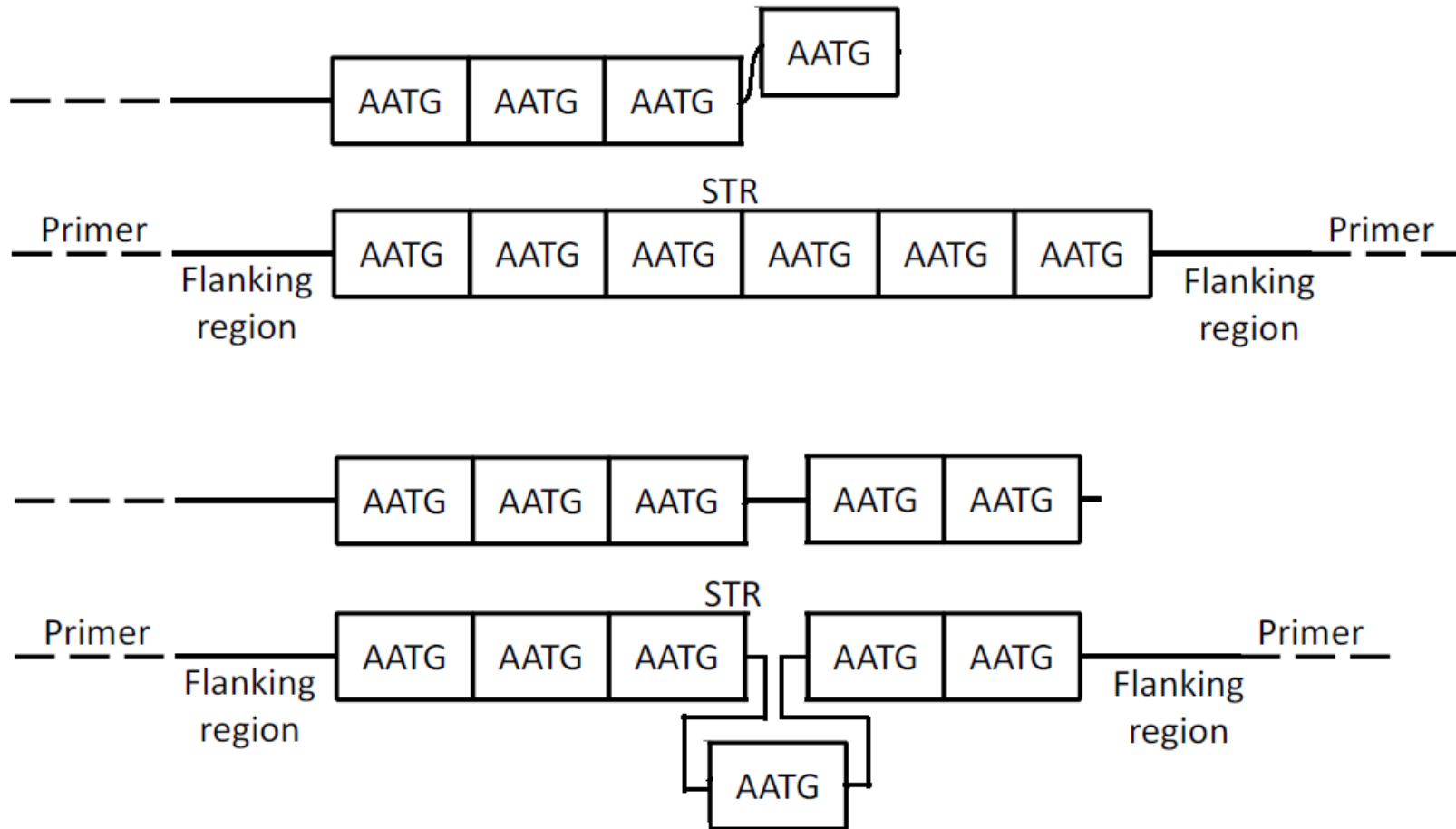
Considering the SSM mechanism, which one is most likely to occur?

- Increase in repeat length
- Decrease in repeat length
- Both equally likely



Source: Microsatellites: simple sequences with complex evolution (Ellengren, 2004).

Contractions are energetically favorable



STR Typing

To effectively interpret DNA evidence, we need to understand STR typing characteristics such as:

- The PCR-CE process
- Anomalies (like drop-ins/drop-outs)
- Peak height variability
- Stutter

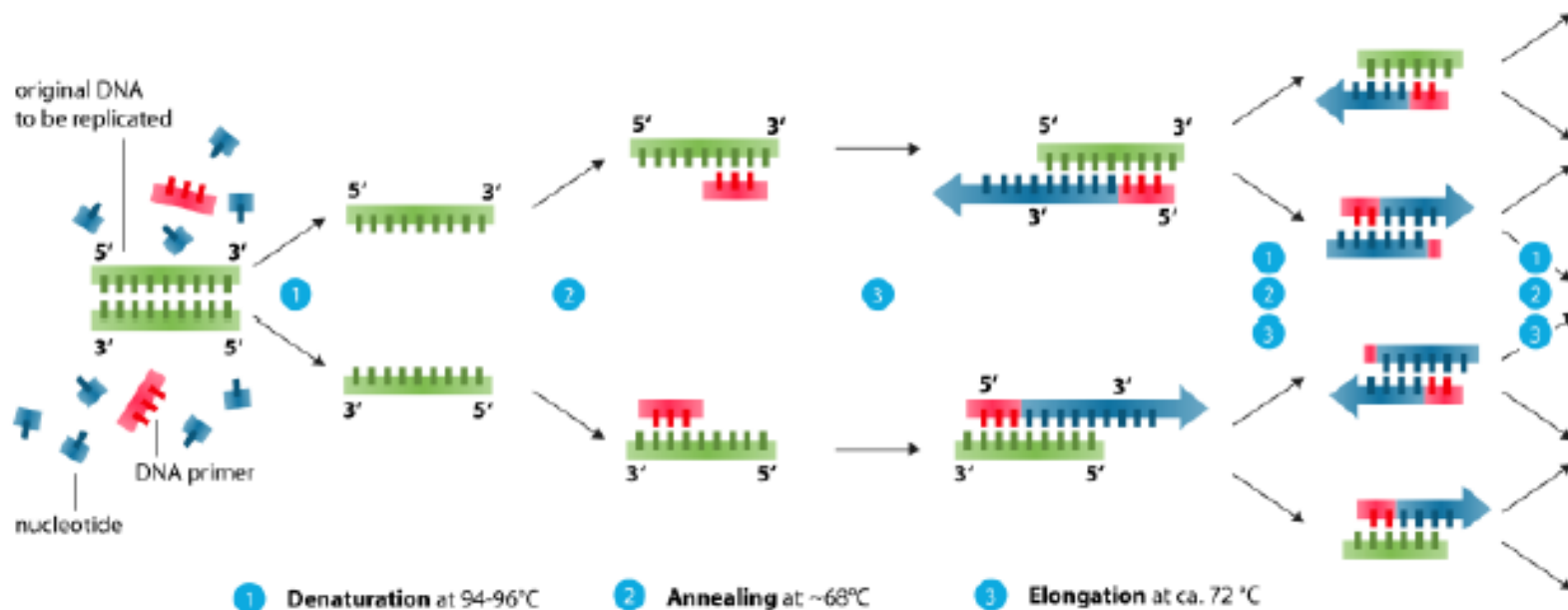
Understanding PCR

To produce an STR profile from a biological sample many identical copies of the DNA molecules within the target region (i.e. the DNA *template*) are needed. PCR (*polymerase chain reaction*) can be used to copy, or amplify, DNA through the following steps:

- **Denaturation:** Melting DNA such that the double-stranded template separates into two single-stranded DNA molecules.
- **Annealing:** Cooling the mixture to let *primers* bind to the strands.
- **Elongation:** DNA polymerase (a special copier molecule) completes missing sequences using available nucleotides.

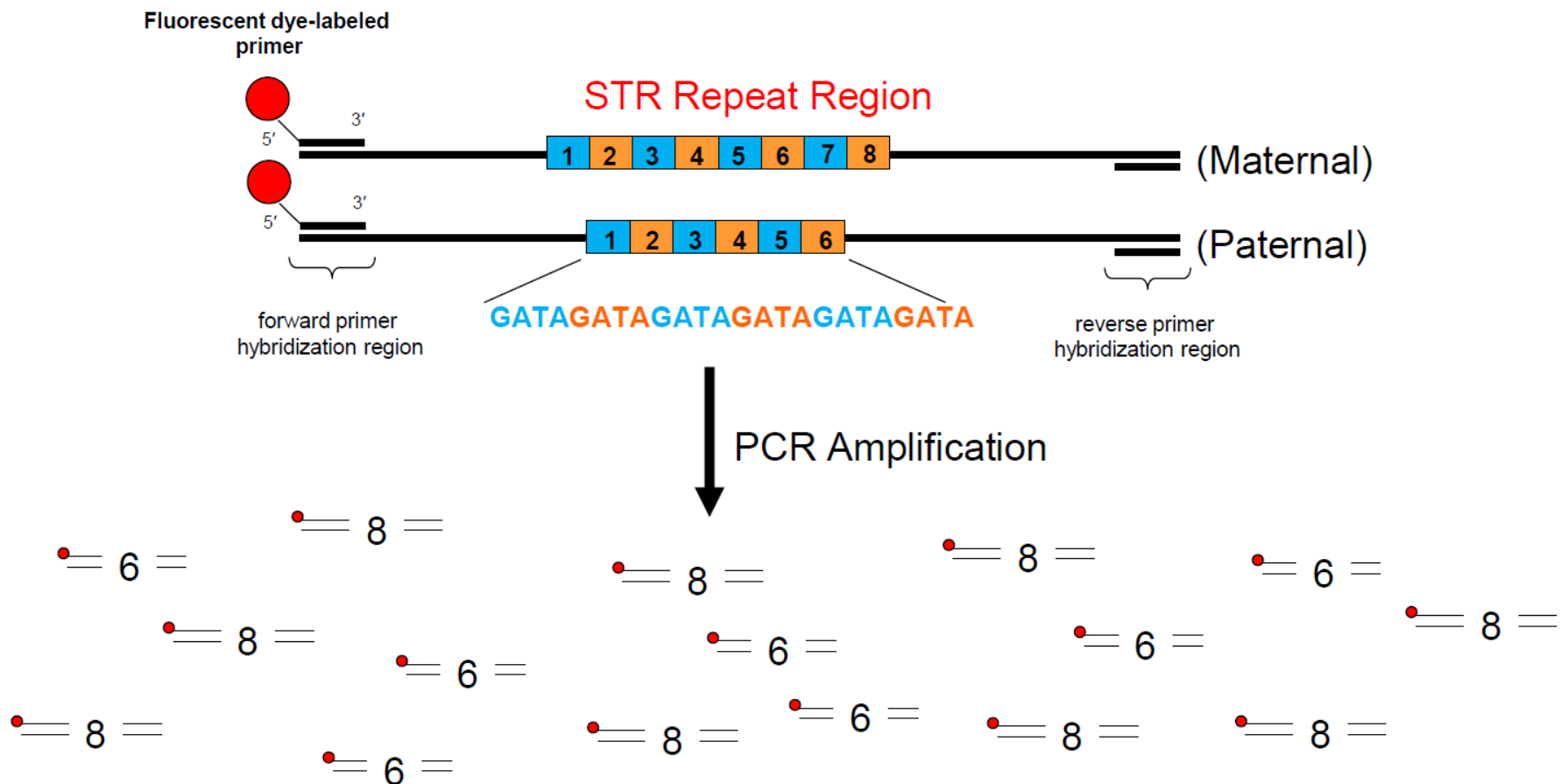
PCR Steps

These basic steps constitute one cycle, so by repeating this process, the DNA target gets amplified to millions of copies.



Source: https://en.wikipedia.org/wiki/Polymerase_chain_reaction

Short Tandem Repeat (STR) Typing

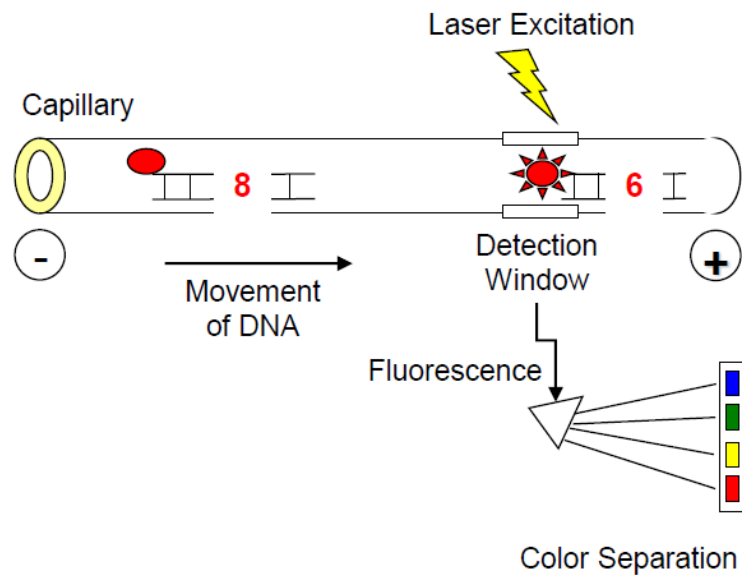


Capillary Electrophoresis

To obtain meaningful results from the PCR process, *capillary electrophoresis* (CE) has traditionally been used, allowing forensic scientists to gain access to the allele numbers contained in a DNA sample.

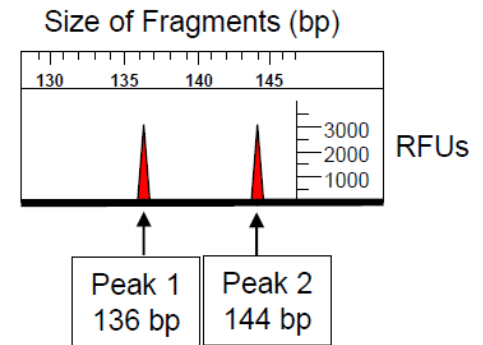
- DNA products are injected into the capillary where they travel in the direction of a positive charge;
- The travel time depends on the fragment size and can thus be used to infer the number of repeats;
- Primers are labeled with fluorescent dye, which will emit visible light at the detector window of the capillary.
- The fluorescence, measured in relative fluorescence units (RFU), is recorded over time and can be visualized with an *electropherogram* (epg).

Capillary Electrophoresis



CCD Panel

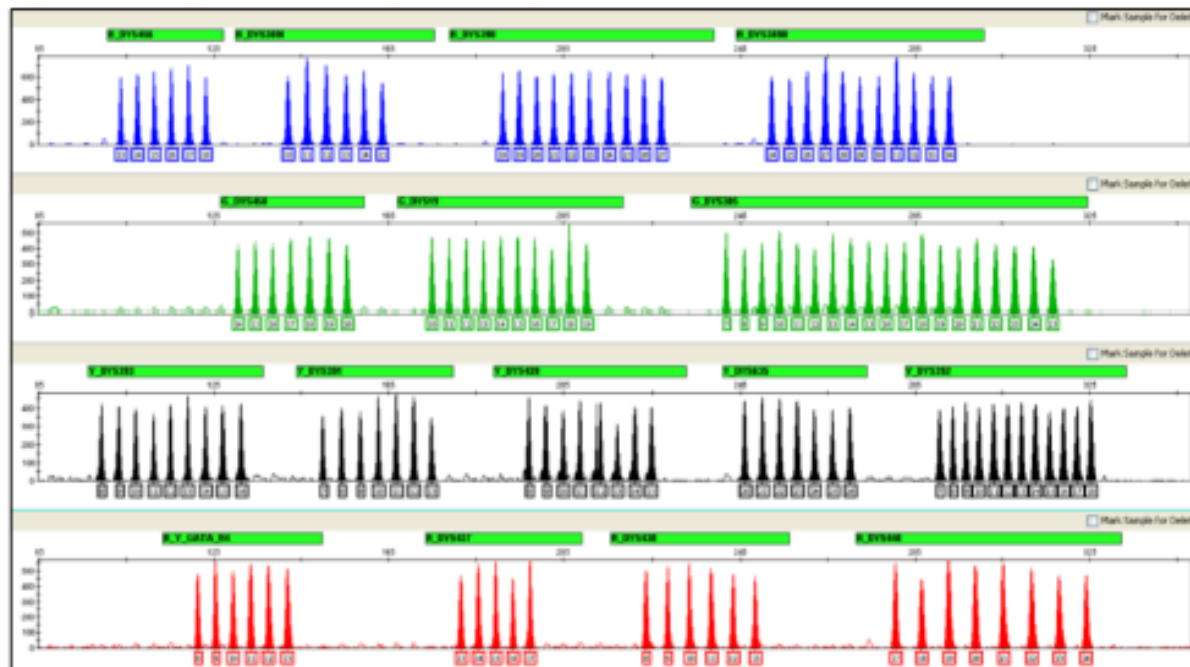
Electropherogram



Marker XYZ = 6, 8

Allelic Ladders

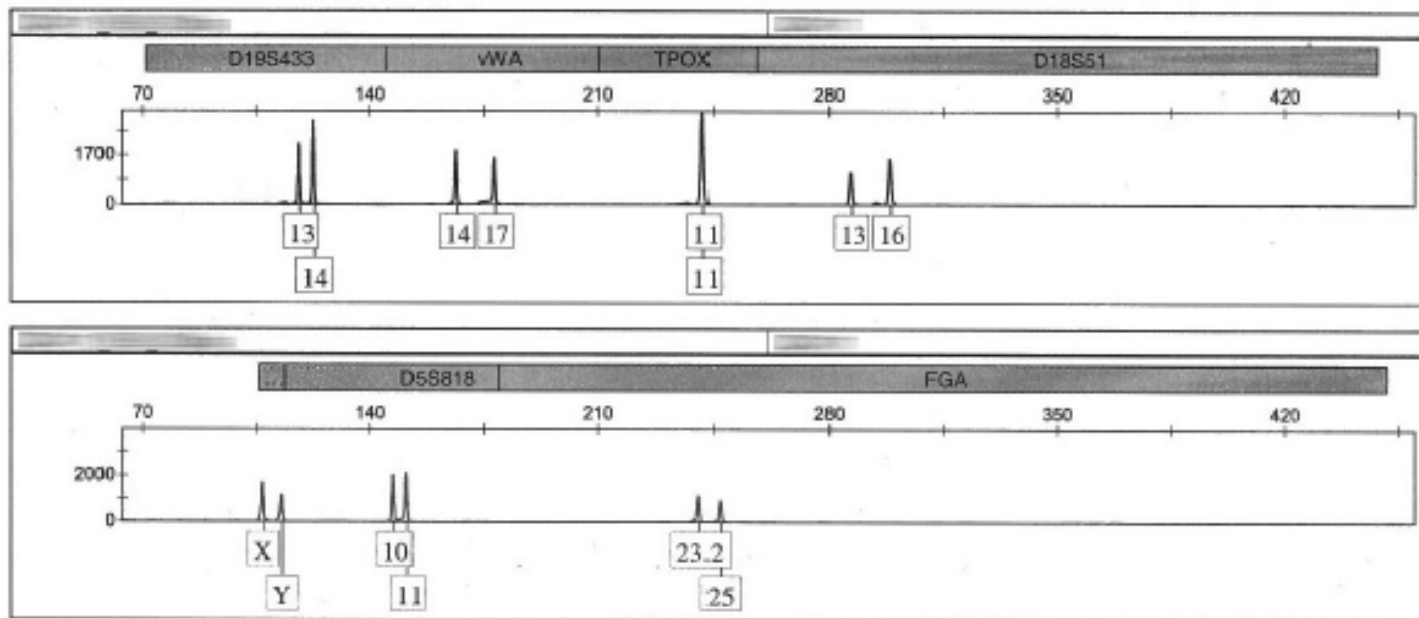
PCR-CE output can be compared to *allelic ladders* to determine allele designations.



Source: *AmpF ℓ STR Yfiler PCR Amplification Kit User Guide*.

Example of an Electropherogram

An epg shows allelic designations, represented by peaks, with integer values indicating the number of complete repeat motifs and additional nucleotides separated by a decimal point.



Source: <https://en.wikipedia.org/wiki/Microsatellite>

STR Classes

STR loci may be categorized in three different classes, based on how well alleles conform to the core repeat pattern:

- **Simple STRs:** only show variation in the number of repeats without additional sequence variation.
- **Compound STRs:** consist of several adjacent repeats of the same repeat unit length.
- **Complex STRs:** contain repeats of variable length as well as sequences.

Sequence Variation

STR loci may be categorized in three different classes, based on how well alleles conform to the core repeat pattern:

Class	Locus	Allele sequence
Simple	CSF1PO	[TCTA] ₈
Simple	Penta D	[AAAGA] ₁₂
Compound	vWA	[TCTA][TCTG] ₄ [TCTA] ₁₃
Compound	D22S1045	[ATT] ₇ ACT[ATT] ₂
Complex	FGA	[TTTC] ₃ TTTTTT[CTTT] ₁₁ CTCC[TTCC] ₂
Complex	D1S1656	[TAGA] ₄ TGA[TAGA] ₁₃ TAGG[TG] ₅

Same-length variants (i.e. *isoalleles*) have the same allele number, but differ at sequence level. This means that traditional DNA typing methods have less discriminatory capability than is potentially available via sequencing techniques.

Locus	Allele number	Allele sequence
D3S1358	15	[TCTA][TCTG] ₃ [TCTA] ₁₁
D3S1358	15	[TCTA][TCTG] ₂ [TCTA] ₁₂
D18S51	20	[AGAA] ₂₀
D18S51	20	[AGAA] ₁₆ GGAA[AGAA] ₃

Anomalies

If DNA profiling technologies were flawless, and no other (human) errors have been introduced, an STR profile would provide a perfect representation.

For good-quality samples, this is a reasonable assumption and STR allele calling is usually pretty straightforward.

However, a number of anomalies may still arise. And more importantly, crime scene profiles rarely belong to this category and usually consist of low template samples that may be contaminated and/or degraded, making them even more prone to typing errors.

Drop-ins

Allelic peaks that do not come from any of the assumed contributors to a DNA sample are termed *drop-ins*.

- Drop-ins may arise from airborne DNA fragments in a laboratory, or due to environmental exposure at the crime scene, and can typically not be reproduced on subsequent analysis of the same DNA extract.
- Verification of the source of drop-ins is not usually possible, although the existence of drop-ins can be confirmed through negative controls.
- As techniques become more sensitive, more drop-ins will occur, and potential difficulties may arise when they are incorrectly classified as analyte signal.

Contamination

Drop-ins are related to the concept of *contamination*.

- Contamination is one of the causes for drop-ins, as a result of DNA that got into a sample during collection or subsequent analysis.
- Databases of lab and scene staff can facilitate the identification of certain kinds of contamination.
- The most dangerous form of contamination is between different evidence samples, from either the same or different crime scenes.
- The observation of a more complete profile resulting from contamination is referred to as gross contamination.

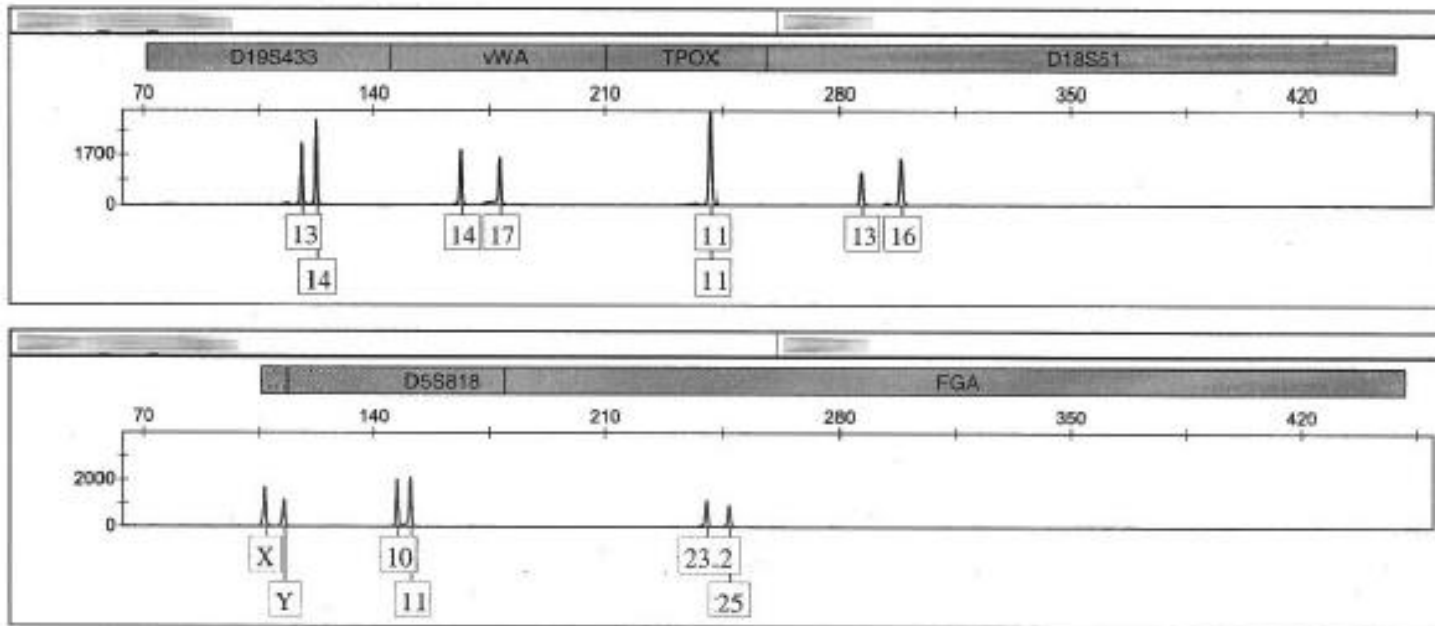
Drop-outs

A *drop-out* occurs when an allele from a contributor to the crime scene sample is not reported in the STR profile.

- This happens when a peak fails to reach the detection threshold, meaning that they cannot be reliably distinguished from background noise.
- Low template DNA samples and degradation increase the drop-out rate, which is believed to be associated with DNA fragment length.

Peak Height Variability

Besides the anomalies already discussed, several other factors play a role in observed variations within STR profiles.



Source: <https://en.wikipedia.org/wiki/Microsatellite>

Template

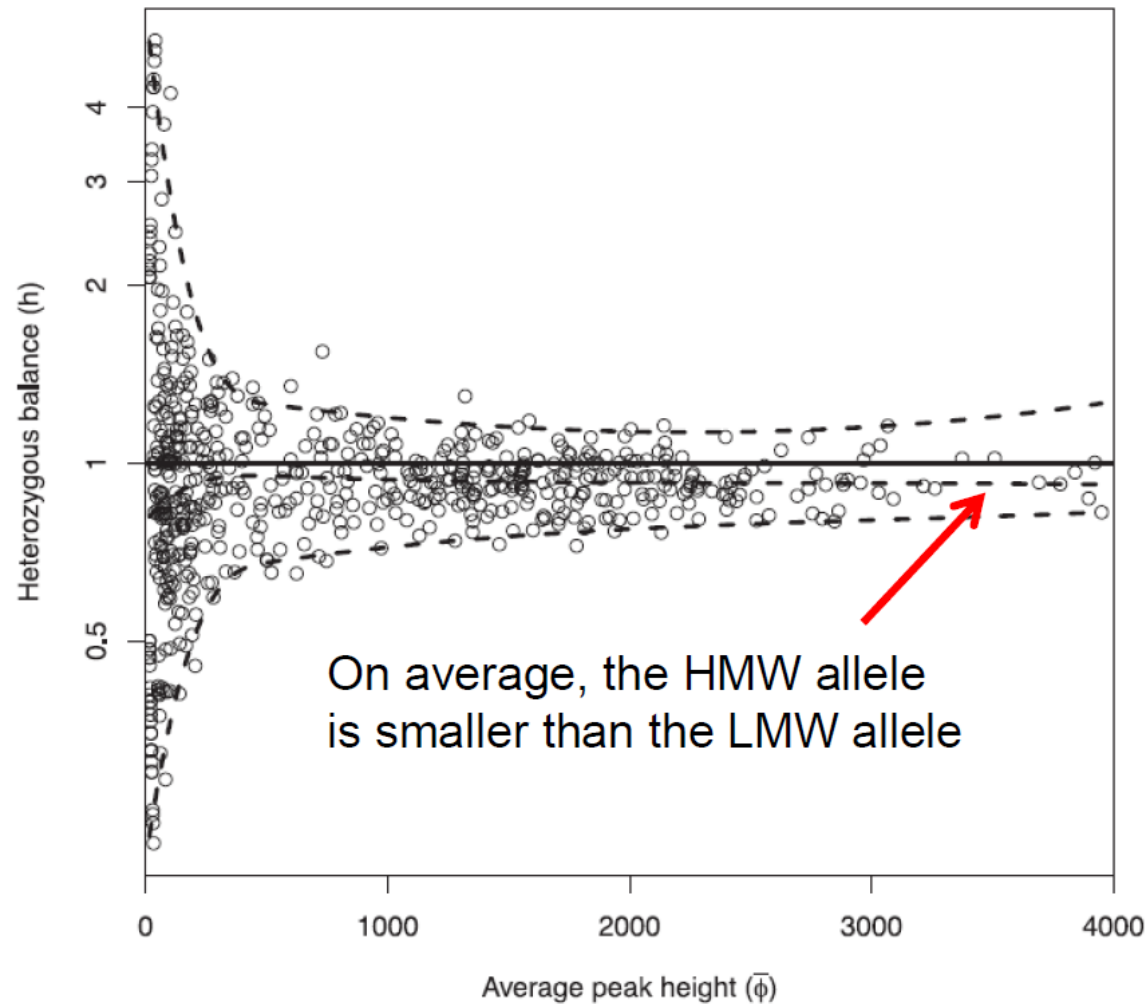
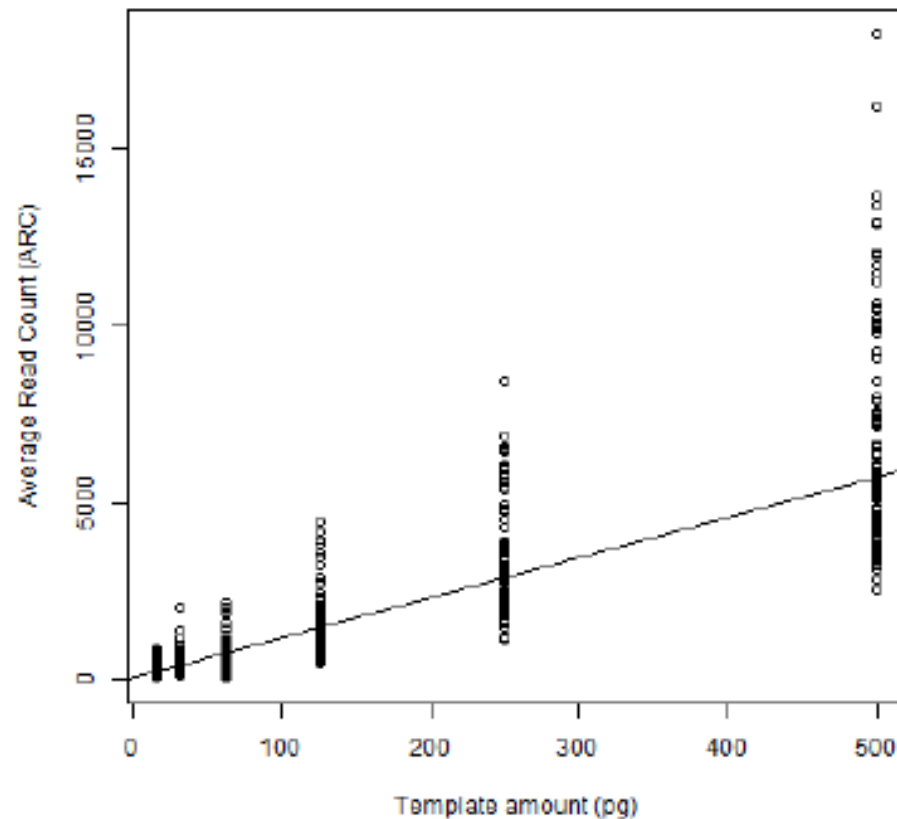


Fig. 1. Heterozygous balance versus average peak height.

Source: Kelly H, Bright J-A, Curran J, Buckleton J. Modelling heterozygote balance in forensic DNA profiles. Forensic Science International: Genetics. 2012; 6:729-734.

Template

In theory, peak heights from a single contributor are expected to be approximately proportional to the amount of undegraded DNA template.



Degradation

DNA evidence is prone to degradation due to a variety of mechanisms and circumstances, including chemical processes and environmental conditions, causing breakage of previously intact DNA molecules.

If breakage occurs in regions where primers anneal, or between the forward and reverse primers, target regions may not amplify efficiently or fail to amplify at all.

Zoom Poll

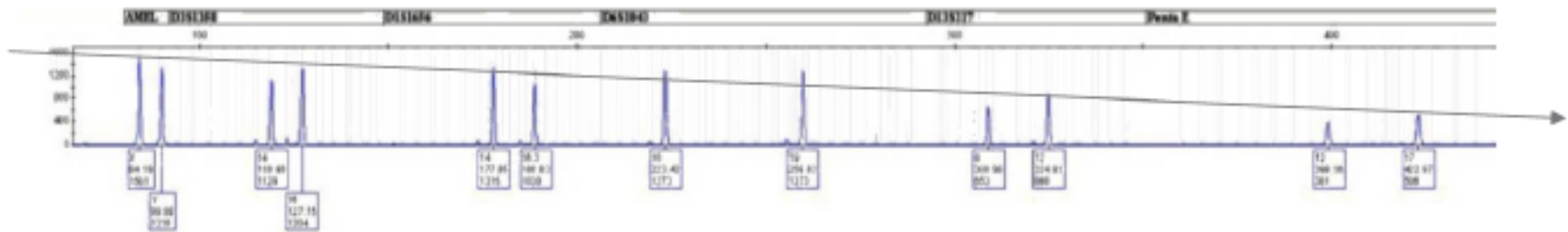
High molecular weight alleles are to drop out than low molecular weight alleles?

- More likely
- Less likely
- Equally likely

Degradation

Studies suggest that degradation leads to peak heights showing a downward trend with increasing molecular weight, supposedly because smaller alleles are more resistant to degradation.

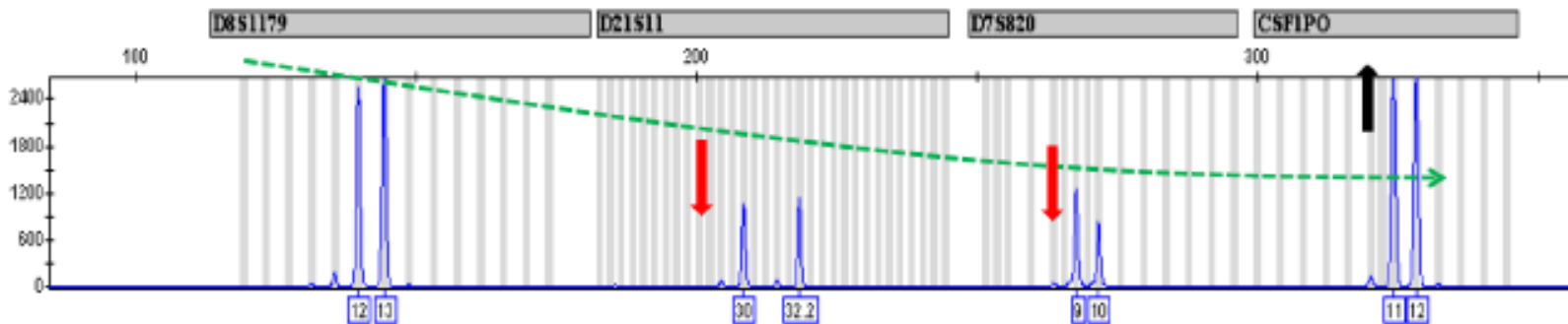
This observation is sometimes referred to as the degradation slope or the ski slope.



Locus Specific Amplification Efficiency

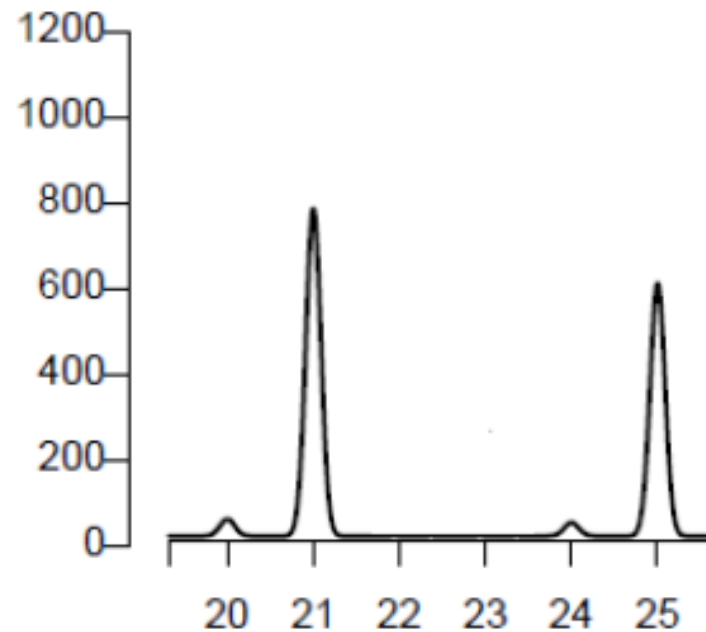
Additional variability arises from differences in amplification efficiency per locus. Observations show that some loci amplify more efficiently than others, and that these differences appear to vary over time.

Amplification bias is thought to be a result of the large variation in target loci length.



Stutter

Since STR typing methods make use of the PCR process, which relies on DNA replication characteristics, replication slippage also exists during DNA amplification of STRs in vitro.



Zoom Poll

Which allele is expected to stutter most?

- [ATAC]8
- [ATAC]12
- [AT]12

Stutter Characteristics

The characteristics shared by mutations and stutter are considerable:

- Rates increase with the number of repeat units (i.e. less stutter for shorter alleles, more stutter for longer alleles);
- Are inversely correlated with repeat unit length (i.e. more stutter for dinucleotide repeats, less stutter for tetranucleotide repeats);
- And typically involve the insertion or deletion of a complete repeat unit.

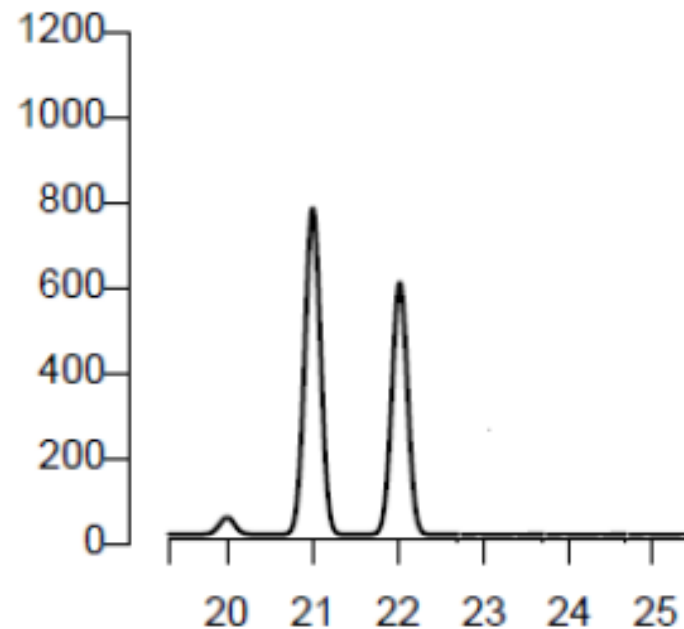
Stutter Difficulties

It is not always possible to distinguish stutter from other molecular artifacts or analyte signal:

- Stutter affected heterozygous genotypes;
- Composite stutter;
- Increase in repeat motif canceled out by a contraction;
- Compound repeats differing one nucleotide in repeat motif.

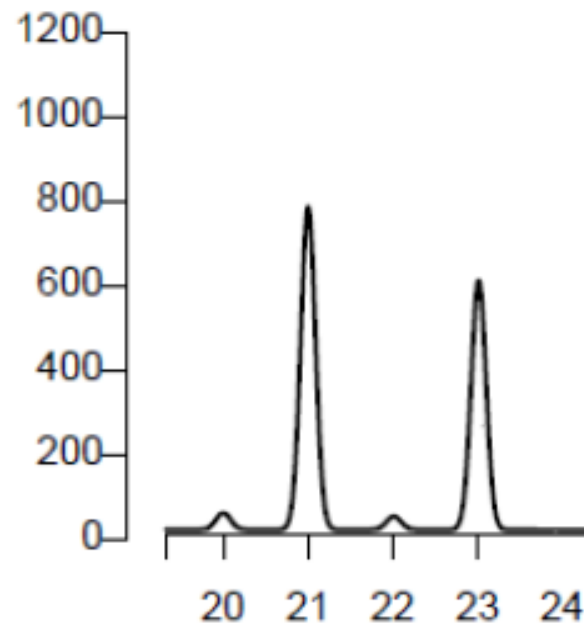
Stutter Affected Heterozygotes

Stutter affected heterozygous genotypes occur when two authentic alleles are separated by one repeat, and the total peak heights are a combination of analyte signal and stutter.



Composite Stutter

Composite stutter arises when the difference between two authentic alleles consist of two repeats and forward stutter of the low molecular weight allele coincides with back stutter of the high molecular weight allele.



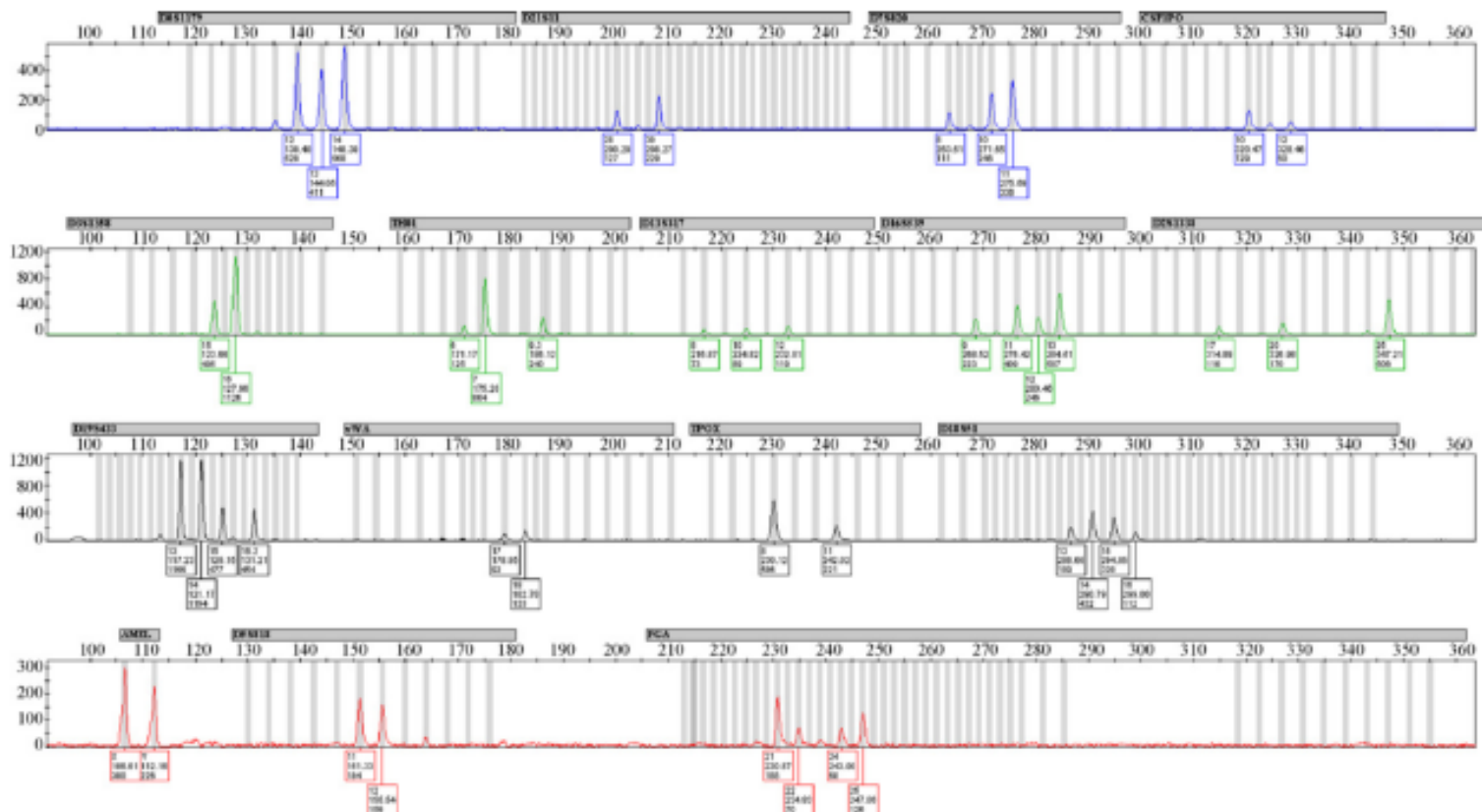
Stutter Expansion and Contraction

In rare situations, an increase in repeat motif may cancel out a repeat contraction. This artifact would not be in stutter position and can only be recognized if the expansion and contraction involve different repeats.

$$[\text{TCTA}][\text{TCTG}]_3[\text{TCTA}]_{11} \Leftrightarrow [\text{TCTA}][\text{TCTG}]_2[\text{TCTA}]_{12}$$

Mixtures

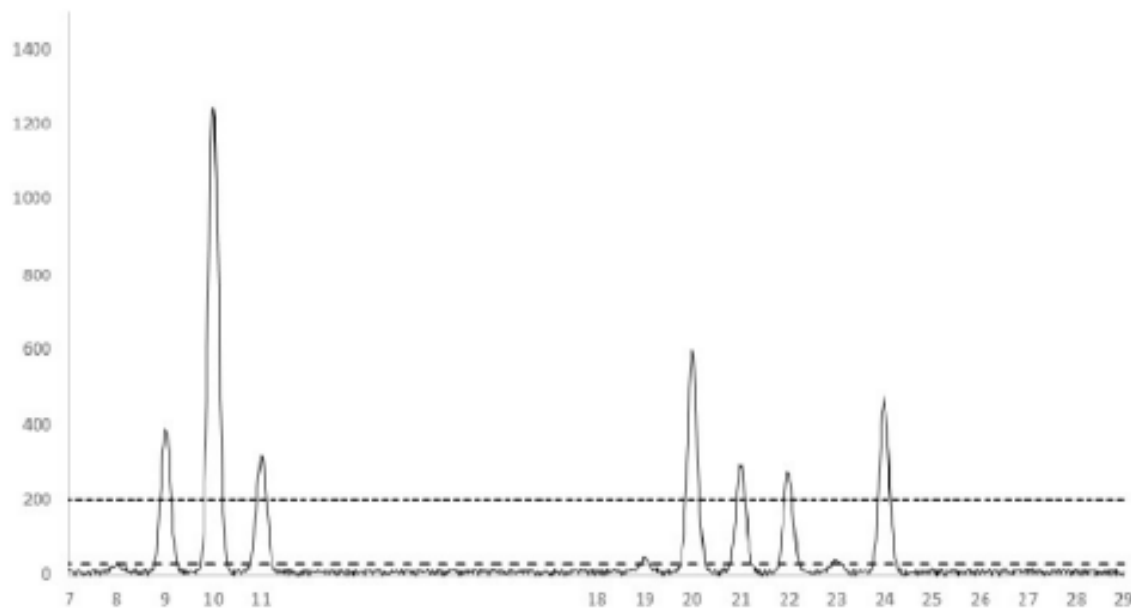
Most forensic stains contain DNA from different individuals. The number of contributors (NoC) is usually unknown.



Source: The interpretation of low level DNA mixtures (Kelly et al., 2012).

Mixtures

Mixing proportions or ratios can range from the contributors being approximately equal to each other to one being in great excess (major vs. minor contributors). Qualitative data cannot distinguish contributors in such case, so a quantitative approach may be preferred if possible.



Number of Contributors

The NoC can technically be any number of contributors. Some guidelines can be used:

- If the number of alleles is known, the maximum allele count (MAC) method can be used to estimate $\text{NoC} \geq \text{MAC}/2$.
- If the heterozygote balance for two peaks is outside a certain range they cannot be a heterozygote.
- If a peak is in stutter position it may be classified as stutter if it falls within an expected range.

These may need modification when multiple effects play a role.



Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Determination of the possible number of genotypes which can contribute to DNA mixtures: Non-computer assisted deconvolution should not be attempted for greater than two person mixtures

Phillip C. Lynch^b, Robin W. Cotton^{a,*}^a Boston University School of Medicine, M.S. Program in Biomedical Forensic Sciences, 72 East Concord St. Room R 806, Boston, MA 02118, USA^b 35 Norfolk Road, Arlington, MA, 02476, USA**Table 2**Computation of the possible number of genotype combinations which represent all possible ways that genotypes from N contributors can combine to produce $\leq 2N$ alleles.

	Number of Alleles L																Total	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		
Number of Contributors	1	1	1															2
N	2	1	4	6	3													14
	3	1	8	29	52	45	15											150
	4	1	13	84	297	600	690	420	105									2,120
	5	1	19	192	1,116	3,933	8,661	11,970	10,080	4,725	945							41,642
	6	1	26	381	3,321	18,080	63,919	150,332	236,978	247,275	163,800	62,370	10,395					956,878
	7	1	34	687	8,484	66,645	346,644	1,231,857	3,052,008	5,316,885	6,483,330	5,415,795	2,952,180	945,945	135,135			25,955,630
	8	1	43	1,155	19,428	210,645	1,529,064	7,687,512	27,472,653	71,004,690	133,874,415	184,033,080	182,338,695	126,756,630	58,648,590	16,216,200	2,027,025	811,819,826

Over 811 million combinations in an 8-person mixture

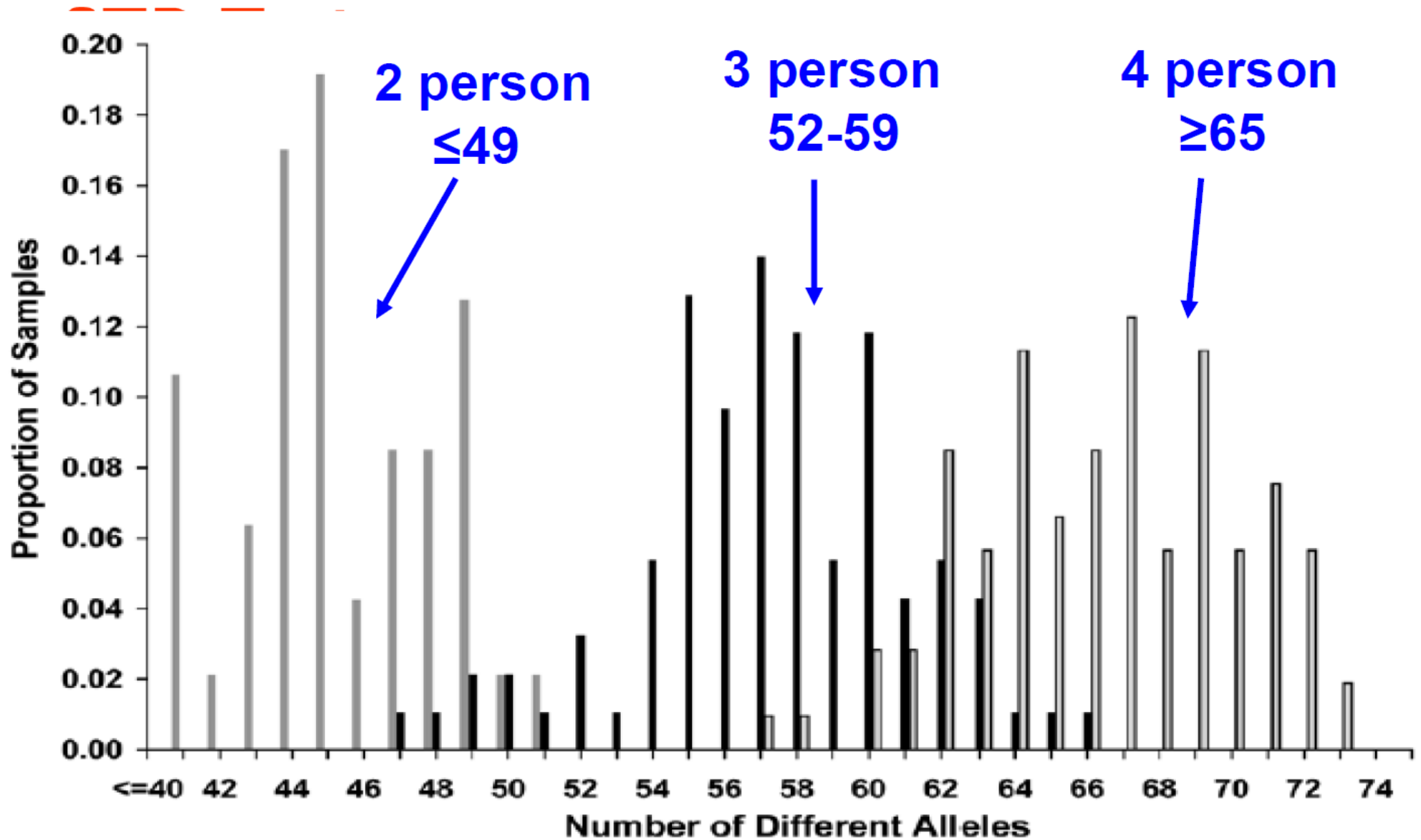


Figure 1. Expected # of different alleles from mixtures.

Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts
Perez et al., Croat Med J. 2011; 52:314-26

Uncertainty in the number of contributors in the proposed new CODIS set

Michael D. Coble^{a,*}, Jo-Anne Bright^{b,c}, John S. Buckleton^{a,b}, James M. Curran^c

^a National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

^b ESR, Private Bag 92021, Auckland 1142, New Zealand

^c University of Auckland Department of Statistics, Private Bag 92019, Auckland 1142, New Zealand



Forensic Science International: Genetics
19 (2015) 207–211

Table 3
Cumulative probability of an N person mixture appearing as a k or fewer person mixture, where $k = 1, \dots, N - 1$ for the Caucasian allele frequencies.

Configuration (# loci)	N contributor mixture	Appearing as k or fewer			
		1	2	3	4
Existing CODIS (13)	5	9.66E-33	2.10E-06	0.4141	0.9897
GlobalFiler (21)		8.21E-61	7.1E-15	0.0048	0.6099
Proposed CODIS (20)		5.82E-54	9.1E-12	0.0592	0.9228
Fusion (22)		3.57E-61	7.8E-14	0.0270	0.8885

Adding more loci – helps up to a point...

Heterozygote Balance

A consequence of all the stochastic variations that have been introduced into the process, is that the two peaks of heterozygous alleles will also show variability, termed the *heterozygote balance*.

The difference is thought to be affected by the number of repeat sequences, since high molecular weight alleles:

- Stutter more;
- And amplify less.

Heterozygote Balance

Understanding the variability in heterozygous balance is important for the interpretation of mixed profiles and low template DNA:

- For LTDNA, peaks may be so imbalanced that it leads to alleles not exceeding the allelic threshold or even a drop-out.
- It may be used to classify combinations of alleles (or genotypes) as possible or impossible when considering a mixture.

Summary

To effectively interpret DNA evidence, phenomena and factors like **mutations, CNVs, contamination, template amount, replicates, amplification efficiency, and degradation** should be considered.

These lead to observations in the form of **stutter, drop-ins, drop-outs, peak height variability and heterozygote balance**, that may need to be incorporated in weight-of-evidence calculations.