Short Tandem Repeat Loci

Workshop

Phoenix, AZ

October 4, 2004

DNA TYPING METHODS

- PCR-based

- Faster results
- Less labor-intensive
- 1 ng DNA
- Degraded DNA

DNA TYPING METHODS

- D1 S80
  - single locus
  - preferential amp.
- DQA1 / PM
  - interpretational difficulties

- PCR-based
- Faster results
- Less labor-intensive
- 1 ng DNA
- Degraded DNA

Need for new typing method

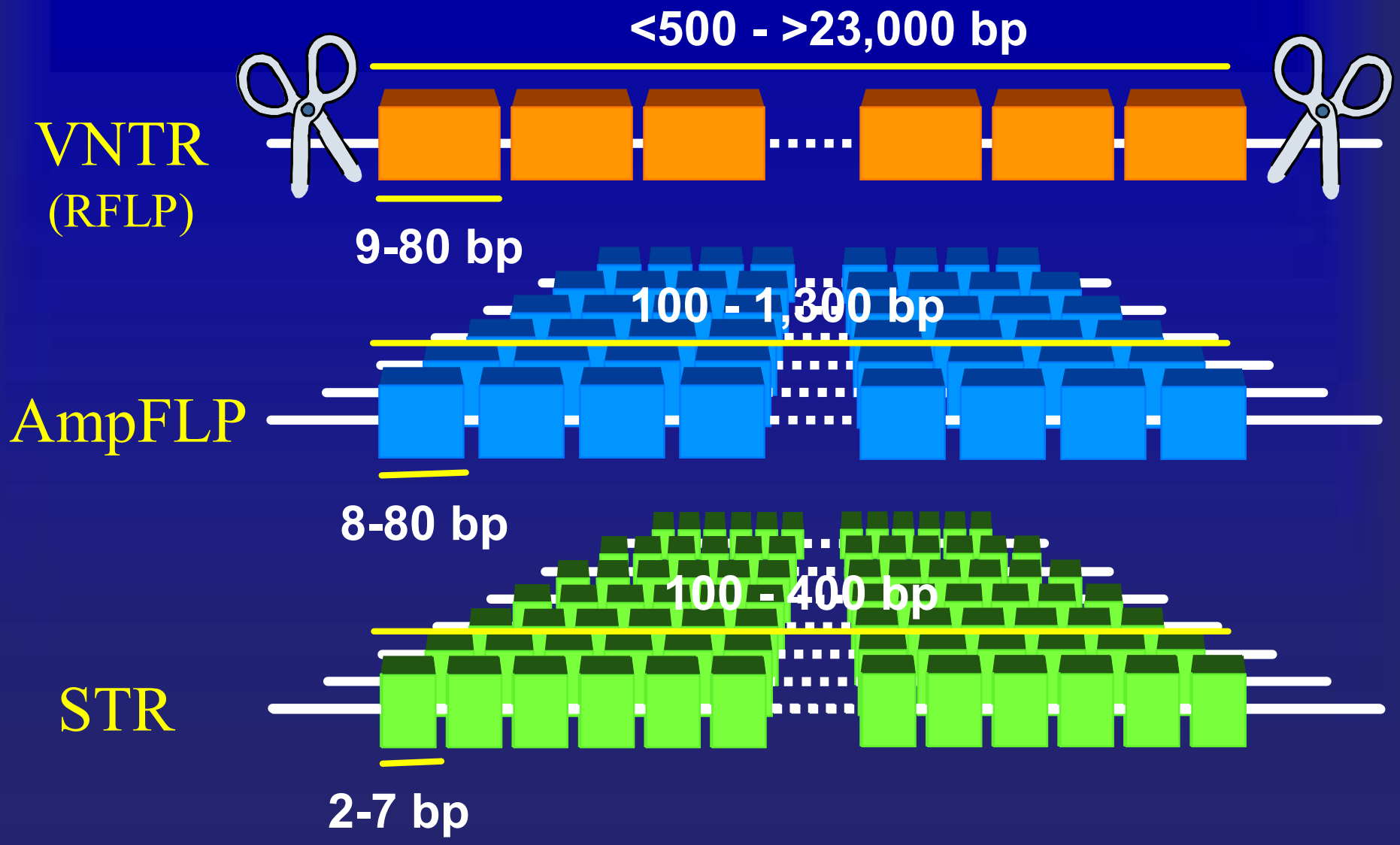
Repetitive DNA contains sequences that occur more than once

REPETITIVE DNA
(20-30% of human genome)

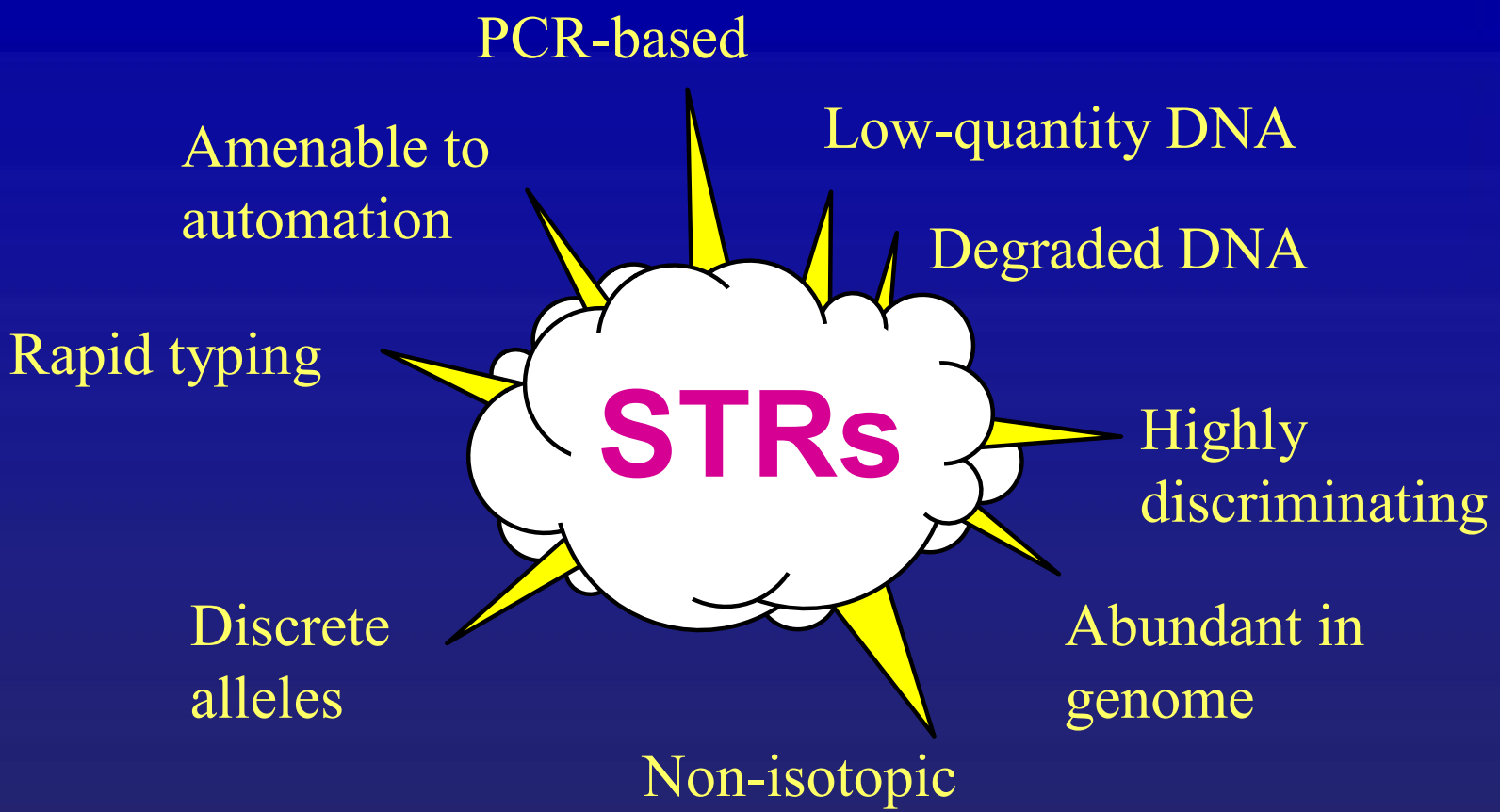
- Interspersed:
- Single repeats scattered throughout the genome
- 10% of genome

- Tandemly Repeated:
- Head-to-tail tandem repetition of a common sequence
- 15-20% of genome

Repeat Polymorphisms



Short Tandem Repeats



SHORT TANDEM REPEATS

- Arrays of short repeats (2-7 bp) that are repeated several times in tandem
- >30,000 in the human genome
- One every 10 kb

LENGTH OF REPEAT UNIT

2	Dinucleotides:	CA
3	Trinucleotides:	AAT
4	Tetranucleotides:	AGAT
5	Pentanucleotides:	AATGT

GENERAL CRITERIA FOR SELECTING STR LOCI

2 Dinucleotide repeats:

Plentiful, but too much stutter

4 Tetranucleotide repeats:

Demonstrate less stutter

5 Pentanucleotide repeats:

Less stutter, but greater bp range

GENERAL CRITERIA FOR SELECTING STR LOCI

4 Tetranucleotide repeats:

PD > 0.85

Size - 100-350 bps

Allele range < 60 bps

Low stutter

TETRANUCLEOTIDE REPEATS

are used in forensics

- Undesired PCR products are greater with di- and trinucleotide repeats
- Some trinucleotides are associated with diseases
- Alleles are more readily resolved
  - 4 bp, not 2 or 3 bp

SHORT TANDEM REPEATS:
Tetramers

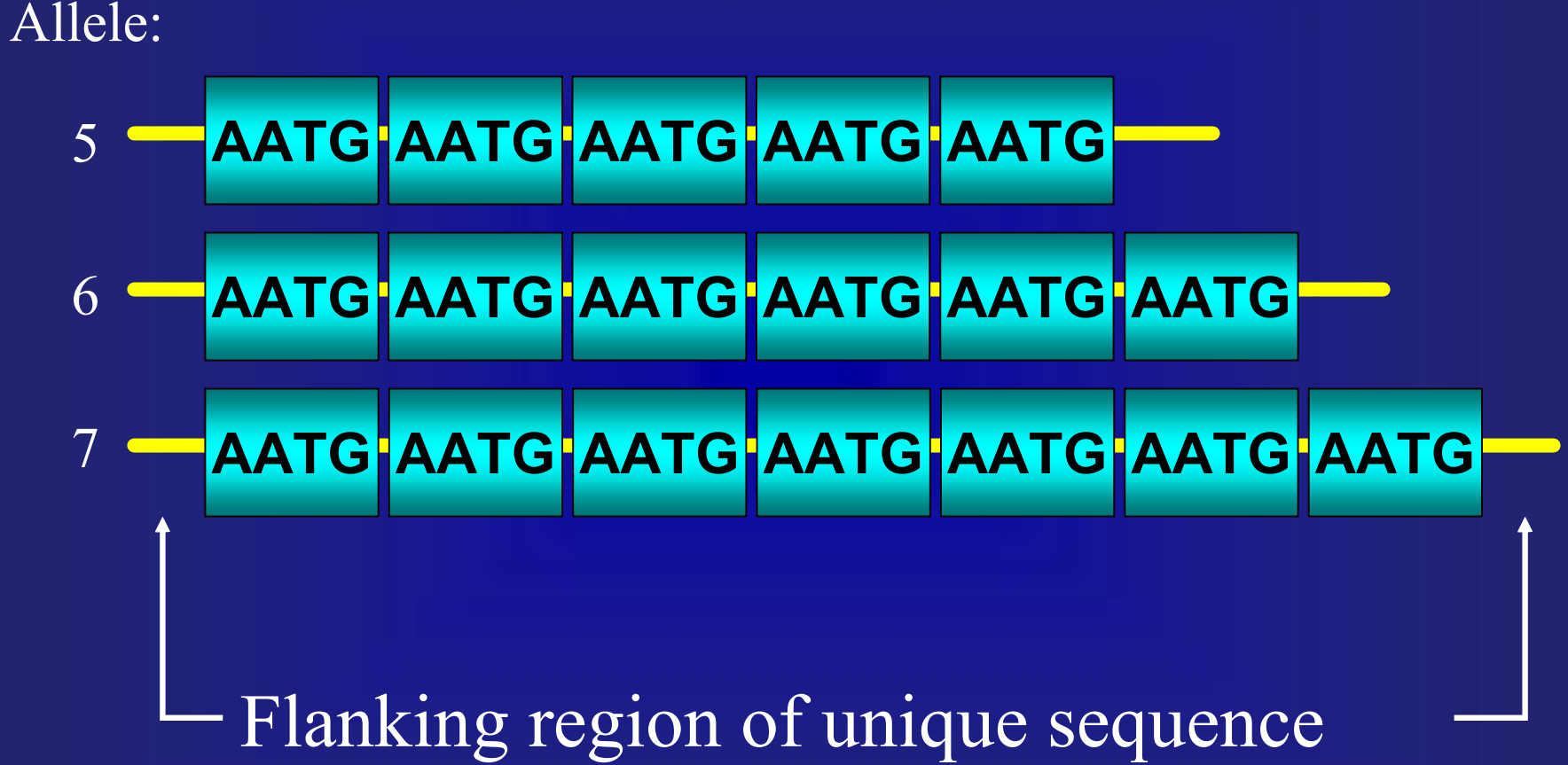
AGAT	AGAT	AGAT	AGAT
AAAG	AAAG	AAAG	AAAG
AATG	AATG	AATG	AATG
AAAT	AAAT	AAAT	AAAT

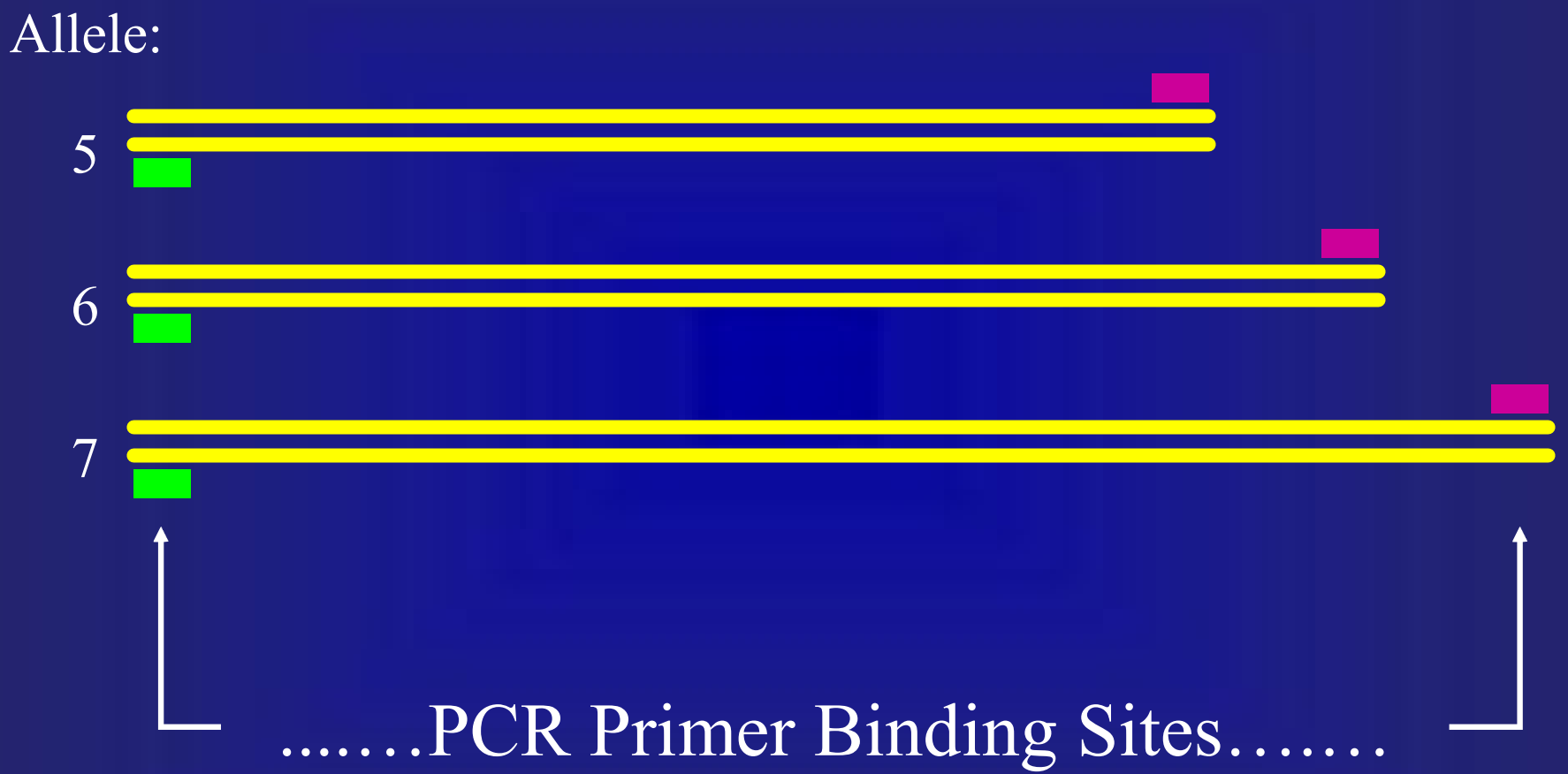
STRs are POLYMORPHIC

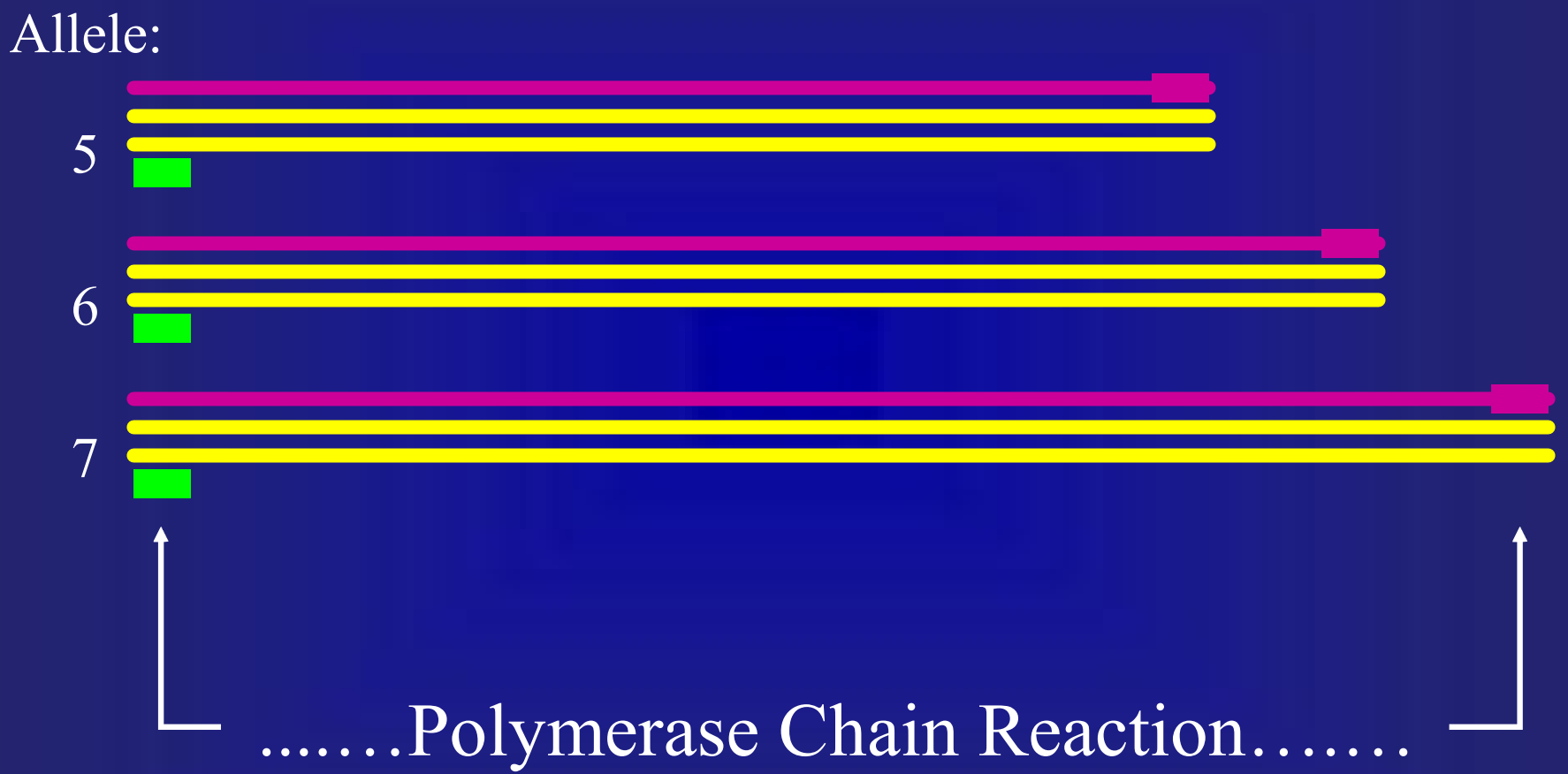
- Usually, variation in the number of tandem repeats that make up the array

- Alleles are designated by the number of repeats that they contain

Sequence variation can affect mobility and thus polymorphic nature - native vs denaturing conditions


Allele:
5
AATG AATG AATG AATG AATG
6
AATG AATG AATG AATG AATG AATG
7
AATG AATG AATG AATG AATG AATG AATG
Flanking region of unique sequence


Allele:
5
6
7
.....PCR Primer Binding Sites.....

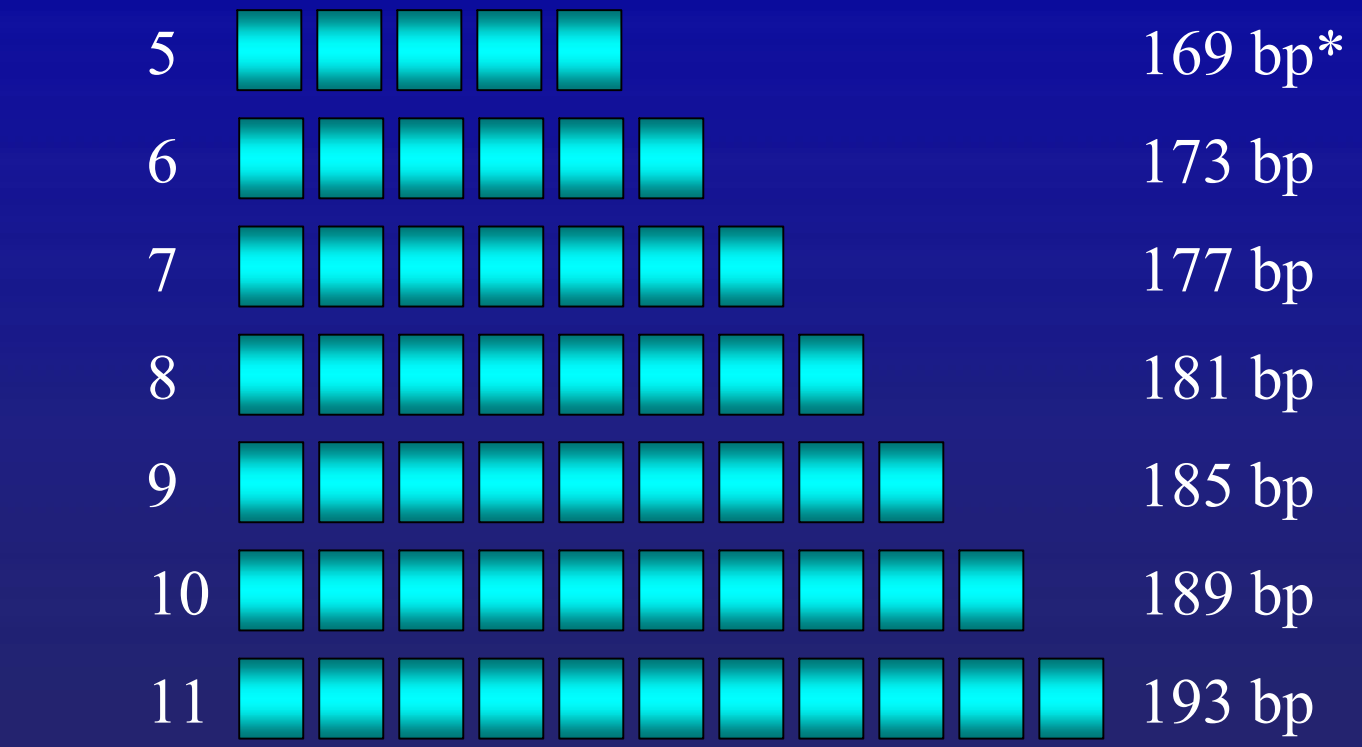

Allele:
5
6
7
.....Polymerase Chain Reaction.....


Allele:
5
6
7
.....Polymerase Chain Reaction.....

TH01, a simple locus

- (AATG)$_{5\text{-}11}$

Operational sizing

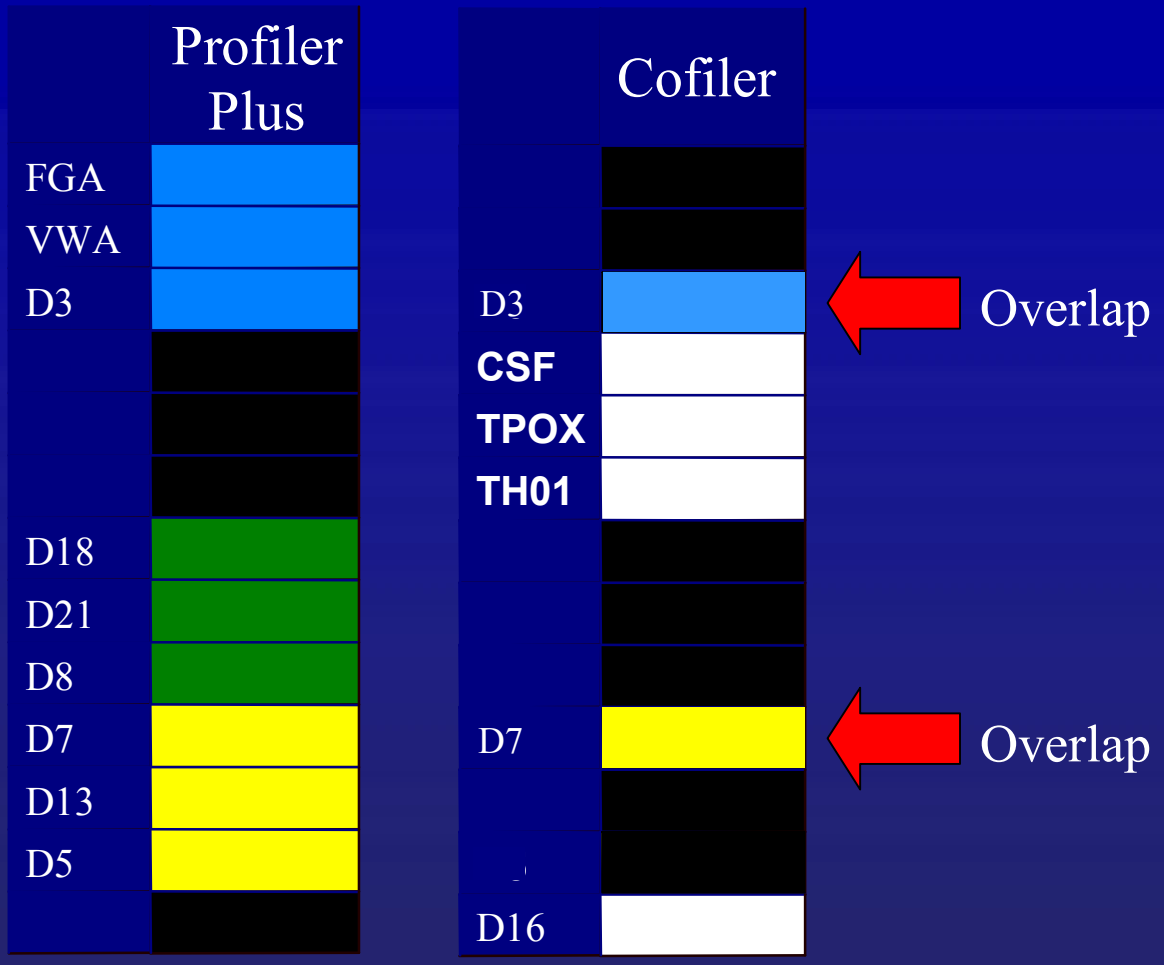


*AmpFlSTR

STR LOCUS TYPES

- SIMPLE
  - TH01: $(AATG)_{5-11}$
- COMPOUND
  - VWA: $TCTA\ (TCTG)_{3-4}(TCTA)_n$
- COMPLEX
  - D21S11: $(TCTA)_{4-6}(TCTG)_{5-6}(TCTA)_3TA\ (TCTA)_3TCA(TCTA)_2TCCATA\ (TCTA)_{8-16}TC$

13 Core Loci in 2 Amplifications



13 Core Loci Plus 2 more STR Loci in 1 Amplification

Penta D
D18S51
D21S11
THO1
D3S1358

FGA
TPOX
D8S1179
VWA
Amelogenin

Penta C
CSF1PO
D16S539
D7S820
D13S317
D5S818

FGA	Intron 3 of fibrinogen alpha chain	4q28
$(TTTC)_3$ TTTT TTCT $(CTTT)_n$ CTCC $(TTCC)_2$		
VWA	Intron 40 of von Wildebrand factor	12p12-12pter
TCTA $(TCTG)_{3-4}$ $(TCTA)_n$		
$D3_{S1358}$	Anonymous	3p
TCTA $(TCTG)_{1-3}$ $(TCTA)_n$		

$D18_{S51}$	Anonymous	18q21.3
$(AGAA)_n$		
$D21_{S11}$	Anonymous	21
$(TCTA)_n\ (TCTG)_n\ [(TCTA)_3TA\ (TCTA)_3\ TCA\ (TCTA)_2\ TCCA\ TA]\ (TCTA)_n$		
$D8_{S1179}$	Anonymous	8
$(TCTR)_n$	[R = A or G]	

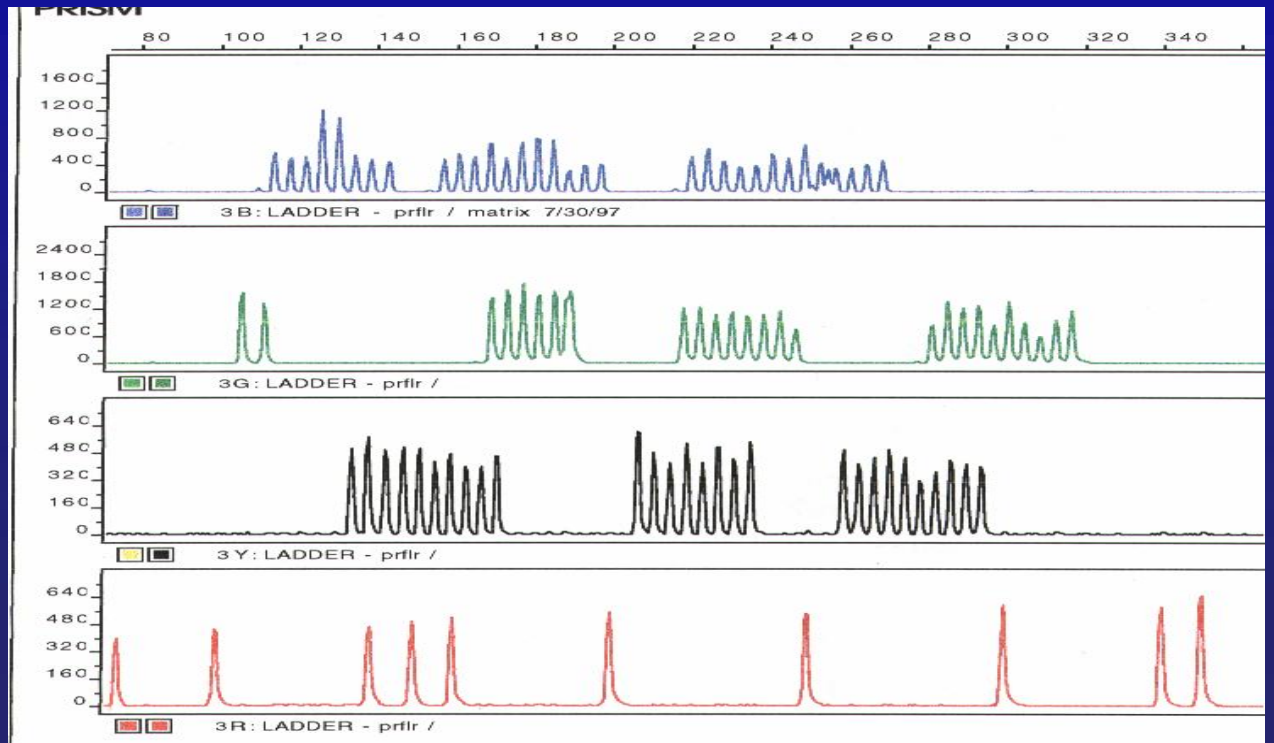
CSF_{1PO}	CSF-1 Receptor	5q33.3-34
$(AGAT)_n$		
TPOX	Intron 20 of thyroid peroxidase	2p23-2pter
$(AATG)_n$		
TH01	Intron 1 of tyrosine hydroxylase	11p15.5
$(AATG)_n$		

$D5_{S818}$	Anonymous	5q21-31
$(AGAT)_n$		
$D13_{S317}$	Anonymous	13q22-31
$(GATA)_n$		
$D7_{S820}$	Anonymous	7q
$(GATA)_n$		

D16S539	Anonymous	16q24-qter
(AGAT)		

ALLELIC LADDERS

Used to designate the alleles when genotyping an unknown sample



ALLELIC LADDERS

- Simple Ladder is made with alleles that are 4-bp apart and that are within the range of common alleles
- Some variants may be included in the ladder (e.g., TH01 9.3)
- *Not all* observed variants are included in the ladder

"OFF-LADDER" VARIANTS

Off-ladder (OL) alleles fail to align with the incremental 4-bp ladder or reside above or below ladder

To classify the variant allele:

Operationally defined nomenclature

Number of repeats, followed by a decimal point number of nucleotides of partial repeat

Above or below ladder use > or < highest or lowest allele in ladder

"OFF-LADDER" VARIANTS

TH01

9	$(AATG)_9$
9.3	$(AATG)_6$ **ATG** $(AATG)_3$
10	$(AATG)_{10}$

"OFF-LADDER" VARIANTS

TH01

9	$(AATG)_9$
9.3	$(AATG)_6$ ATG $(AATG)_3$
10	$(AATG)_{10}$

VWA

16	$(TCTA)_2 (TCTG)_4 (TCTA)_{10}$
16.2	$(TCTA)_2 (TCTG)_4$ TA $(TCTA)_{10}$

Designing STR Systems

- Primers
- Test primers for compatility:
  specificity, balance, primer dimer,
  full A addition
- Test PCR components
  MgCl2, KCl, dNTPs, template
- Thermal cycler conditions
- Post-PCR analytical procedures
  fluors, electrophoresis, detection
  systems, interpretation

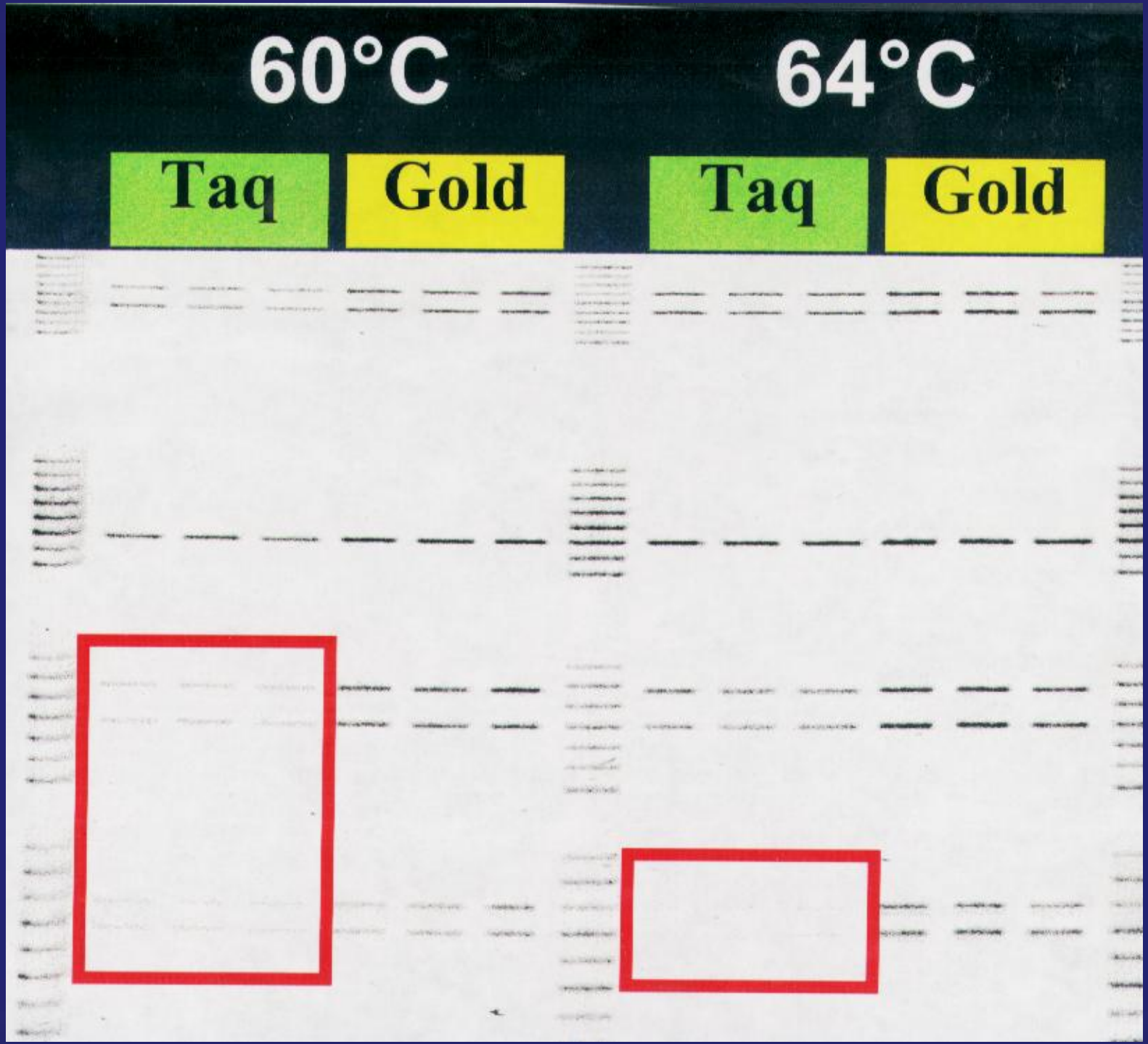
STR Primer Design

- No size overlap for loci with same fluor
- Similar Tm for primers
- Avoid hairpin structures
- Minimal complementary binding at 3' end
- Maximal 3' A addition
- Test multiplex at low Tanneal for specificity

Specificity

AmpliTaq Gold

- Hot start
- Simplifies PCR set up
- Reduces primer design requirements
- Decreases primer-dimer formation
- Facilitates multiplexing
- Automation

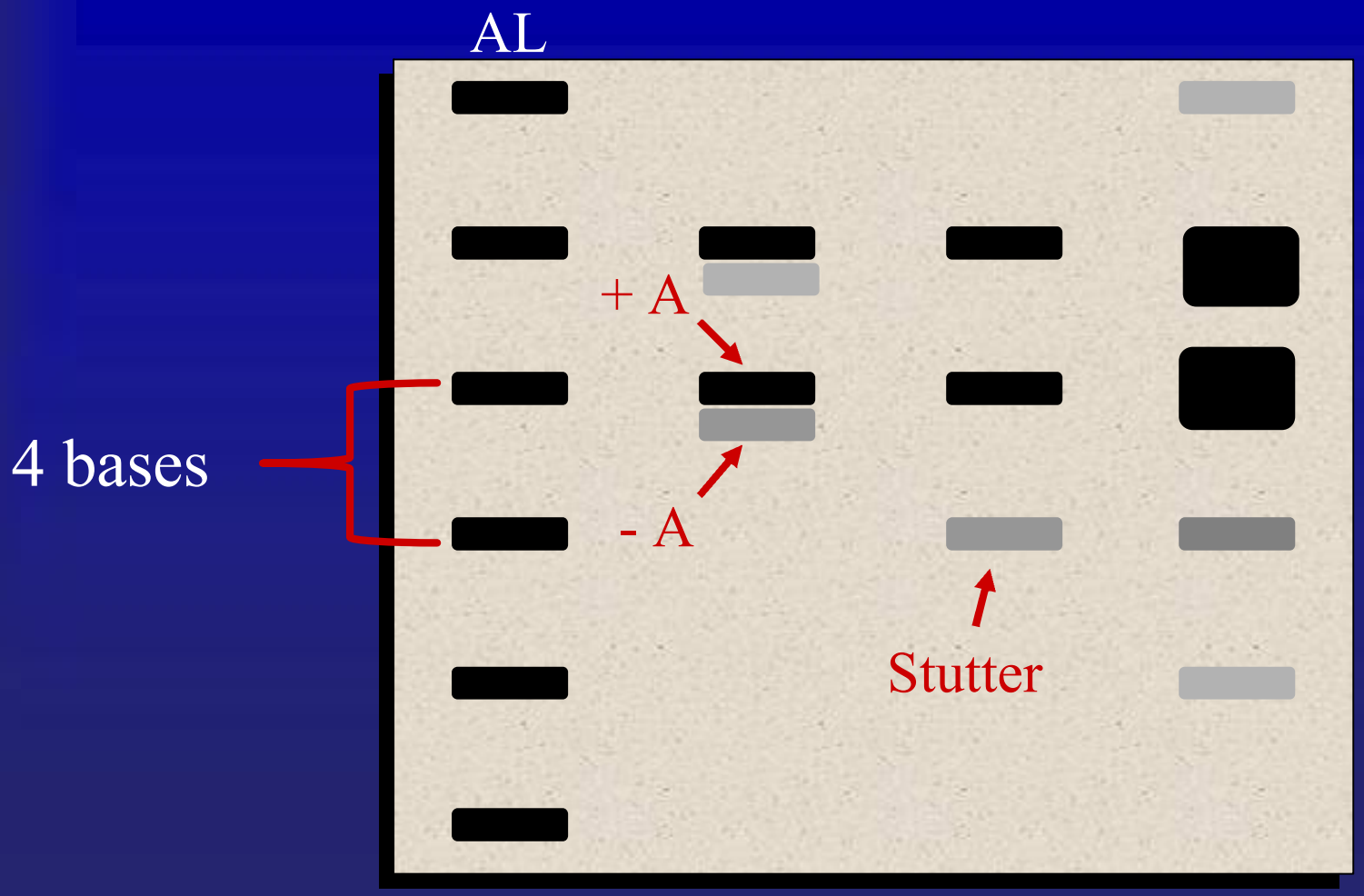

60°C
64°C
Taq
Gold
Taq
Gold

UNDESIRED PCR PRODUCTS

Undesired!
But inherent.

"Minus A" or "Plus A"
Stutter

The difference between "Minus A" and Stutter

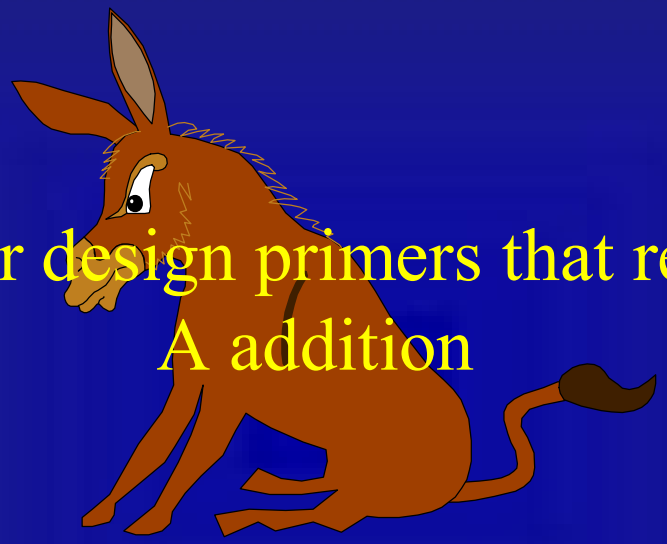


Non-Templated Nucleotide Addition ("Plus A")

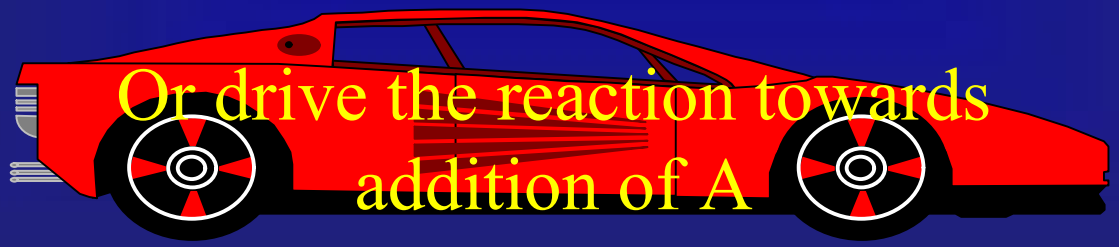
- Many DNA polymerases add an extra nucleotide to the blunt end of a DNA product *without* the use of a template
- Usually A is added
- The resulting band/peak is 1-bp larger in size than that predicted from the DNA sequence and primer locations

Non-Templated Nucleotide Addition ("Plus A")

There are two ways to deal with+/-A



Either design primers that resist A addition



Or drive the reaction towards addition of A

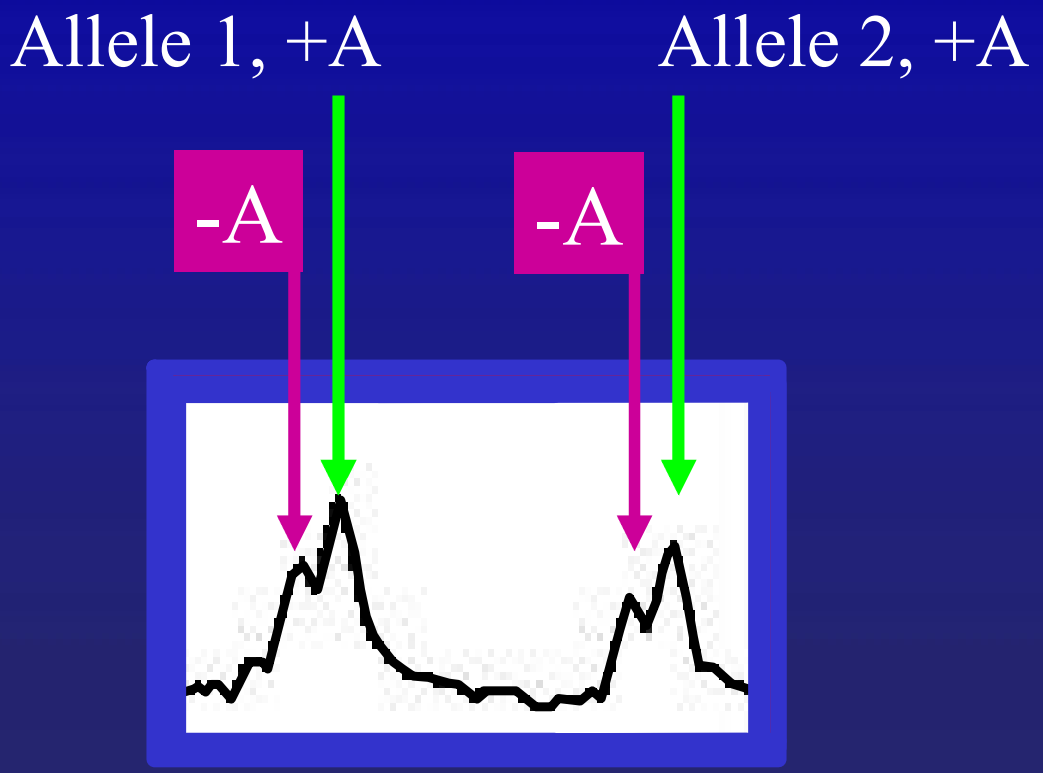
If you can't beat 'em, join 'em

One cannot effectively prevent +A. However, one can increase +A through primer design...

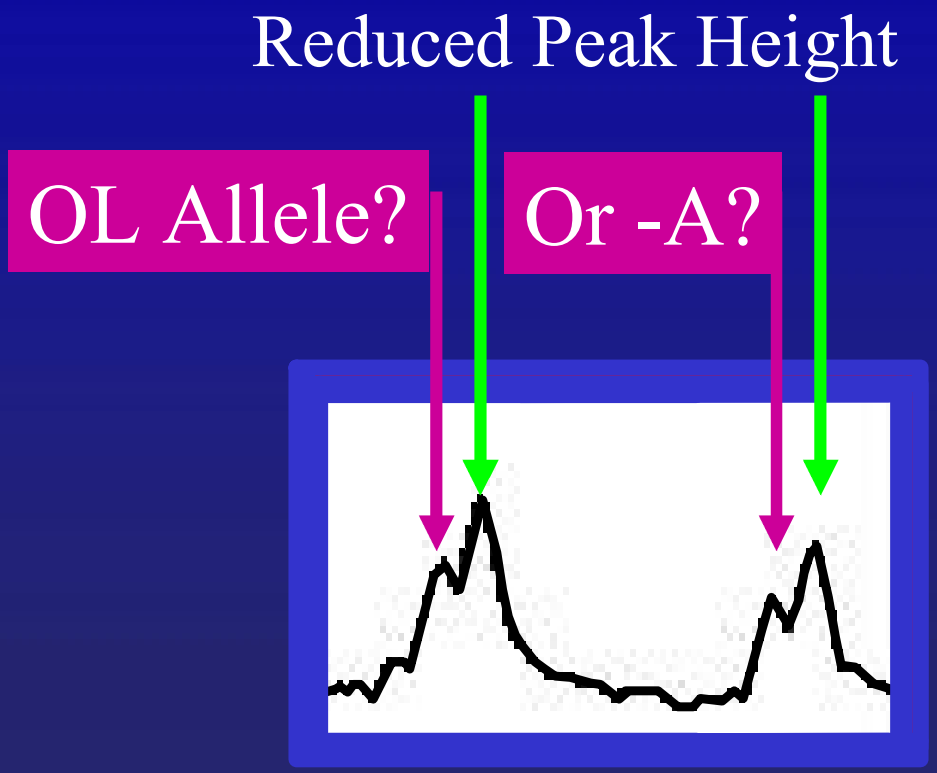
“Plus A” is the desired product.

Ladder bands / peaks are +A!

Inefficient adenylation results in 2 bands or peaks (or odd-shaped peaks)



Inefficient adenylation results in 2 bands or peaks (or odd-shaped peaks)



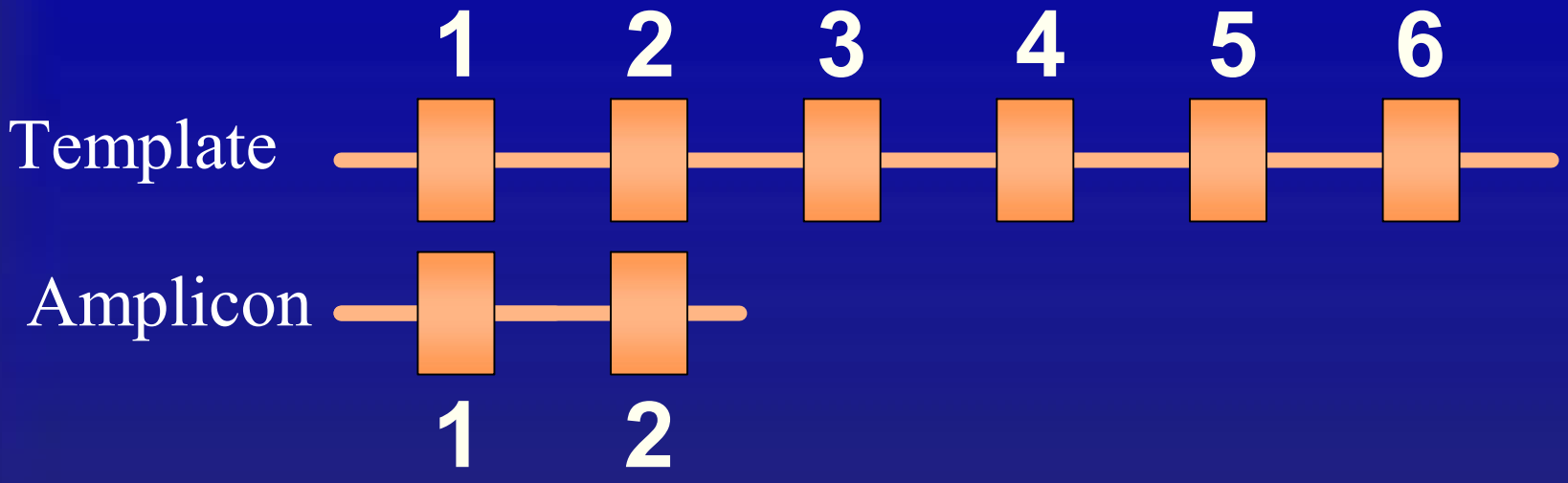
ADDRESSING “MINUS A”PEAKS

- Repeat final extension step with additional enzyme
- Use knowledge of adenylation efficiency at a given locus
- Use database information to aid in interpretation

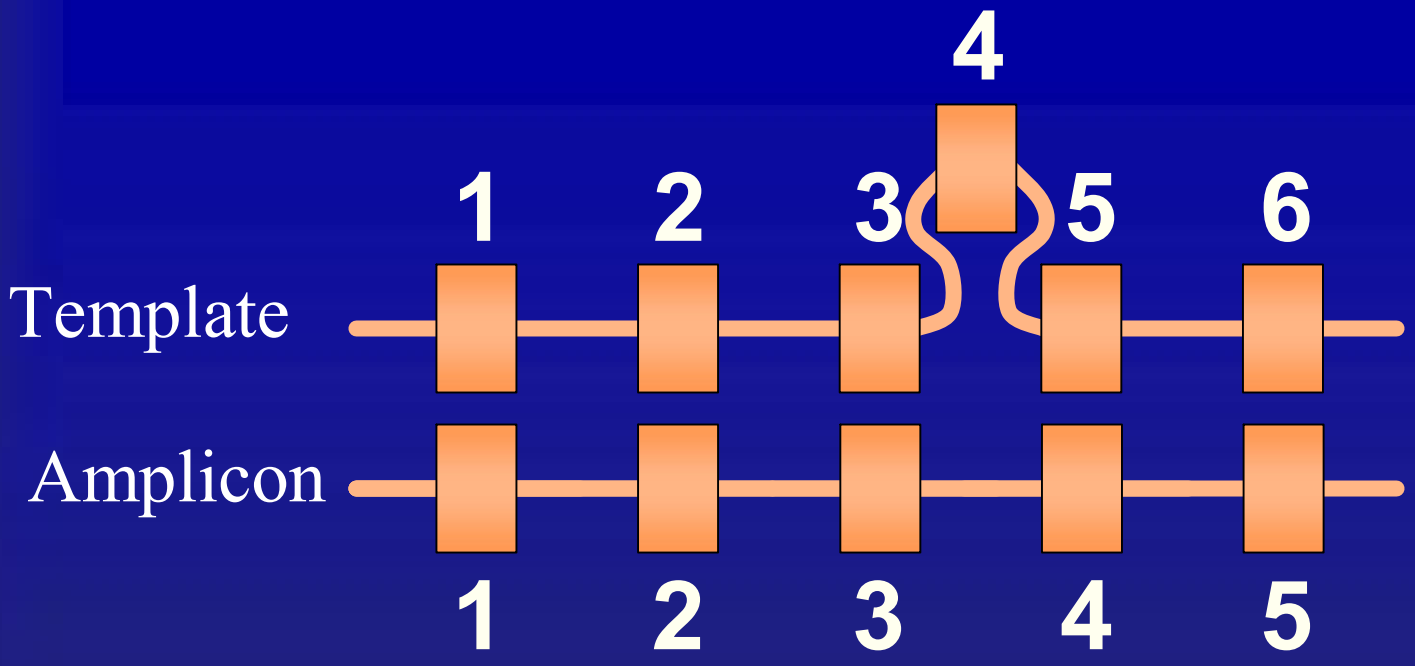
STUTTER

- Characteristic of repetitive DNA amplification
- Tetramers have less stutter than dinucleotide repeats
- Usually one repeat unit less in size
- Increases with magnesium
- 0-14% of allele

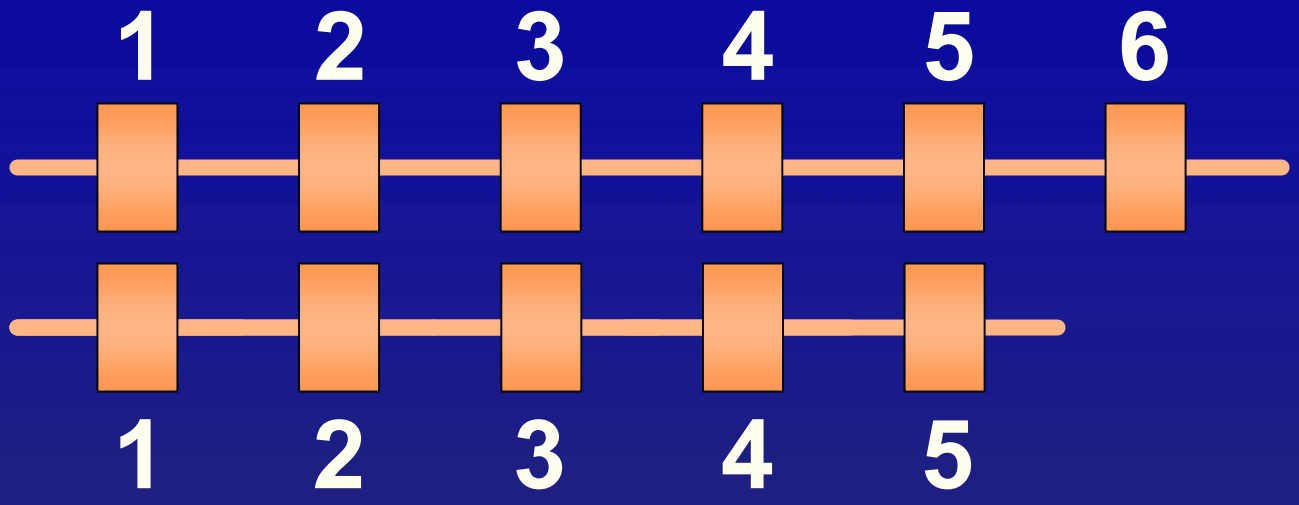
HOW STUTTER OCCURS

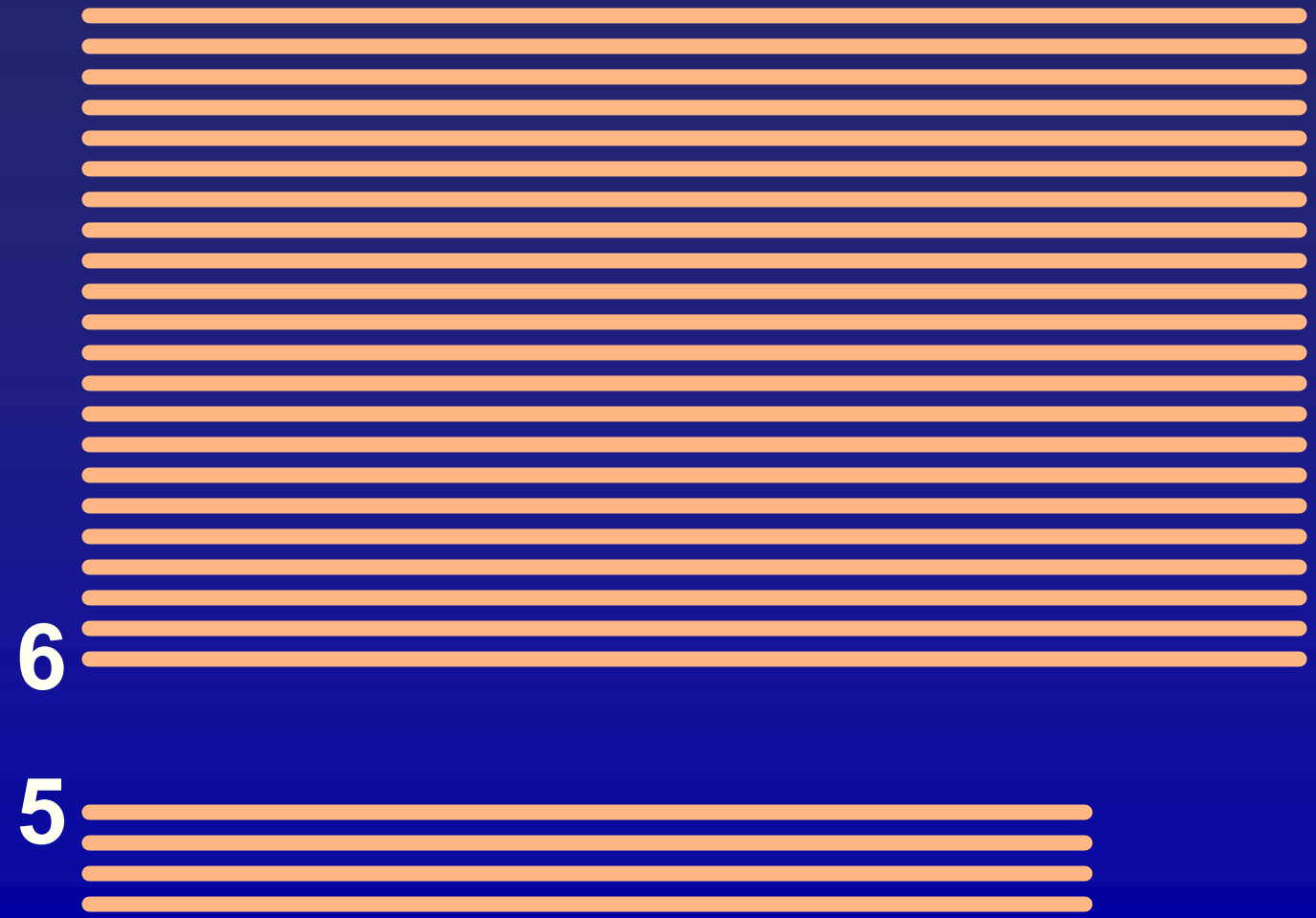


HOW STUTTER OCCURS



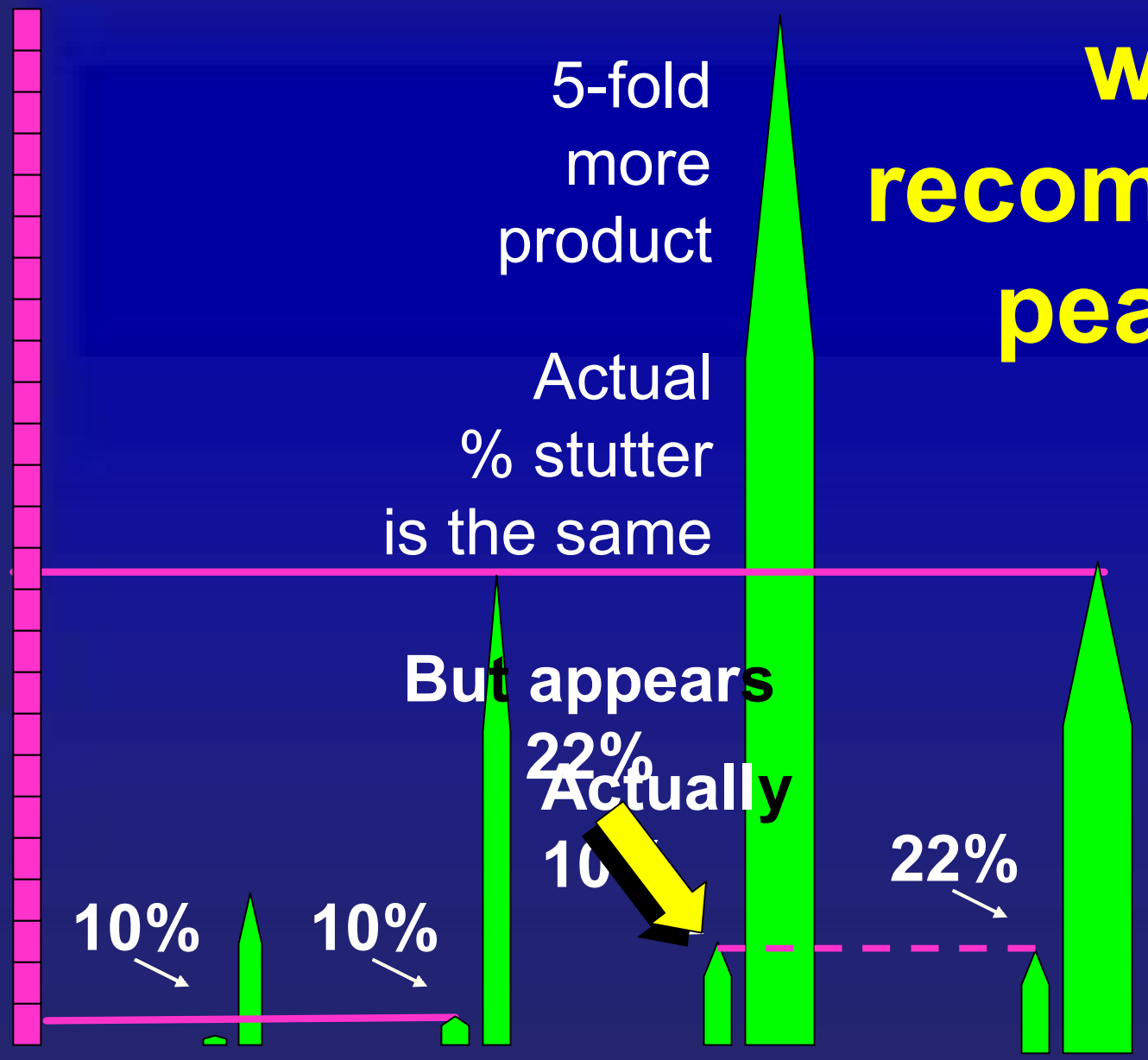
HOW STUTTER OCCURS





TRUE ALLELE PREDOMINATES

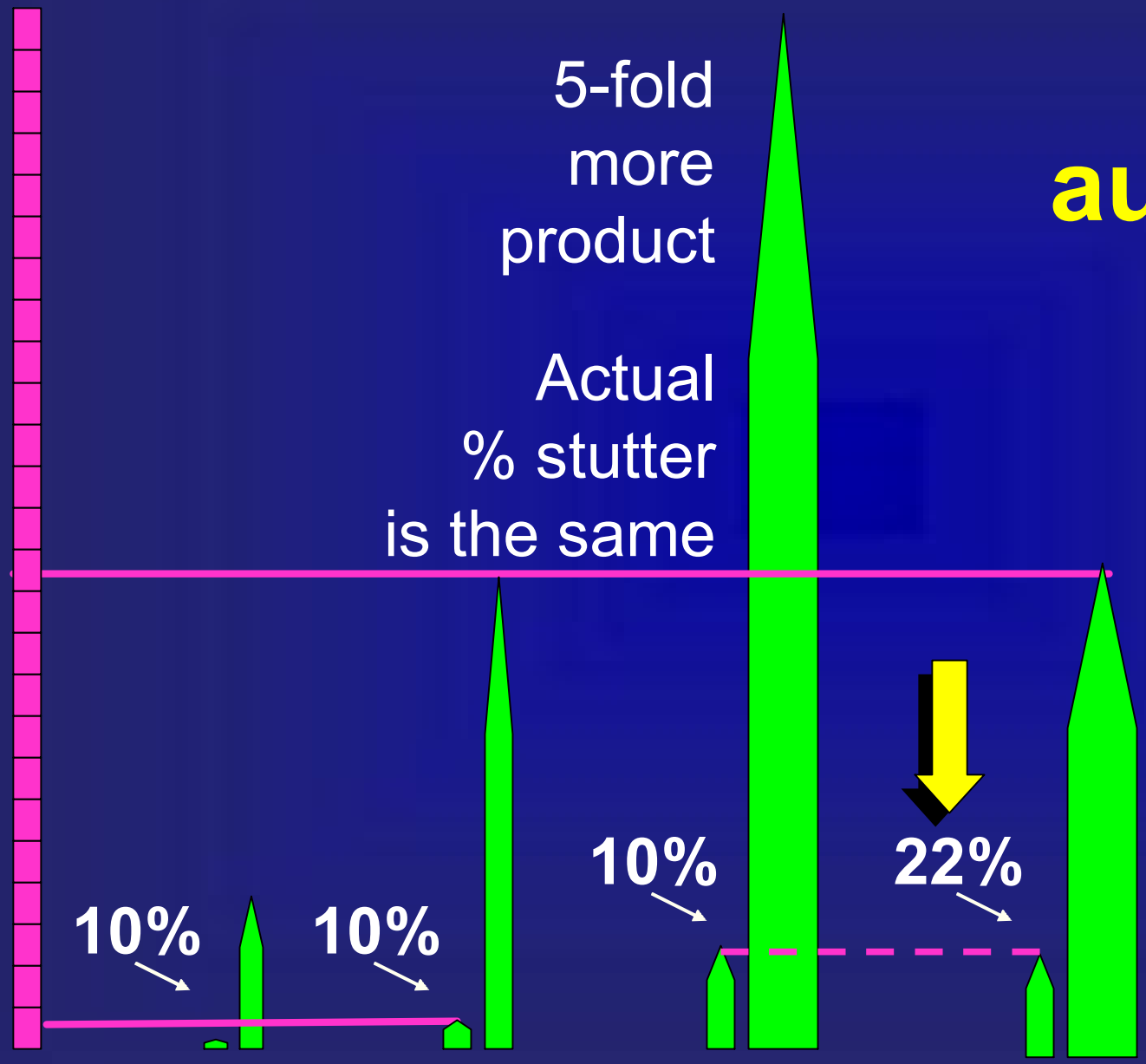
Minimize stutter by working within the recommended peak height range



Threshold of detection is exceeded for main peak

…but not for stutter peak

Percent stutter is falsely augmented



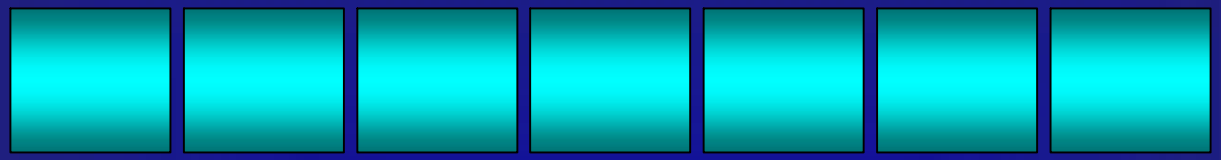
Threshold of detection is exceeded for main peak

…but not for stutter peak

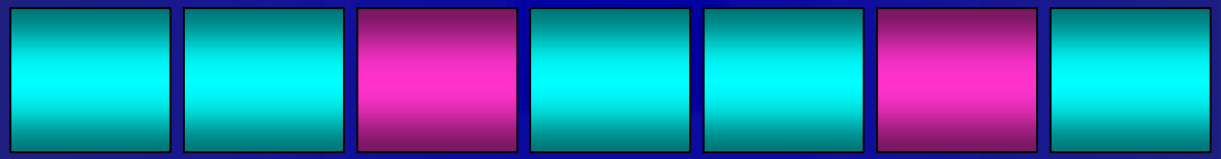
STUTTER TRENDS

- Different loci exhibit different degrees of stutter
- Loci with “compound” or “complex” structures exhibit less stutter than those with “simple” structure (e.g., VWA)
- Also variant alleles at a locus may have less stutter (e.g., VWA)

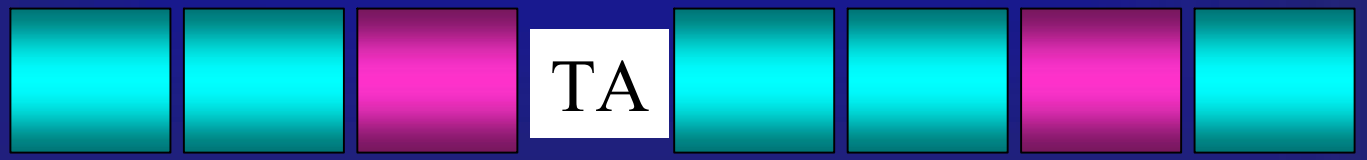
Simple:



Compound:



Complex:



Allele 16

16 $(TCTA)_2(TCTG)_4(TCTA)_{10}$

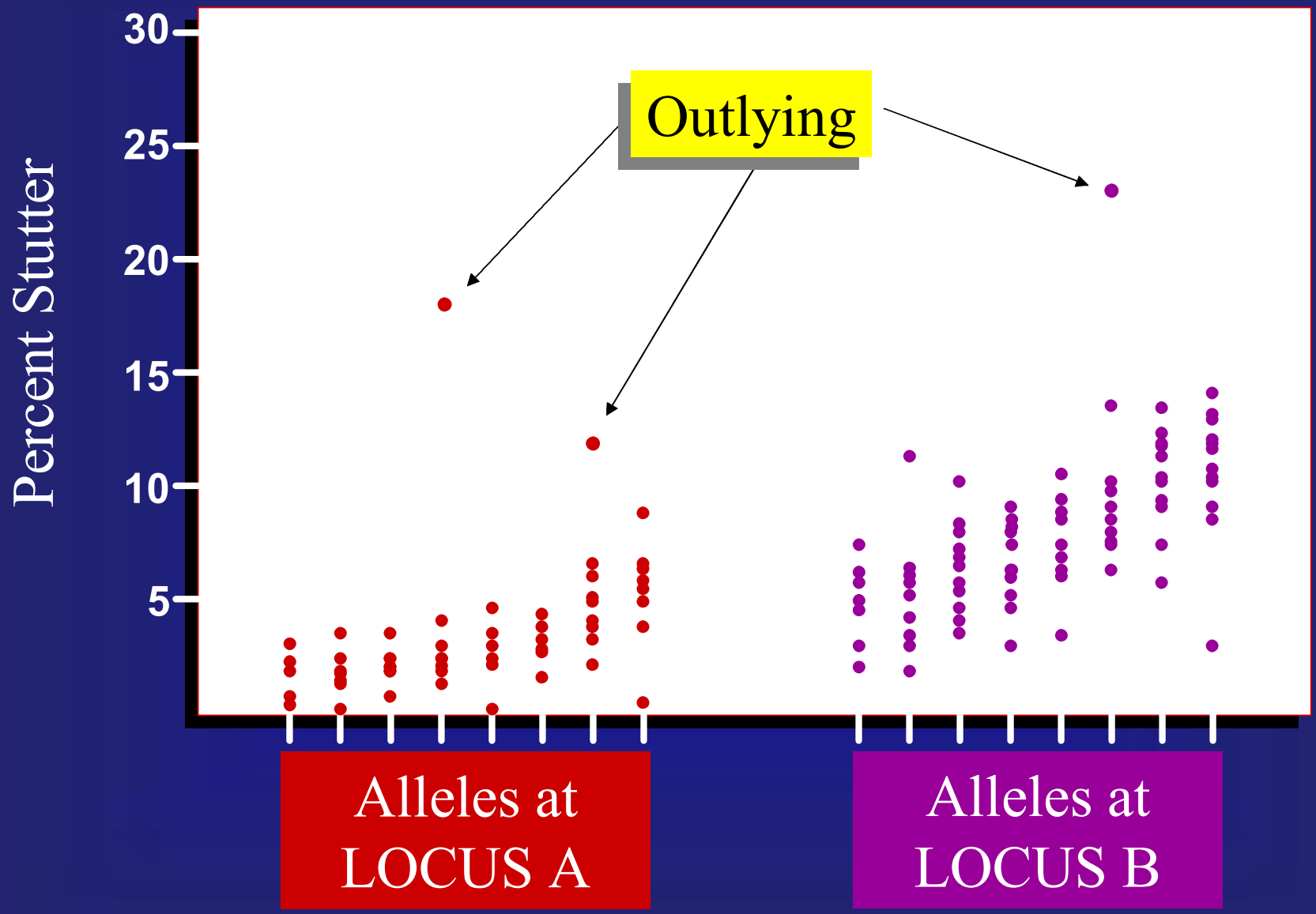
16 $(TCTA)_2(TCTG)_4(TCTA)_4(TCTG)(TCTA)_5$

Less prone to stutter

STUTTER TRENDS

- Different loci exhibit different degrees of stutter
- Loci with “compound” or “complex” structures exhibit less stutter than those with “simple” structure

- At any given locus, smaller alleles typically exhibit a lower percent stutter than larger alleles


Percent Stutter
30
25
20
15
10
5
Outlying
Alleles at LOCUS A
Alleles at LOCUS B

STUTTER TRENDS

- Different loci exhibit different degrees of stutter
- Loci with “compound” or “complex” structures exhibit less stutter than those with “simple” structure
- At any given locus, smaller alleles typically exhibit a lower percent stutter than larger alleles

- General guidelines

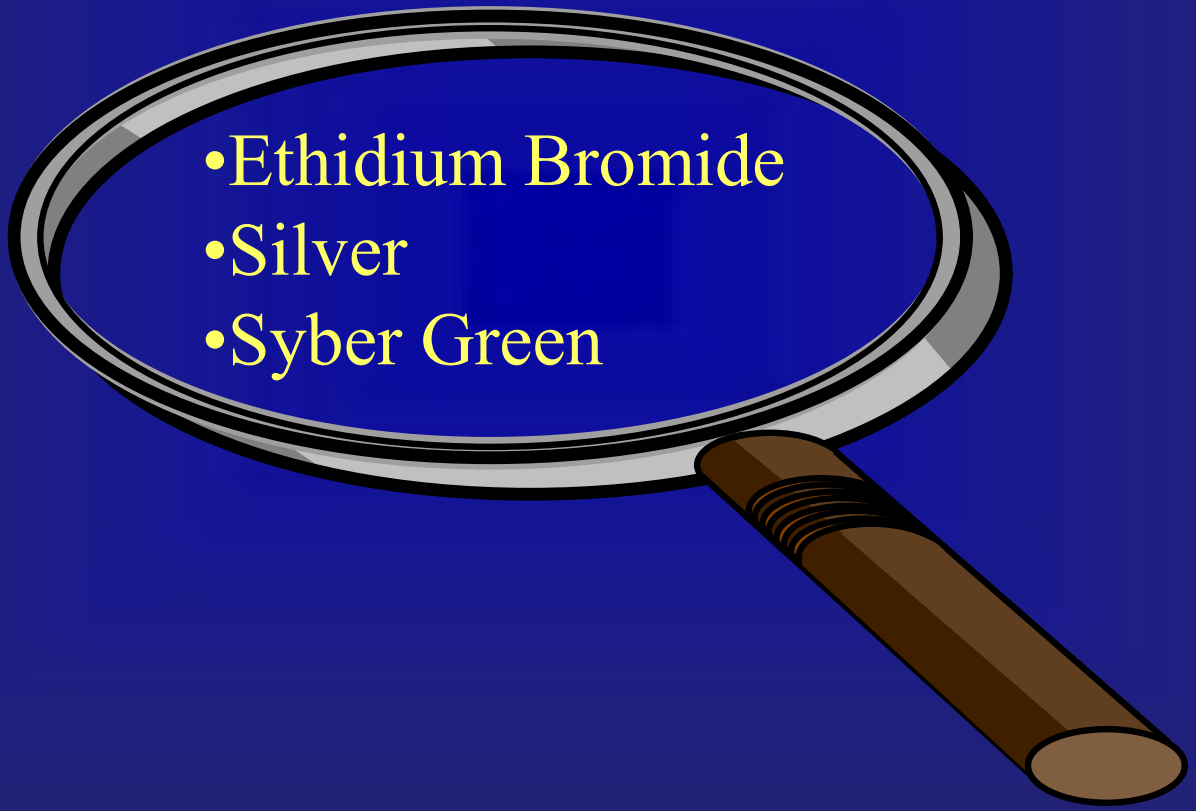
The presence of stutter or -A is not necessarily a problem

- Stutter is a feature of STR amplification
- Low-level stutter is manageable, when peaks are within recommended range
- Mixture interpretation
- Samples with -A can be extended
- If +A failure occurs, both alleles at a heterozygous locus tend to exhibit -A
- Adenylation failure does not preclude correct enotyping

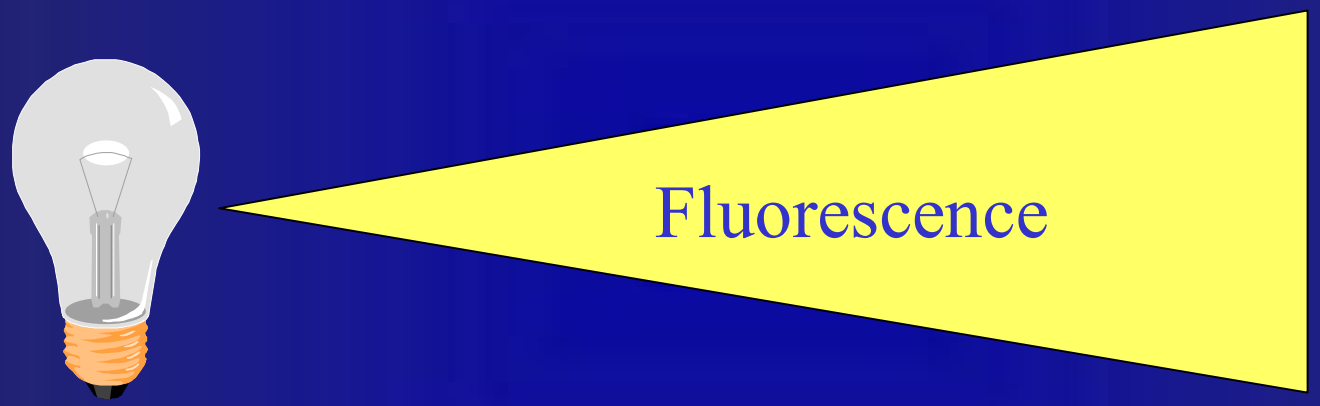
Resolution
Typing

- SA-32 Gel Apparatus, FMBIO
  - Scan gel after run
- ABI Prism 377, 310, 3100, 3700,…
  - Detection window

Detection Methods

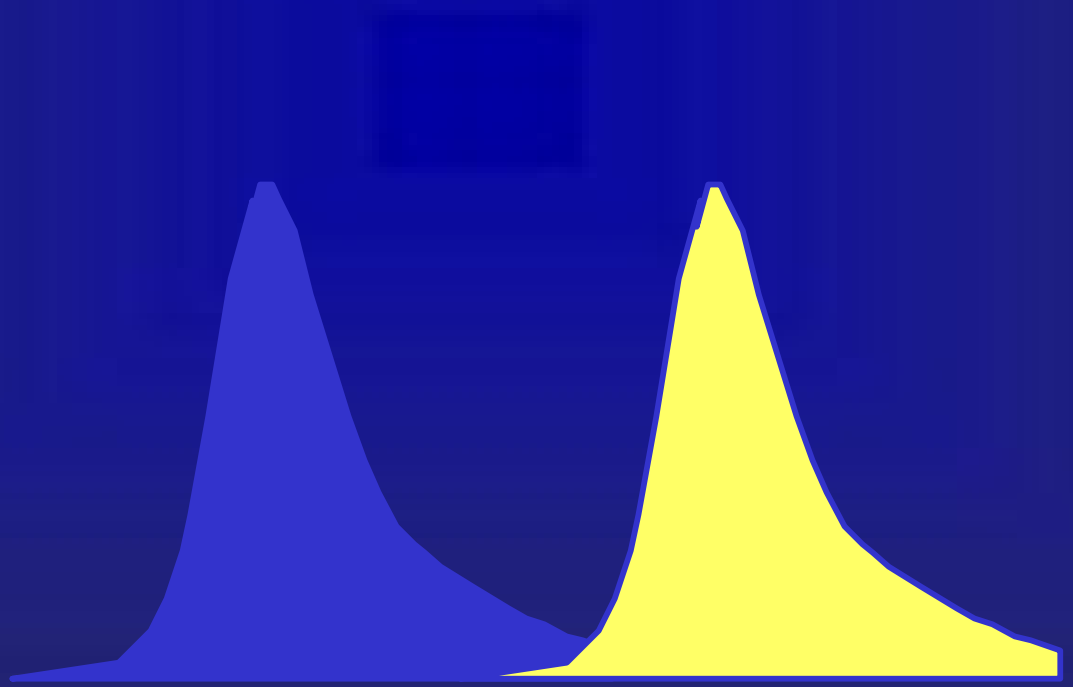


Detection Methods



Label one of the primers for each locus

Fluorescence results when a fluorophore or fluorescent dye absorbs incident light and in response emits light at a different wavelength

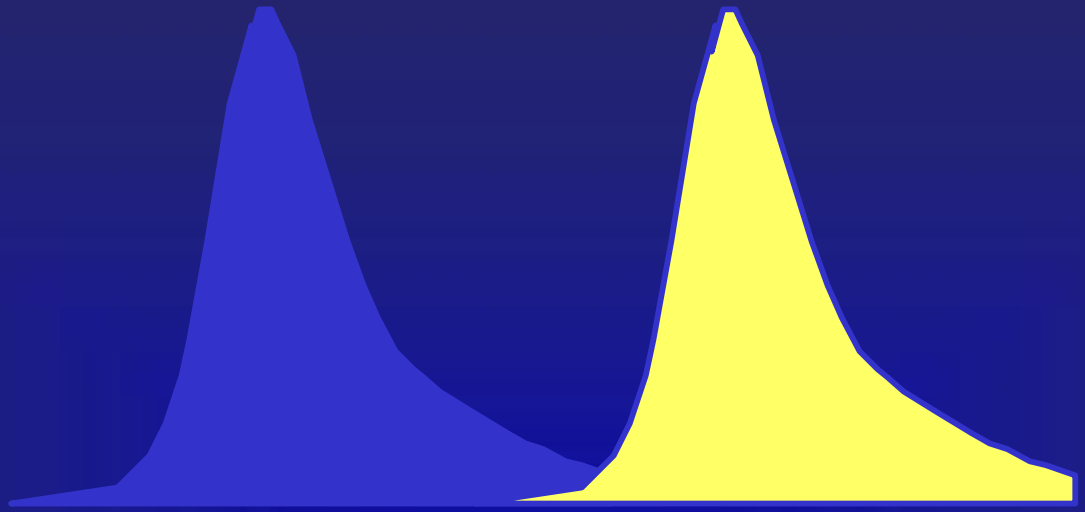


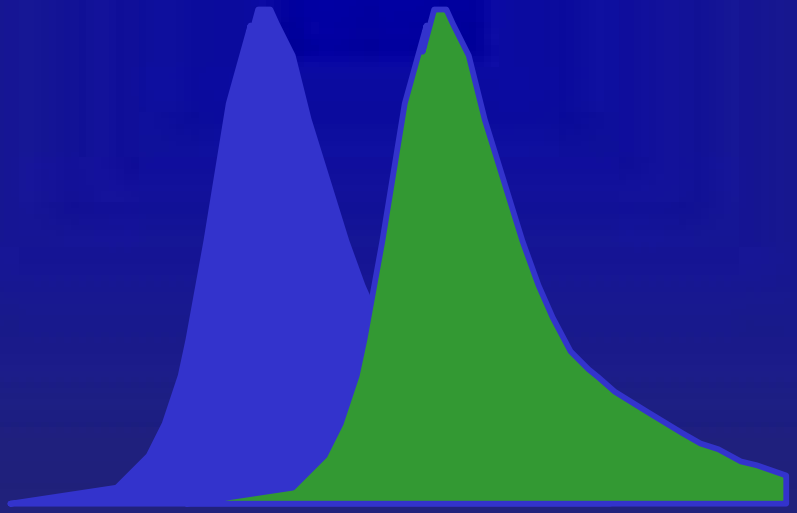
Molar Extinction Coefficient is

a measure of a dye's ability to absorb light

Stokes Shift is

The difference in wavelength between the fluorescence excitation maximum and fluorescence emission maximum


excitation
emission


excitation
emission

Spectral Criteria

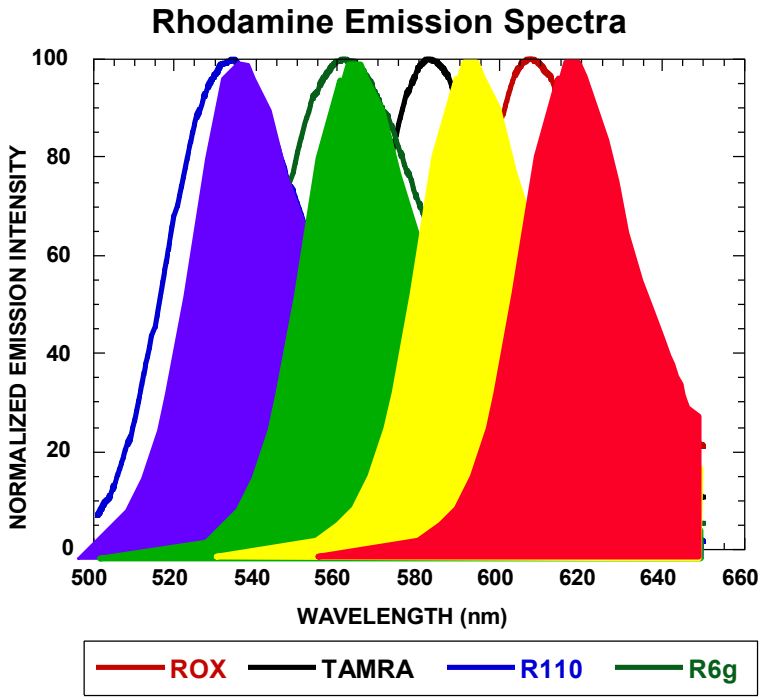
The dye's spectral excitation and spectral emission wavelengths should be compatible with excitation source and filter wavelengths

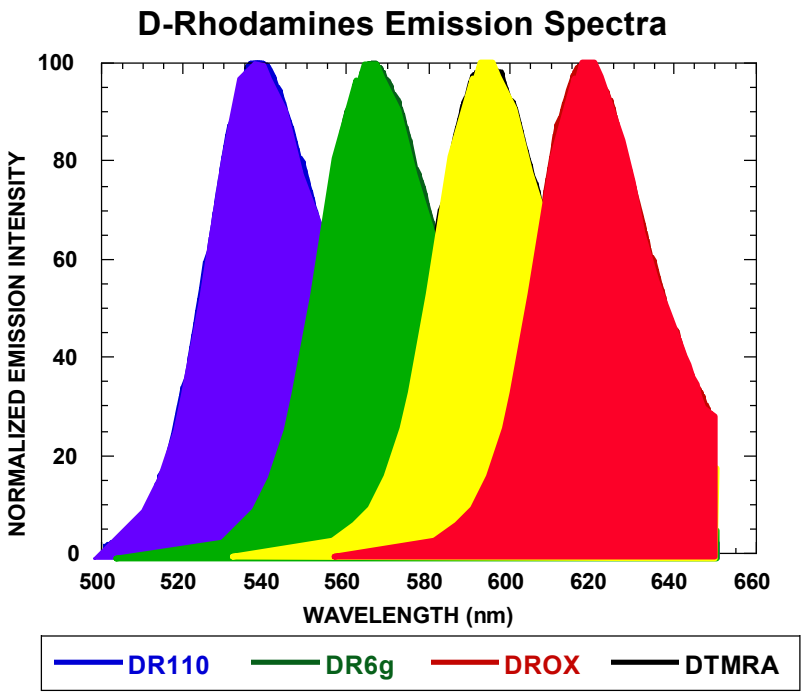
Spectral Bandwidth

Spectral overlap and the ability of an instrument to distinguish one dye from another

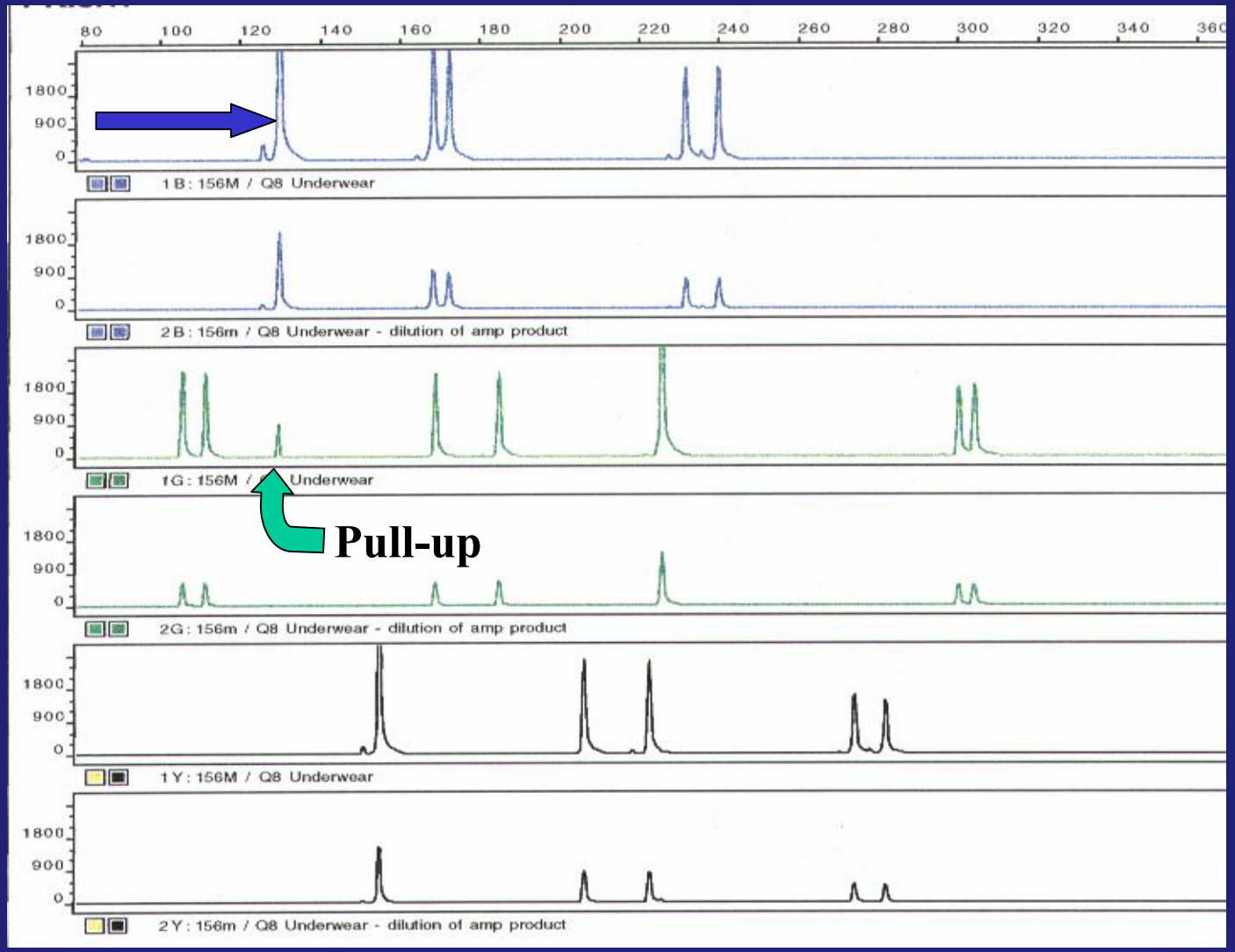
PULL-UP

Comparison of Dichlororhodamine and Rhodamine dye spectra





Pull-up



Due to signal and spectral overlap

Fluorescence Quantum Yield is

The efficiency with which the excited molecule is able to convert absorbed light to emitted light

Dyes also can affect
Mobility of DNA molecules

SUMMARY

- STRs used today in forensics generally are characterized by tandem repeats of 4(or 5)-bp units
- STR polymorphism is due to variation in the number of repeated units
- There are different repeat unit sequences (AATG, TCTA, CTTT, etc.)
- STR loci are either simple, compound or complex

SUMMARY

- Alleles are designated by comparison with an allelic ladder
- Off-ladder variants of tetramers usually have 1, 2, or 3 extra bases within the repeat array
- Off-ladder variants also can occur above and below the extremities of the ladder

SUMMARY

- STR loci "associated with" genes are typically in non-coding regions (i.e., introns or flanking regions)
- Many STR loci are "anonymous," or not currently known to reside within a known gene

SUMMARY

- STRs are amplified by the PCR. PCR primers bind to unique sequences that flank the repeat region
- Alleles that align with the allelic ladder have “plus A”
- “Minus A” may occur at some loci but is easily addressed

SUMMARY

- Stutter is inherent to STR (tetramers) amplification of some loci
- Loci with a simple repeat structure tend to stutter more
- Loci vary in the extent to which they exhibit stutter
- Small alleles show less stutter than larger alleles
- Stutter is generally predictable and general guidelines can be used for profile interpretation

SUMMARY

- Interpreting electropherograms is faciltated, especially with software

THE END