10-13-2007

Evaluating Forensic DNA Evidence - A Day-Long Workshop

Dan E. Krane
Wright State University - Main Campus, dan.krane@wright.edu

William C. Thompson

Forensic DNA profiling workshop

Dan E. Krane, Wright State University, Dayton, Ohio
William C. Thompson, University of California, Irvine, CA

Forensic Bioinformatics
(www.bioforensics.com)

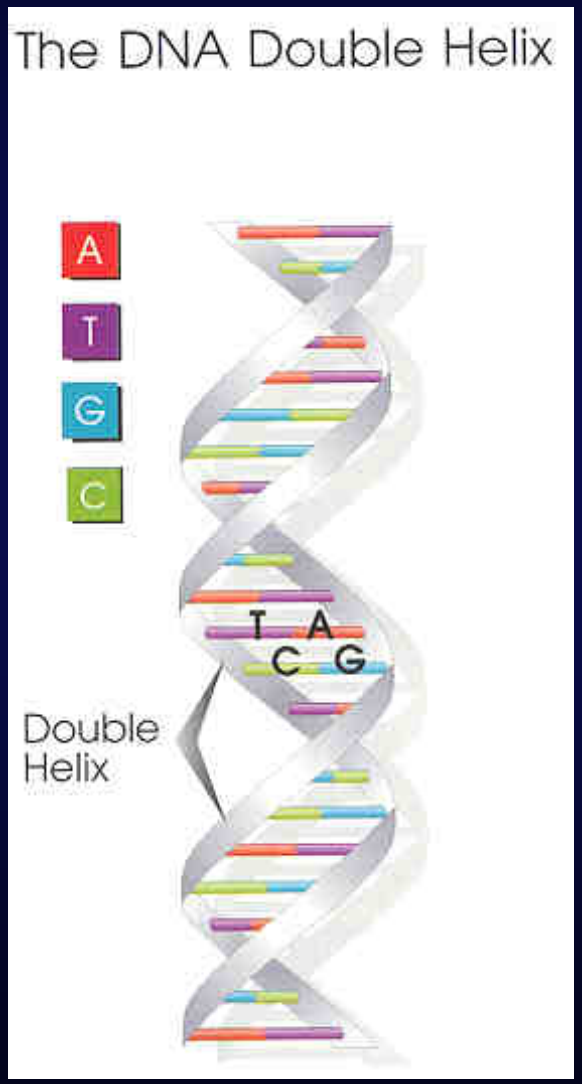
I: Overview of what DNA tests can do for:

A. Prosecution

B. Defense

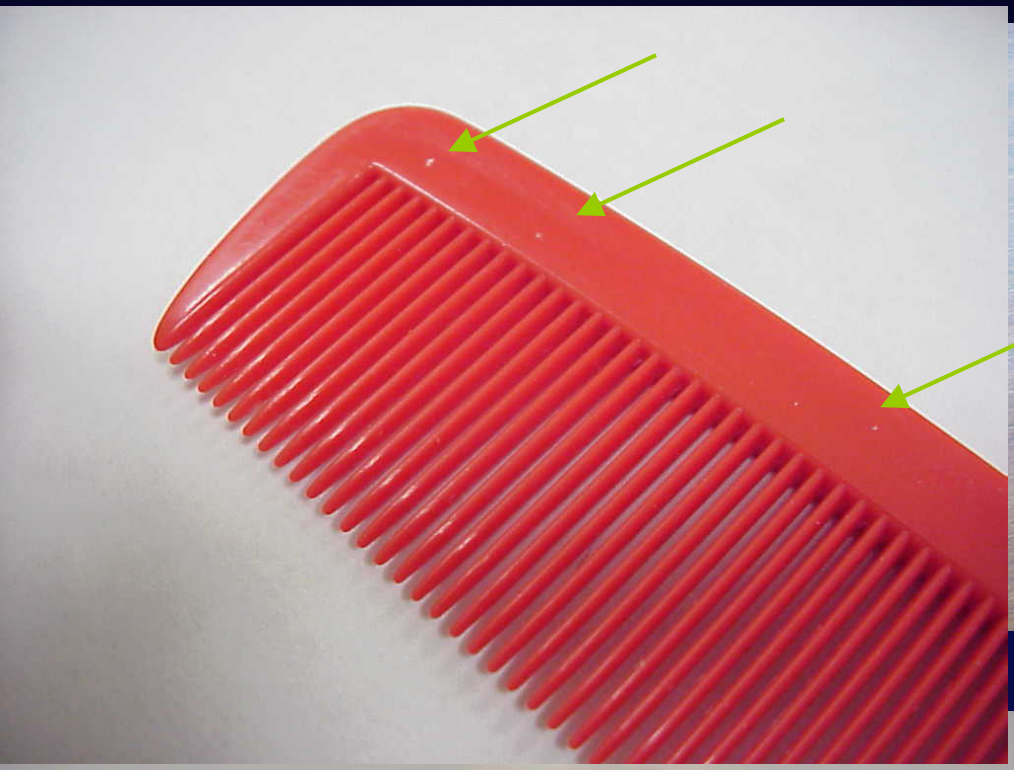
C. Post-conviction testing

DNA Technology in Court



- Criminal Prosecution
  - Unprecedented sensitivity and specificity for typing biological samples
  - Growing use of databanks and dragnets to identify suspects
  - Rapidly becoming cheaper and faster

Possible DNA Sources



DNA Technology in Court

- Criminal Defense
  - Unprecedented sensitivity and specificity for typing biological samples
  - Potential support for alternative theories of the case

DNA Technology in Court

- Post-conviction exonerations (208 in US) based on DNA evidence have revealed problems with the justice system

Sources of Error

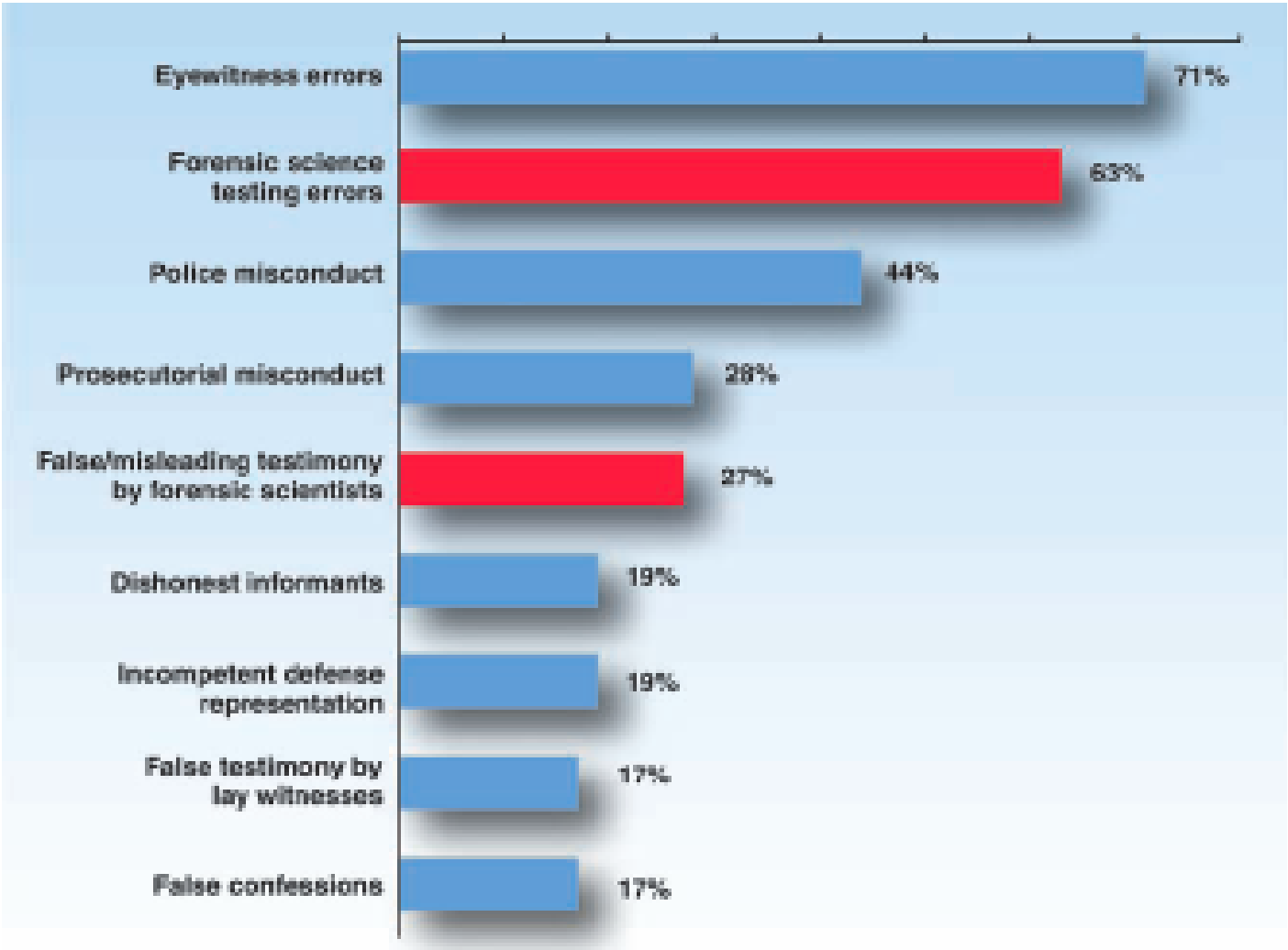
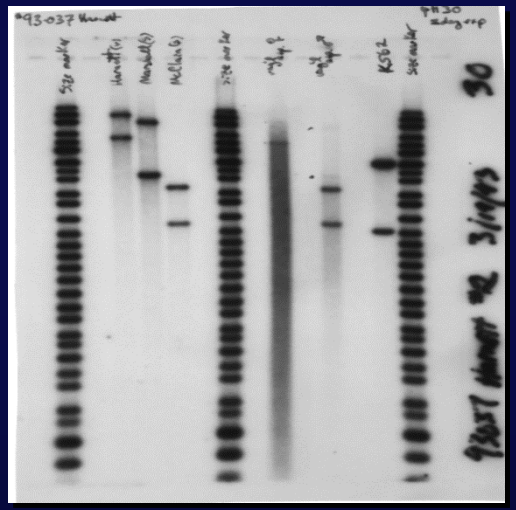


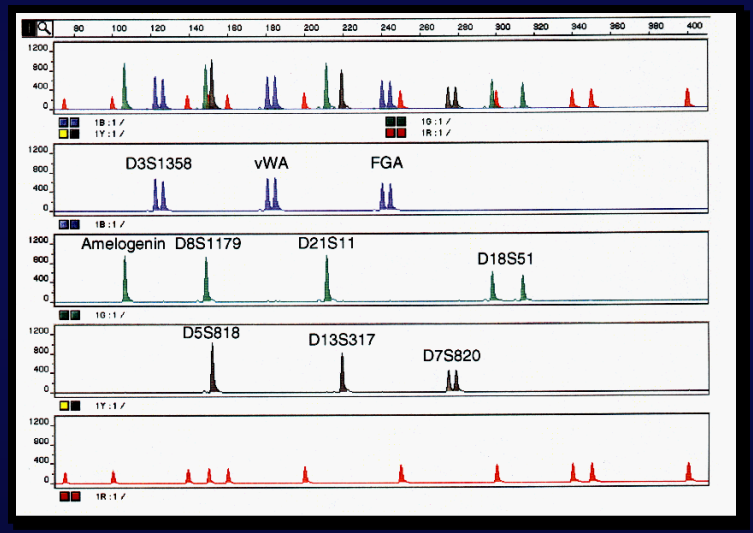
Fig. 1. Factors associated with wrongful conviction in 86 DNA exoneration cases, based on case analysis data provided by the Innocence Project, Cardozo School of Law (New York, NY), and computed by us. Percentages exceed 100% because more than one factor was found in many cases. Red bars indicate factors related to forensic science.

- Saks & Koehler, *Science* (2005)

II: The evolution of DNA technology

Three generations of DNA testing



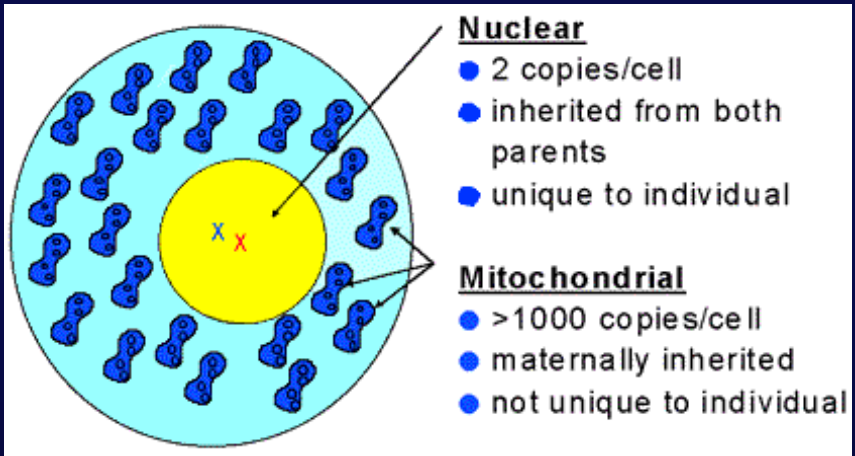


RFLP
AUTORAD
Allele = BAND

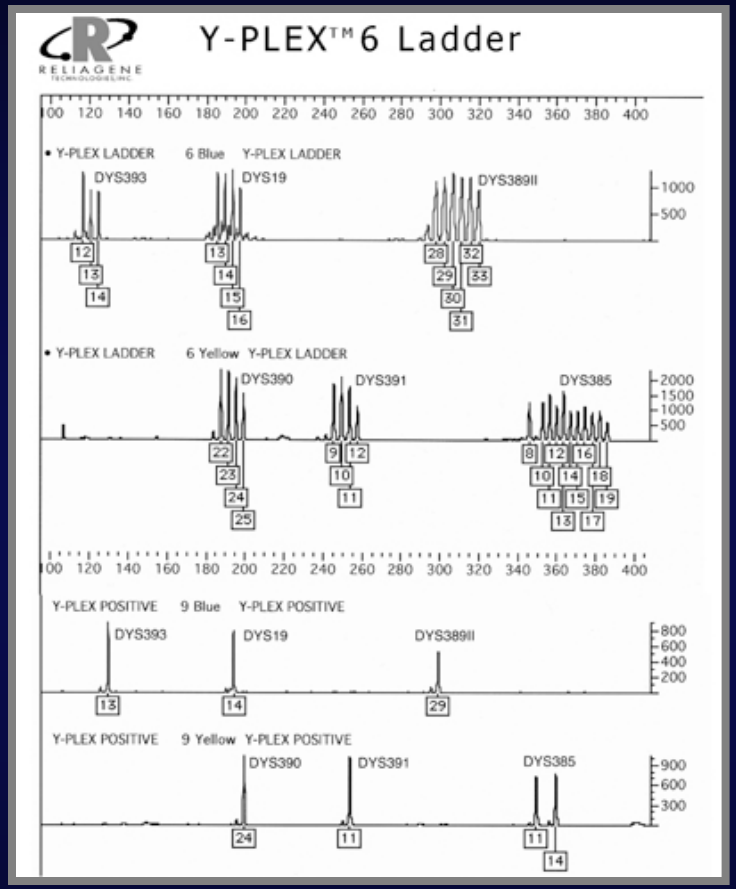
DQ-alpha
TEST STRIP
Allele = BLUE DOT

Automated STR
ELECTROPHEROGRAM
Allele = PEAK

Two relatively new DNA tests

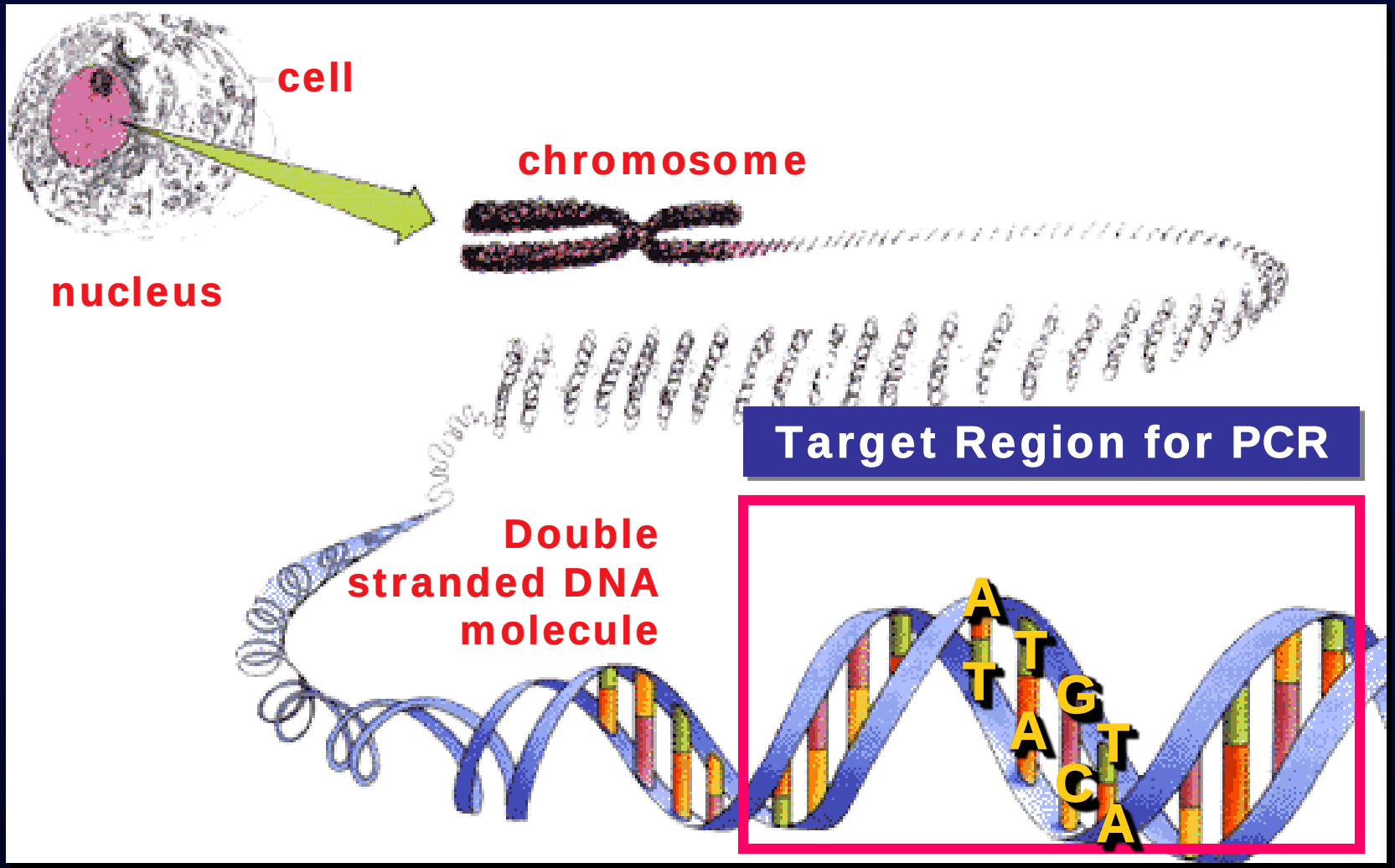


Mitochondrial DNA
mtDNA sequence
Sensitive but not discriminating



Y-STRs
Useful with mixtures
Paternally inherited

DNA in the Cell



DNA content of biological samples:

Type of sample	Amount of DNA
Blood	30,000 ng/mL
stain 1 cm^2 in area	200 ng
stain 1 mm^2 in area	2 ng
Semen	250,000 ng/mL
Postcoital vaginal swab	0 - 3,000 ng
Hair	
plucked	1 - 750 ng/hair
shed	1 - 12 ng/hair
Saliva	5,000 ng/mL
Urine	1 - 20 ng/mL

Basic terminology: Genetics

- DNA Polymorphism (“many forms”)
  - Regions of DNA which differ from person to person
- Locus (plural = loci)
  - Site or location on a chromosome
- Allele
  - Different variants which can exist at a locus
- DNA Profile
  - The combination of alleles for an individual

Basic terminology: Technology

- Amplification or PCR (Polymerase Chain Reaction)
  - A technique for ‘replicating’ DNA in the laboratory (‘molecular Xeroxing’)
  - Region to be amplified defined by PRIMERS
  - Can be ‘color coded’
- Electrophoresis
  - A technique for separating molecules according to their size

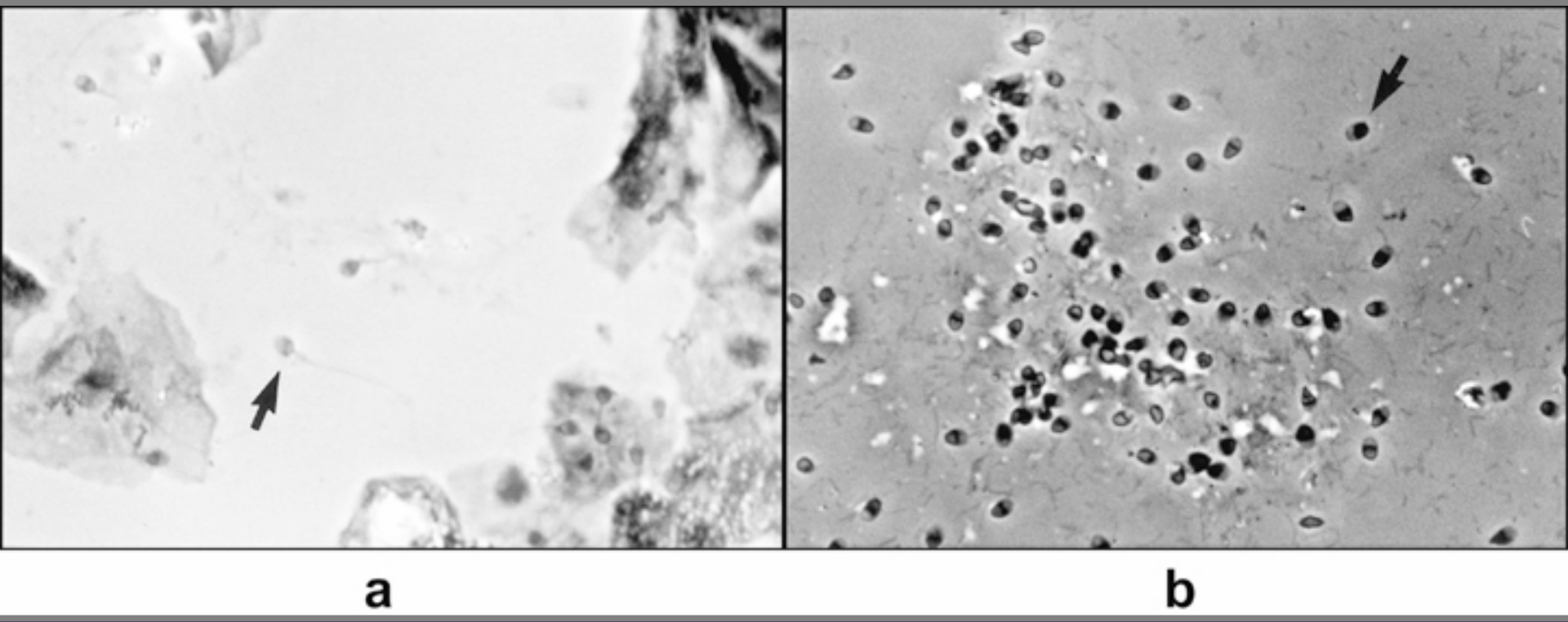
Automated STR Test

Crime Scene Samples & Reference Samples

- Extract and purify DNA

Differential extraction in sex assault cases separates out DNA from sperm cells

Differential Extraction of Semen Stain



Female Extract

Male Extract

Graphic from Inman & Rudin, An Introduction fo Forensic DNA Analysis. CRC Press.

Extract and Purify DNA

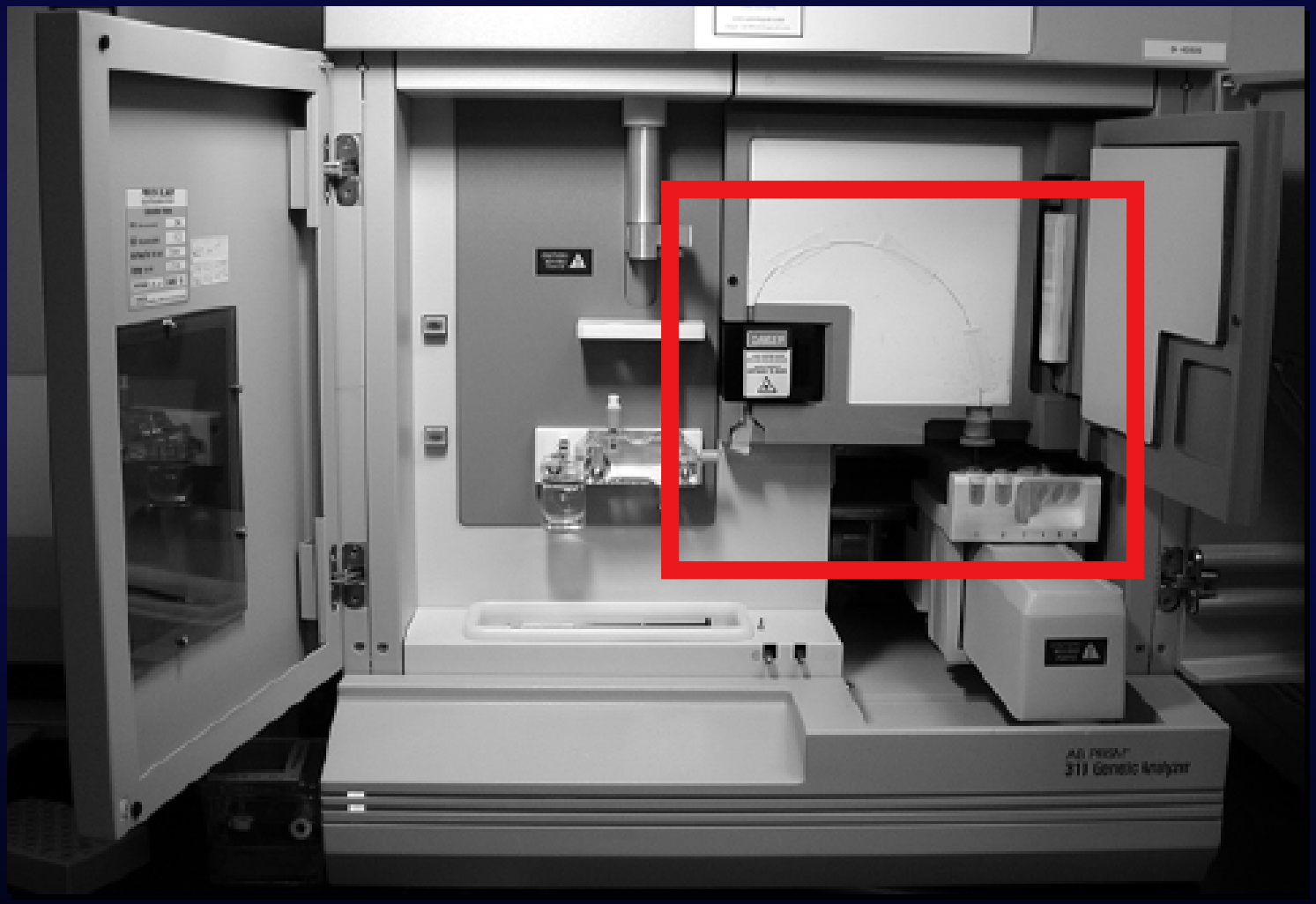
- Add primers and other reagents

PCR Amplification

- DNA regions flanked by primers are amplified

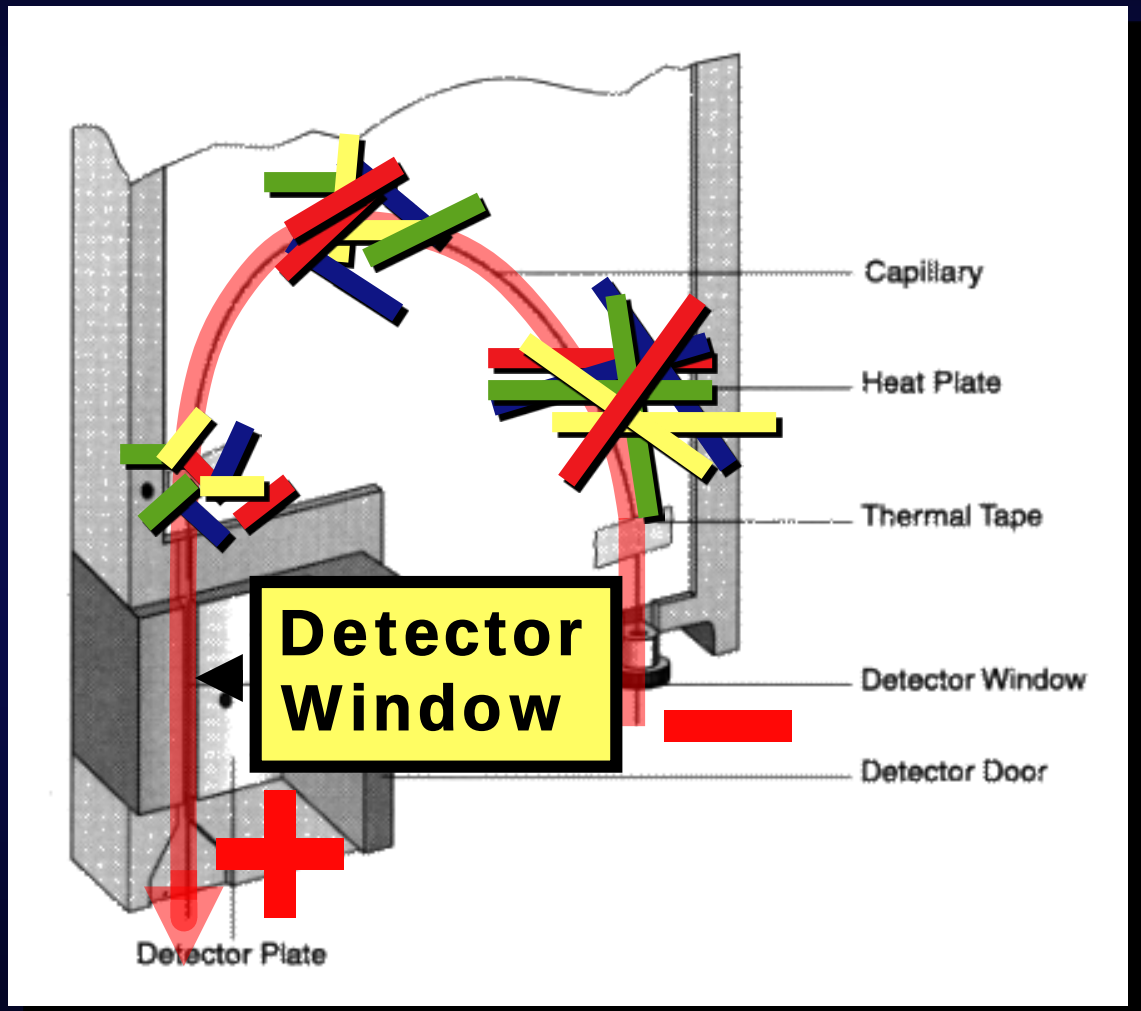
Groups of amplified STR products are labeled with different colored dyes (blue, green, yellow)

The ABI 310 Genetic Analyzer

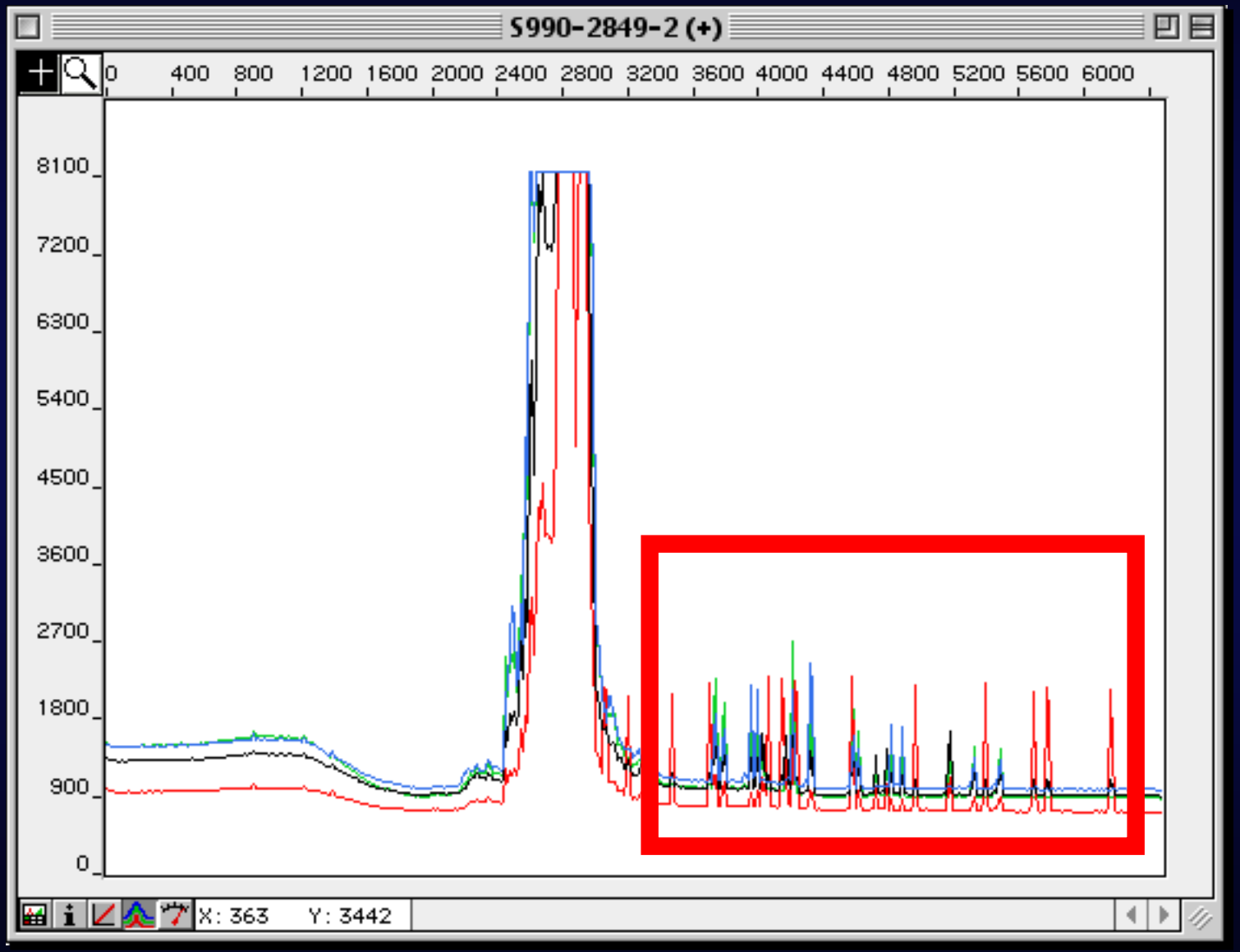


ABI 310 Genetic Analyzer: Capillary Electrophoresis

- Amplified STR DNA injected onto column
- Electric current applied
- DNA pulled towards the positive electrode
- DNA separated out by size:
  - Large STRs travel slower
  - Small STRs travel faster
- Color of STR detected and recorded as it passes the detector



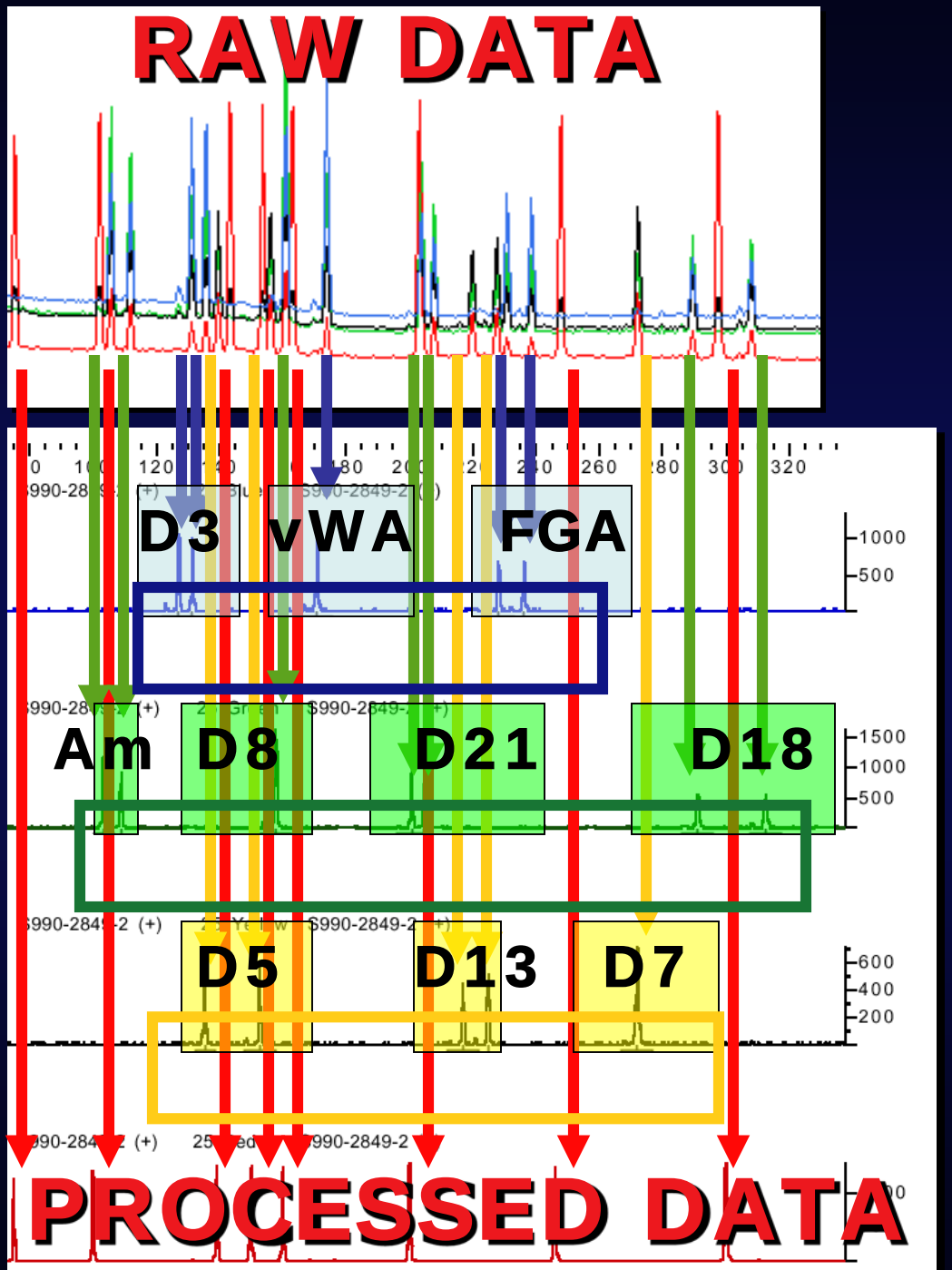
Profiler Plus: Raw data



- GENESCAN divides the raw data into a separate electropherogram for each color:
  - Blue
  - Green
  - Yellow
  - Red

- GENOTYPER identifies the different loci and makes the allele calls

- The type of this sample is:
  - D3: 16, 17
  - vWA: 15, 15
  - FGA: 21,23
  - Amelogenin: X, Y
  - D8: 16, 16
  - D21: 28, 29
  - D18: 14, 19
  - D5: 8, 12
  - D13: 11, 13
  - D7: 10 10



STR

- Short tandem repeat
- Describes a type of DNA polymorphism in which:
  - a DNA sequence repeats
  - over and over again
  - and has a short (usually 4 base pair) repeat unit
- A length polymorphism -- alleles differ in their length

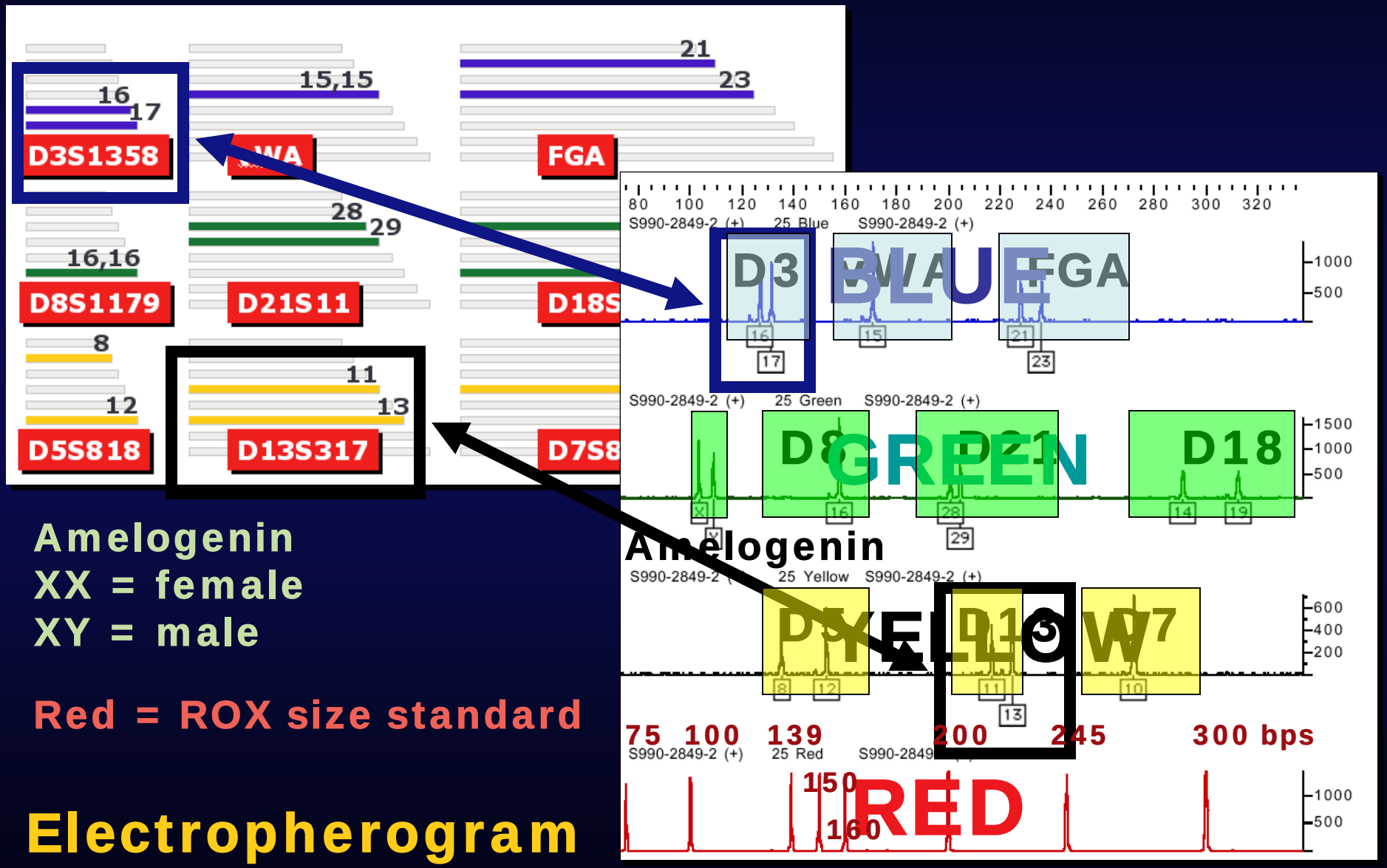
3 repeats: AATG AATG AATG

4 repeats: AATG AATG AATG AATG

5 repeats: AATG AATG AATG AATG AATG

6 repeats: AATG AATG AATG AATG AATG AATG

Reading an electropherogram
Peaks correspond to alleles



Amelogenin
XX = female
XY = male

Red = ROX size standard

Electropherogram

Reading an electropherogram

NUMBER OF PEAKS

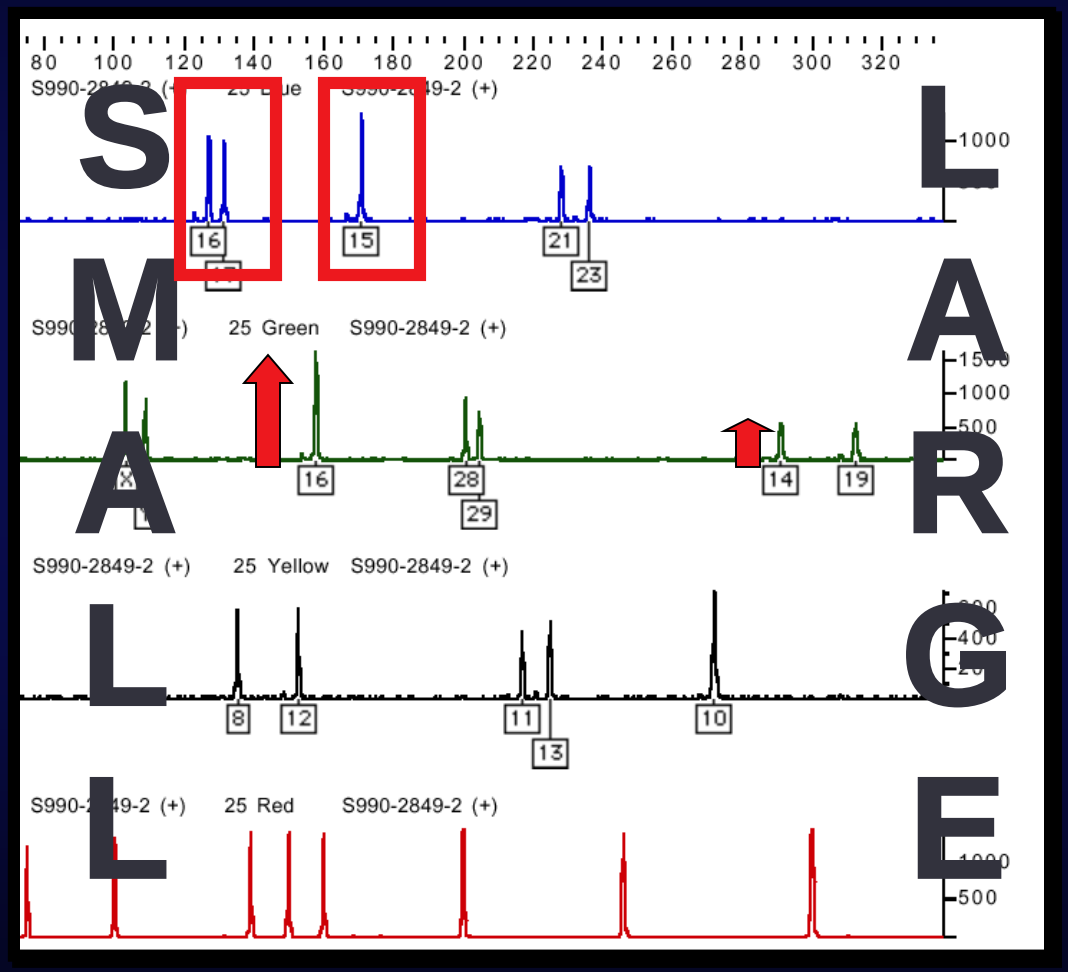
- 1 peak = homozygous
- 2 peaks = heterozygous
- 3 or more peaks = mixed sample (?)

POSITION OF PEAK

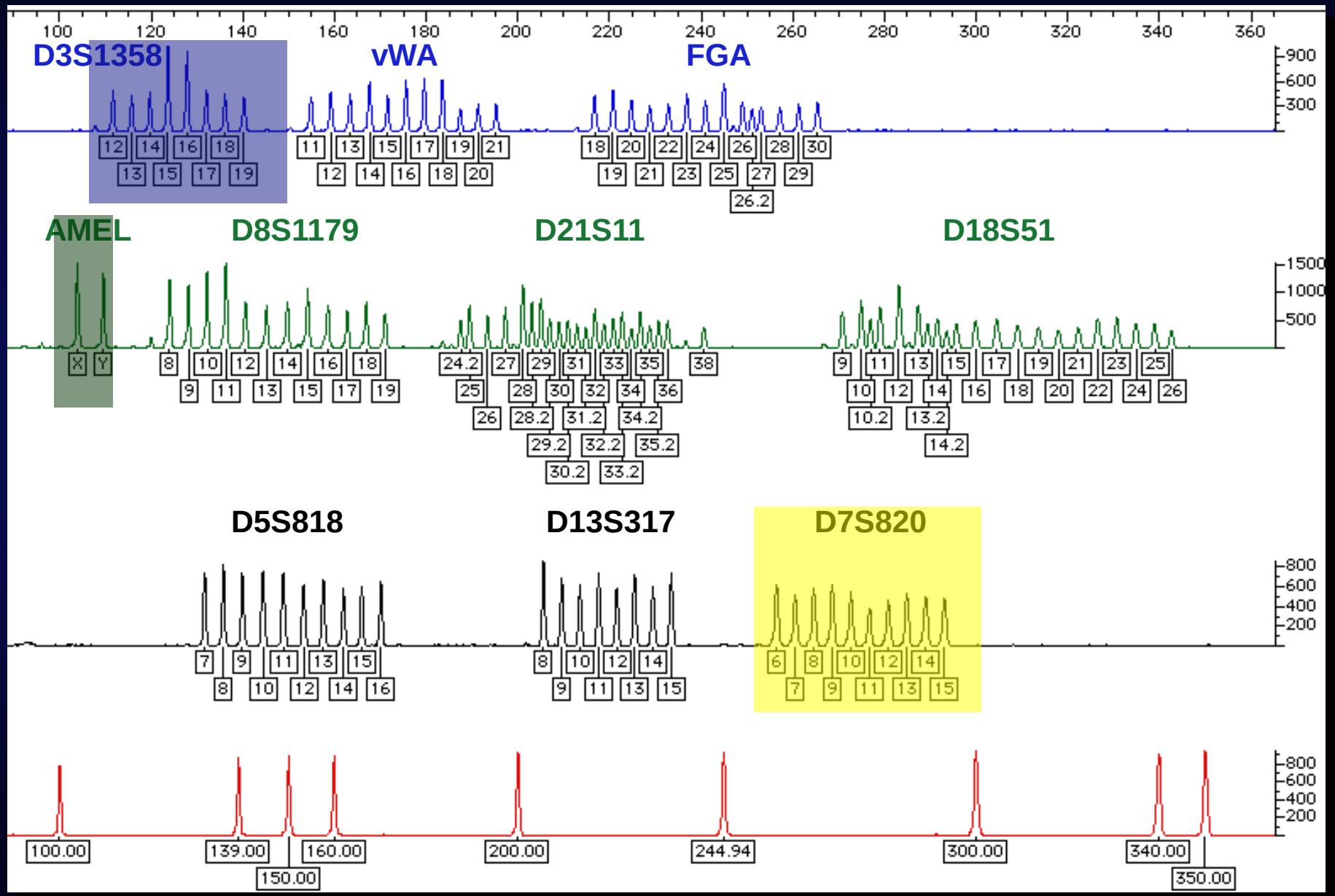
- Smaller alleles on left
- Larger alleles on right

HEIGHT OF PEAK

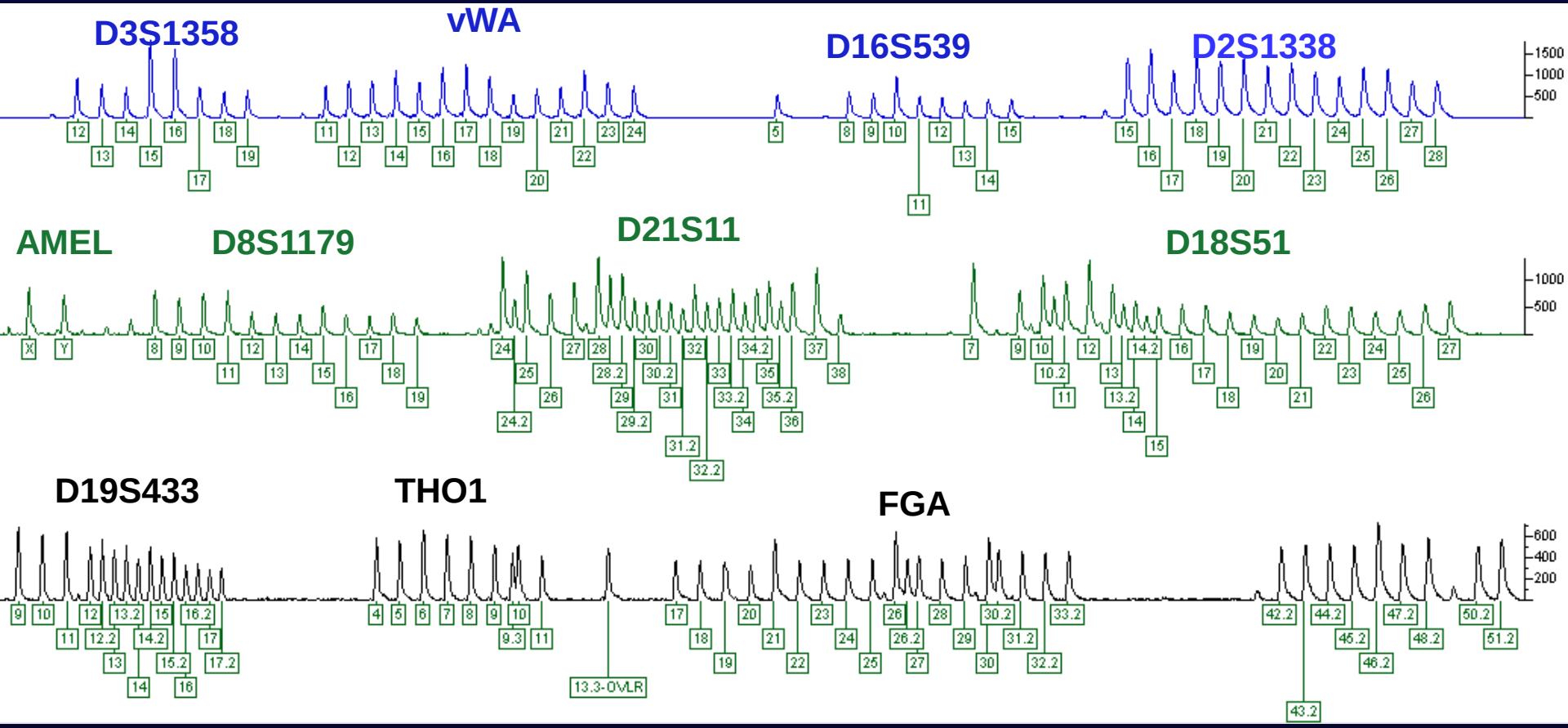
- Proportional to amount of allele (approx)



Profiler Plus



SGM+


D3S1358
vWA
D16S539
D2S1338
AMEL
D8S1179
D21S11
D18S51
D19S433
THO1
FGA
1500
1000
500
1000
500
600
400
200

Statistical estimates: the product rule

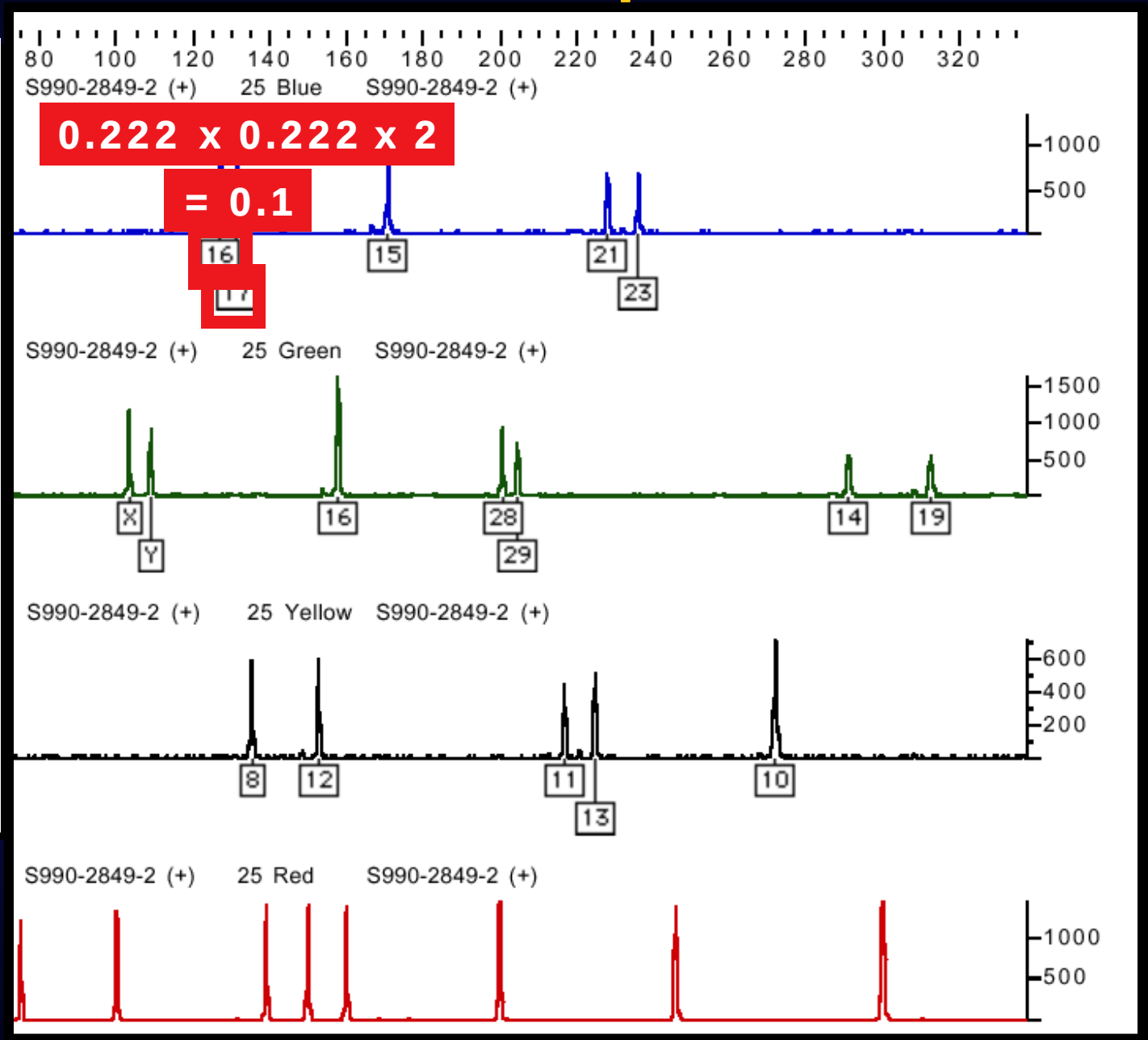
Allele Frequencies

Locus D3S1358
Race Caucasian
(N = 203)

Allele	Frequency
12	0.012
13	0.012
14	0.140
[redacted]	[redacted]
19	0.012

Locus vWA
Race Caucasian
(N = 196)

Allele	Frequency
11	0.012
12	0.012
13	0.012
14	0.102
15	0.082



Statistical estimates: the product rule

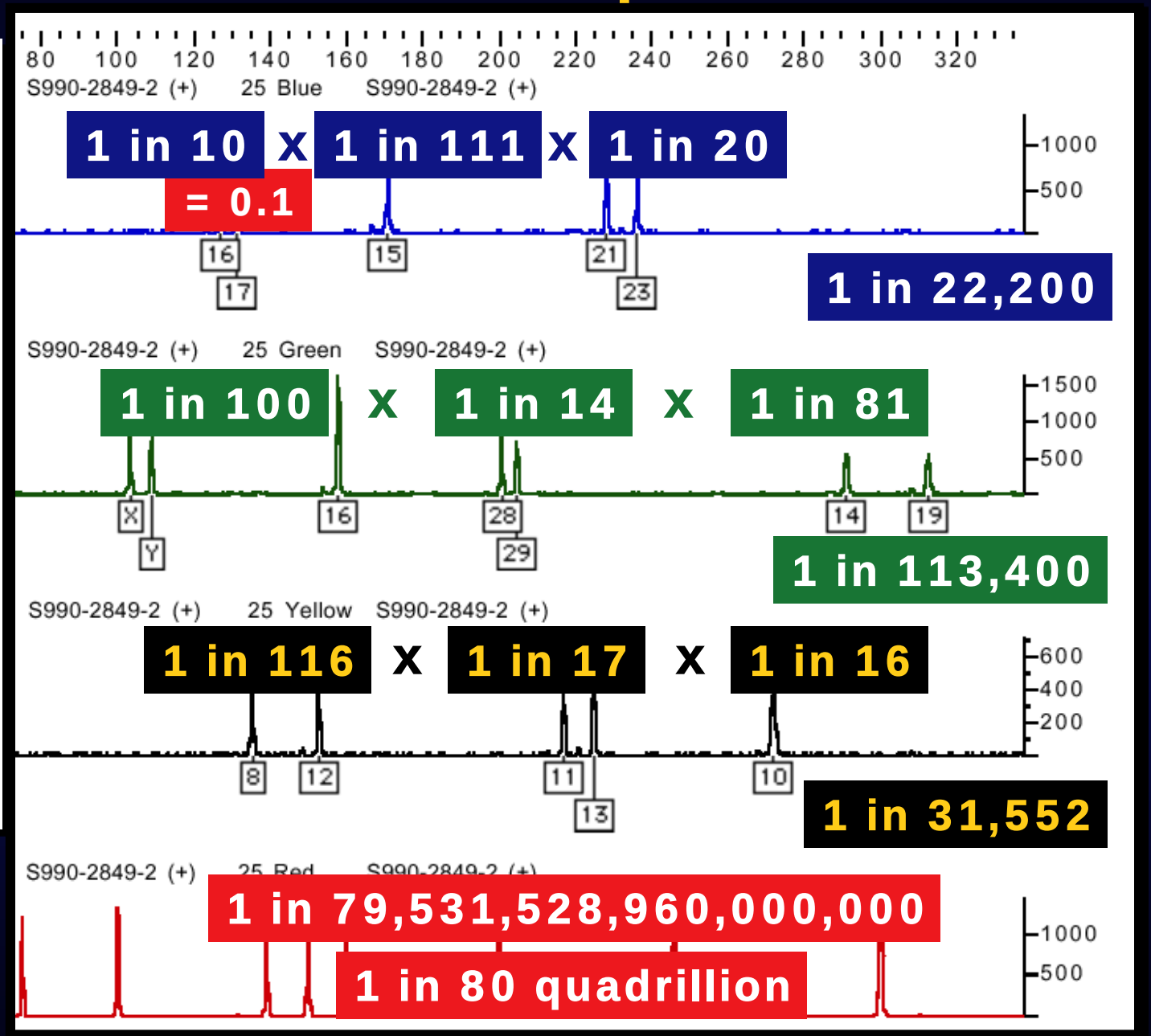
Allele Frequencies

Locus D3S1358
Race Caucasian
(N = 203)

Allele	Frequency
12	0.012
13	0.012
14	0.140
15	0.246
16	0.222
17	0.222
18	0.163
19	0.012

Locus vWA
Race Caucasian
(N = 196)

Allele	Frequency
11	0.012
12	0.012
13	0.012
14	0.102
15	0.082



What more is there to say after you have said: “The chance of a coincidental match is one in 80 quadrillion?”

What more is there to say after you have said: “The chance of a coincidental match is one in 80 quadrillion?”

- Two samples really do have the same source
- Samples match coincidentally
- An error has occurred

The Debate Over Statistics

- Initial assumptions of statistical independence
- Were challenged by academic experts and National Research Council (1992)
- Creating a controversy that led some courts to exclude DNA evidence under the Frye (general acceptance) standard
- The controversy prompted validation research and improved methods
- That were endorsed by the NRC (1996)

Statistical Fallacies

- The “prosecutor’s fallacy”
  - Equates frequency of matching DNA profile with probability suspect is “not the source” or probability “someone else” is the source
- The “defense attorney’s fallacy”
  - Assumes probability of guilt is 1/m, where m is number of matching profiles in some population

Coincidence or Crime?

- SF Chronicle Headline, Dec 12, 2002: Double lottery winners beat odds of 1 in 24,000,000,000,000

  "...they won the jackpot -- not once, but twice, on the same day. An hour after winning $126,000 in the Fantasy Five game, they won $17 million in SuperLotto Plus.

  That's never been done before, lottery officials said Wednesday, maybe because the odds of its happening are 1 in 24 trillion -- which is a 24 followed by 12 zeros."

- Does this episode prove the California lottery is a fraud?

Important terminology

- Frequency (F)—
  - the rate at which a profile occurs in some population
  - E.g., the *frequency* of this DNA profile among US Caucasians is 1 in 1 billion
- Random Match Probability (RMP)—
  - The *chance* that a randomly chosen, unrelated individual would have the same DNA profile as the evidence
  - RMP is what the jury needs to know
  - RMP is *not necessarily* the same as F

More terminology

- Single-Opportunity Search
  - Comparison process where there is one opportunity for a coincidental match
  - E.g., what is the probability *you* will share my birthday?
    - F = 1/365
    - RMP = 1/365
  - For a single opportunity search RMP = F
- Multiple-Opportunity Search
  - Comparison process where there is more than one opportunity for a coincidental match
  - E.g., What is the probability *someone in the room* will share my birthday?
  - F = 1/365
  - RMP = $1-(1-F)^N$ where N=number in the room

Even more terminology

- Birthday Problem
  - Multiple opportunities for a multiple opportunity search
  - What is the probability that *any two people* in the room will share a birthday?
    - F = 1/365
    - RMP > 1/2 when N>22; approaches certainty when N>60

Database Searches and the Birthday Problem

- Suppose the probability of a random match between any two DNA profiles is between 1 in 10 billion and 1 in 1 trillion
- What is the probability of finding a match between two such profiles in a database of:
  - 1,000
  - 100,000
  - 1,000,000

Approximate likelihood of finding a matching pair of DNA profiles in a database of *unrelated* individuals

Database Size	Profile Frequency		
	1 in 10 billion	1 in 100 billion	1 in 1 trillion
1000	1 in 20,000	1 in 200,000	1 in 2 million
10,000	1 in 200	1 in 2000	1 in 20,000
100,000	1 in 2.5	1 in 20	1 in 200
1,000,000	1 in 1	1 in 1	1 in 2.5

Last bit of terminology

- Ascertainment Bias
  - The elevated probability of a coincidental match in a multiple-opportunity search
  - Error arising from assuming RMP=F where a multiple-opportunity search makes RMP<F
- Key Issues for DNA Evidence
  - Is a database search a multiple-opportunity search?
  - If so, how to deal with ascertainment bias when characterizing the evidentiary value of a cold hit?

Cold Hit Statistics

- NRC I—test additional loci and report F for those loci only
  - Presumes ascertainment bias is a serious problem
- NRC II—report FxN, where N is the number of profiles in the database
  - e.g., if F=1 in 1 billion; N=1 million; then tell jury RMP=1 in 1000
- Friedman, Balding, Donnelly, Weir (and prosecutors everywhere)—ascertainment bias is *not* a problem, so just tell the jury F

Balding / Donnelly Position

- A DNA database search is *not* a multiple-opportunity search, it is a multitude of single-opportunity searches
- Although there are multiple opportunities to match *someone,* there is only a single opportunity to match *your client,* therefore RMP=F *for the defendant*
- Is this position generally accepted?
- What is the relevant question?

Problems with Balding/Donnelly Position

- Some database searches *do* create multiple opportunities to incriminate *the same person*
  - e.g., suspect's profile searched against multiple items of evidence from multiple unsolved crimes
- B/D assume probability of guilt in a cold hit case may be low, notwithstanding tiny value of F, because *prior probability* is low
  - Will jurors understand (and share) this assumption?
- Failure to consider probability of error

The False Positive Fallacy

"If the probability of a false positive is one in a thousand that means there are 999 chances in 1000 we have the right guy."

- Not necessarily true; probability of "having right guy" depends on strength of all the evidence
- If prior odds of guilt are 1:1000 and odds of a false positive are 1:1000, then chances of "having the right guy" are 50:50 (even odds)
  - See, Thompson, Taroni & Aitken, JFS, 2003.

Inadvertent Transfer of DNA

- **Primary transfer** -- from individual to an object or another person
  - R. van Oorschot & M. Jones, DNA fingerprints from fingerprints. Nature, 387: 767 (1997).
- **Secondary transfer** -- from the point of primary transfer to a second object or person
  - "...in some cases, material from which DNA can be retrieved is transferred from object to hand." *Id*.

Quantities of DNA

- Optimum amount of template: 0.5 to 2.0 ng
- 6 to 7 pg of DNA in each diploid human cell
- Our bodies are made of many billions if not trillions of cells
- pg = picogram (milligram, microgram, nanogram, picogram)
- SGM+ and Profiler Plus test kits are *designed* to fail with less than 100 pg to minimize these problems

DNA content of biological samples:

Type of sample	Amount of DNA
Blood	30,000 ng/mL
stain 1 cm^2 in area	200 ng
stain 1 mm^2 in area	2 ng
Semen	250,000 ng/mL
Postcoital vaginal swab	0 - 3,000 ng
Hair	
plucked	1 - 750 ng/hair
shed	1 - 12 ng/hair
Saliva	5,000 ng/mL
Urine	1 - 20 ng/mL

Taylor & Johnson Studies (1)

A kisses B on cheek

C touches B's cheek with a glove

DNA consistent with A and B found on glove

Taylor & Johnson Studies (2)

A wipes his own face with a damp towel

B wipes her face with same towel

C touches B's face with glove

DNA consistent with A and B found on glove

Pennsylvania v. McNeil

- Woman abducted on street and raped by a stranger wearing a mask
- McNeil lives in the neighborhood
- Laboratory reports DNA matching his profile in vaginal swab, cervical swab and semen stain on victim’s panties

Investigation of the Sexual Assault of Danah H[redacted].
The defendant is Joseph McNeil.

SUMMARY OF AMPLITYPE POLYMARKER/HLA DQ ALPHA TYPING:

Item Description	PR#	LDLR	GYPA	HBGG	D7S8	GC	HLA DQ ALPHA
1) Blood of Joseph McNeil	2186246	AB	AB	AA	AB	AB	2, 4.2/4.3
2) Blood of Danah H[redacted]	2089574	AB	AA	AC	AA	BC	1.1, 1.1
3) Vaginal Swab (sperm fraction) taken from Danah H[redacted]	2089574	AB	AB	(A)C	AB	A(B)C	4.1,(2,4.2/4.3
4) Cervix Swab (sperm fraction) taken from Danah H[redacted]	2089574	AB	AB	(A)C	AB	A(B)C	4.1,(2,4.2/4.3
5) Seminal Stain from panties crotch (area BPPA sperm fraction)	2089575	Inc.	Inc.	Inc.	Inc.	Inc.	NR
6) Seminal Stain from panties crotch (area BPPA Ecell fraction)	2089575	AB	AB	AA	AB	AB	2, 4.2/4.3

Conclusion:

1.) The DNA detected in the sperm fraction of the Vaginal and Cervix swabs (#'s 3 & 4 above) originate from more than one source. Joseph McNeil (#1 above) cannot be excluded as a possible secondary donor to the sperm fraction of the vaginal and cervix swabs.

2.) Joseph McNeil is included as a source of the seminal stains detected in the panties crotch area (#'s 5 & 8 above). The frequency of this combination of genetic polymorphisms is 1 in 16,900 in the African American Population (an exclusion of 99.99% of the African American Population).

Lewis Brenner

Brian Pflegor

9/24/99 Conclusion:

Conclusion:
1.) The DNA detected in the sperm fraction of the Vaginal and Cervix swabs (#'s 3 & 4 above) originate from more than one source. Joseph McNeil (#1 above) cannot be excluded as a possible secondary donor to the sperm fraction of the vaginal and cervix swabs.

2.) Joseph McNeil is included as a source of the seminal stains detected in the panties crotch area (#'s 5 & 8 above). The frequency of this combination of genetic polymorphisms is 1 in 16,900 in the African American Population (an exclusion of 99.99% of the African American Population).

Lewis Brenner Brian Pfleegor

2/7/00 Conclusion:

Conclusion:
1.) Joseph McNeil (#2 above) is excluded as a source of the DNA detected in the vaginal swabs, vulva swabs, and the bloodstains in the crotch area of the panties (#'s 3 - 6 above).

Lewis Brenner Brian Pfleegor

9/24/99

1) Blood of Joseph McNeil	2186246	AB	AB	AA	AB	AB	2, 4.2/4.3
2) Blood of Danah Ha[redacted]	2089574	AB	AA	AC	AA	BC	1.1, 1.1
3) Vaginal Swab (sperm fraction) taken from Danah H[redacted]	2089574	AB	AB	(A)C	AB	A(B)C	4.1,(2,4.2/4.3
4) Cervix Swab (sperm fraction) taken from Danah H[redacted]	2089574	AB	AB	(A)C	AB	A(B)C	4.1,(2,4.2/4.3
5) Seminal Stain from panties crotch (area BPPA sperm fraction)	2089575	Inc.	Inc.	Inc.	Inc.	Inc.	NR
6) Seminal Stain from panties crotch (area BPPA Ecell fraction)	2089575	AB	AB	AA	AB	AB	2, 4.2/4.3
7) Seminal Stain from panties crotch (area BPPB sperm fraction)	2089575	Inc.	Inc.	Inc.	Inc.	Inc.	NR
8) Seminal Stain from panties crotch (area BPPB Ecell fraction)	2089575	AB	AB	AA	AB	AB	2, 4.2/4.3

2/7/00

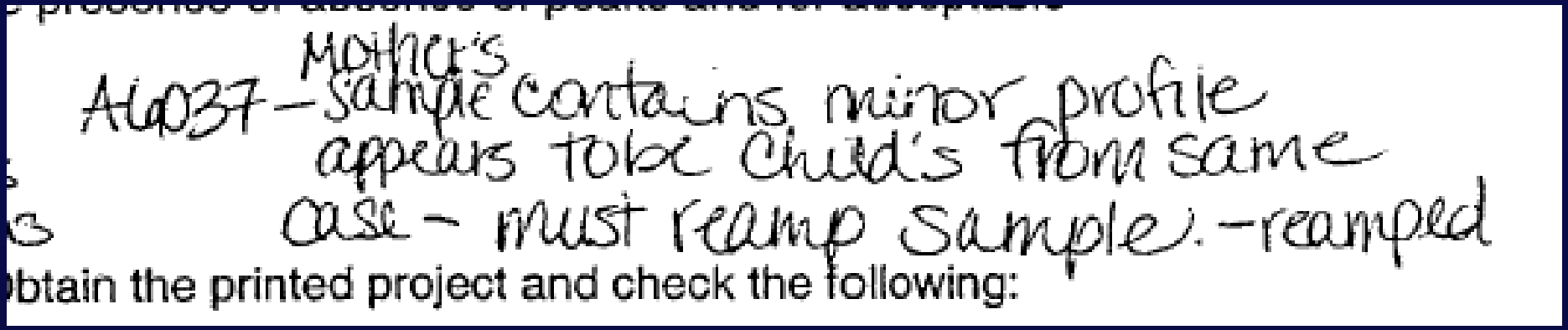
Item Description	PR#	LDLR	GYPA	HBGG	D7S8	GC	DQ ALPHA
1) Blood of Danah Ha[redacted]	2089574	AB	AB	AA	AB	AB	2, 4.2/4.3
2) Blood of Joseph McNeil	2186246	AB	AA	AC	AA	BC	1.1, 1.1
3) Vaginal Swab (sperm fraction) taken from Danah Ha[redacted]	2089574	AB	AB	(A)C	AB	A(B)C	4.1,(2,4.2/4.3)
4) Cervix Swab (sperm fraction) taken from Danah H[redacted]	2089574	AB	AB	(A)C	AB	A(B)C	4.1,(2,4.2/4.3)
5) Bloodstain from panties crotch (area BPPA)	2089575	AB	AB	AA	AB	AB	2, 4.2/4.3
6) Bloodstain from panties crotch (area BPPB)	2089575	AB	AB	AA	AB	AB	2, 4.2/4.3

Documenting errors:
DNA Advisory Board Quality Assurance Standards for Forensic DNA Testing Laboratories, Standard 14

[Forensic DNA laboratories must] “follow procedures for corrective action whenever proficiency testing discrepancies and/or casework errors are detected” [and] “shall maintain documentation for the corrective action.”

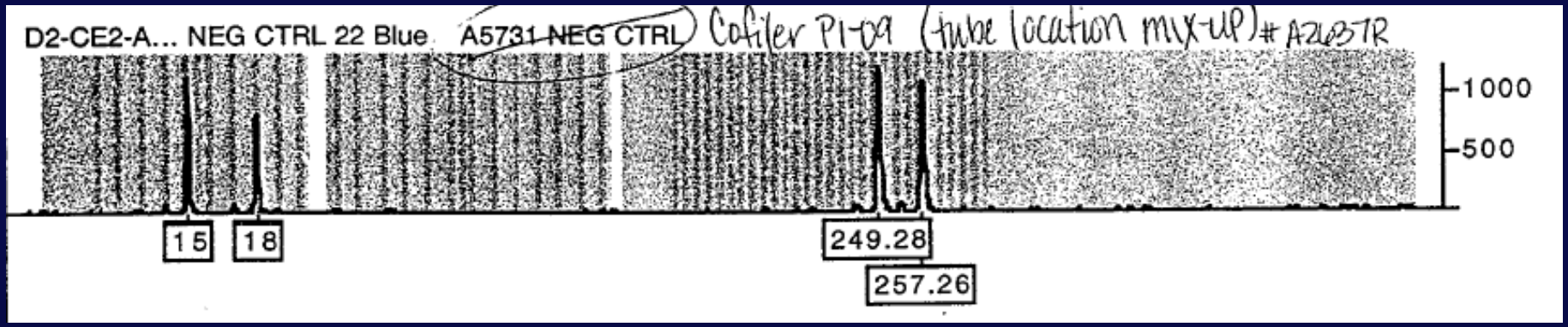
Documenting errors

Cross contamination:



Documenting errors

Positive result in negative control:



Documenting errors

Positive result in negative control, due to tube swap:

Conclusion:

When the data was analyzed as a whole before separating into mini projects, it was clear to see that the tubes in positions D2 and D4 were switched on accident. On the Injection list, the D2 position was labeled as the negative control (A5731) however the data shows a profile consistent with PI-09 in COfiler. The D4 position that was labeled PI-09 (A2637R) shows a profile consistent with that of a negative control.

Corrective Action for future analyses BS.-

Consecutively label 310 tubes, as well as, number w/ DNA numbers.

Check tubes carefully when placing in 310 tube racks.

02274

Documenting errors

Analyst contamination:

Contamination of Known Reference Sample:

On March 7, 2003 it was discovered that the known reference sample for victim Gilbert Osorno, Lab #PE02-00459-02, was contaminated with the Analyst, Vickie Kump's, DNA profile. It could not be determined how or at what stage of the analysis the contamination occurred. The analyst repeated the analysis using the same lab protocols as used previously resulting in a single profile of the victim.

Documenting errors

Separate samples combined in one tube . . . .

Sample Handling Problem During Extraction of Evidence Items for PE96-00286 and PE02-00405

On April 29, 2002 during the extraction procedure for evidence items in Lab Number PE96-00286-03 the extract for A3822 was inadvertently added to the Centricon that already contained the extract for sample A3821. The problem was

Documenting errors

Separate samples combined in one tube . . . .

Sample Handling Problem During Extraction of Evidence Items for PE96-00286 and PE02-00405

On April 29, 2002 during the extraction procedure for evidence items in Lab Number PE96-00286-03 the extract for A3822 was inadvertently added to the Centricon that already contained the extract for sample A3821. The problem was

. . . . leading to corrective action:

Corrective Action

Try to minimize interruptions during any sample transfer procedures. Signs were made for the analysts to put at the end of their work areas indicating that they cannot be interrupted at the time the sign is out. Also, phone calls will not be answered during any sample handling and transfer steps.

Documenting errors

Samples mixed up

1/13/2003

The case file includes 3 sets of data.

Set #1: Item #1 and Item #2 – Known reference samples

Set #2: Item #3 and Item #4 (DATA NOT USED)

Set #3: Item #3 and Item #4

} evidence items - sample switch during processing

Set #2 (DATA NOT USED) was obtained and analyzed, but when compared to the results of Julie Marquez, the data did not match. It was determined that Set#2 results did not match due to a sample mix-up of DNA numbers. Therefore, Item#3 and Item #4 was extracted, quantitated, amplified and analyzed (Set#3) again. The data for Set#3 matched with Julie Marquez results and was used for the CTS#02-576 data.

Documenting errors

Suspect doesn't match himself . . . .

On 5/16/01, we had reported another case (PA01-00171) involving Mr. Phillips. The genetic profiles generated for Mr. Phillips in these two cases differed at all loci tested. Clearly, at least one had to be erroneous. I left a note for Vickie Kump (who was preparing to generate the draft

. . . . but then, staff is "'always' getting people's names wrong":

Case Processing Form" (see attached #4) released [illegible]

from Terry Phillips. I immediately asked Paul why he had not noted of followed up on the difference between the name "Lucio Flores" on both the sample envelope and swab, and "Terry Phillips" on the appointment slip and COC forms. He responded that he did not expect to have to do someone else's job for them, and that they [collection room staff] are 'always' getting people's names wrong, so he only goes by the case number not the name. In addition, he later told me that during pre-extraction sample processing, he fills out the sample identifiers on the case processing form prior to (not during) sample examination.

LOOKING AT A DNA REPORT

Components of a DNA report

- The samples tested
  - **Evidence samples (crime scene)**
  - **Reference samples (defendant, suspect)**
- The lab doing the testing
- The test used:
  - **SGM+, Profiler Plus, Cofiler, Identifiler, mtDNA**
- The analyst who did the testing
- Results and conclusions:
  - **Table of alleles**
  - **Narrative conclusions**

Table of alleles

TABLE OF RESULTS

ITEM	DESCRIPTION	D3S1358	vWA	FGA	AMEL	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820
1	Reference From [redacted]	15,18	18,20	26,28	X,X	10,13	30,31	12,17	12,12	11,12	10,12
2	Reference From [redacted]	15,16	15,16	19,26	X,Y	12,13	31,31	16,21	11,12	11,12	10>11
3	Neck Swab	15,16 (18)	15,16 (18,20)	19,26 (28)	X,Y	12,13 (10)	31,31 (30)	16,21 (12,17)	11,12	11,12	10,11 (12)
4	Chest Swab	15,16 (17<18)	15,16 (18,20)	19,26 (28)	X>Y	12,13 (10)	31,31 (30)	16,21 (12,17)	11,12	11,12	10,11
	Extraction Blanks	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Key: NA = No activity. < = Less than.
( ) = Weak results for types in parenthesis. X,X = Female DNA.
> = Greater than. X,Y = Male DNA.

- Some labs include more information than others
- Usually includes information about mixed samples
- May also include:
  - **Indication of low level results**
  - **Indication of results not reported**
  - **Relative amounts of different alleles (in mixed samples)**
- No standard format

Narrative conclusions

CONCLUSIONS

1. The neck and chest swabs (items 3 and 4) have an elevated level of amylase 1 present in the extracts. These results strongly indicate saliva on the swabs.

2. The genetic marker results in the DNA extracted from the neck and chest swabs (items 3 and 4) are a mixture of at least two persons. The results indicate a major (or stronger donor) and a secondary (or weaker donor). [redacted] is, in my opinion, the major DNA donor on items 3 and 4. Due to the presence of weak typing results at some loci, it is possible that minor components of the mixture have dropped out in the larger loci. As a result, [redacted] cannot be excluded as a contributor to the secondary DNA profile obtained from the neck and chest swabs (items 3 and 4). In addition, a weak amount of D3S1358 type 17 was detected on item 4 which could not have originated from [redacted] or [redacted]. It is unclear as to whether this allele is artifactual in origin or from another donor.

- Indicates which samples match
- Includes a statistical estimate
- Identifies samples as mixed
- May include an ‘identity statement’ i.e., samples are from the same source to a scientific degree of certainty (FBI)
- May allude to problems (e.g. interpretative ambiguity, contamination)

Looking beneath the report

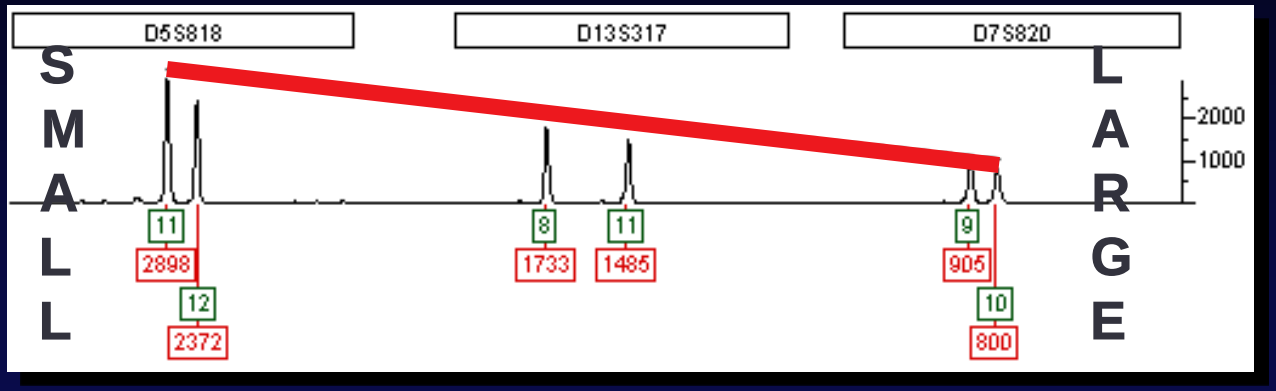
The science of DNA profiling is sound.

But, not all of DNA profiling is science.

Sources of ambiguity in STR interpretation

- Degradation
- Allelic dropout
- False peaks
- Mixtures
- Accounting for relatives
- Threshold issues -- marginal samples

Degradation

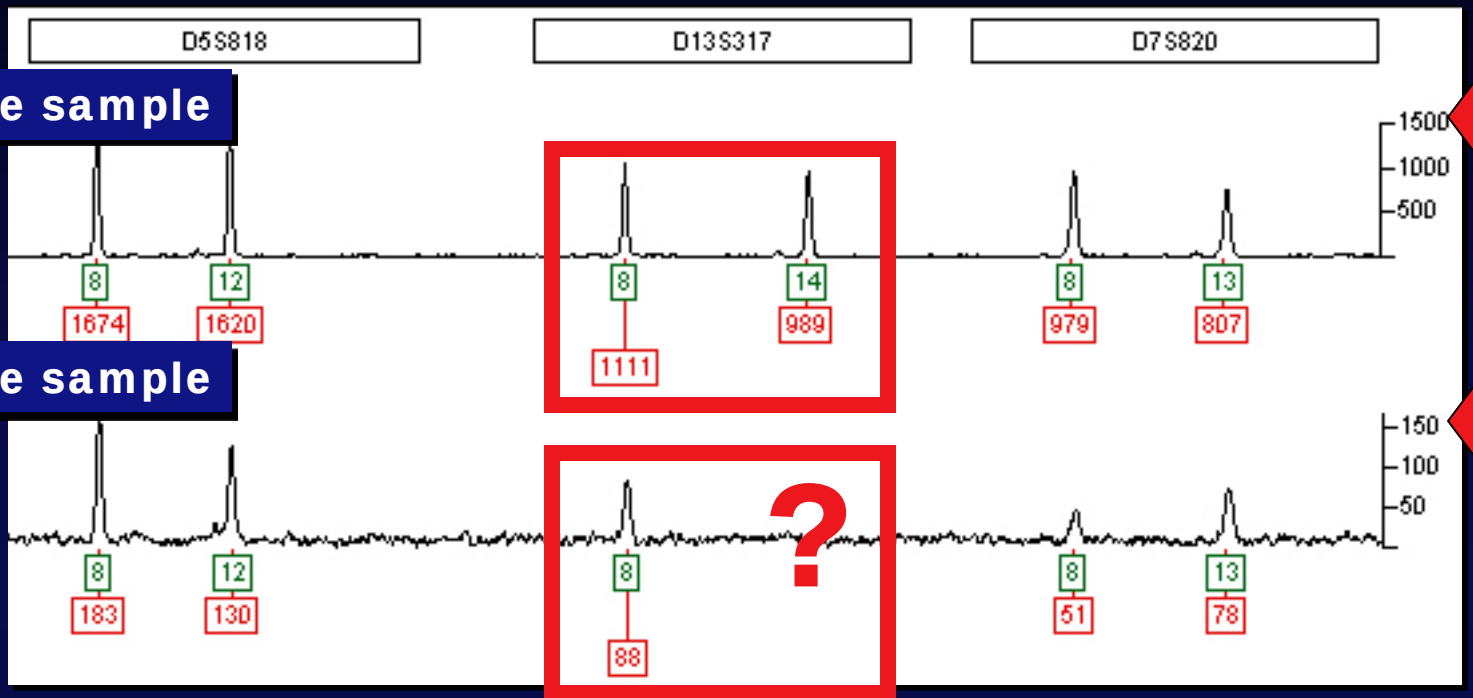


- When biological samples are exposed to adverse environmental conditions, they can become degraded
  - **Warm, moist, sunlight, time**
- Degradation breaks the DNA at random
- Larger amplified regions are affected first
- Classic ‘ski-slope’ electropherogram
- Peaks on the right lower than peaks on the left

Sources of ambiguity in STR interpretation

- Degradation
- Allelic dropout
- False peaks
- Mixtures
- Accounting for relatives
- Threshold issues -- marginal samples

Allelic Dropout



- Peaks in evidence samples all very low
  - **Mostly below 150 rfu**
- Peaks in reference sample much higher
  - **All well above 800 rfu**
- At D13S817:
  - **Reference sample: 8, 14**
  - **Evidence sample: 8, 8**
- 14 allele has dropped out -- or has it?
- Tend to see with ‘marginal samples’

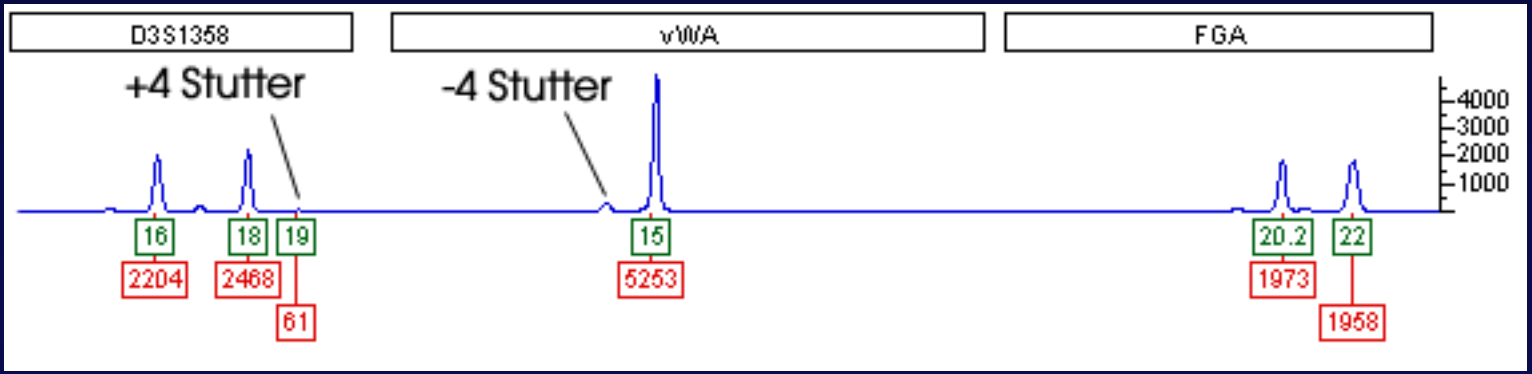
Sources of ambiguity in STR interpretation

- Degradation
- Allelic dropout
- False peaks
- Mixtures
- Accounting for relatives
- Threshold issues -- marginal samples

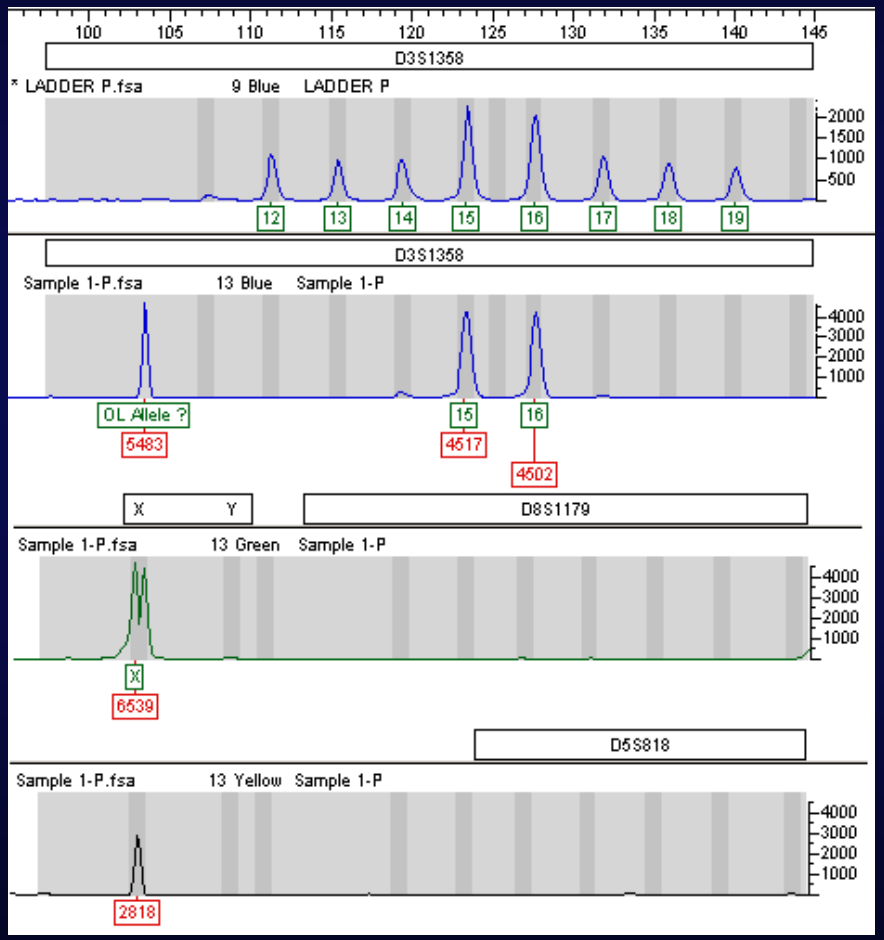
Not all signal comes from DNA associated with an evidence sample

- Stutter peaks
- Pull-up (bleed through)
- Spikes and blobs

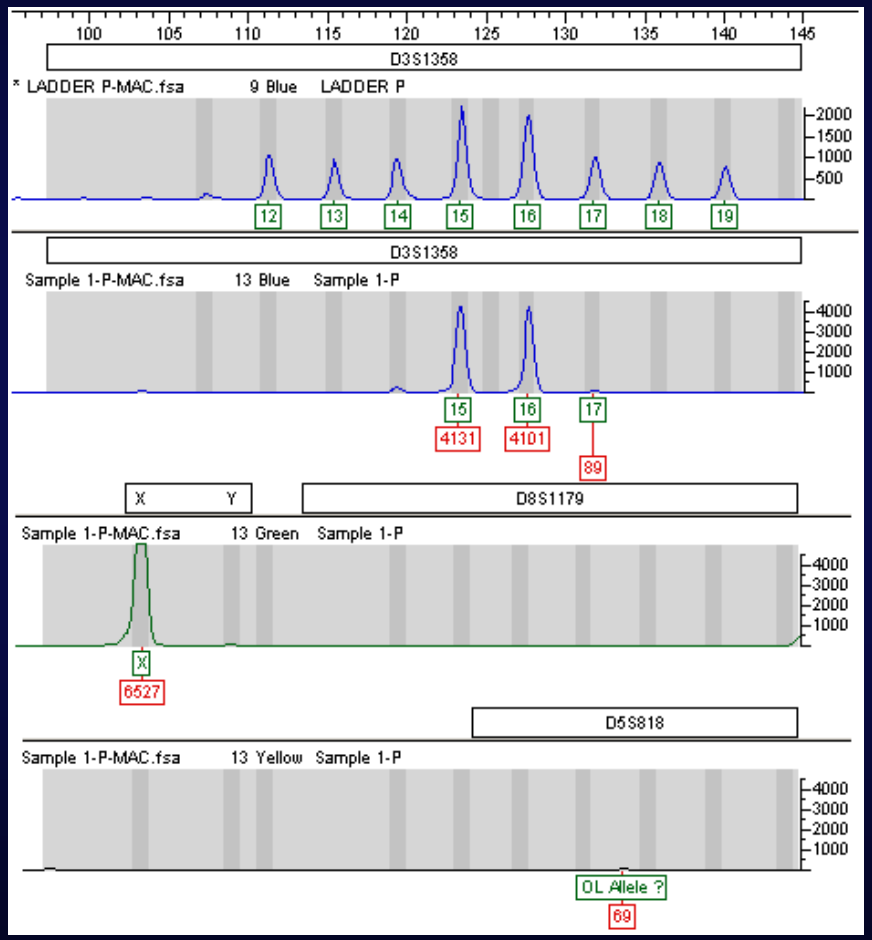
Stutter peaks



Pull-up (and software differences)

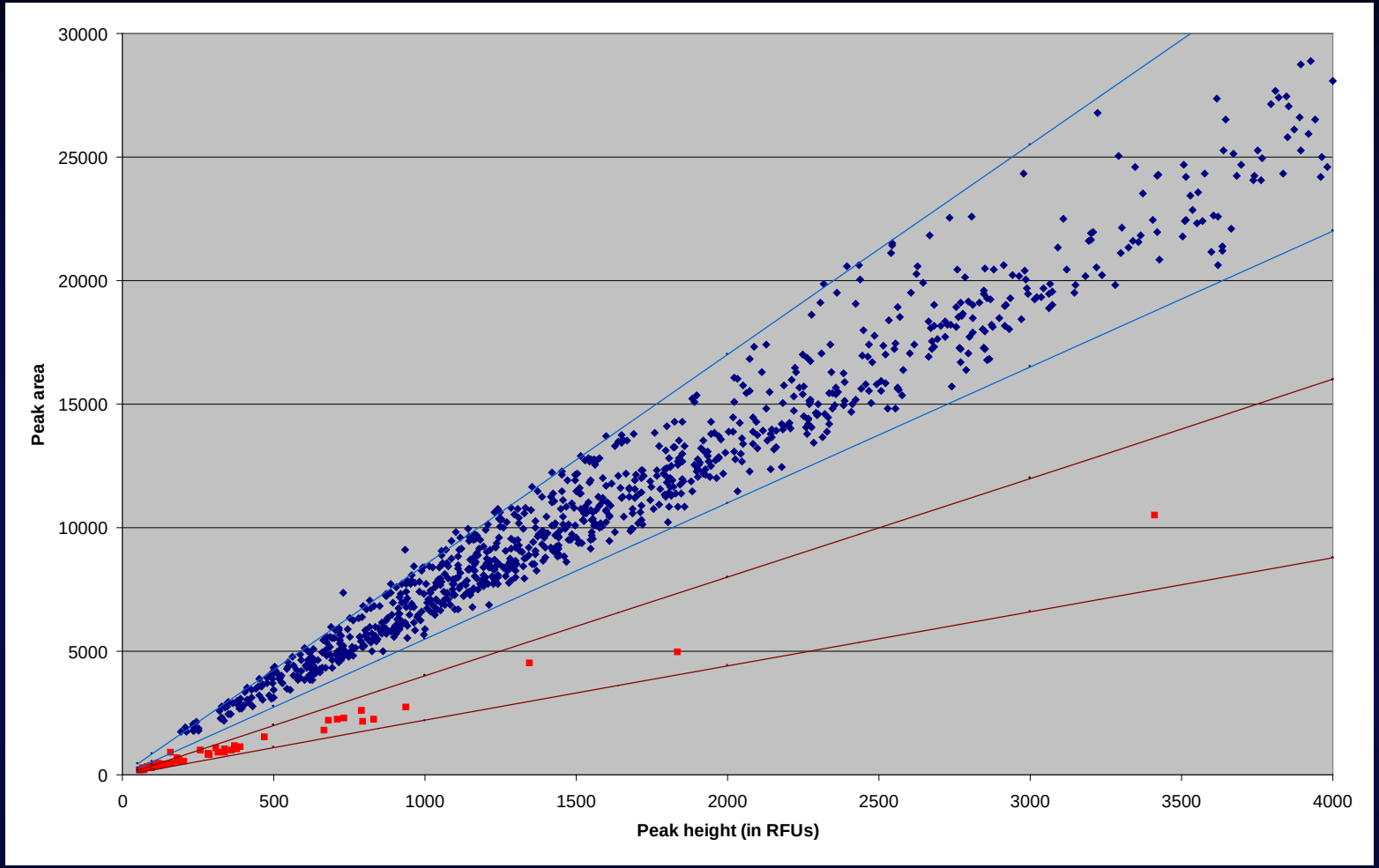


Advanced



Classic

Spikes and blobs

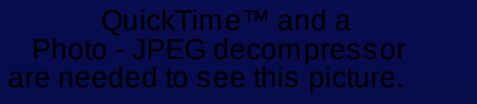


- 89 samples (references, pos controls, neg controls)
- 1010 “good” peaks
- 55 peaks associated with 24 spike events
- 95% boundaries shown

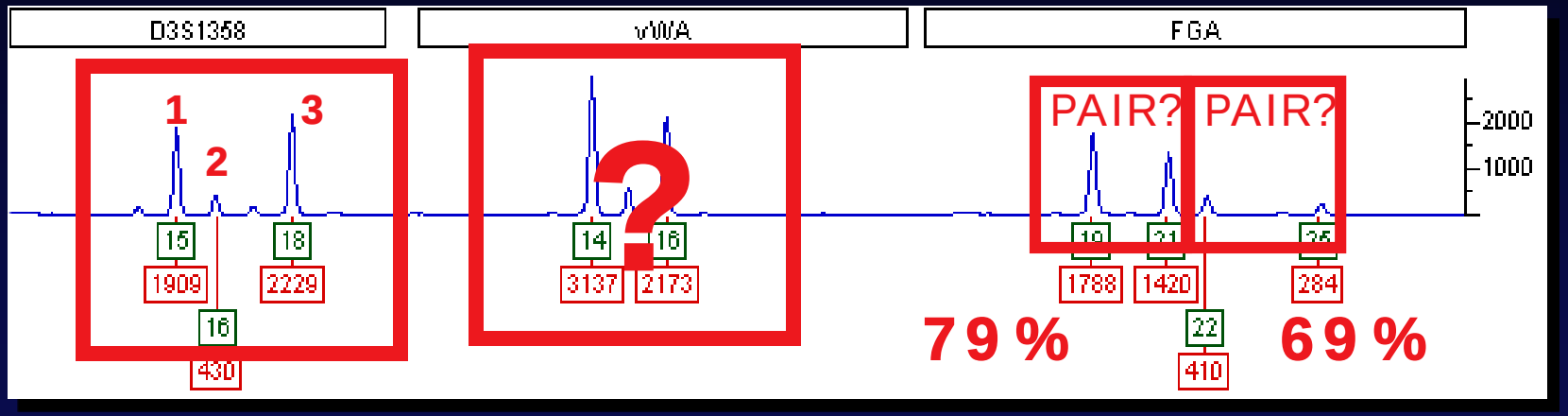
Sources of ambiguity in STR interpretation

- Degradation
- Allelic dropout
- False peaks
- Mixtures
- Accounting for relatives
- Threshold issues -- marginal samples

Mixed DNA samples



Mixtures



- More than two alleles at a locus may indicate a mixture
- Number of contributors often unclear because of sharing alleles
- Some labs rely on ‘peak height ratio’ to pair peaks up (one peak 70% of another peak)
- May be arbitrary: factors other than the quantity of DNA can effect peak height
- Statistics used in mixture cases: may make debatable assumptions

How many contributors to a mixture? If analysts can discard a locus?

Maximum # of alleles observed in a 3-person mixture	# of occurrences	Percent of cases
2	0	0.00
	3,398	0.00
3	78	0.00
	26,788,540	18.28
4	4,967,034	3.39
	112,469,398	76.75
5	93,037,010	63.49
	7,274,823	4.96
6	48,532,037	33.12

There are 146,536,159 possible different 3-person mixtures of the 959 individuals in the FB I database (Paoletti et al., November 2005 *JFS*).

How many contributors to a mixture? If analysts can discard a locus?

Maximum # of alleles observed in a 3-person mixture	# of occurrences	Percent of cases
2	0	0.00
	8,151	0.02
3	310	0.00
	11,526,219	25.53
4	2,498,139	5.53
	32,078,976	71.07
5	29,938,777	66.32
	1,526,550	3.38
6	12,702,670	28.14

There are 45,139,896 possible different 3-person mixtures of the 648 individuals in the MN BCI database (genotyped at only 12 loci).

How many contributors to a mixture?

Maximum # of alleles observed in a 4-person mixture	# of occurrences	Percent of cases
4	13,480	0.02
5	8,596,320	15.03
6	35,068,040	61.30
7	12,637,101	22.09
8	896,435	1.57

There are 57,211,376 possible different 4-way mixtures of the 194 individuals in the FB I Caucasian database (Paoletti et al., November 2005 *JFS*). (35,022,142,001 4-person mixtures with 959 individuals.)

Sources of ambiguity in STR interpretation

- Degradation
- Allelic dropout
- False peaks
- Mixtures
- Accounting for relatives
- Threshold issues -- marginal samples

What contributes to overlapping alleles between individuals?

- Identity by state

  -- many loci have a small number of detectable alleles (only 6 for TPOX and 7 for D13, D5, D3 and TH01)

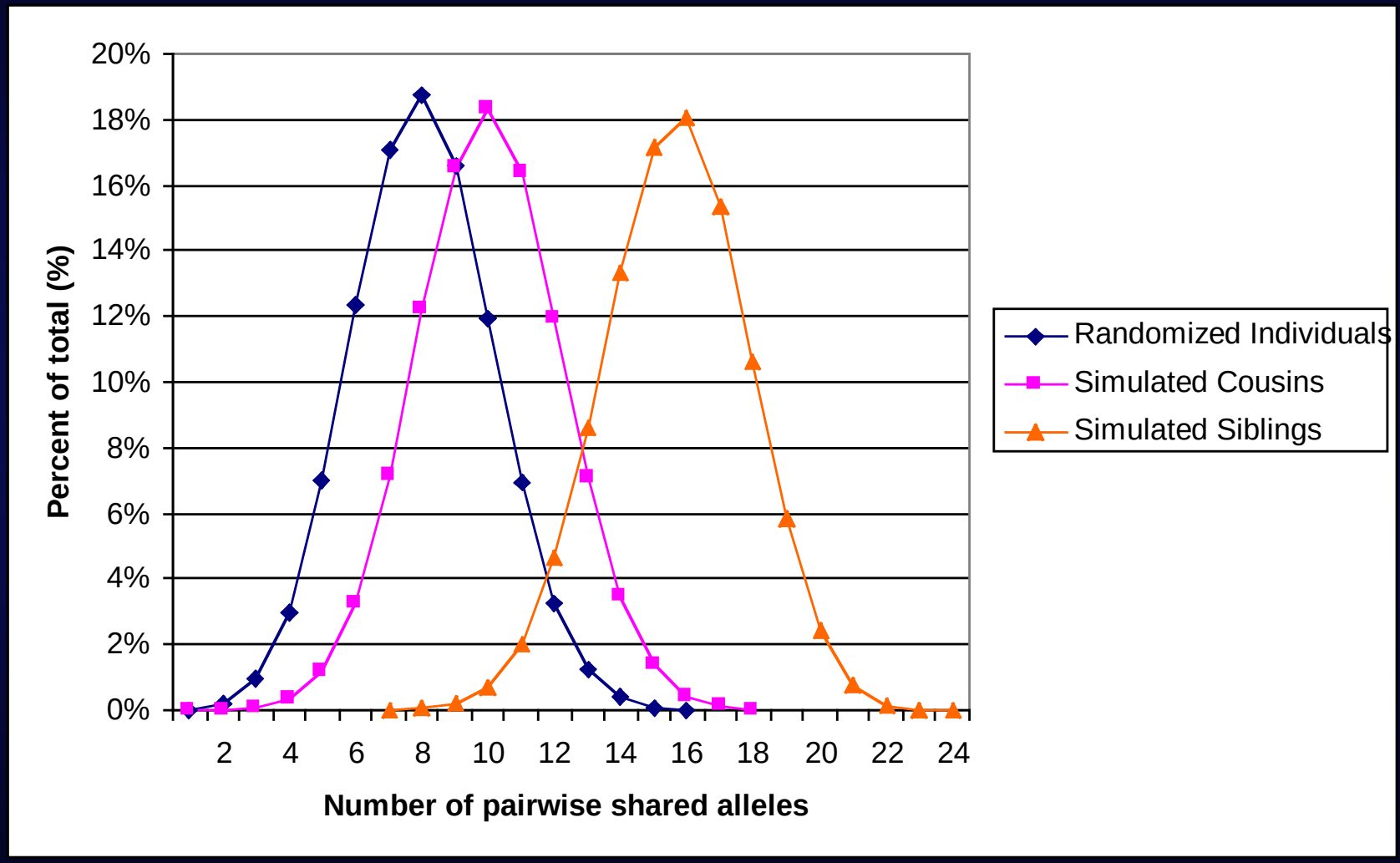
  -- some alleles at some loci are relatively common

- Identity by descent

  -- relatives are more likely to share alleles than unrelated individuals

  -- perfect 13 locus matches between siblings occur at an average rate of 3.0 per 459,361 sibling pairs

Allele sharing between individuals



Allele sharing in databases

- Original FBI dataset's mischaracterization rate for 3-person mixtures (3.39%) is more than two σ above the average observed in five sets of randomized individuals
- Original FBI dataset has more shared allele counts above 19 than five sets of randomized individuals (3 vs. an average of 1.4)

Sources of ambiguity in STR interpretation

- Degradation
- Allelic dropout
- False peaks
- Mixtures
- Accounting for relatives
- Threshold issues -- marginal samples

Where do peak height thresholds come from (originally)?

- Applied Biosystems validation study of 1998
- Wallin et al., 1998, “TWGDAM validation of the AmpFISTR blue PCR Amplification kit for forensic casework analysis.” *JFS* 43:854-870.

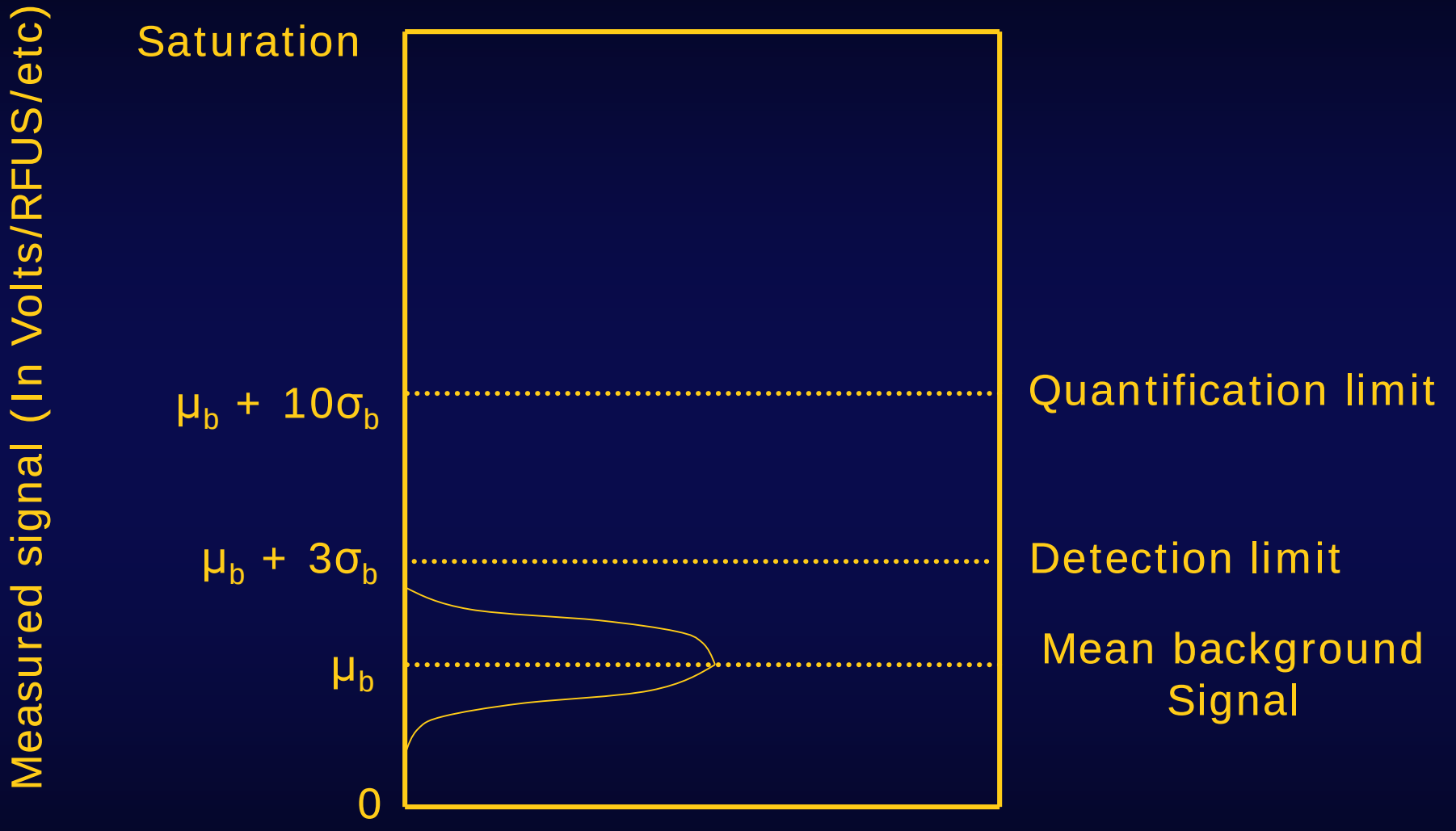
Where do peak height thresholds come from (originally)?

PCR products were examined on both the 377 DNA Sequencer and the 310 Genetic Analyzer. The results of 0.25 to 1.0 ng were clearly typable with peak heights of approximately 150 RFU and greater (data not shown). At 0.125 ng and less, the peak heights in both samples were not significantly above the background (< 150 RFU) or were undetectable. At 0.0313 ng specifically, peaks were extremely low or undetectable, and thus, DNA quantities as low as approximately 35 pg did not produce a typable result. Based on these results, we employed a peak height threshold of 150 RFU, below which peaks were interpreted with caution. Laboratories should determine a minimum peak height threshold for their instruments using high quality, single source genomic DNA samples which provides them with the desired sensitivity while not allowing for detection of low copy DNA. This is particularly important as the overall sensitivity of the assay may vary between laboratories.

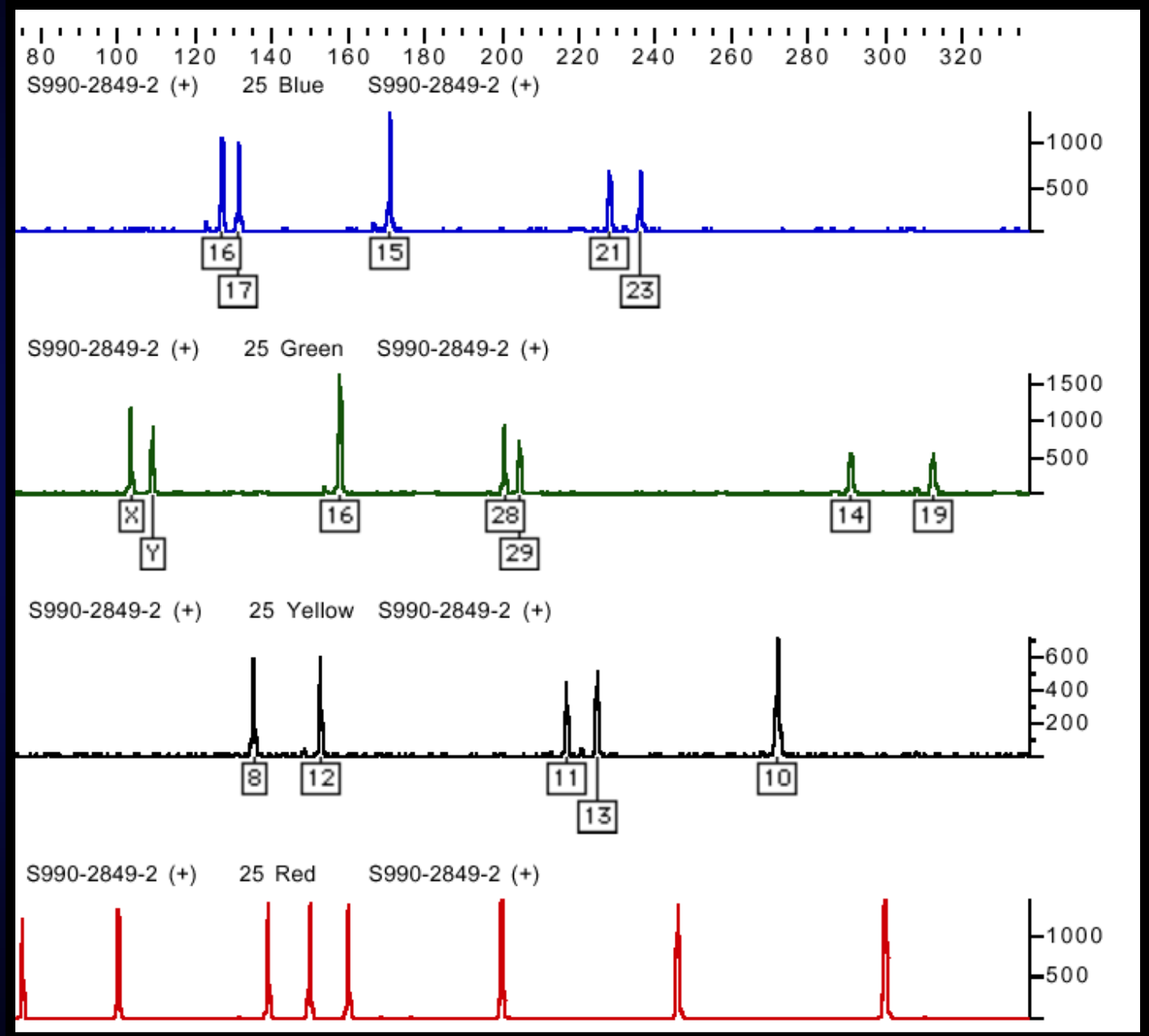
Where do peak height thresholds come from?

- “Conservative” thresholds established during validation studies
- Eliminate noise (even at the cost of eliminating signal)
- Can arbitrarily remove legitimate signal
- Contributions to noise vary over time (e.g. polymer and capillary age/condition)
- Analytical chemists use LOD and LOQ

Signal Measure



Many opportunities to measure baseline



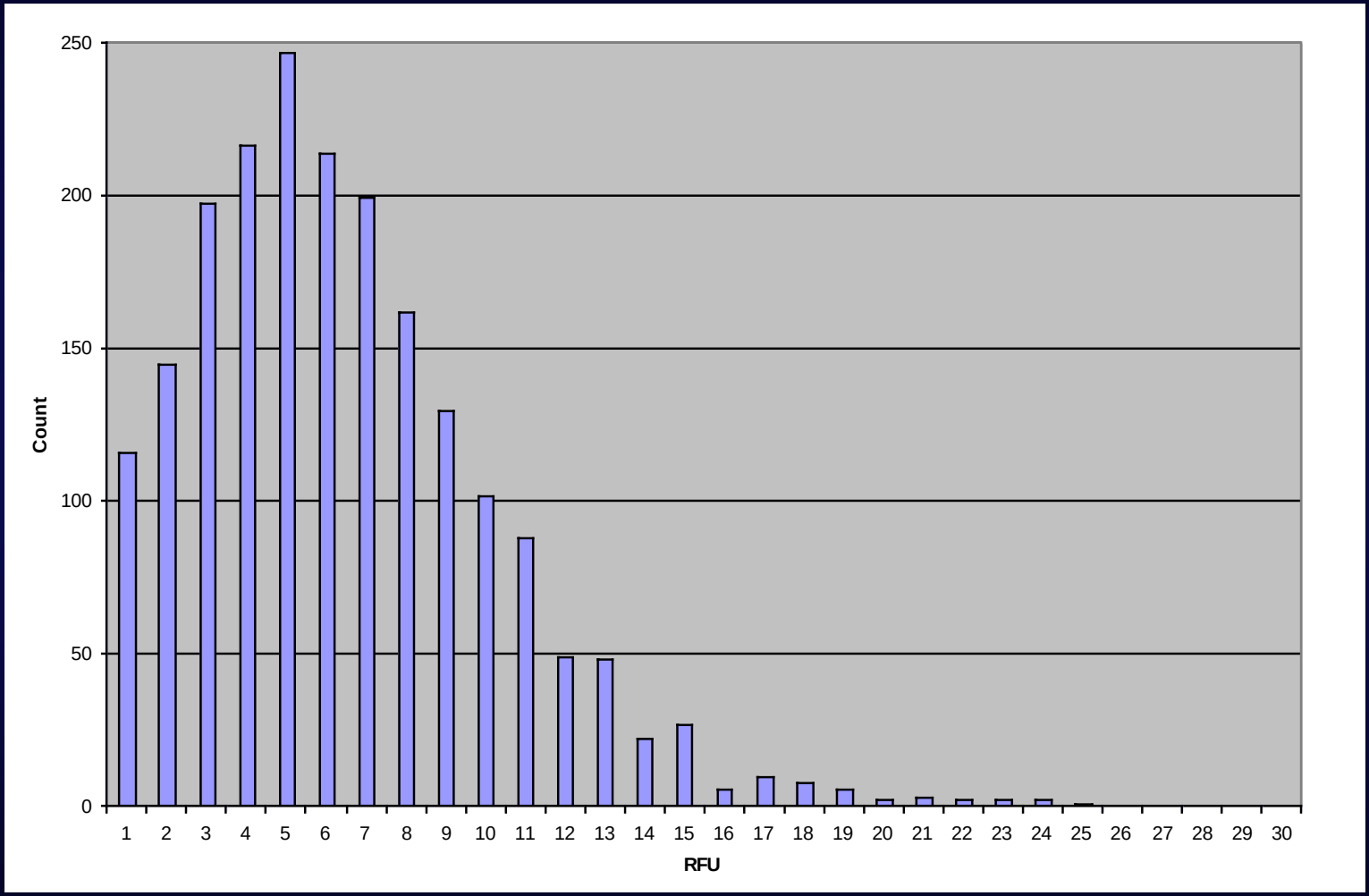
Measurement of baseline in control samples:

- Negative controls: 5,932 data collection points (DCPs) per run (σ = 131 DCPs)
- Reagent blanks: 5,946 DCPs per run (σ = 87 DCPs)
- Positive controls: 2,415 DCP per run (σ = 198 DCPs)

Measurement of baseline in control samples:

- Negative controls: 5,932 data collection points (DCPs) per run (σ = 131 DCPs)
- Reagent blanks: 5,946 DCPs per run (σ = 87 DCPs)
- Positive controls: 2,415 DCP per run (σ = 198 DCPs)
- DCP regions corresponding to size standards and 9947A peaks (plus and minus 55 DCPs to account for stutter in positive controls) were masked in all colors

RFU levels at all non-masked data collection points

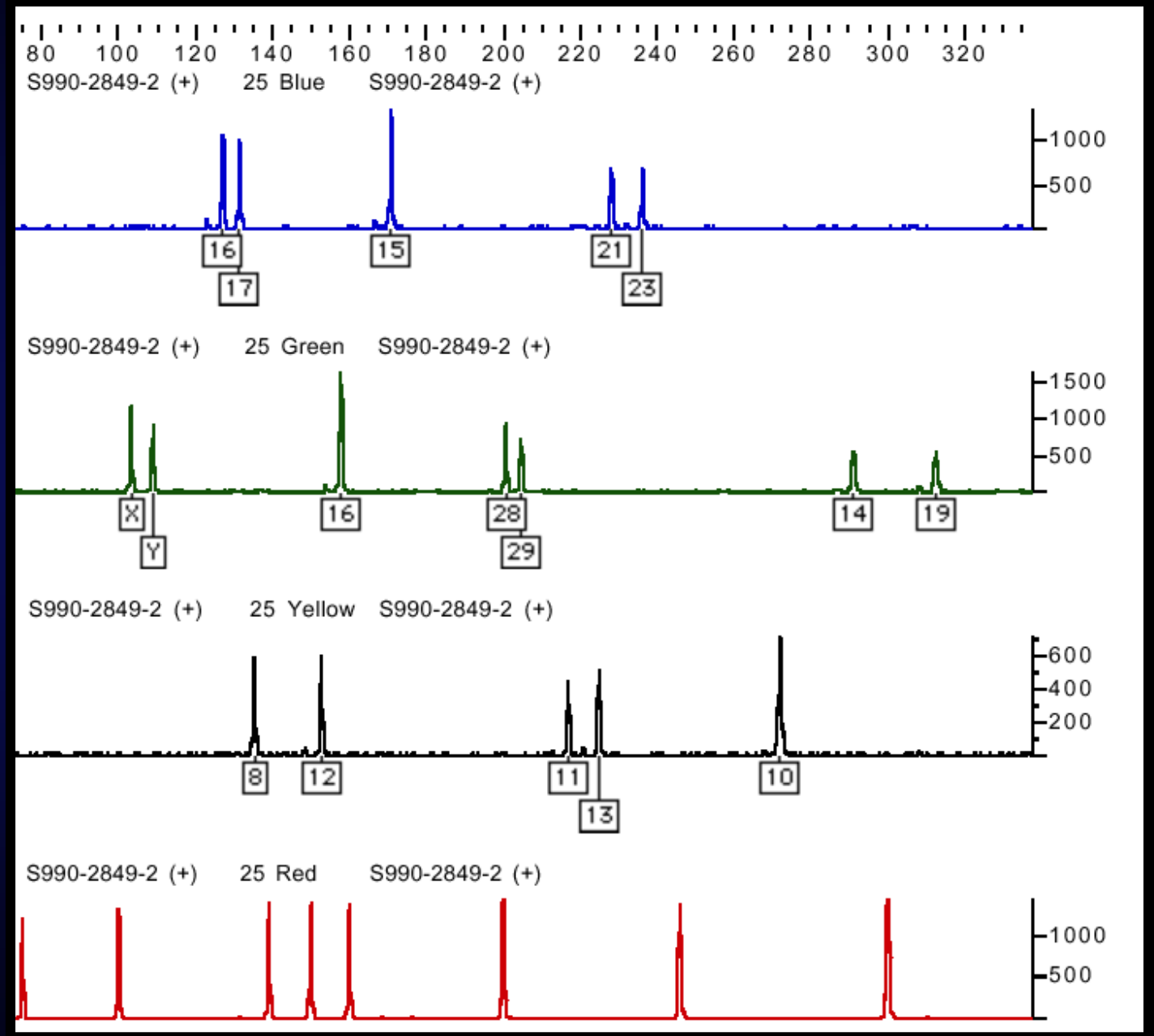


Variation in baseline noise levels

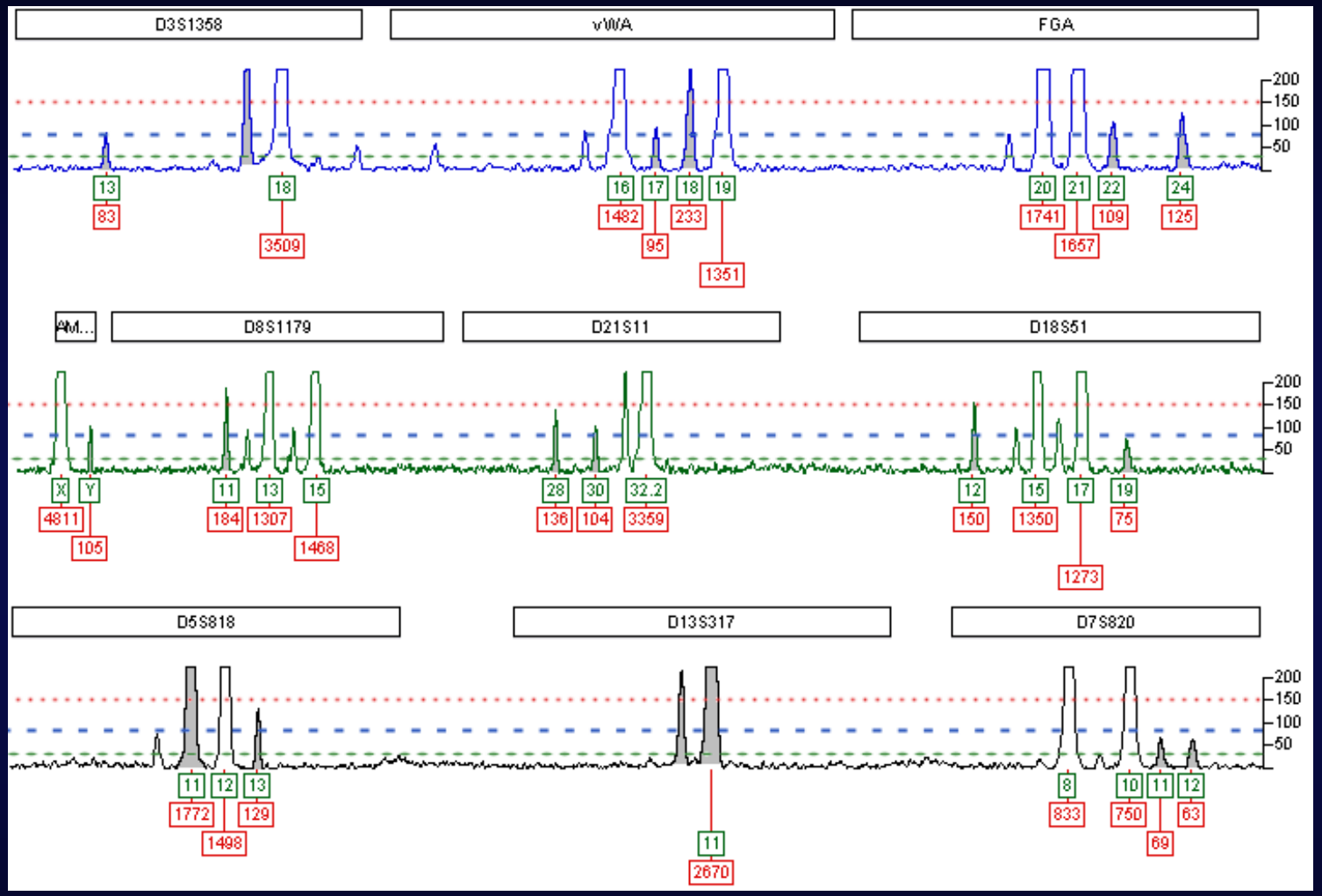
Positive Control		μ_b	σ_b	$\mu_b + 3\sigma_b$	$\mu_b + 10\sigma_b$
	Maximum	6.7	6.9	27.4	75.7
	Average	5.0	3.7	16.1	42.0
	Minimum	3.7	2.4	10.9	27.7
Negative Control		μ_b	σ_b	$\mu_b + 3\sigma_b$	$\mu_b + 10\sigma_b$
	Maximum	13.4	13.2	53.0	145.4
	Average	5.4	3.9	17.1	44.4
	Minimum	4.0	2.6	11.8	30.0
Reagent Blank		μ_b	σ_b	$\mu_b + 3\sigma_b$	$\mu_b + 10\sigma_b$
	Maximum	6.5	11.0	39.5	116.5
	Average	5.3	4.0	17.3	45.3
	Minimum	4.0	2.6	11.8	30.0
All three controls averaged		μ_b	σ_b	$\mu_b + 3\sigma_b$	$\mu_b + 10\sigma_b$
	Maximum	7.1	7.3	29.0	80.1
	Average	5.2	3.9	16.9	44.2
	Minimum	3.9	2.5	11.4	28.9

Average (μ_b) and standard deviation (σ_b) values with corresponding LODs and LOQs from positive, negative and reagent blank controls in 50 different runs. BatchExtract: ftp://ftp.ncbi.nlm.nih.gov/pub/forensics/

Doesn’t someone either match or not?

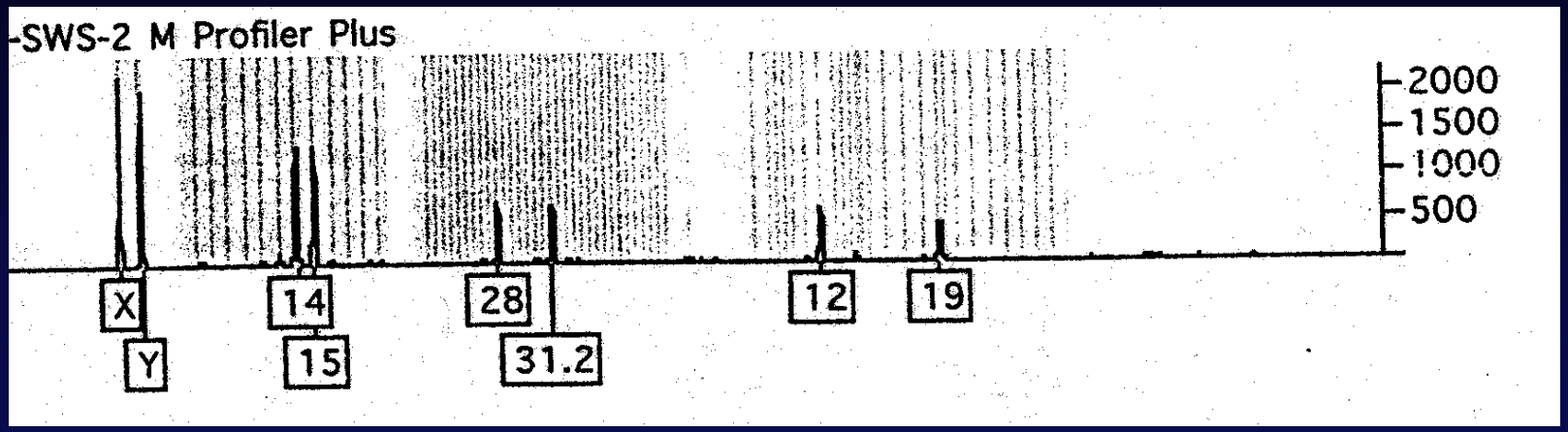


Lines in the sand: a two-person mix?

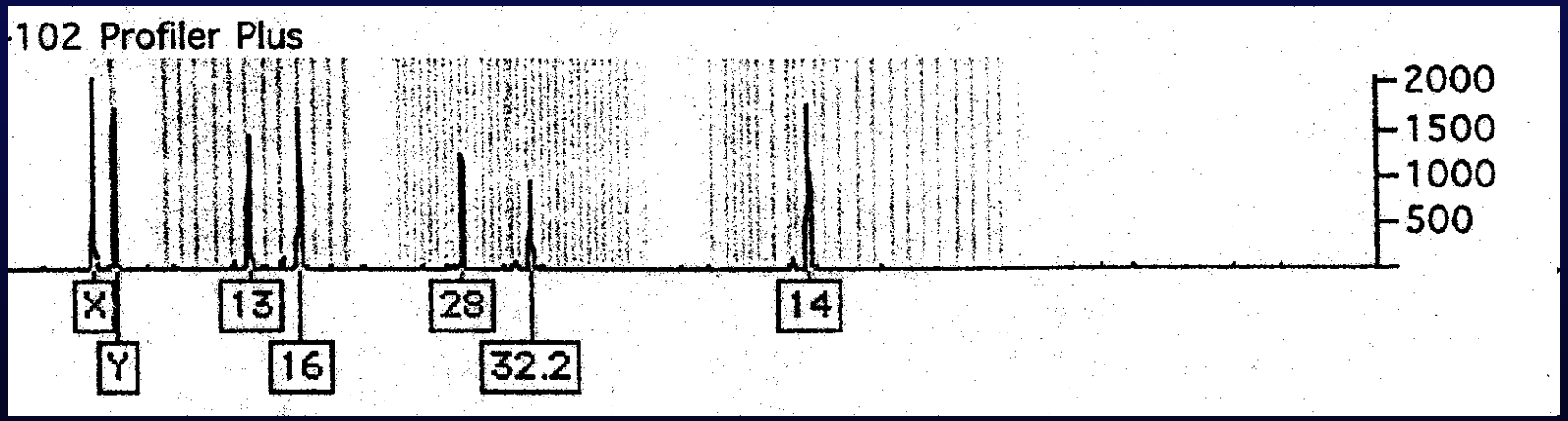


Two reference samples in a 1:10 ratio (male:female). Three different thresholds are shown: 150 RFU (red); LOQ at 77 RFU (blue); and LOD at 29 RFU (green).

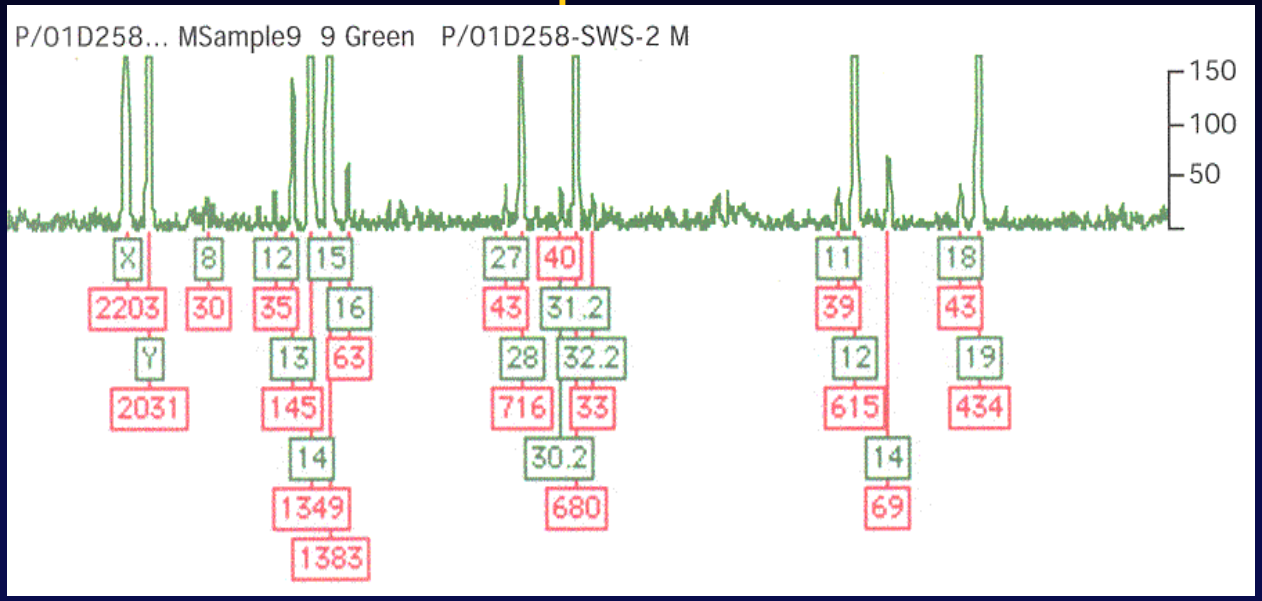
Vaginal Swab—male fraction (showing defendant's profile)



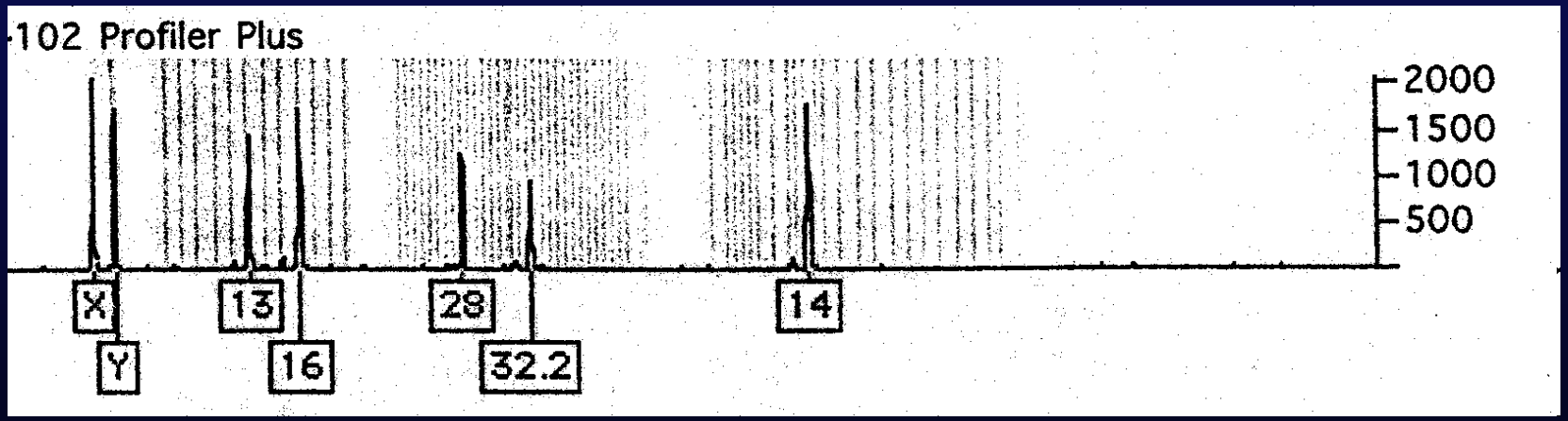
Profile of second man (could he also be a contributor?)



Vaginal swab—close examination of electronic data shows evidence of second profile



Could second man be the source?



Observer effects, aka expectation effects

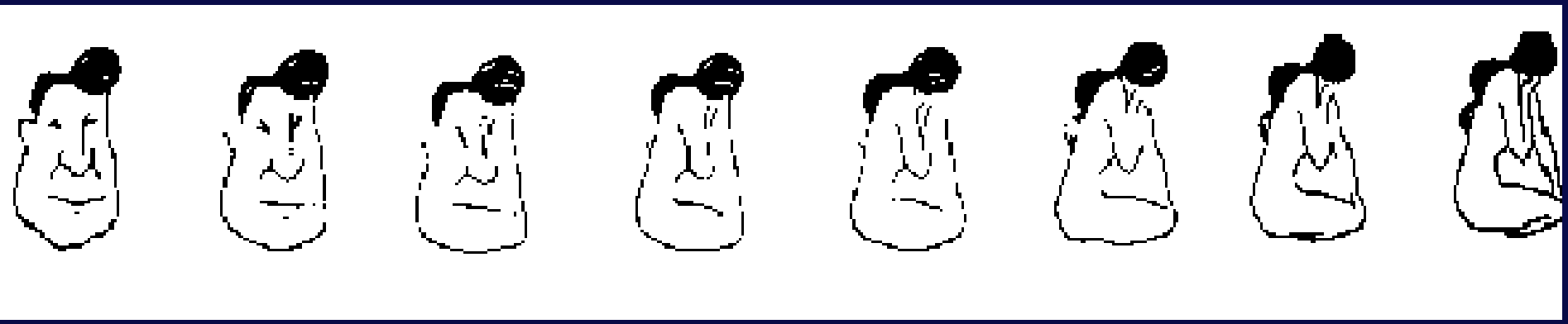
- *--the tendency to interpret data in a manner consistent with expectations or prior theories (sometimes called "examiner bias")*
- Most influential when:
  - Data being evaluated are ambiguous or subject to alternate interpretations
  - Analyst is motivated to find a particular result

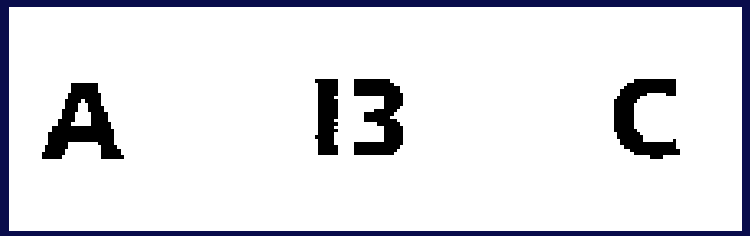
Context and Expectations Influence Interpretation of Data

Context and Expectations Influence Interpretation of Data

Context and Expectations Influence Interpretation of Data

Context and Expectations Influence Interpretation of Data

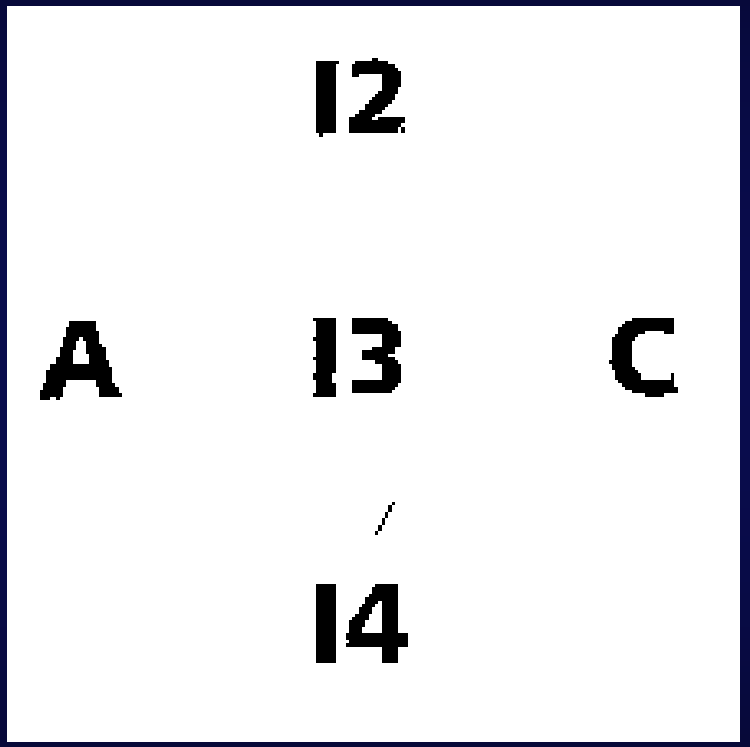



A 13 C

12

13

14


12
A
13
C
14

Analyst often have strong expectations about the data

DNA Lab Notes (*Commonwealth v. Davis*)

- *"I asked how they got their suspect. He is a convicted rapist and the MO matches the former rape...The suspect was recently released from prison and works in the same building as the victim...She was afraid of him. Also his demeanor was suspicious when they brought him in for questioning...He also fits the general description of the man witnesses saw leaving the area on the night they think she died...So, I said, you basically have nothing to connect him directly with the murder (unless we find his DNA). He said yes."*

Analyst often have strong expectations about the data

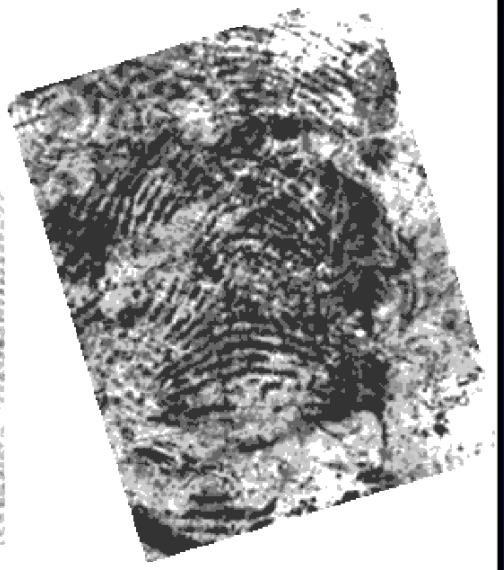
DNA Lab Notes

- *"Suspect-known crip gang member--keeps 'skating' on charges-never serves time. This robbery he gets hit in head with bar stool--left blood trail. Miller [deputy DA] wants to connect this guy to scene w/DNA ..."*
- *"Death penalty case! Need to eliminate Item #57 [name of individual] as a possible suspect"*

Analysts' expectations may lead them to:

- Resolve ambiguous data in a manner consistent with expectations
- Miss or disregard evidence of problems
- Miss or disregard alternative interpretations of the data
- Thereby undermining the scientific validity of conclusions
  - **See, Risinger, Saks, Thompson, & Rosenthal, *The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*. 93 California Law Review 1 (2002).**

Compare The Prints



FBI's Explanation of Mayfield Error

- "Confirmation Bias"
- "[B]ecause the initial examiner was a highly respected supervisor with many years of experience, it was concluded that subsequent examinations were incomplete and inaccurate. To disagree was not an expected response."
  - Robert B. Stacey, A report on the erroneous fingerprint individualization in the Madrid Train Bombing Case. 54 J.Forensic Identification 706 (2004).
  - See, Thompson & Cole, Lessons from the Brandon Mayfield Case. The Champion, April 2005

What is LCN?

- DNA profiling performed at or beneath the stochastic threshold
- Typically less than 0.5 ng of DNA template
- Typically involves modifications of the testing methodology (e.g. increased polymerase; additional rounds of amplification; skipping quantitation)
- Consensus profiles

Quantitating and Amplifying DNA

- Determine the quantity of DNA in samples to be amplified. See Chapter 4 for more details on DNA quantitation.
- Amplify DNA samples using the AmpFℓSTR SGM Plus kit reagents (see Chapter 5). The recommended range of input DNA is 1.0–2.5 ng.

Note A useful initial experiment is to amplify a range of input DNA for each of several samples in order to establish the range of input DNA (as determined by your laboratory's quantitation system) that provides optimal results. For example, amplify 0.5, 1.0, 1.5, 2.0, 2.5 ng, and 5.0 ng of input DNA for each sample.

Applied Biosystems SGM Plus User's Manual p.1-14

“The PCR amplification parameters have been optimized to produce similar peak heights within and between loci. The peak height generated at a locus for a heterozygous individual should be similar between the two alleles. The kit is also designed to generate similar peak heights between loci labeled with the same dye so that each locus will have approximately the same sensitivity.”

Applied Biosystems SGM Plus User’s Manual p.1-13

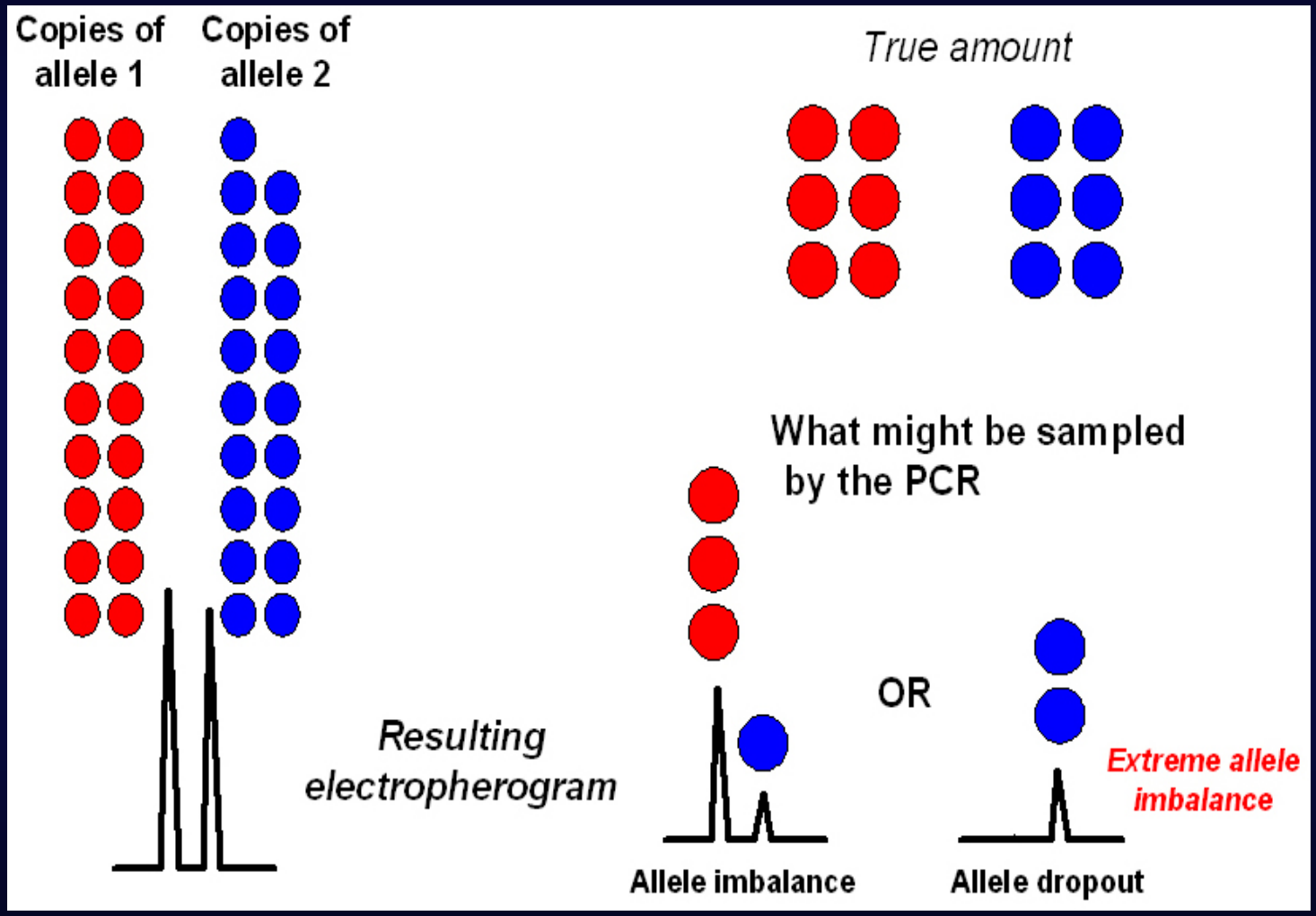
What is LCN?

- DNA profiling performed at or beneath the stochastic threshold
- Typically less than 0.5 ng of DNA template
- Typically involves modifications of the testing methodology (e.g. increased polymerase; additional rounds of amplification; skipping quantitation)
- Consensus profiles

Stochastic effects

- Ultimately due to poor statistical sampling of underlying template
- The four horsemen of stochasticism
  - Exaggerated stutter
  - Exaggerated peak height imbalance (0 to 100%)
  - Allelic drop-out (extreme peak height imbalance)
  - Allelic drop-in (contamination)

Stochastic sampling effects



Stochastic effects

- Ultimately due to poor statistical sampling of underlying template
- The four horsemen of stochasticism
  - Exaggerated stutter (up to 50%)
  - Exaggerated peak height imbalance (0 to 100%)
  - Allelic drop-out (extreme peak height imbalance)
  - Allelic drop-in (contamination)

Comparison of STR Kit Amplification SOP with LCN

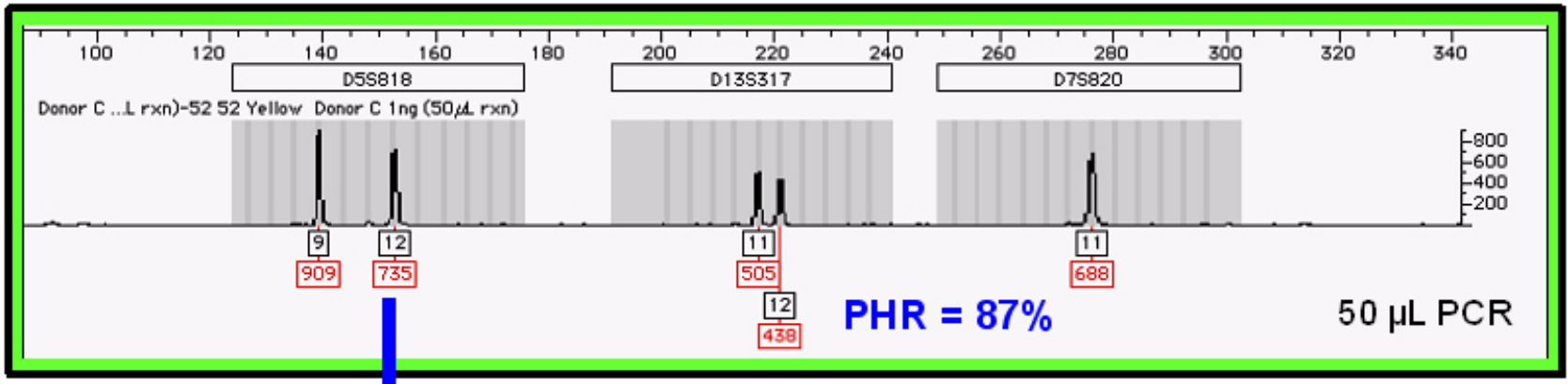
Using the Same DNA Donor

Input DNA

Data from Debbie Hobson (FBI) – LCN Workshop AAFS 2003

SOP

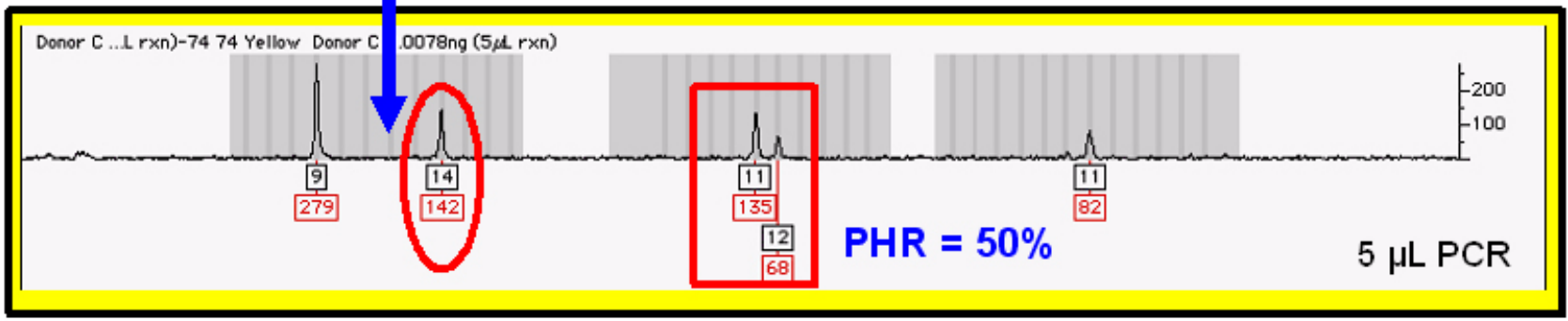
1ng



Allele Drop Out

LCN

8pg



Allele Drop In

Heterozygote Allele Imbalance

Impact of DNA Amount into Multiplex PCR

DNA amount
(log scale)

100 ng

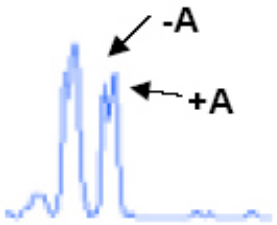
10 ng

1 ng

0.1 ng

0.01 ng

High levels of DNA create interpretation challenges (more artifacts to review)



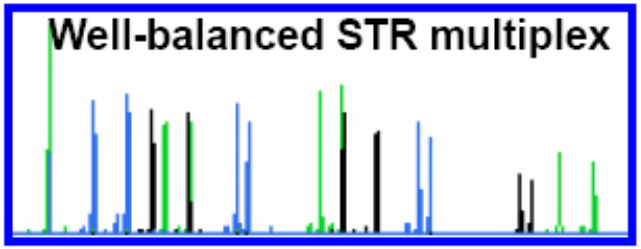
Too much DNA

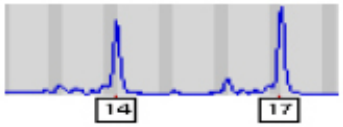
- Off-scale peaks
- Split peaks (+/-A)
- Locus-to-locus imbalance

2.0 ng

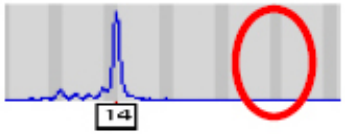
STR Kits Work Best in This Range

0.5 ng





100 pg template



5 pg template

Too little DNA

- Heterozygote peak imbalance
- Allele drop-out
- Locus-to-locus imbalance

Stochastic effects when amplifying low levels of DNA can produce allele dropout

How helpful is quantitation?

- Optimum amount of template: 0.5 to 2.0 ng
- 6 to 7 pg of DNA in each diploid human cell
- In a mixed sample containing 0.5 ng of template, less than 0.5 ng comes from each contributor

Assume sample is a **1:1 mixture** of two sources:

Amount of DNA	Total Cells in sample	~ # of cells from each component
1 ng	152	76
0.5 ng	76	38
0.25 ng	38	19
0.125 ng	19	10
0.0625 ng	10	5

Robin Cotton, AAFS 2003 LCN Workshop
"Are we already doing low copy number (LCN) DNA analysis?"

Assume sample is a **1:9 mixture** of two sources:

Amount of DNA	~ # of cells from **major component**	~ # of cells from **minor component**
1ng	137	15
0.5ng	68	8
0.25ng	34	4
0.125ng	17	2
0.0625ng	9	1

Robin Cotton, AAFS 2003 LCN Workshop
"Are we already doing low copy number (LCN) DNA analysis?"

Consensus profiles

- Alleles are not reported unless they are seen in at least two runs
- Considering two runs serves as a safeguard against allelic drop-in (contamination)
- Considering three or more runs begins to safeguard against drop-out
- If a sample is being split four or more times, shouldn’t conventional tests be done?

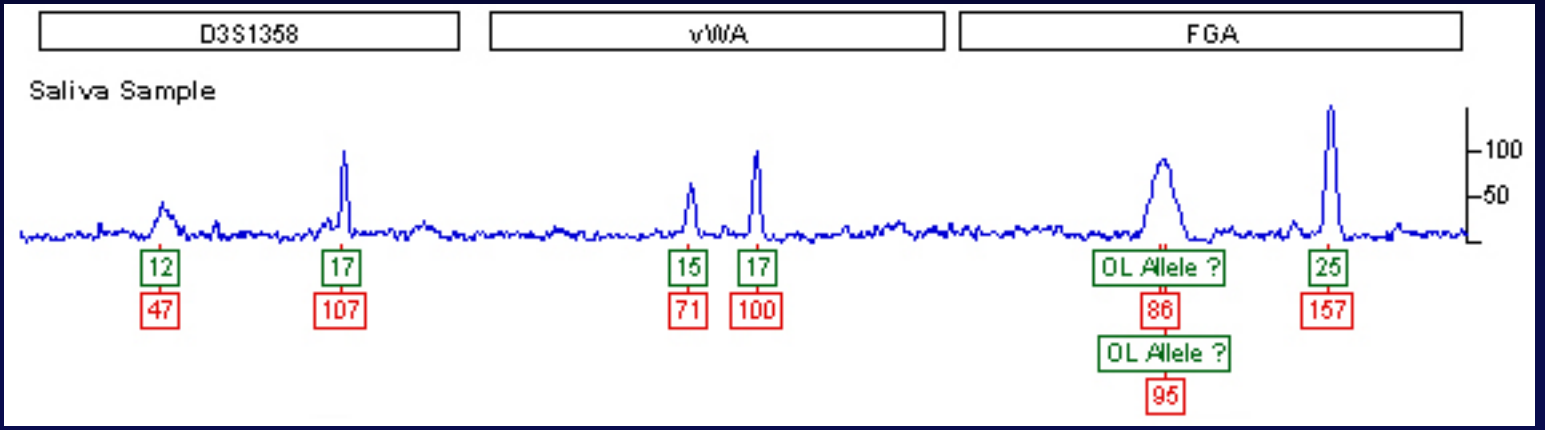
Consensus profiles

Runs used to make consensus	D3	vWA	D16	D2	D8	D21	D18	D19	THO1	FGA
1+2+3	16 17	17	10 13	20	10 13	28 30		12 13 14 15	9.3	23 24
1+2	16 17	17	13	20	10 13	30		12 13 14 15		
1+3	16 17		13	20	10 13	30		13 14 15		
2+3	16 17		10 13	20	10 13	28 30		13 14 15	9.3	23 24

Sources of ambiguity in STR interpretation

- Degradation
- Allelic dropout
- False peaks
- Mixtures
- Accounting for relatives
- Threshold issues -- marginal samples

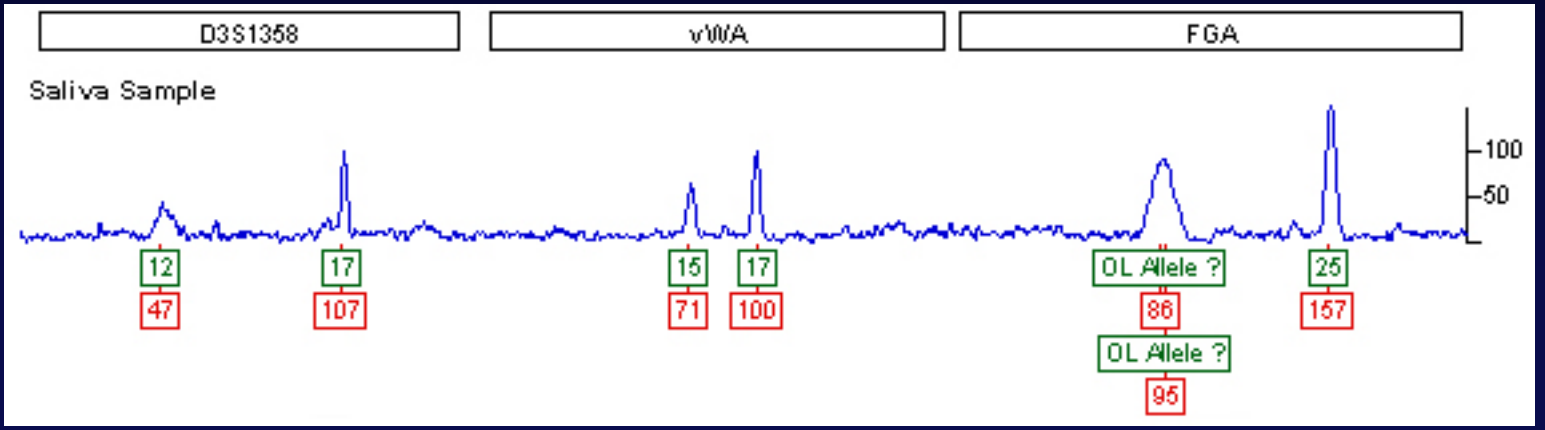
Opportunities for subjective interpretation?



Can “Tom” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25

Opportunities for subjective interpretation?

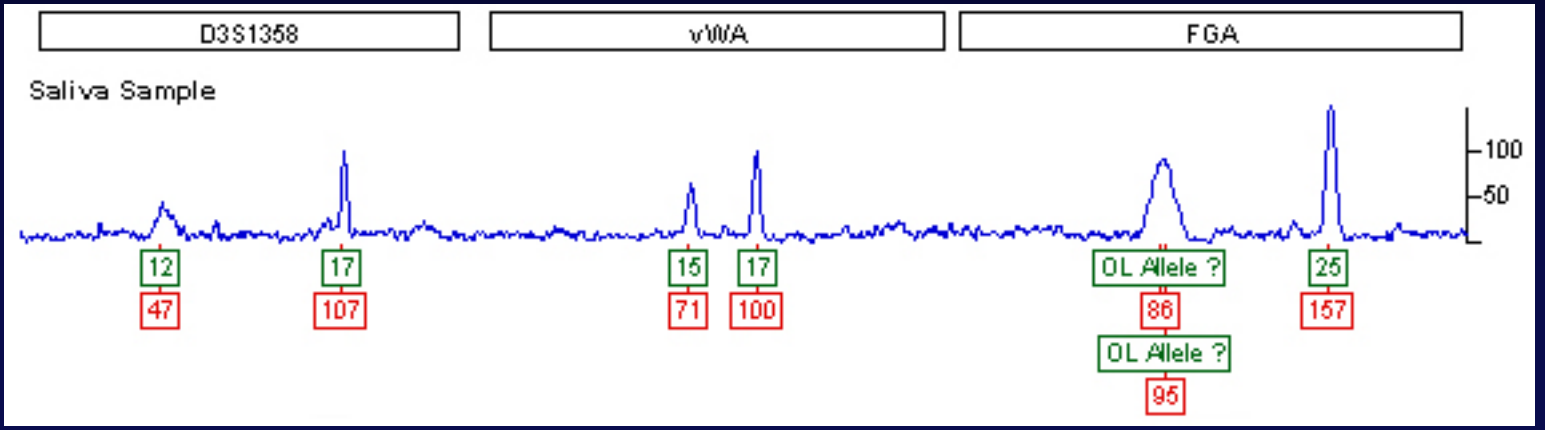


Can “Tom” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25

No -- the additional alleles at D3 and FGA are “technical artifacts.”

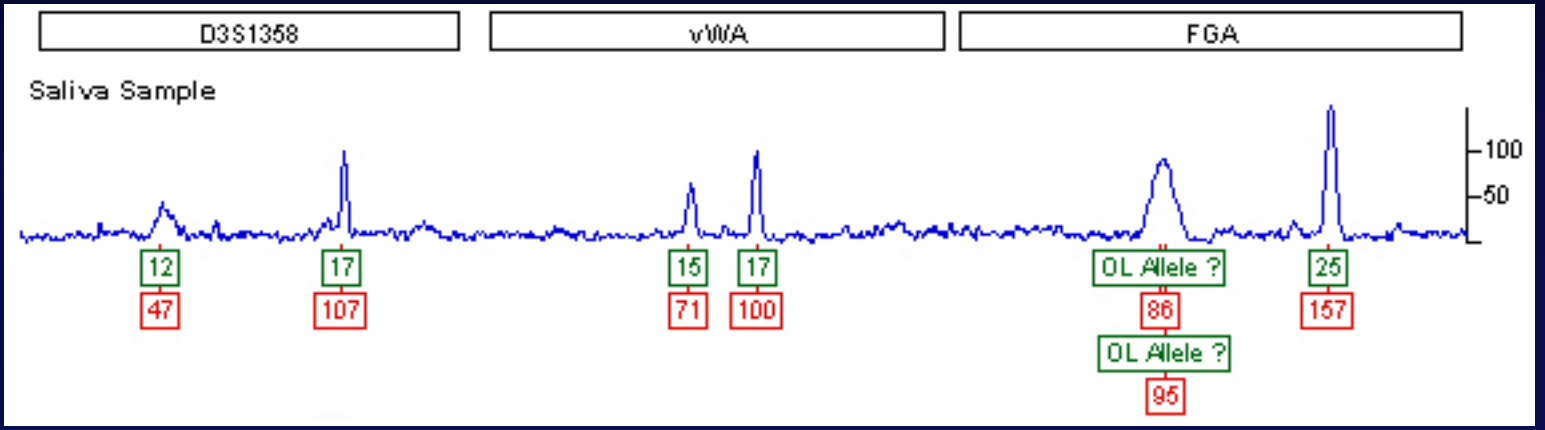
Opportunities for subjective interpretation?



Can “Dick” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25
Dick	12, 17	15, 17	20, 25

Opportunities for subjective interpretation?

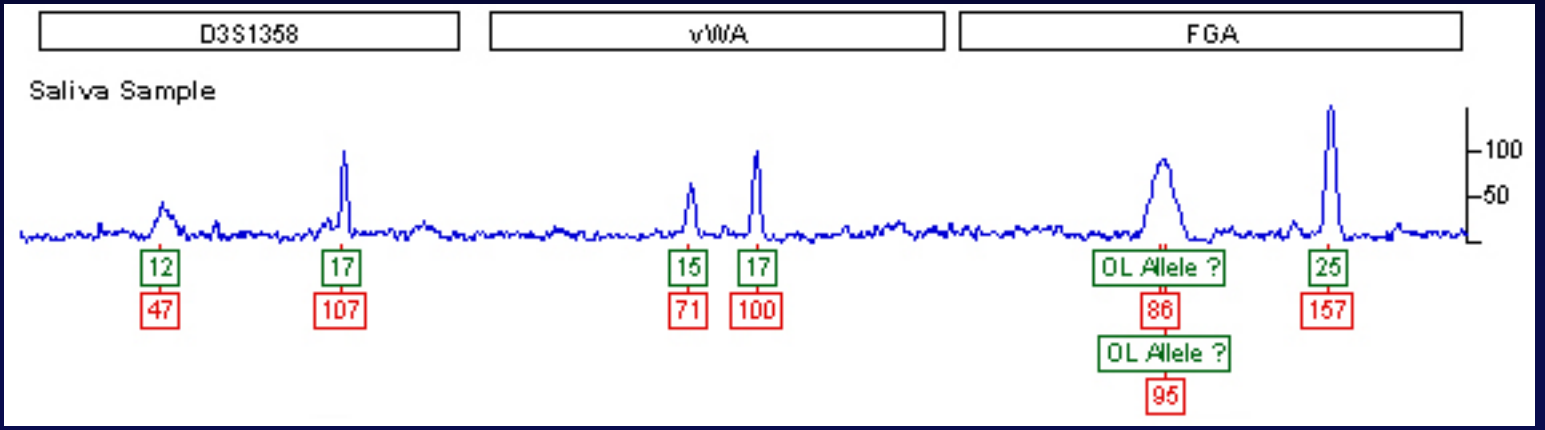


Can “Dick” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25
Dick	12, 17	15, 17	20, 25

No -- stochastic effects explain peak height disparity in D3; blob in FGA masks 20 allele.

Opportunities for subjective interpretation?

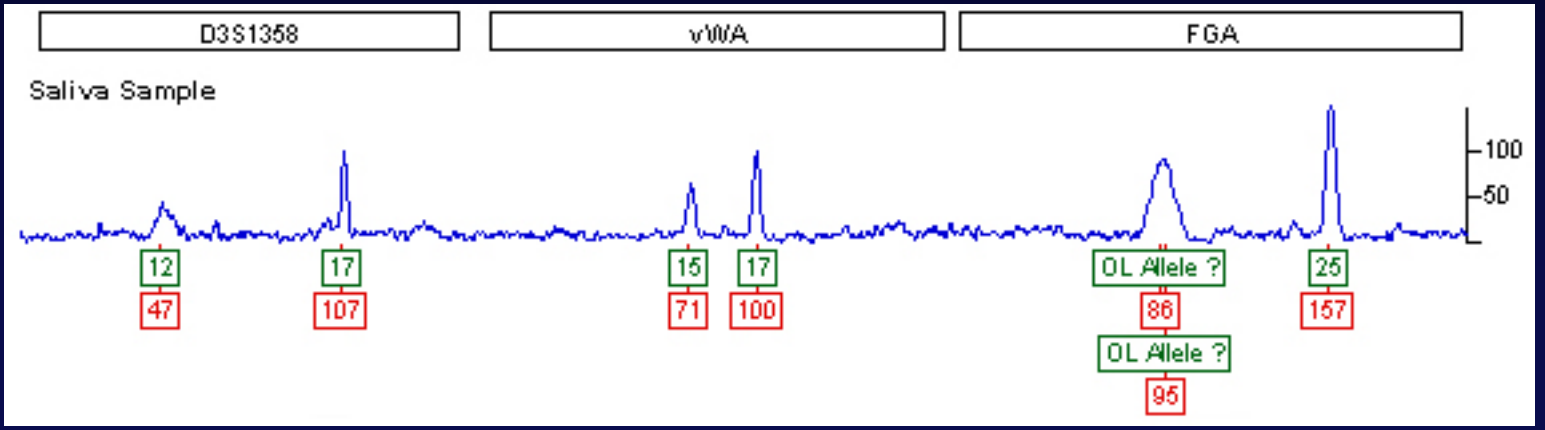


Can “Harry” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25
Dick	12, 17	15, 17	20, 25
Harry	14, 17	15, 17	20, 25

No -- the 14 allele at D3 may be missing due to “allelic drop out”; FGA blob masks the 20 allele.

Opportunities for subjective interpretation?

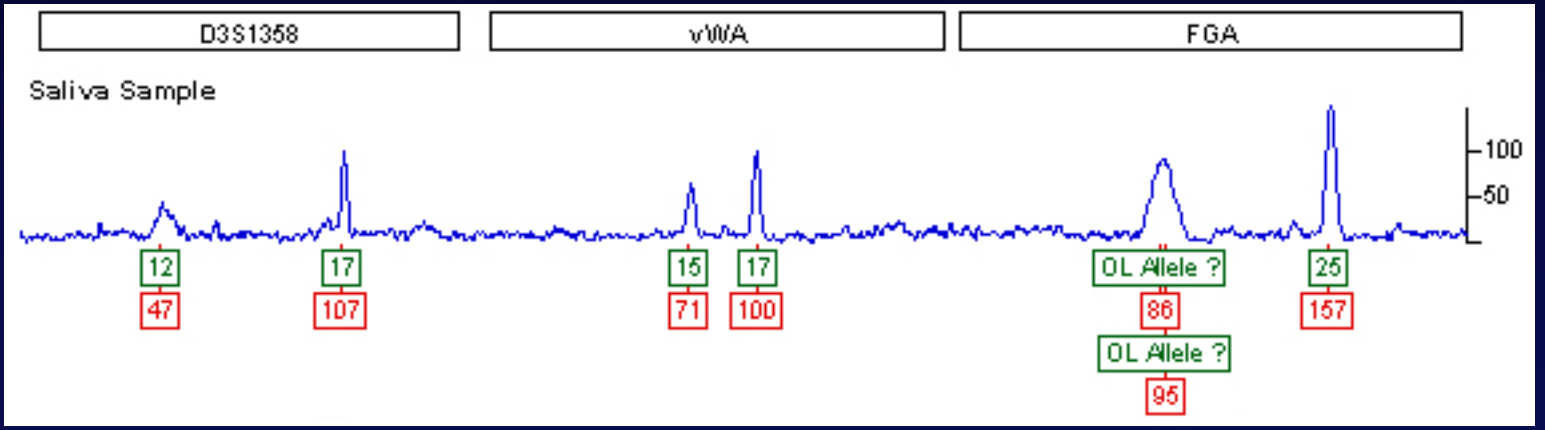


Can “Sally” be excluded?

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25
Dick	12, 17	15, 17	20, 25
Harry	14, 17	15, 17	20, 25
Sally	12, 17	15, 15	20, 22

No -- there must be a second contributor; degradation explains the “missing” FGA allele.

Subjective interpretation and statistics



Frequency estimates (for Tom):

p^2 x 2pq x p^2

Suspect	D3	vWA	FGA
Tom	17, 17	15, 17	25, 25
Dick	12, 17	15, 17	20, 25
Harry	14, 17	15, 17	20, 25
Sally	12, 17	15, 15	20, 22

Partial Profile Statistics

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Forensic Science International 160 (2006) 90–101

Forensic Science International

www.elsevier.com/locate/forsciint

DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures

P. Gill [a,*], C.H. Brenner [b], J.S. Buckleton [c], A. Carracedo [d], M. Krawczak [e], W.R. Mayr [f], N. Morling [g], M. Prinz [h], P.M. Schneider [i], B.S. Weir [j]

There are three kinds of alleles in a crime stain profile:

A. alleles that are unmistakeable;
B. alleles that may be masked by an artefact such as a stutter;
C. alleles that have dropped out completely and are therefore not detected.

In particular it is problematical to apply the method when there are loci which, under the hypothesis being considered of the suspect at hand, appear to have alleles in category C. We have seen many instances in which laboratories do just this, usually by omitting from the RMNE calculation the inconvenient loci. Such a calculation implies, certainly incorrectly, that among the "random men" considered for comparison by the calculation only the same loci would be used for inculpation/exculpation as those being considered for the present suspect. It fails to acknowledge that choosing the omitted loci is suspect-centric and therefore prejudicial against the suspect. (If, on the other hand, a locus is eliminated from analysis simply because it is a poor result showing no alleles at all, then of course there is no prejudice in ignoring it.)

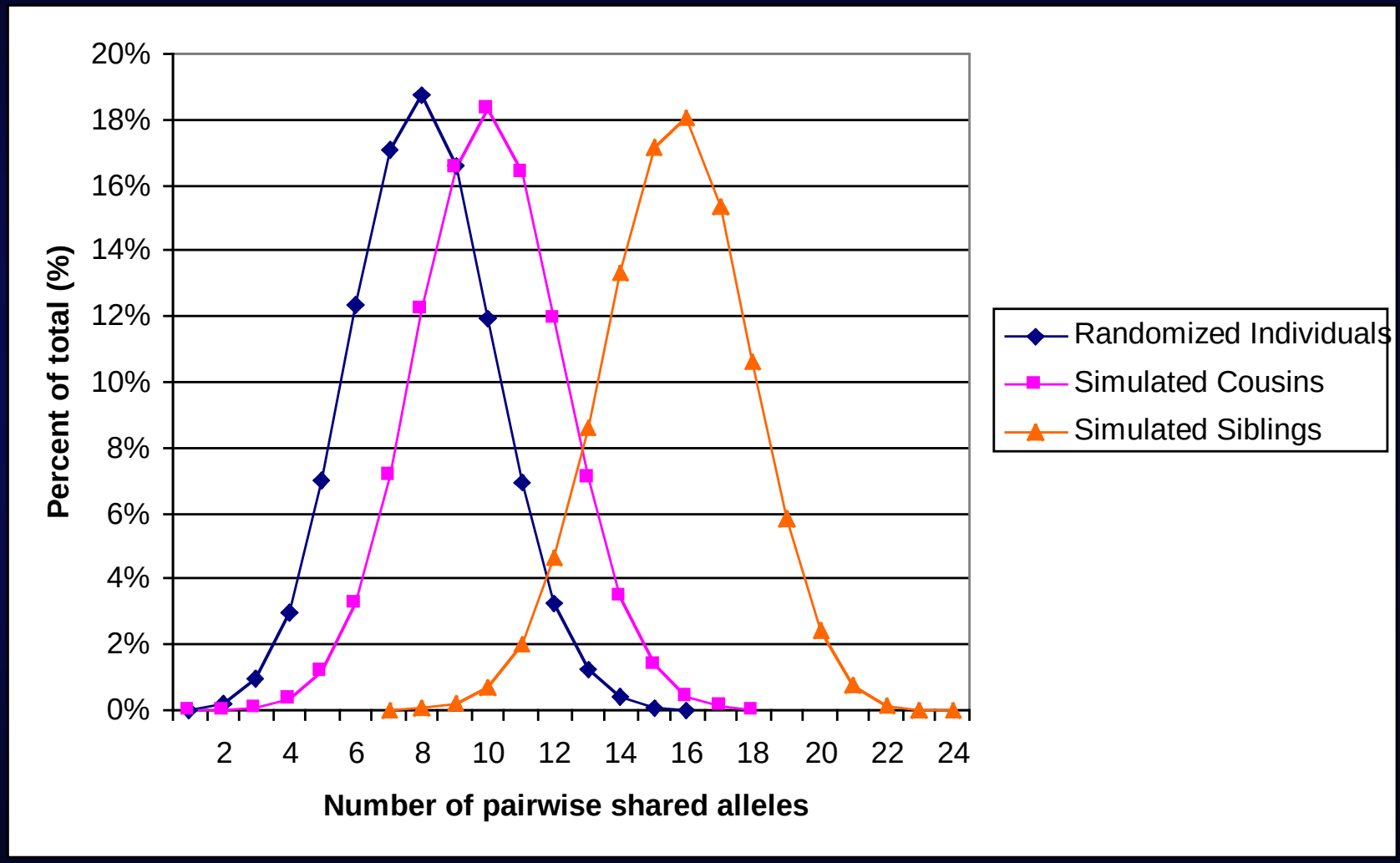
Familial searching

- Database search yields a close but imperfect DNA match
- Can suggest a relative is the true perpetrator
- Great Britain performs them routinely
- Reluctance to perform them in US since 1992 NRC report
- Current CODIS software cannot perform effective searches

Three approaches to familial searches

- Search for rare alleles (inefficient)
- Count matching alleles (arbitrary)
- Likelihood ratios with kinship analyses

Pair-wise similarity distributions



Is the true DNA match a relative or a random individual?

- Given a closely matching profile, who is more likely to match, a relative or a randomly chosen, unrelated individual?

- Use a likelihood ratio

$$LR = \frac{P(E \mid relative)}{P(E \mid random)}$$

Is the true DNA match a relative or a random individual?

- What is the likelihood that a relative of a single initial suspect would match the evidence sample perfectly?
- What is the likelihood that a single randomly chosen, unrelated individual would match the evidence sample perfectly?

$$LR = \frac{P(E \mid relative)}{P(E \mid random)}$$

Probabilities of siblings matching at 0, 1 or 2 alleles

$$P(E \mid sib) = \begin{cases} \dfrac{P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 0 \\ \dfrac{P_b + P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 1 \\ \dfrac{1 + P_a + P_b + P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 2 \end{cases}$$

HF = 1 for homozygous loci and 2 for heterozygous loci;
P_a is the frequency of the allele shared by the evidence sample and the individual in a database.

Probabilities of parent/child matching at 0, 1 or 2 alleles

$$P(E \mid parent/child) = \begin{cases} 0, & if \quad shared = 0 \\ \frac{P_b}{2}, & if \quad shared = 1 \\ \frac{P_a + P_b}{2}, & if \quad shared = 2 \end{cases}$$

HF = 1 for homozygous loci and 2 for heterozygous loci; P_a is the frequency of the allele shared by the evidence sample and the individual in a database.

Other familial relationships

Cousins:

$$P(E \mid cousins) = \begin{cases} \dfrac{6 \cdot P_a \cdot P_b \cdot HF}{8}, & if \quad shared = 0 \\ \dfrac{P_b + 6 \cdot P_a \cdot P_b \cdot HF}{8}, & if \quad shared = 1 \\ \dfrac{P_a + P_b + 6 \cdot P_a \cdot P_b \cdot HF}{8}, & if \quad shared = 2 \end{cases}$$

Grandparent-grandchild; aunt/uncle-nephew-neice;half-sibings:

$$P(E \mid GG / AUNN / HS) = \begin{cases} \dfrac{2 \cdot P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 0 \\ \dfrac{P_b + 2 \cdot P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 1 \\ \dfrac{P_a + P_b + 2 \cdot P_a \cdot P_b \cdot HF}{4}, & if \quad shared = 2 \end{cases}$$

HF = 1 for homozygous loci and 2 for heterozygous loci;
P_a is the frequency of the allele shared by the evidence sample and the individual in a database.

Familial search experiment

- Randomly pick related pair or unrelated pair from a synthetic database
- Choose one profile to be evidence and one profile to be initial suspect
- Test hypothesis:
  - H_0: A relative is the source of the evidence
  - H_A: An unrelated person is the source of the evidence

Paoletti, D., Doom, T., Raymer, M. and Krane, D. 2006. Assessing the implications for close relatives in the event of similar but non-matching DNA profiles. *Jurimetrics*, 46:161-175.

Hypothesis testing using an LR threshold of 1 with prior odds of 1

		True state	
		Evidence from Unrelated individual	Evidence from sibling
Decision	Evidence from unrelated individual	~ 98% [Correct decision]	~4% [Type II error; false negative]
	Evidence from sibling	~ 2% [Type I error; false positive]	~ 96% [Correct decision]

Is the true DNA match a relative or a random individual?

- What is the likelihood that a close relative of a single initial suspect would match the evidence sample perfectly?
- What is the likelihood that a single randomly chosen, unrelated individual would match the evidence sample perfectly?

$$LR = \frac{P(E \mid relative)}{P(E \mid random)}$$

Is the true DNA match a relative or a random individual?

- What is the likelihood that the source of the evidence sample was a relative of an initial suspect?

$$P(sib \mid E) = \frac{P(E \mid sib) \cdot P(sib)}{P(E \mid sib) \cdot P(sib) + P(E \mid random) \cdot P(random)}$$

Prior odds:

$$P(sib) = \frac{s}{popsize}$$

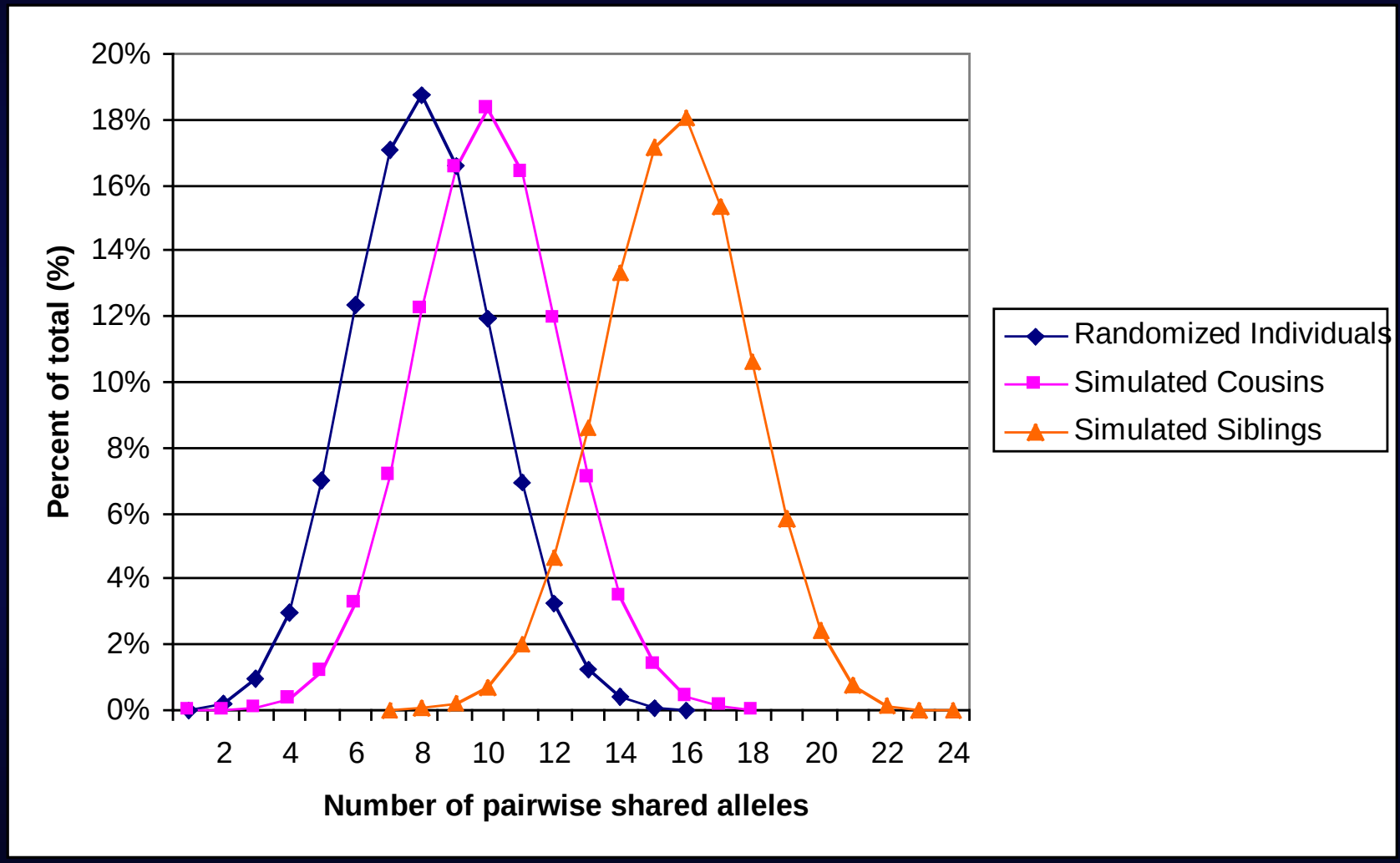
$$P(random) = \frac{popsize - s}{popsize}$$

Is the true DNA match a relative or a random individual?

- This more difficult question is ultimately governed by two considerations:
  - What is the size of the alternative suspect pool?
  - What is an acceptable rate of false positives?

$$LR = \frac{P(E \mid sib)}{P(E \mid random)}$$

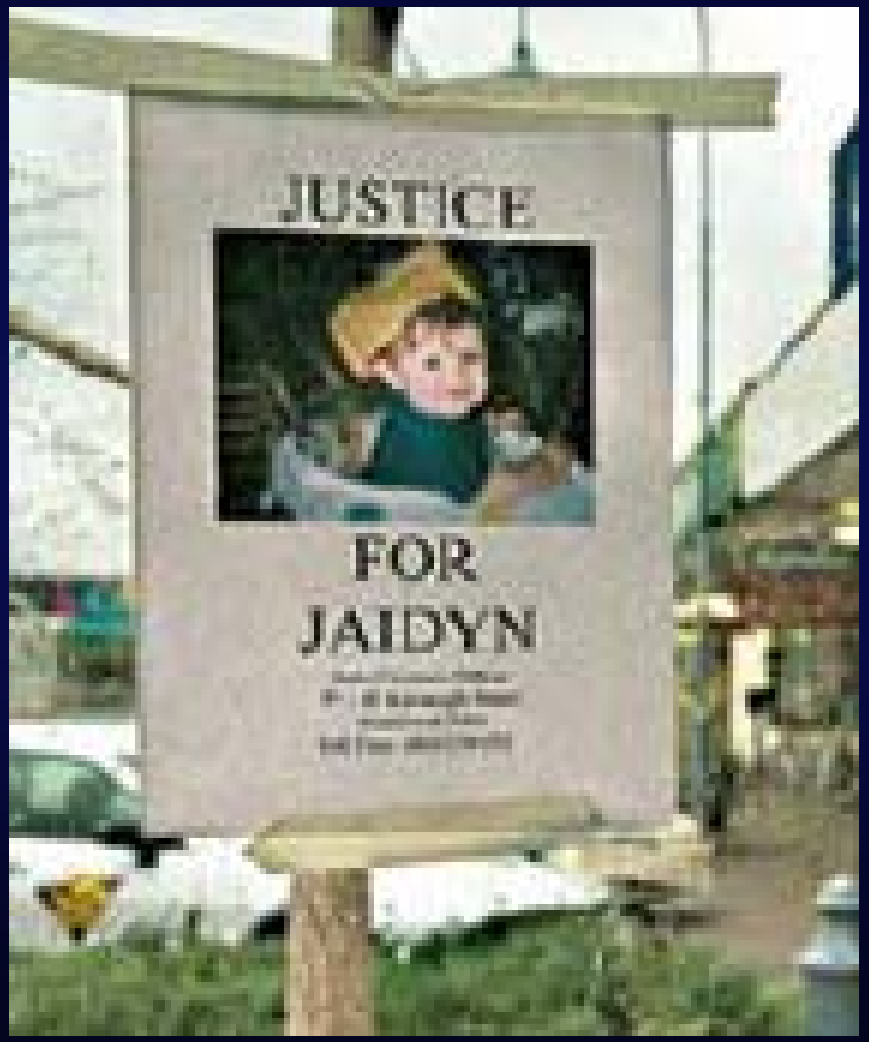
Pair-wise similarity distributions



III: What can go wrong and where problems might occur

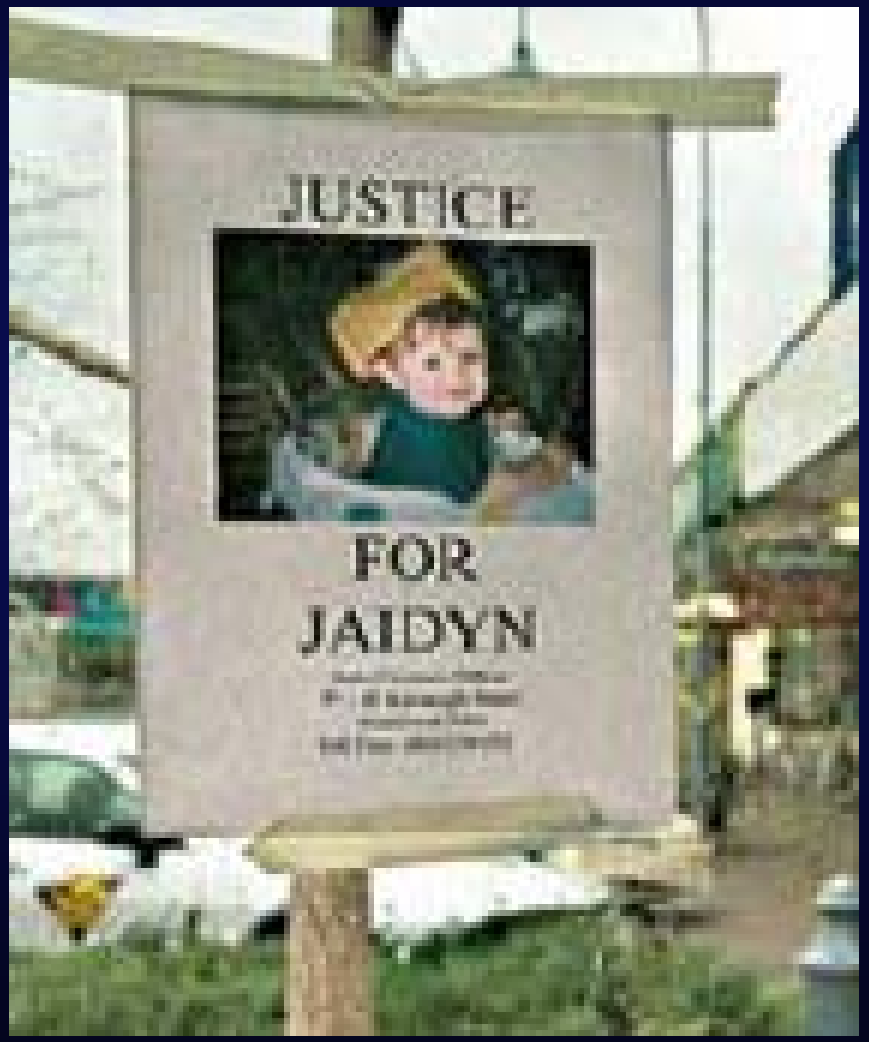
Victorian Coroner's inquest into the death of Jaidyn Leskie

- Toddler disappears in bizarre circumstances: found dead six months later
- Mother's boy friend is tried and acquitted.
- Unknown female profile on clothing.
- Cold hit to a rape victim.
- RMP: 1 in 227 million.
- Lab claims "adventitious match."

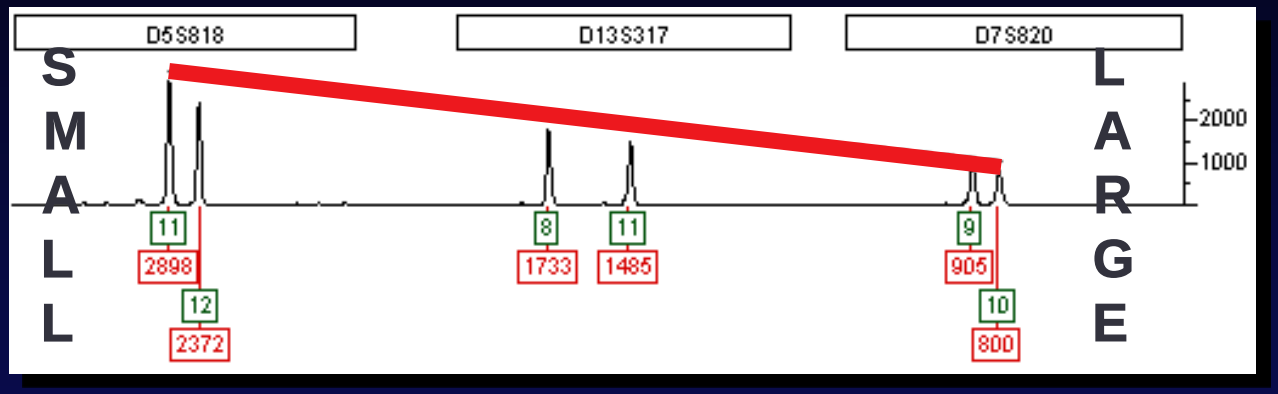


Victorian Coroner's inquest into the death of Jaidyn Leskie

- Condom with rape victim's DNA was processed in the same lab 1 or 2 days prior to Leskie samples.
- Additional tests find matches at 5 to 7 more loci.
- Review of electronic data reveals low level contributions at even more loci.
- Degradation study further suggests contamination.



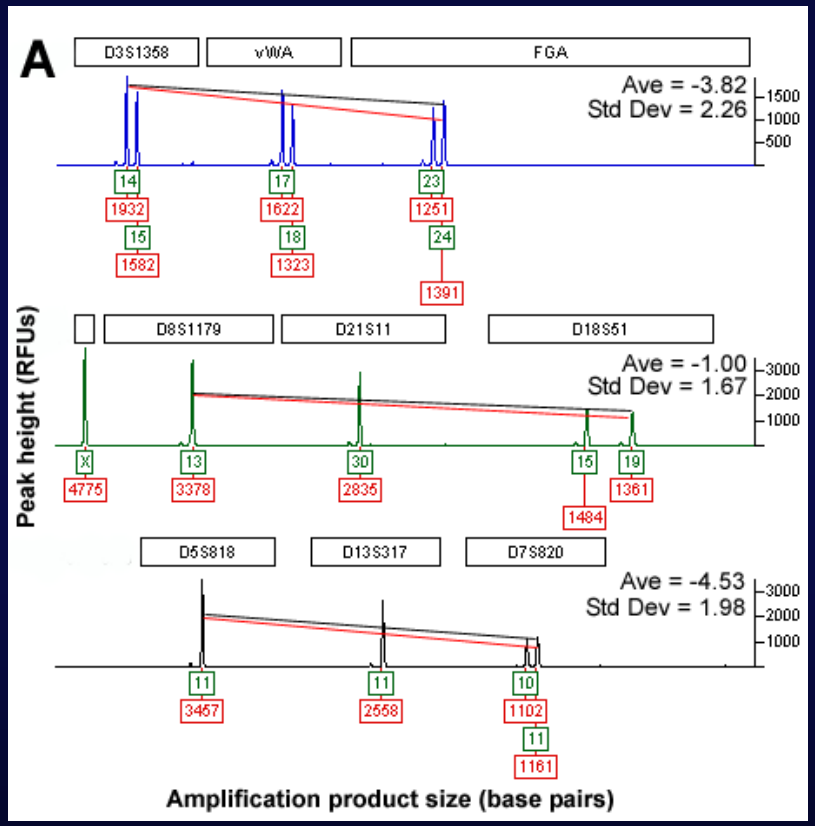
Degradation, inhibition



- When biological samples are exposed to adverse environmental conditions, they can become degraded
  - **Warm, moist, sunlight, time**
- Degradation breaks the DNA at random
- Larger amplified regions are affected first
- Classic ‘ski-slope’ electropherogram
- Degradation and inhibition are unusual and noteworthy.

Degradation, inhibition

The Leskie Inquest, a practical application

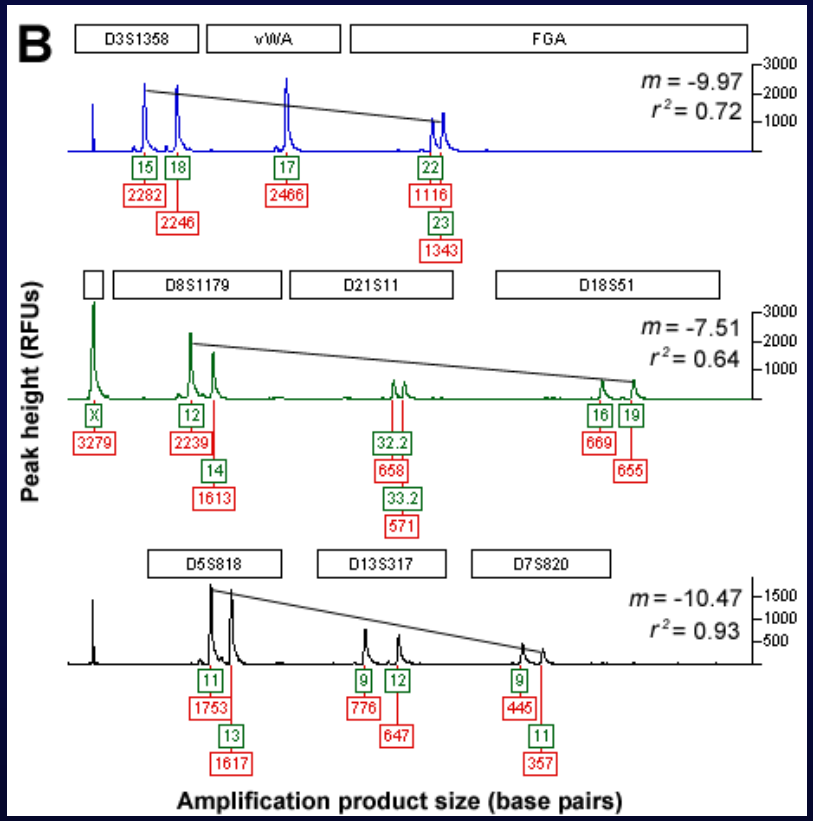


- Undegraded samples can have “ski-slopes” too.
- How negative does a slope have to be to an indication of degradation?
- Experience, training and expertise.
- Positive controls should not be degraded.

Degradation, inhibition

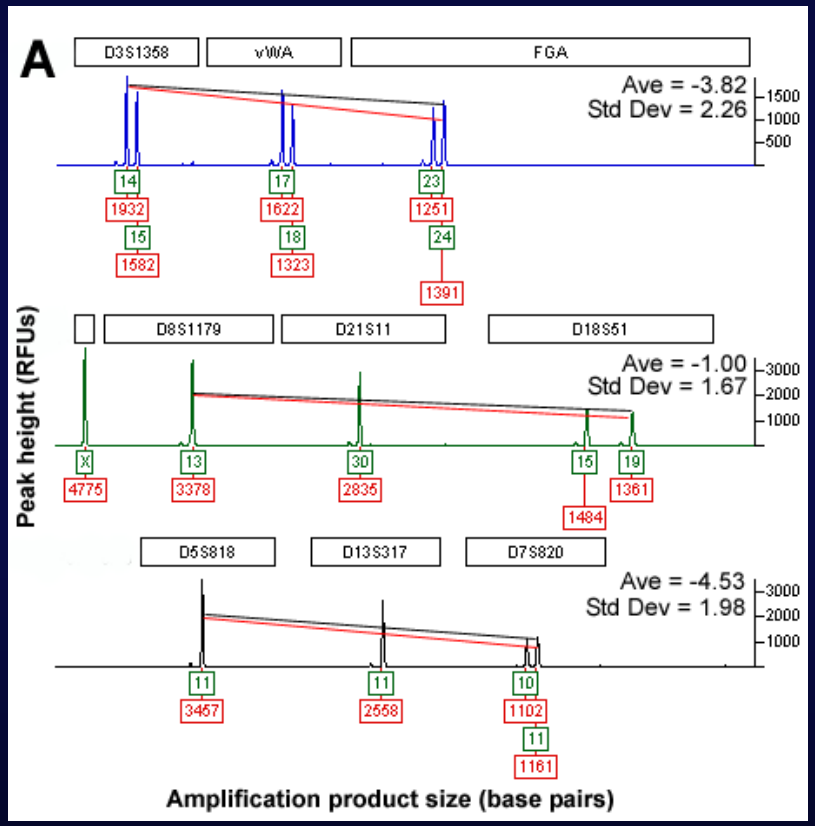
The Leskie Inquest

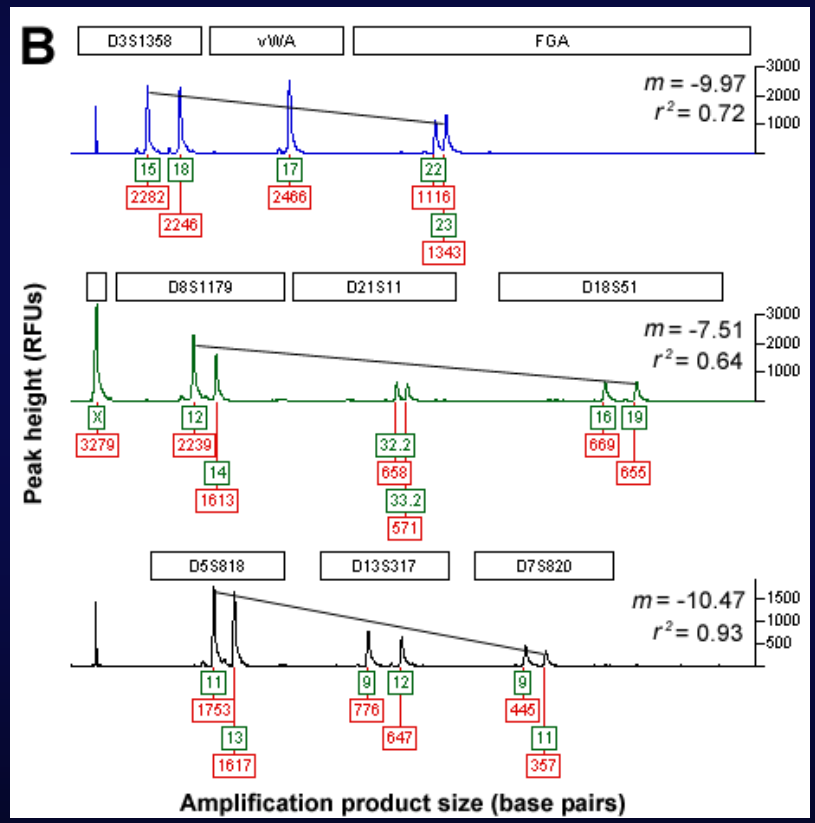
- DNA profiles in a rape and a murder investigation match.
- Everyone agrees that the murder samples are degraded.
- If the rape sample is degraded, it could have contaminated the murder samples.
- Is the rape sample degraded?



Degradation, inhibition

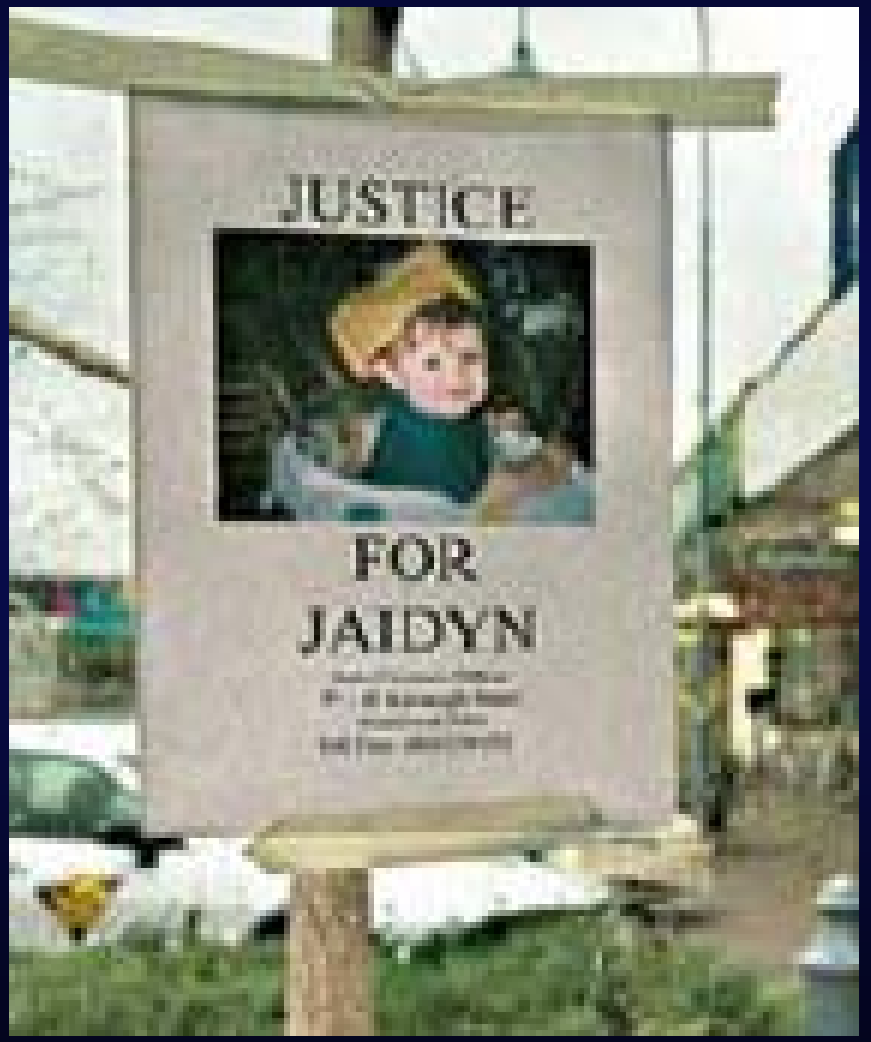
The Leskie Inquest





Victorian Coroner's inquest into the death of Jaidyn Leskie

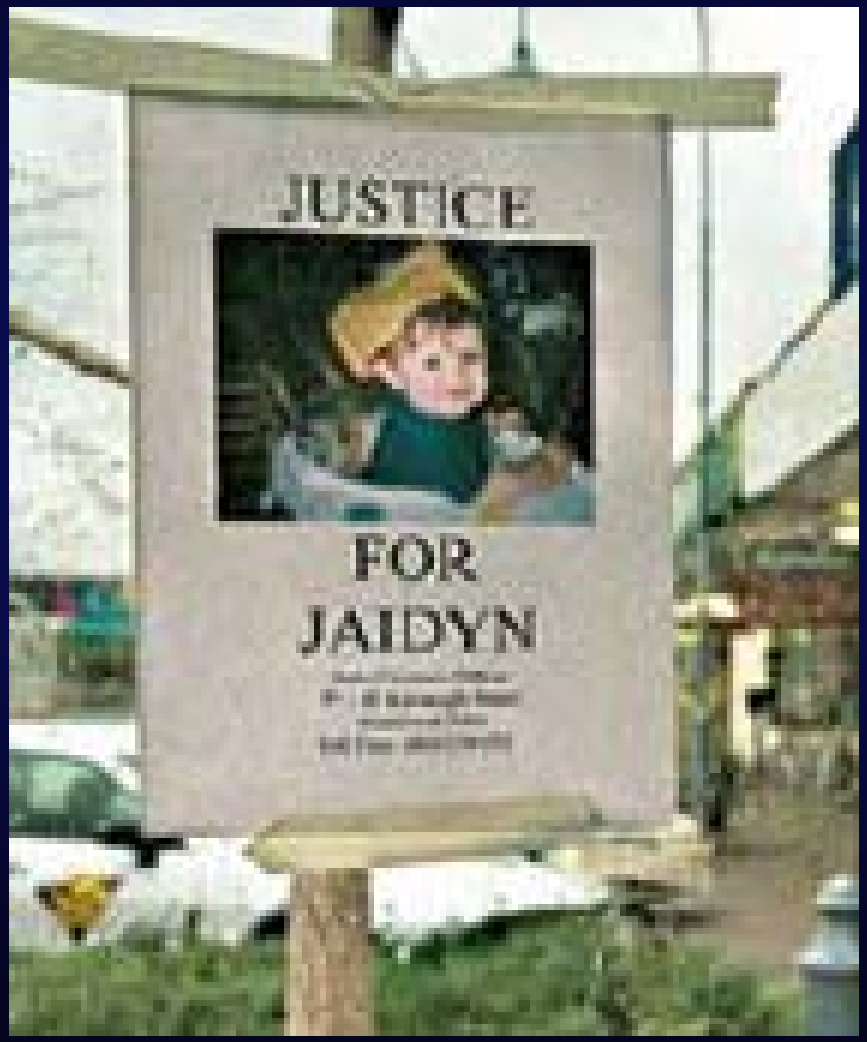
"8. During the conduct of the preliminary investigation (before it was decided to undertake an inquest) the female DNA allegedly taken from the bib that was discovered with the body was matched with a DNA profile in the Victorian Police Forensic Science database. This profile was from a rape victim who was subsequently found to be unrelated to the Leskie case."



Victorian Coroner's inquest into the death of Jaidyn Leskie

"8. The match to the bib occurred as a result of contamination in the laboratory and was not an adventitious match. The samples from the two cases were examined by the same scientist within a close time frame."

www.bioforensics.com/articles/Leskie_decision.pdf



The science of DNA profiling is sound.

But, not all of DNA profiling is science.

This is especially true in situations involving: small amounts of starting material, mixtures, relatives, and analyst judgment calls.

Steps in Preparing a DNA Case

- Obtain all lab reports
- Red flags:
  - unfamiliar techniques
  - equivocal matches (profile “similar but cannot be definitively included nor excluded”);
  - contingent matches (profile included “if…” or “but…”;
  - partial/incomplete profiles;
  - mixtures;
  - unusually modest statistics; no statistics; likelihood ratios

Steps in Preparing a DNA Case

- Initial discovery
  - Full history of all samples from collection to current disposition
  - Complete DNA lab notes (bench notes)
  - Electronic data
  - Analysts' credentials, proficiency test record
  - Lab's incidence reports; unexpected event files; accreditation files
- Obtain expert assistance for initial review

Steps in Preparing a DNA Case

- Initial evaluation of case
  - Identify possible lines of attack
  - Additional/alternative experts needed
  - Needs for follow-up discovery—e.g., validation; proficiency problems; error problems
- Consider advisability of additional testing
  - Replications; untested items; other experiments
- Final evaluation of strategy
  - Consider ways to blunt/deflect prosecution (or defense) testimony
- Prepare exhibits, lines of examination, motions in limine; notices of objection, etc.

Resources

- Internet
  - **Forensic Bioinformatics Website: http://www.bioforensics.com/**
  - **Applied Biosystems Website: http://www.appliedbiosystems.com/ (see human identity and forensics)**
  - **STR base: http://www.cstl.nist.gov/biotech/strbase/ (very useful)**
- Books
  - **'Forensic DNA Typing' by John M. Butler (Academic Press)**
- Scientists
  - **Larry Mueller (UC Irvine)**
  - **Simon Ford (Lexigen, Inc. San Francisco, CA)**
  - **William Shields (SUNY, Syracuse, NY)**
  - **Mike Raymer and Travis Doom (Wright State, Dayton, OH)**
  - **Marc Taylor (Technical Associates, Ventura, CA)**
  - **Keith Inman (Forensic Analytical, Haywood, CA)**
- Testing laboratories
  - **Technical Associates (Ventura, CA)**
  - **Forensic Analytical (Haywood, CA)**
- Other resources
  - **Forensic Bioinformatics (Dayton, OH)**