Bruce McCord, Professor of Analytical/Forensic Chemistry

Department of Chemistry and Biochemistry, Florida International University, Miami, FL

http://faculty.fiu.edu/~mccordb/

mccordb@fiu.edu

FIU
FLORIDA INTERNATIONAL UNIVERSITY
Miami's public research university

IFRI International Forensic Research Institute
Florida International University

Promega

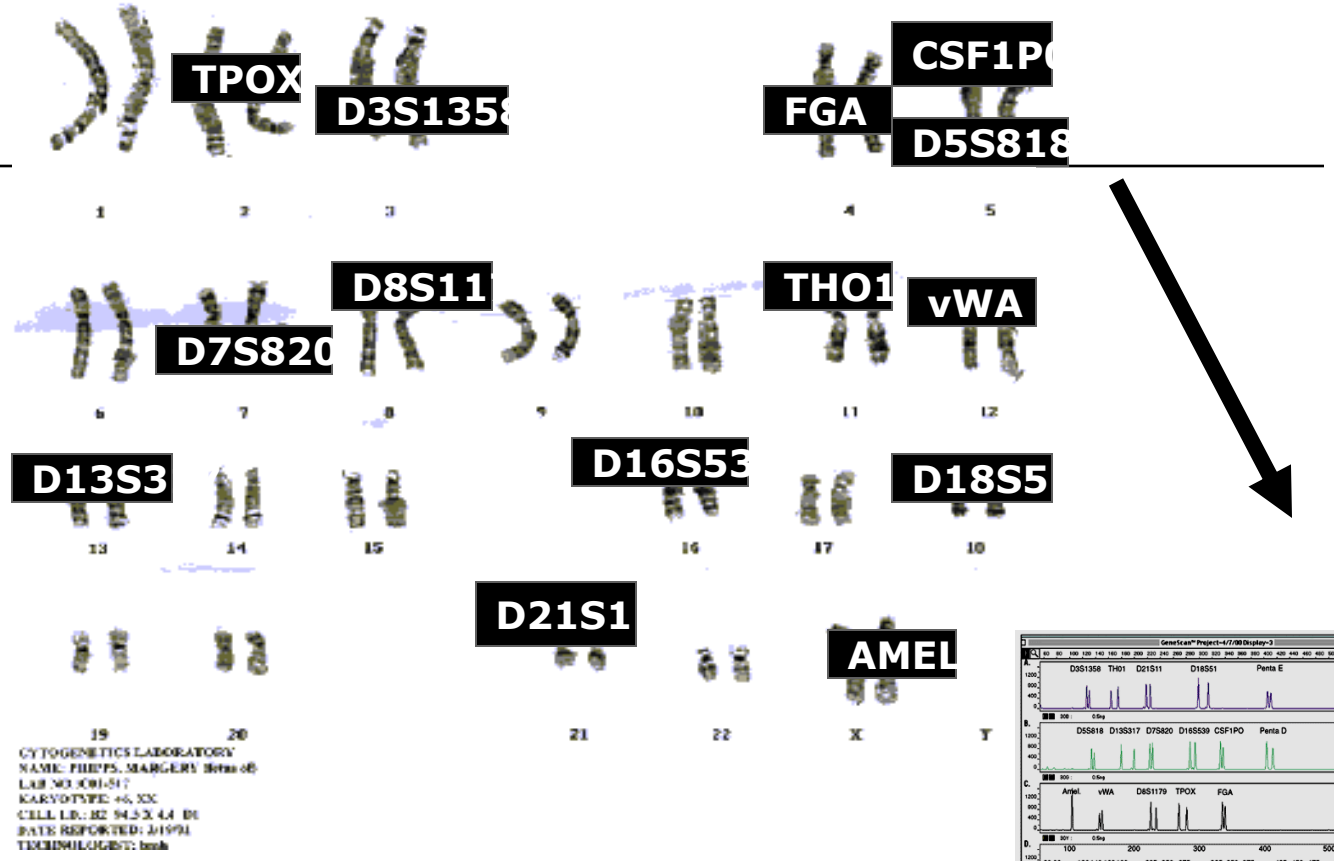
Forensic MAGAZINE

Capillary electrophoresis: analysis and related interpretational issues

Outline

- Overview of STR analysis by CE
  - Separation
  - Injection
  - Detection
- Setting thresholds
  - Analytical figures of merit
  - Sensitivity studies
  - Interpretation of low level DNA

With 23 pairs of Chromosomes you can get Specific

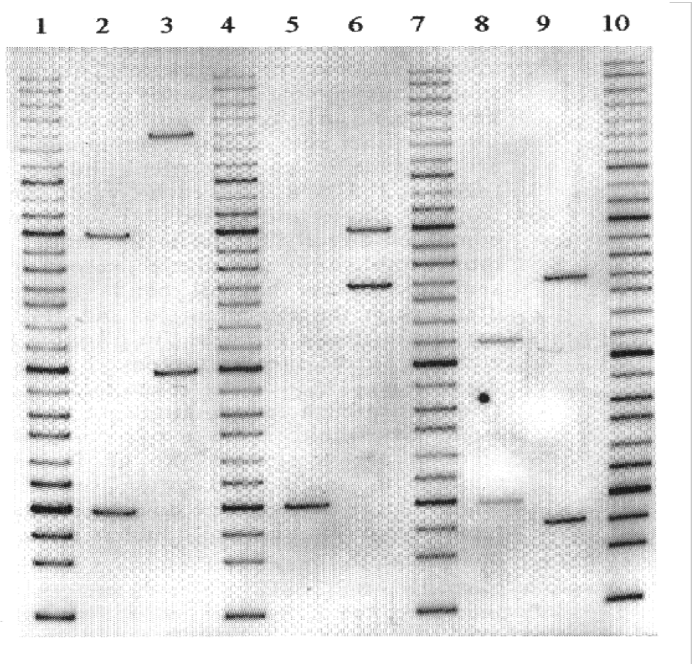


Melannie McCord

The Random Match Probability for this profile in the FBI Caucasian population is 1 in 1.56 quadrillion (10^{15})

But How To Process All This Data?

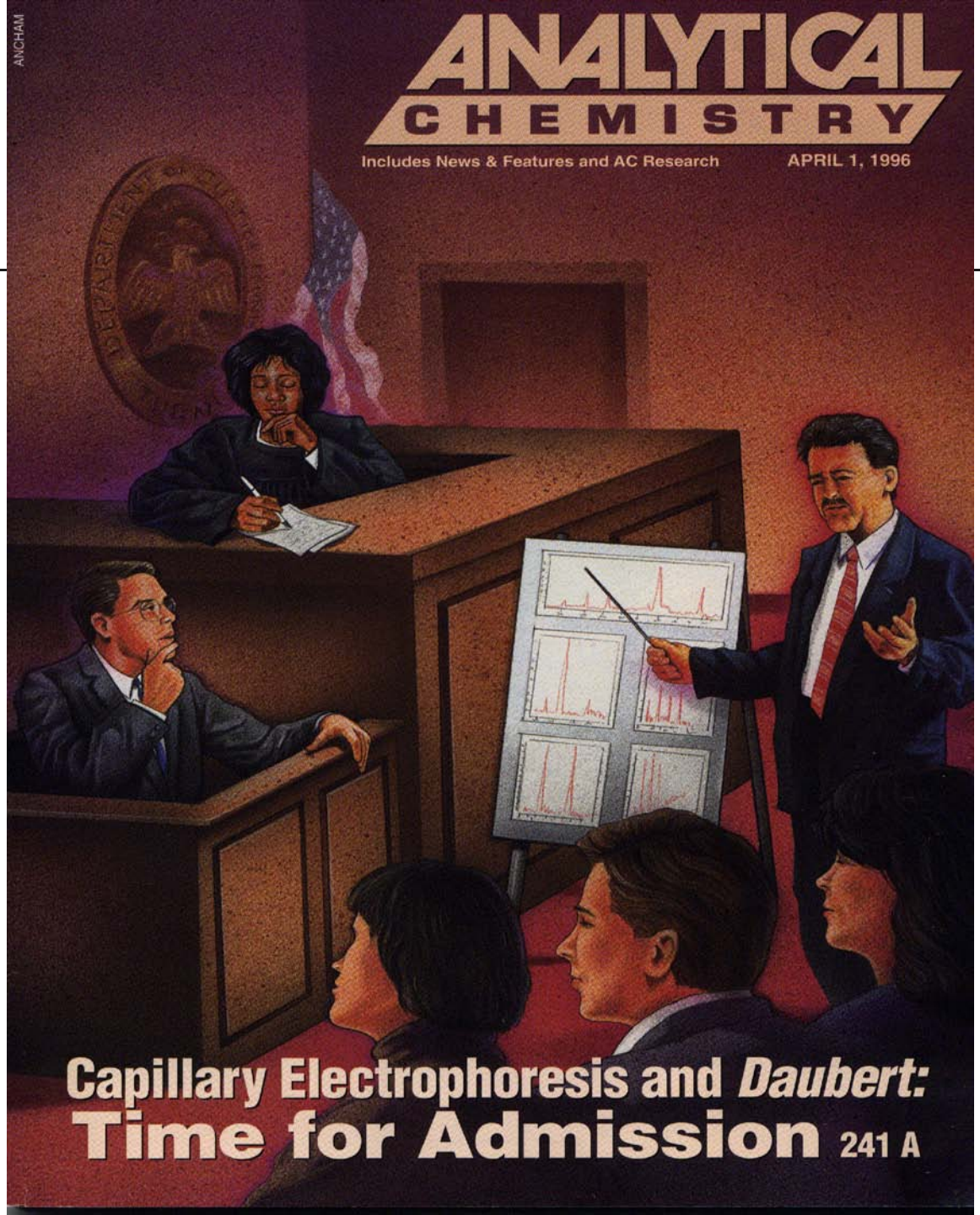
Hundreds of thousands of samples?



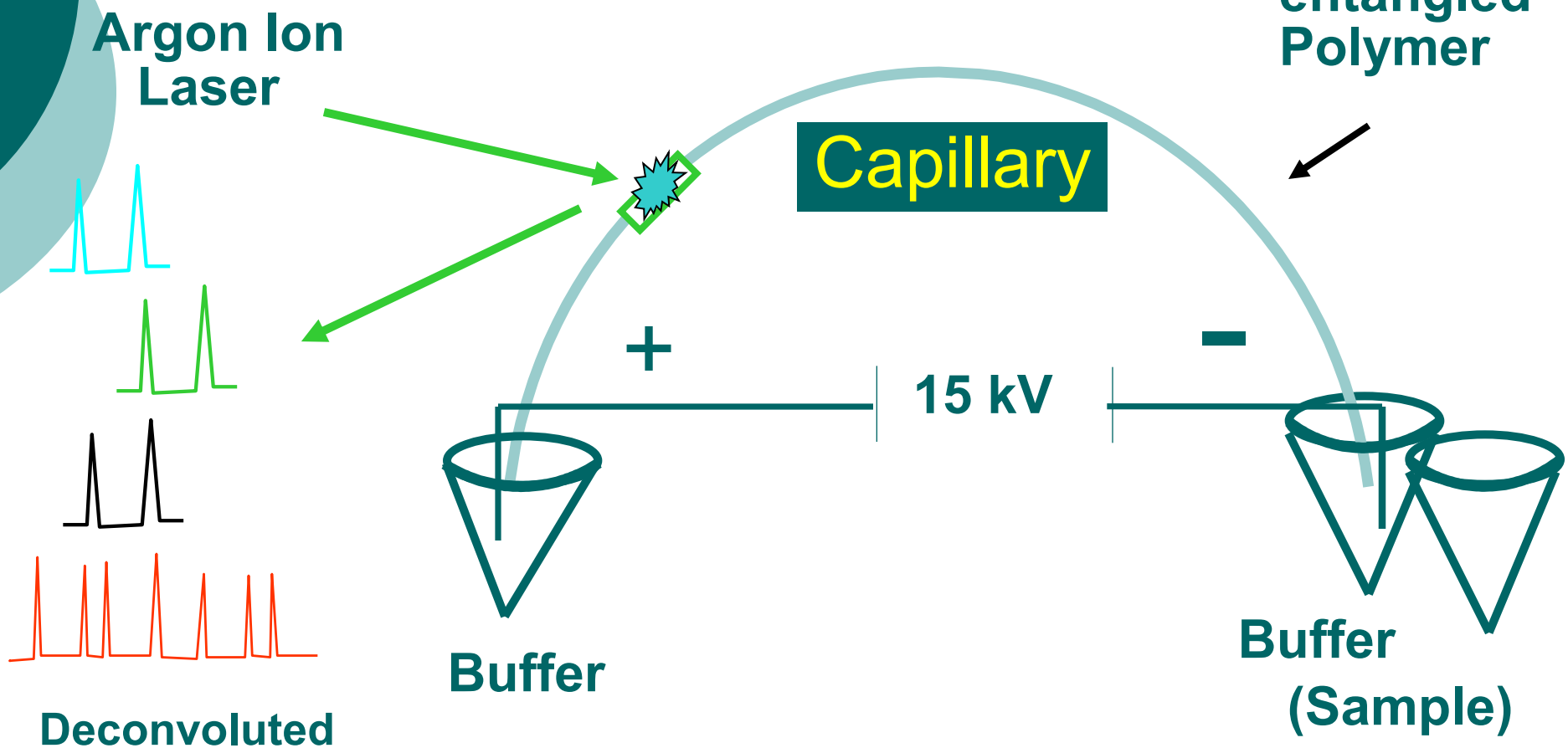
Silver Stained Slab Gel?

Lab Floors like a Darkroom!

Fingers stained like Some strange election


ANALYTICAL
CHEMISTRY
Includes News & Features and AC Research
APRIL 1, 1996
ANCHAM
Capillary Electrophoresis and *Daubert:*
Time for Admission 241 A

Capillary Electrophoresis System

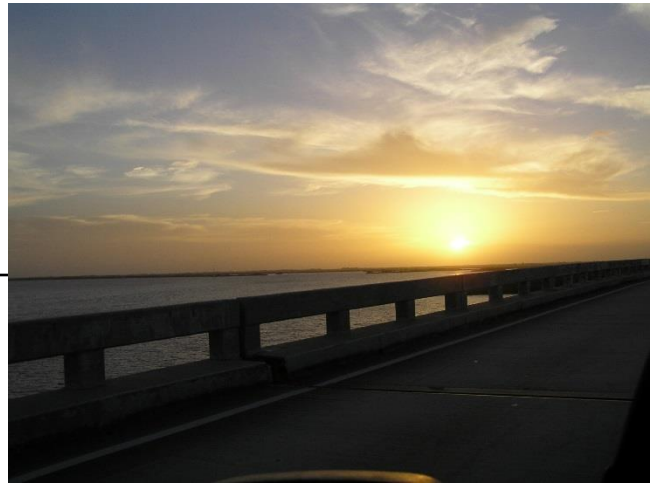


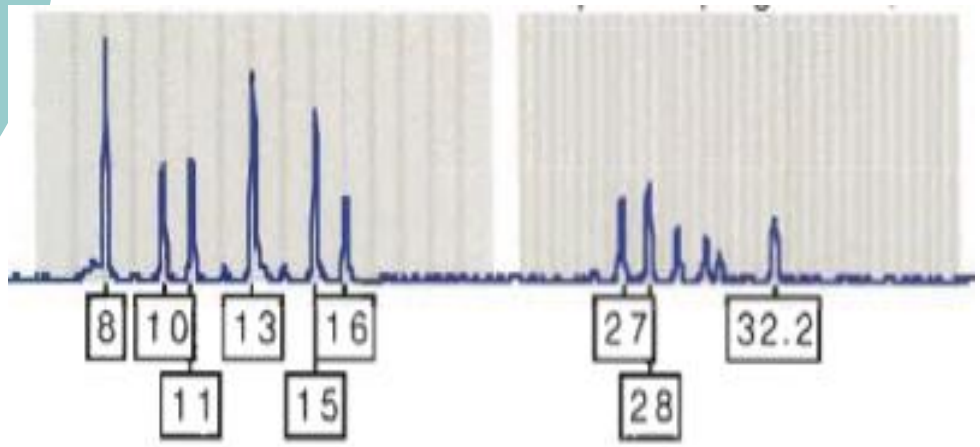
Fast, automated injection and analysis

Process Involved in CE Analysis

- Injection
  - electrokinetic injection process (formamide, water)
  - importance of sample stacking
- Separation
  - Capillary – 50um fused silica, 47 cm (36 cm to detector)
  - POP-4 polymer – Polydimethyl acrylamide
  - Buffer - TAPS pH 8.0
  - Denaturants – urea, pyrolidinone
- Detection
  - fluorescent dyes with excitation and emission traits
  - CCD with defined virtual filters produced by assigning certain pixels

Separation

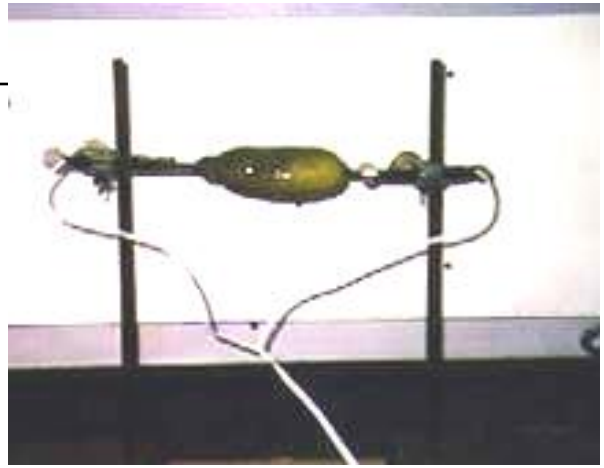




The development of refillable capillaries revolutionized DNA analysis and lead to the early sequencing of the human genome.

Electrophoresis Theory

"Ok here's my recipe idea called the electric pickle. Attach the hot lead to a screw and shove it in. The neutral lead goes in the other end. Turn out the lights and plug it in It glows and sizzles. The juicy ones work best"
www.voltnet.com/cook



$P = VI = I^2R$ **Pickle cooks**

$v_{ep} = \mu_{ep}V$ **Ions move through pickle faster at high voltage**

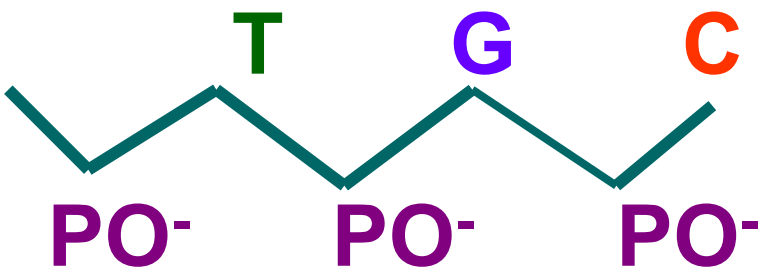
$\mu_{ep} = q/6\pi\eta r$ **Small ions with high charge move fastest**

DNA and Electrophoresis

"From a practical point of view it is disappointing that electrophoresis cannot be used to fractionate or analyze DNA on the basis of size" Olivera, Biopolymers 1964, 2, 245

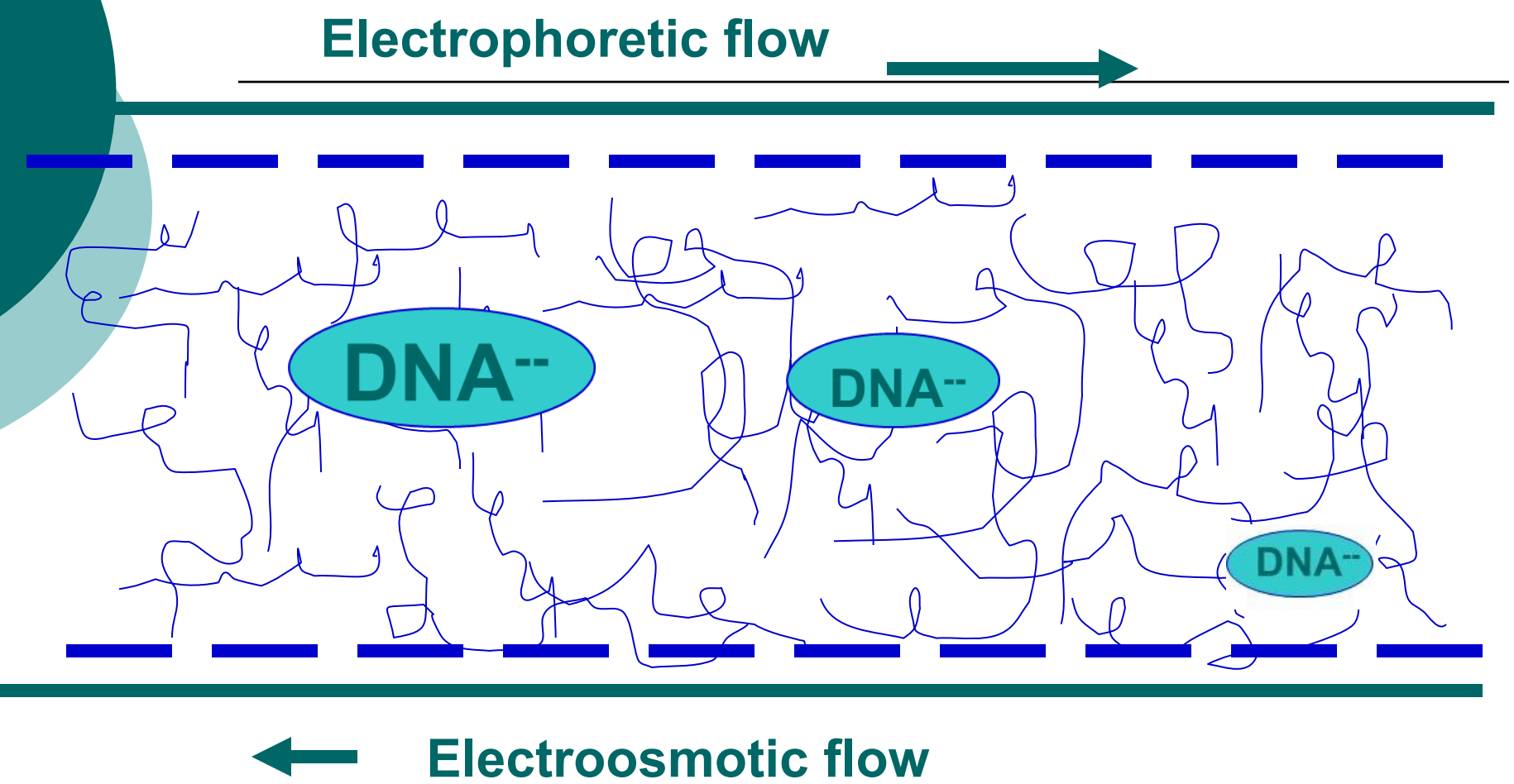
$$\mu_{ep} = q/6\pi\eta r$$

small ions with high charge move fastest



As size increases so does charge!

Separation Mechanism



Electroosmotic flow is a bulk flow that is created by the build up of charge on a capillary wall. POP polymer minimizes this effect

$SiOH \rightarrow SiO^- + H^+$

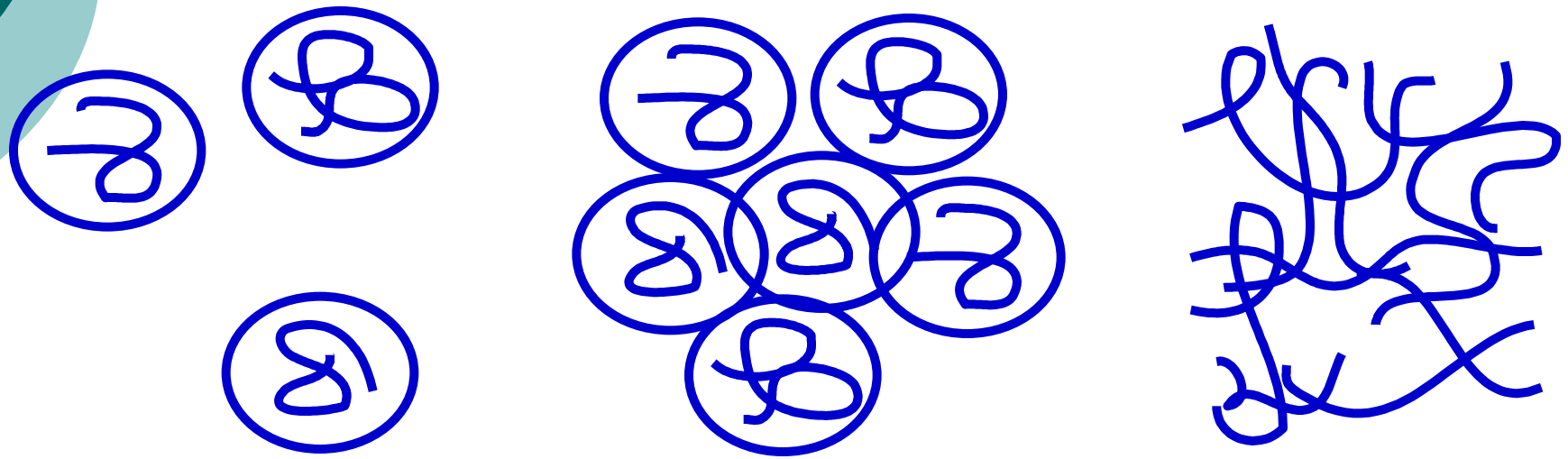
So what are sieving buffers?

They are gels - very similar to polyacrylamide

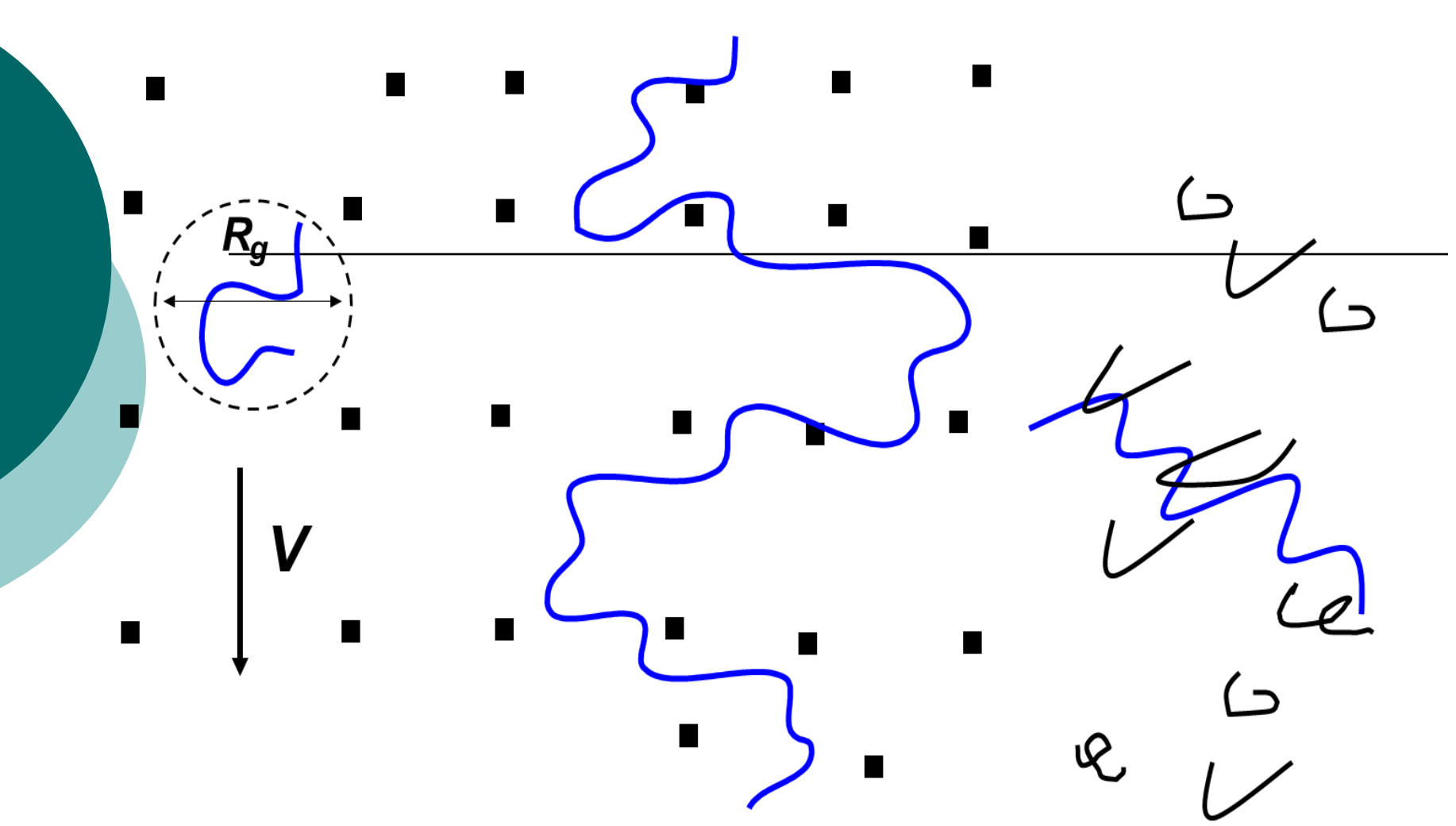
They are not gels - they flow

Actually these are known as entangled linear polymers and there are many common applications

Transient Pores Are Formed Above the Entanglement Threshold.



C* < *C* ***C* = *C**** ***C* > *C****



Ogston Sieving

$\mu \sim \mu_0 e^{-NC}$

Reptation

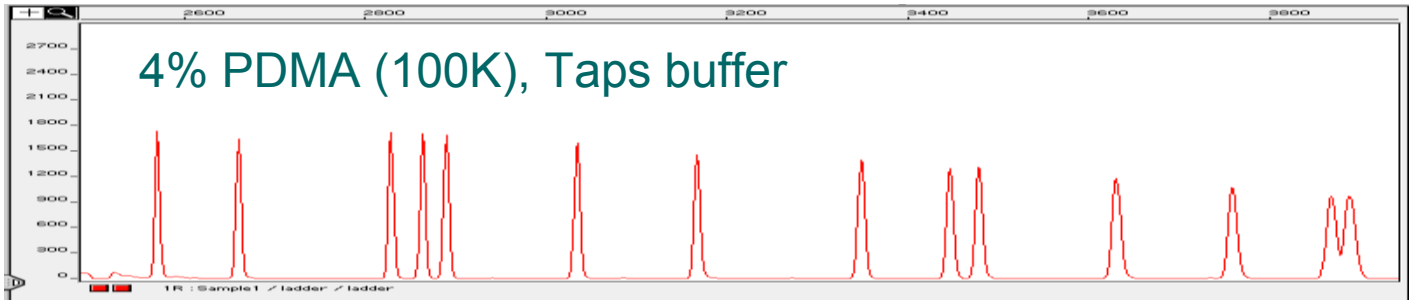
$\mu \sim 1/N$

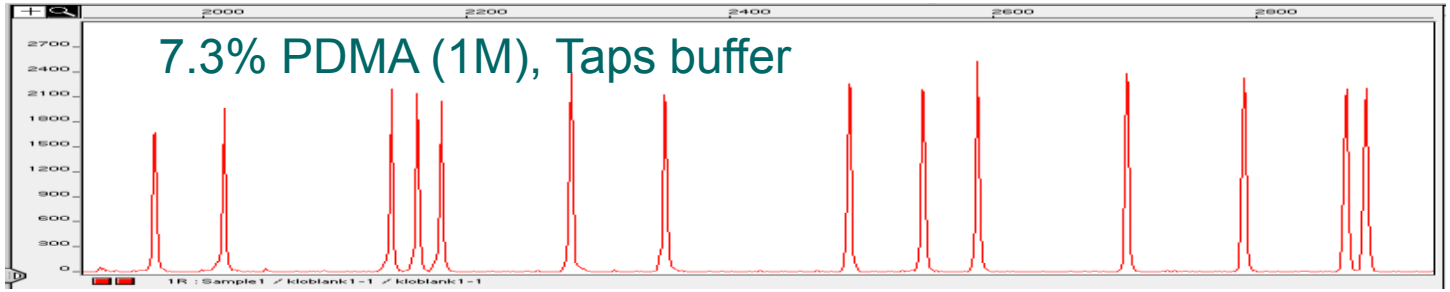
Entanglement

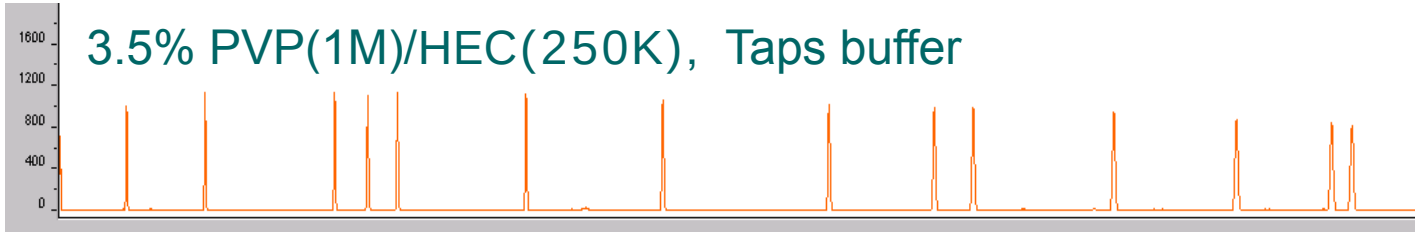
$\mu \sim f(1/CN)$

Synthesis of DNA Sieving Buffers

Effect of Concentration and Molecular Weight on Resolution

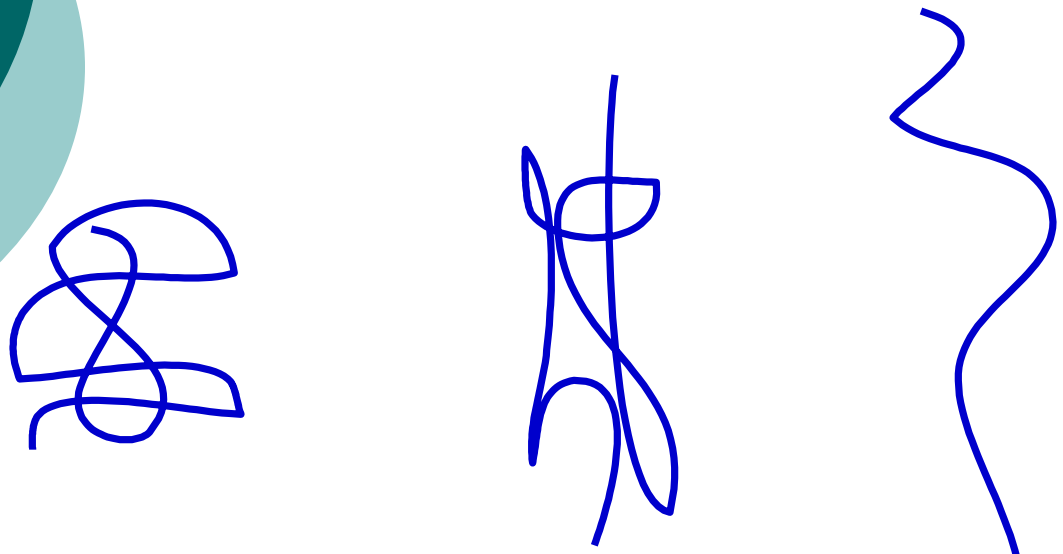






Boulas, S. ; Blas, M.; Cabrices, O.; McCord, B., *Electrophoresis*, **2008** *29(23)* 4695-4703.

The electric field strength and gel can fluence the shape of the DNA molecule.

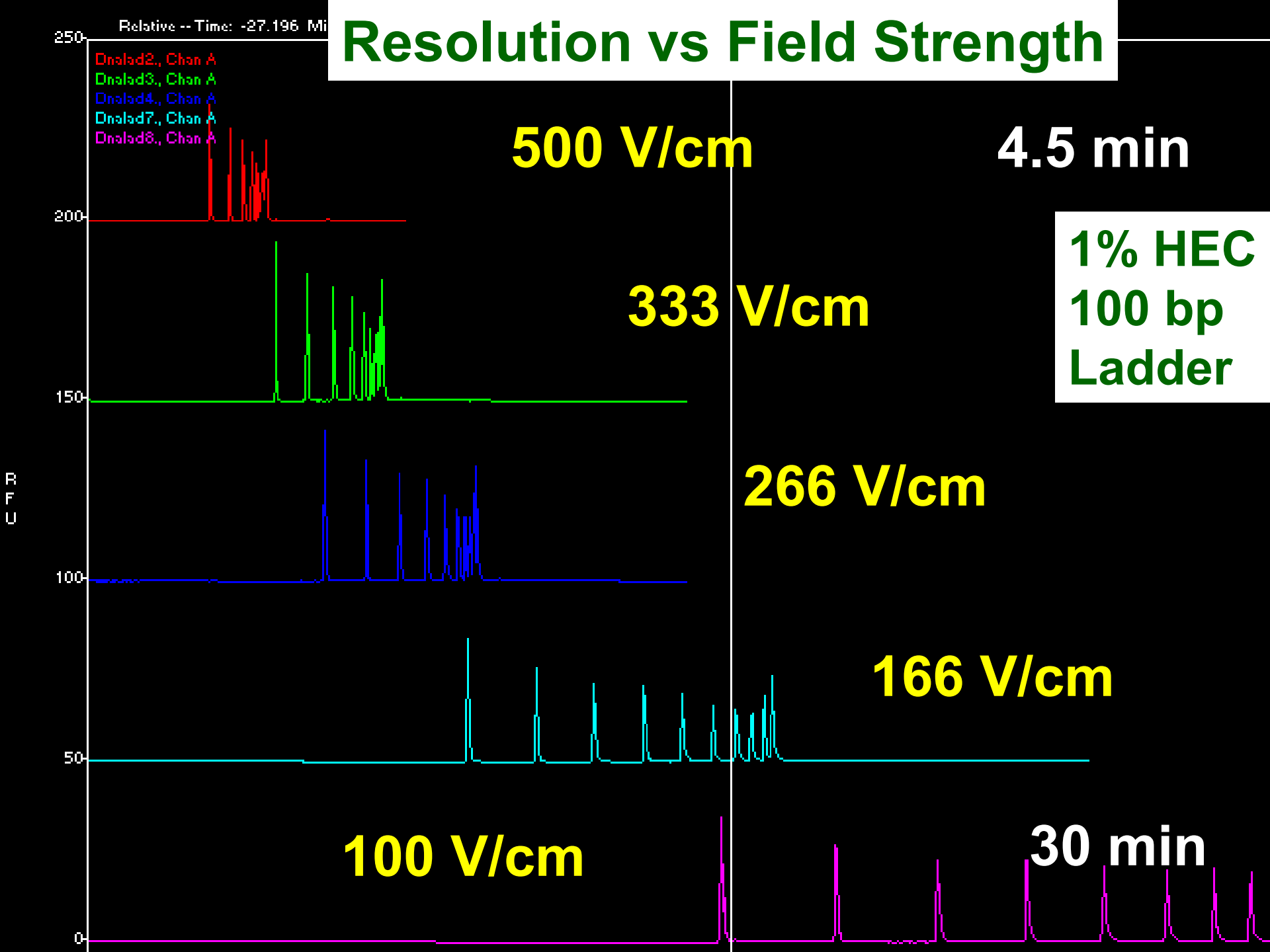


low

moderate

high

Follow the dancing DNA


Resolution vs Field Strength
Relative -- Time: -27.196 Mi
Dnalad2., Chan A
Dnalad3., Chan A
Dnalad4., Chan A
Dnalad7., Chan A
Dnalad8., Chan A
500 V/cm
4.5 min
1% HEC
100 bp
Ladder
333 V/cm
266 V/cm
166 V/cm
100 V/cm
30 min
250
200
150
100
50
0
RFU

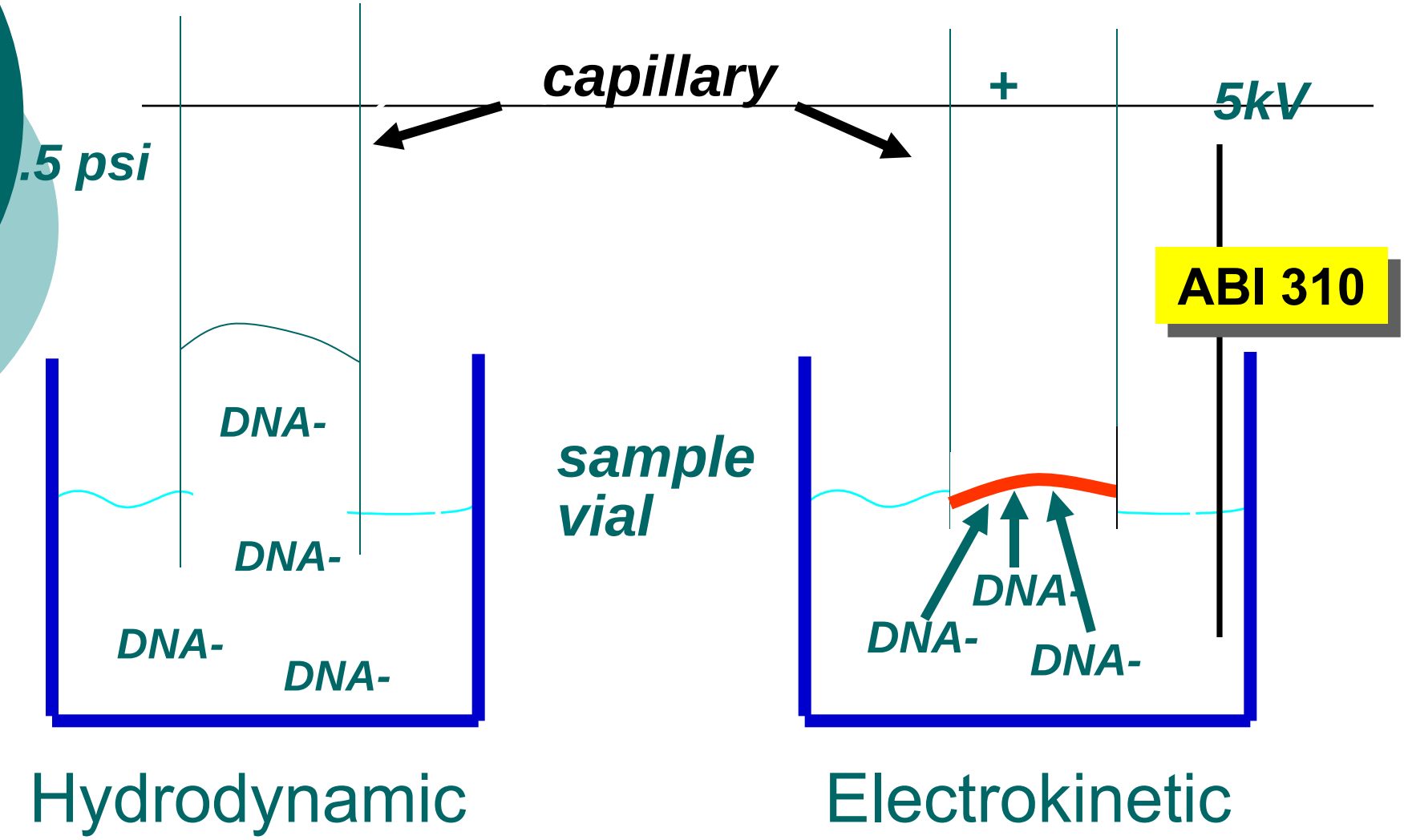
Separation Issues

- Electrophoresis buffer –
  - Urea for denaturing and viscosity
  - Buffer for consistent pH
  - Pyrolidinone for denaturing DNA
  - EDTA for DNA stability and chelating metals
- Polymer solution –
  - Entangled to separate DNA
  - High molecular weight for good resolution
  - Mimimum concentration/viscosity for easy refilling (POP4, POP6, etc.
  - Coats capillary wall to minimize adsorption
- Run temperature –
  - 60 ^{o}C helps reduce secondary structure improves precision.
  - (Temperature control affects DNA sizing)
- Electric field –
  - affects orientation and diffusion of DNA

Injection

Injection is actually improved in CE through dilution of the sample.

Injection Methods for CE

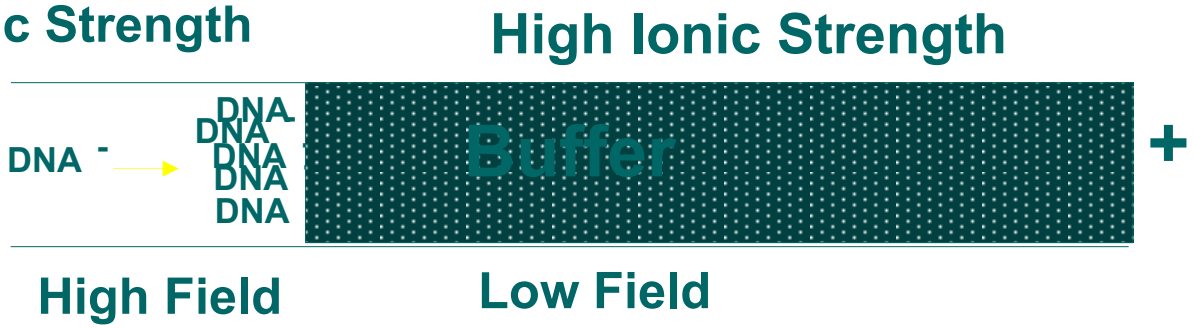


Ulfelder K. J.; McCord, B. R. (1996) Capillary Electrophoresis of DNA, In *Handbook of Capillary Electrophoresis* (Landers, J., ed.), CRC Press: NY, pp. 347-378.

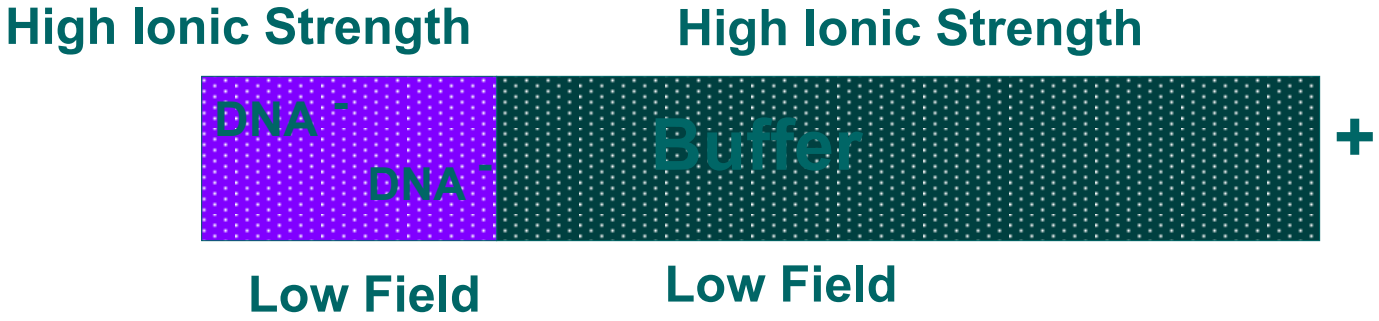
Butler, J.M. (1997) Effects of sample matrix and injection on DNA separations. *Analysis of Nucleic Acids by Capillary Electrophoresis* (Heller, C., ed.), Vieweg: Germany, Chapter 5, pp. 125-134

Stacking Effects

(a) Stacking with Low Ionic Strength



(b) Regular EK Injection



Ion Mobility Effects

The Injection of DNA by voltage is described by

$[DNA_{inj}] = E(\pi r^2)[DNA_{sam}](\mu_{ep} + \mu_{eof})$

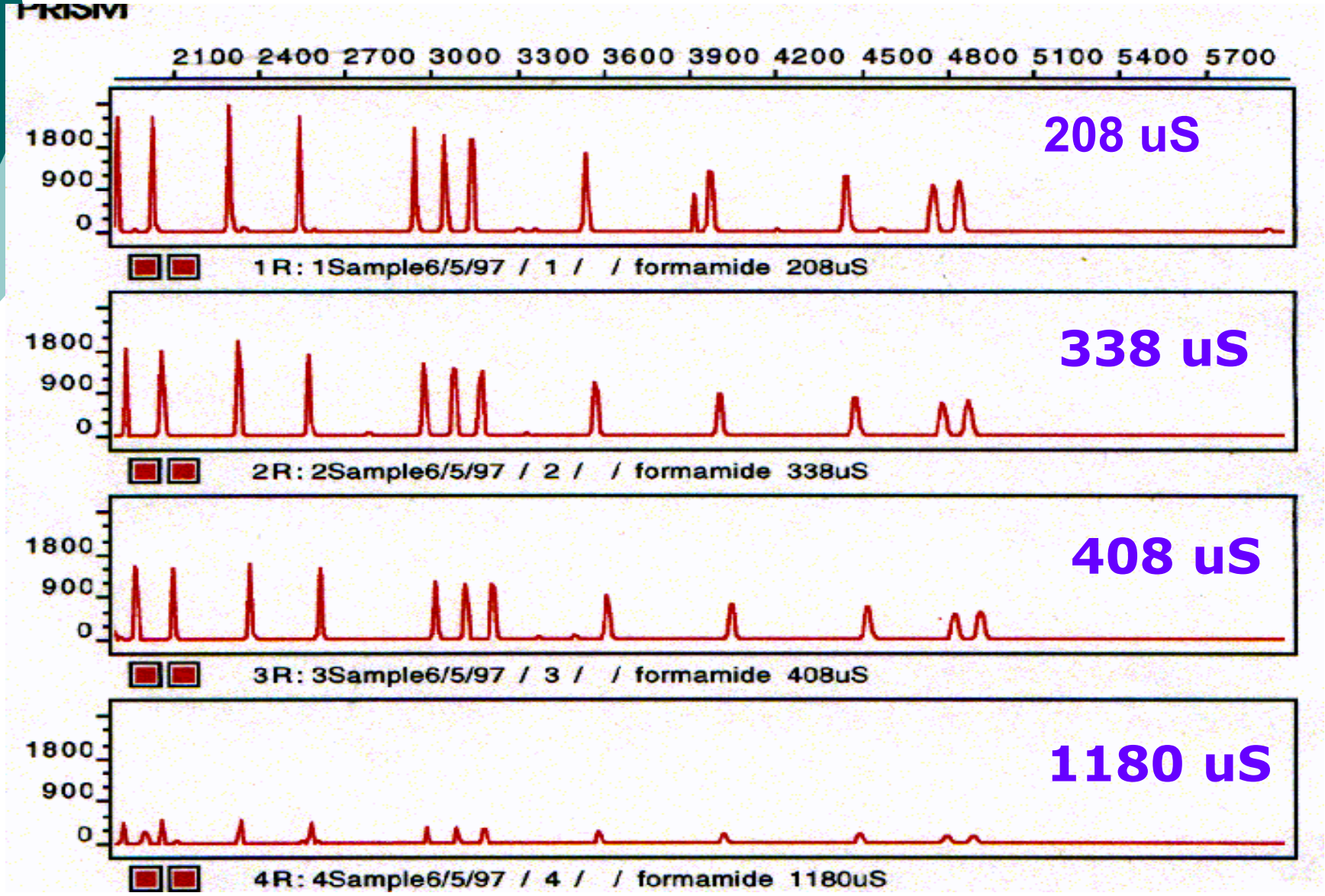
However this equation assumes no interfering ions are present.

Cl^- ions and other interferents will compete with DNA

$\{DNA_{inj}\} = [DNA_{inj}]/[\text{other ions}_{inj}]$

Ions such as Cl^- have a higher charge/mass ratio and μ_{ep} is greater

Effect of Poor Formamide on DNA Concentration

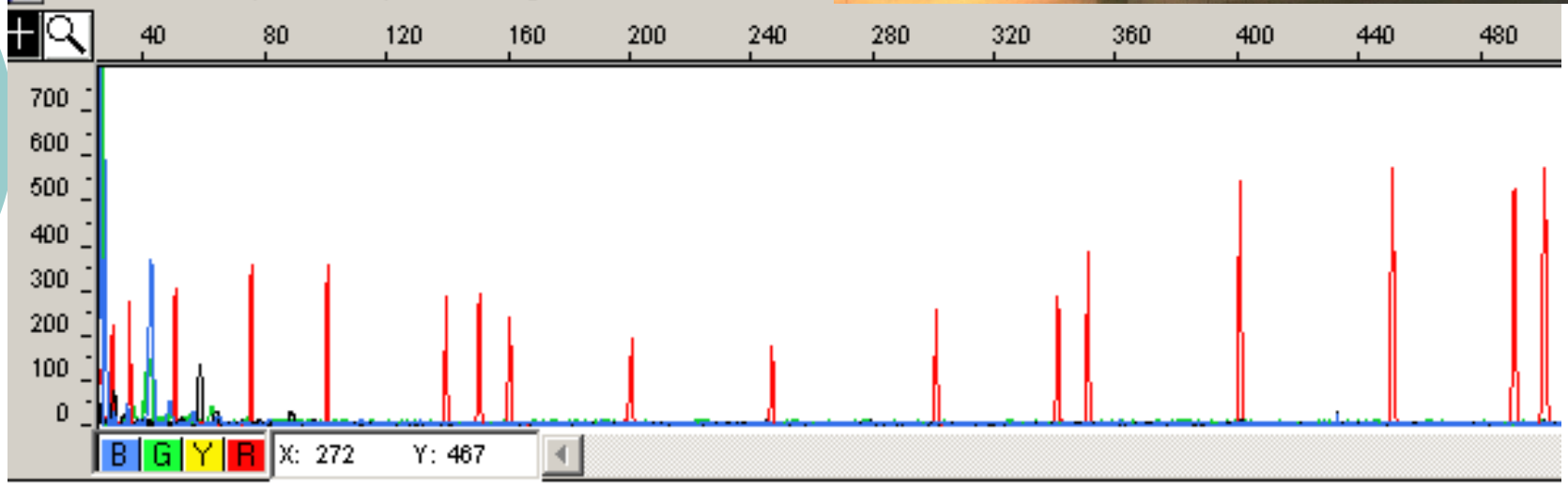


Golden Gate Effect

Attributed to poor formamide

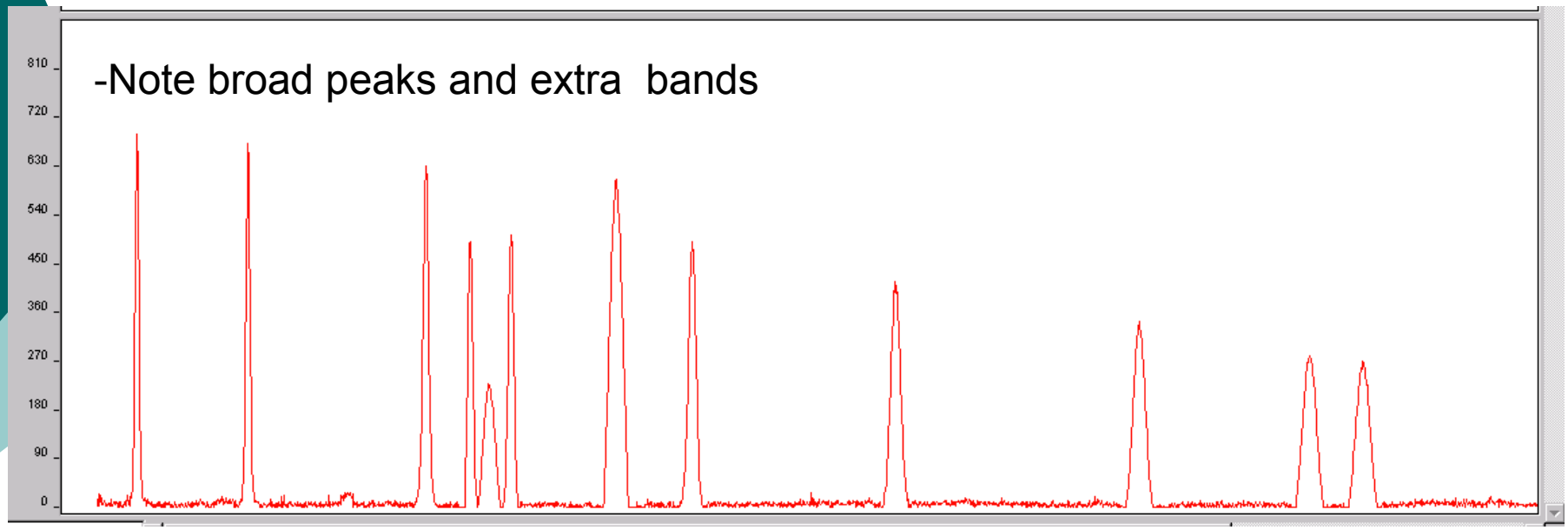
GeneScan 3.7 - [RC_08_09_04_E04_29Sep2004658_1

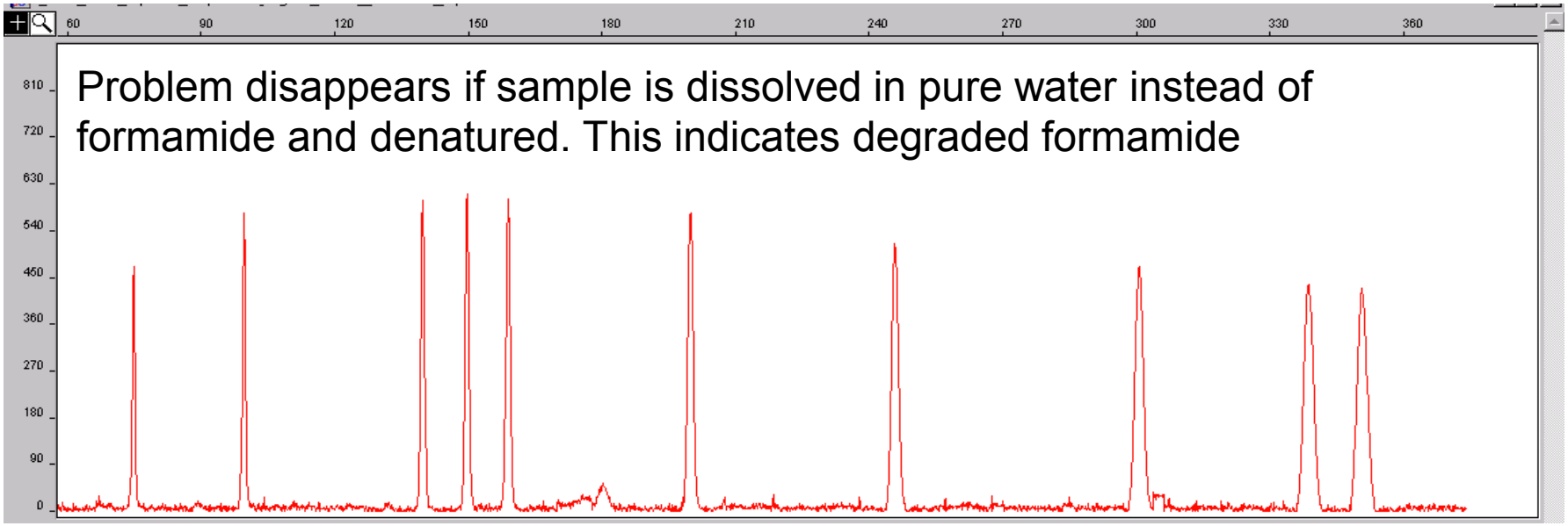
File Edit Project Sample Settings View Windows H



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
R, 1 •	7.29	75.00	380	3201	2735
R, 2 •	8.18	100.00	379	3119	3067
R, 3 •	9.53	139.00	307	3008	3575
R, 4 •	9.86	150.00	309	3242	3699
R, 5 •	10.19	160.00	248	3144	3820
R, 6 •	11.51	200.00	205	2684	4318
R, 7	13.01	246.09	192	1950	4877
R, 8 •	14.67	300.00	273	2818	5501
R, 9 •	15.83	340.00	299	3191	5938
R, 10 •	16.17	350.00	406	4336	6062
R, 11 •	17.67	400.00	566	6049	6627
R, 12 •	19.08	450.00	595	6718	7156

Effect of bad Formamide





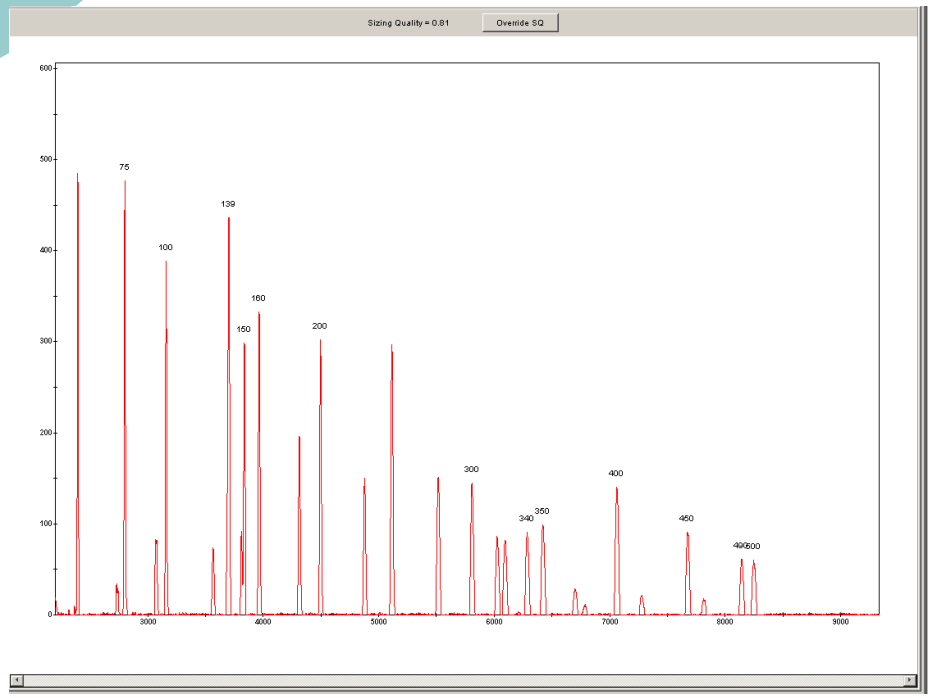
Shadow peaks

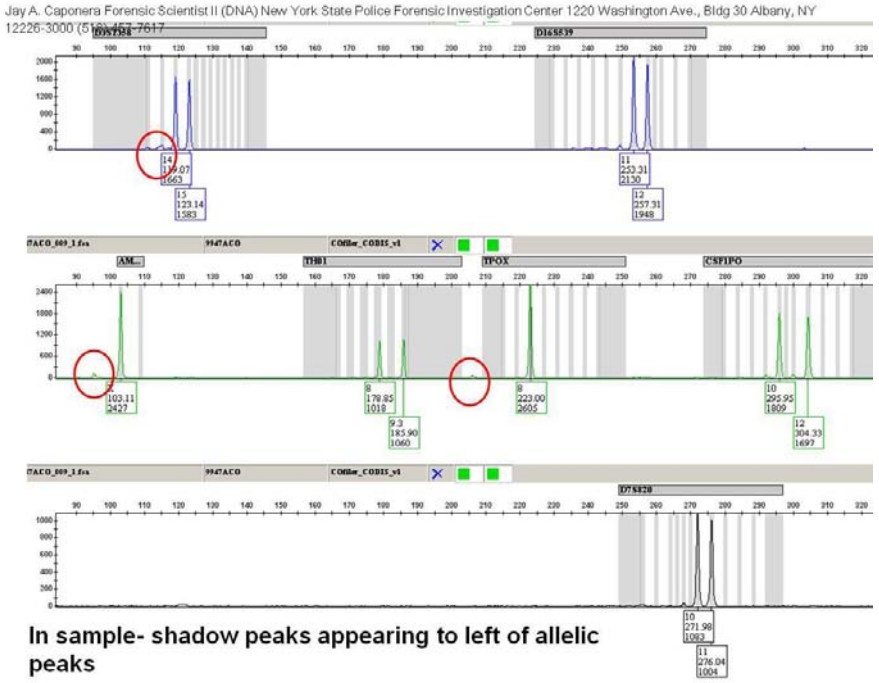
Shadow peaks result from bad formamide, incomplete denaturation or from rehybridization.

---

dsDNA migrates faster than ssDNA and the extra peaks appear ahead of the main peaks

They are most visible in the size standard but can appear in other dye lanes





In sample- shadow peaks appearing to left of allelic peaks

Measuring Formamide Conductivity

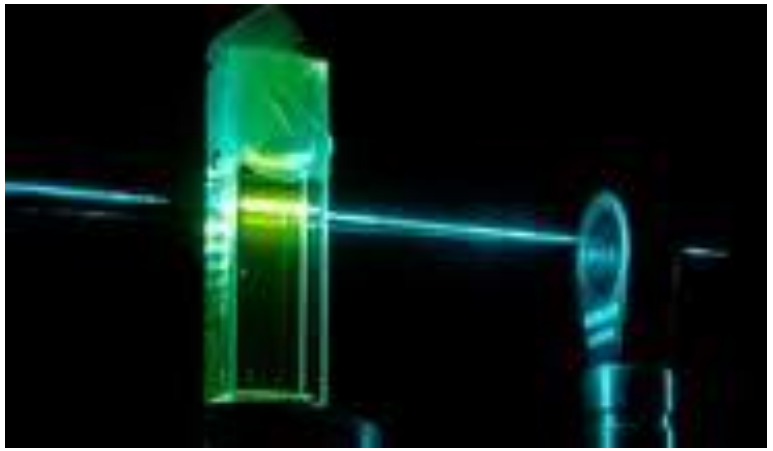
(not this way)

The key is to measure the bottle when it comes in or buy the good stuff and immediately pipette it out into small tubes with or without ROX already added. Then freeze the tubes.

Do not ever open a cold bottle of formamide. Water will condense inside and aid in the formation of conductive formic acid.

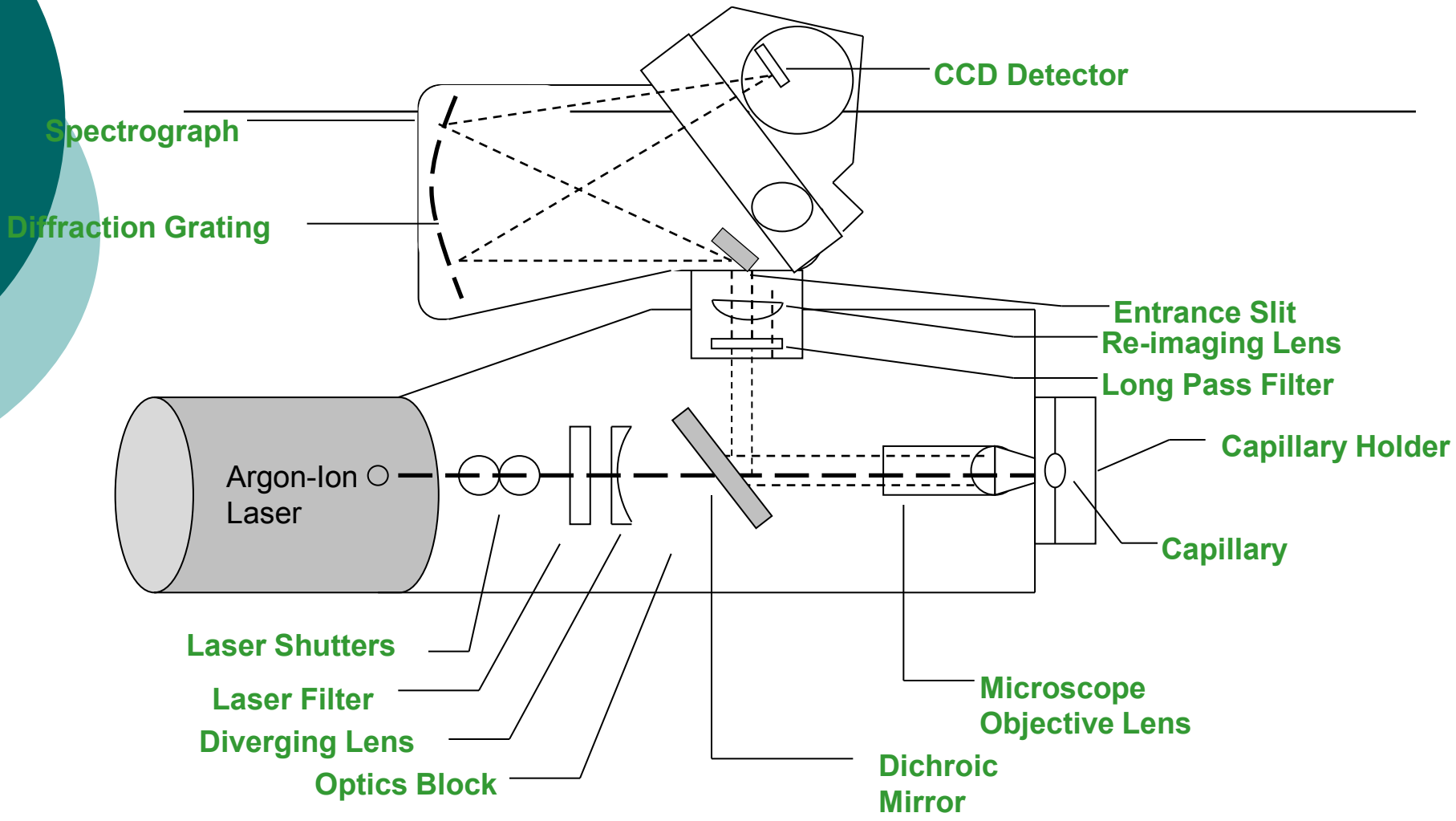
Another way to check for bad formamide is to run a heat denatured sample in distilled water. If it looks different the formamide is contaminated.

Detection



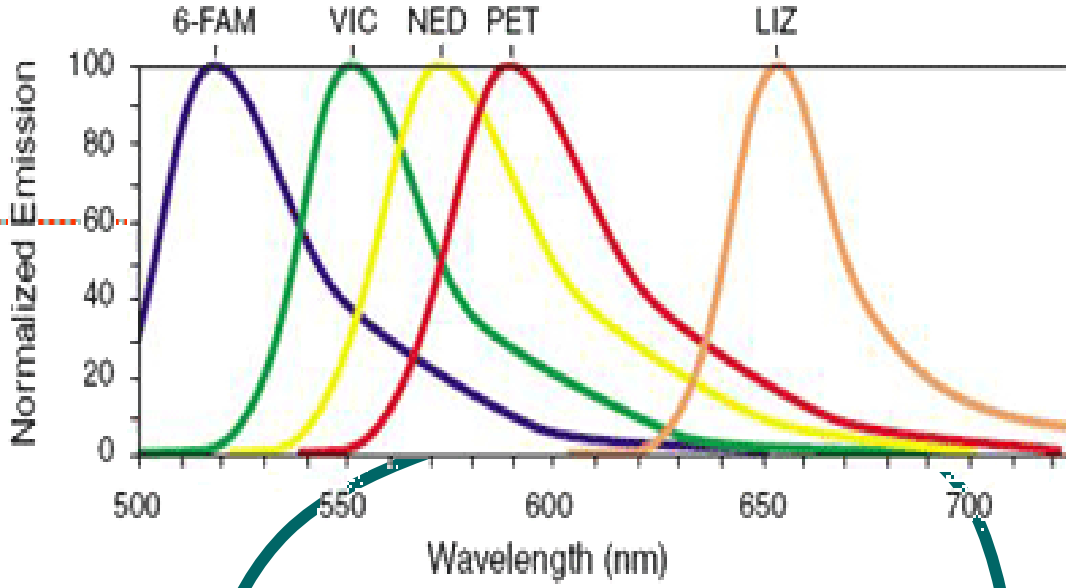
Laser induced fluorescence provides exquisite sensitivity and specificity

ABI 310 confocal detector



Watts, D. Genotyping STR Loci using an automated DNA Sequencer. In *Forensic DNA Profiling Protocols;* Lincoln, P.J.; Thomson, J. Eds.; Humana Press Inc.: Totowa, NJ, 1998; Vol. 48, pp 193-208.

OPTICS



Laser (488nm)

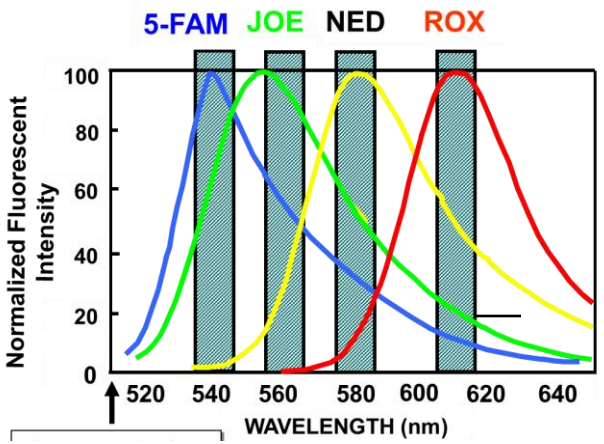
Lens

Charged coupled device

capillary

Grating

Matrix calculations



$I_{540} = bx_b + gy_b + yz_b + rw_b$ intensity of blue
$I_{560} = bx_g + gy_g + yz_g + rw_g$ intensity of green
$I_{580} = bx_y + gy_y + yz_y + rw_y$ intensity of yellow
$I_{610} = bx_r + gy_r + yz_r + yw_r$ intensity of red

Where
b is the %blue labeled DNA
g is the %green labeled DNA, etc.

x,y,z,w are the numbers in the matrix (sensitivity to each color)

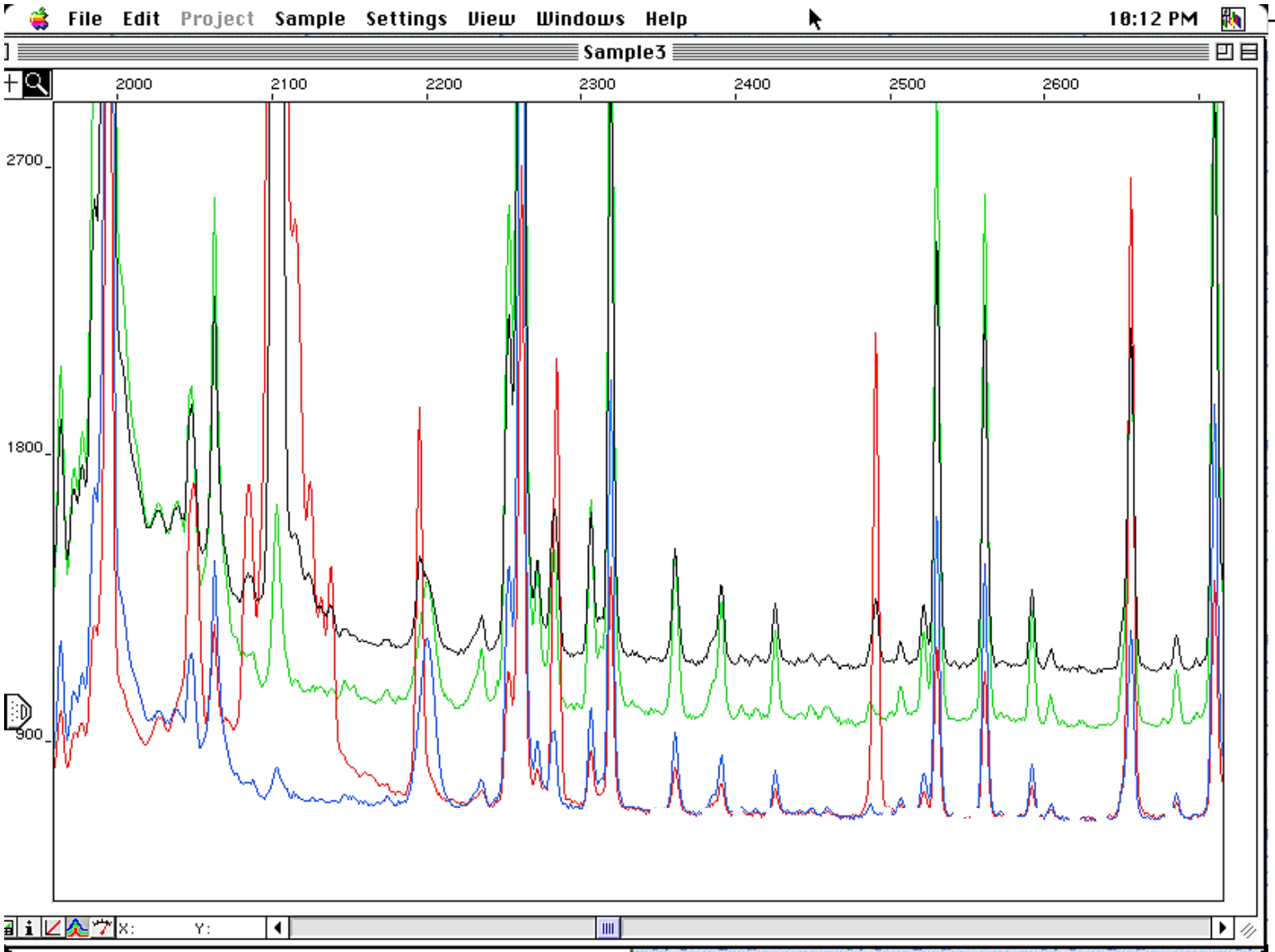
POP4STRMODF

Reactions

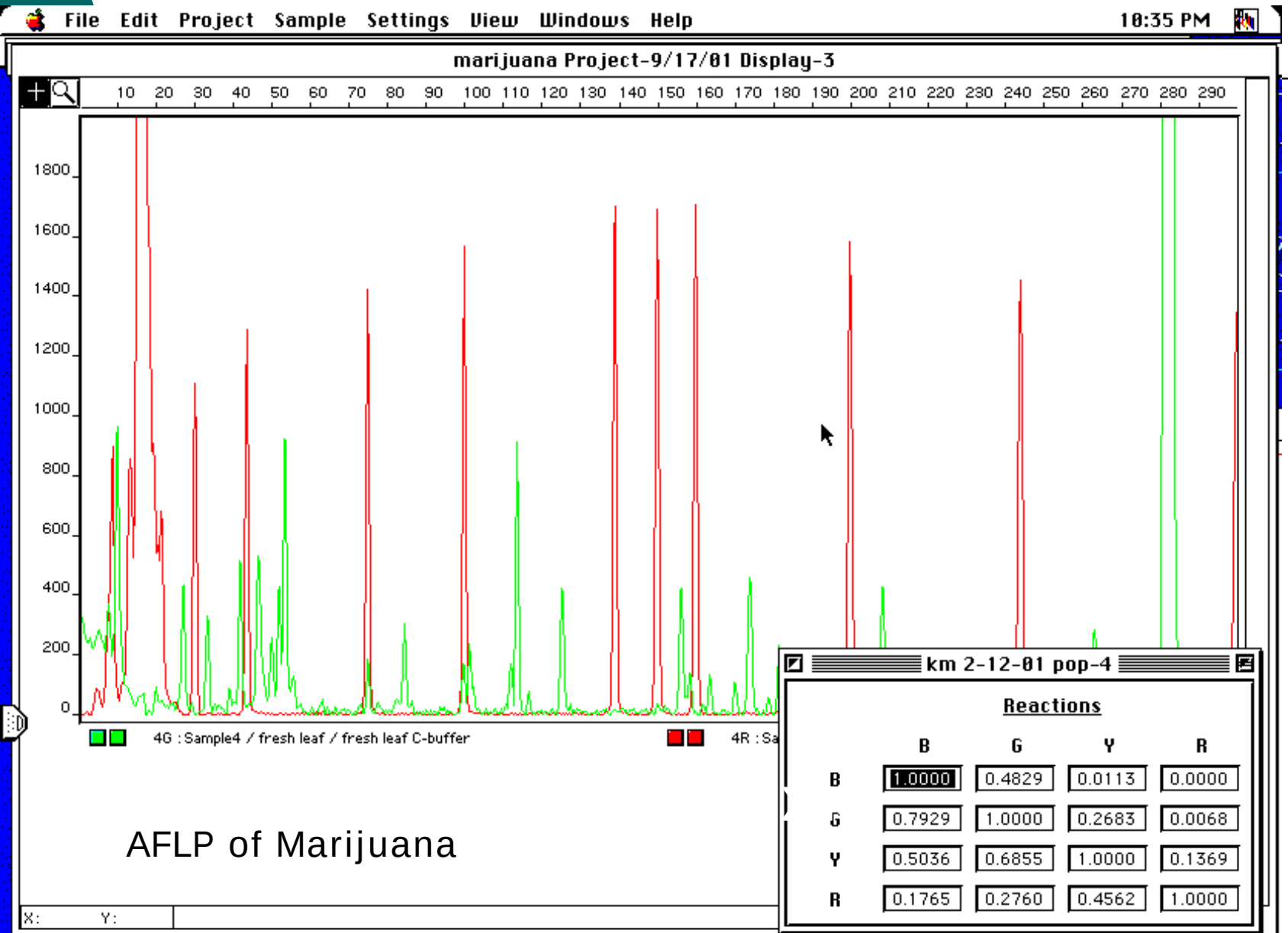
	B	G	Y	R
B	1.0000	0.8502	0.1380	0.0009
G	0.8300	1.0000	0.7622	0.0051
Y	0.6416	0.8324	1.0000	0.1102
R	0.4493	0.6484	0.7851	1.0000

If you solve xyzw for each dye individually
Then you can determine dye contribution for any mixture

Uncorrected Raw Data (AFLP of a Marijuana Sample)



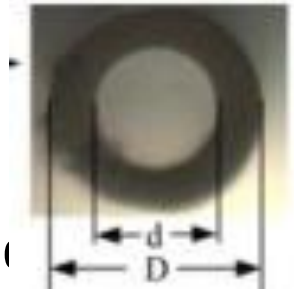
Corrected Result



AFLP of Marijuana

Issues with the Optical System

- Argon Ion lasers outgas and eventually loose intensity; **take note of laser power and monitor it over time**



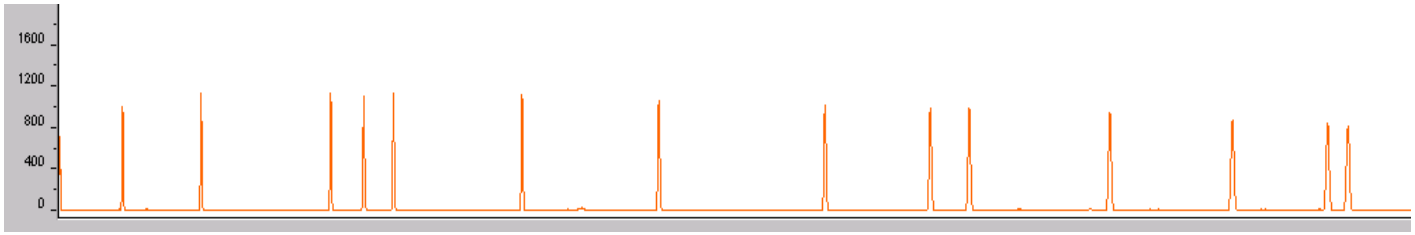
- Fluorescence expression:

  $I_f = I_0 k \varepsilon d C \varphi$ - changes in input intensity: I_0

  - changes in capillary diameter: d

  - cleanliness of capillary, optics: k

  - All these things directly affect peak RFUs, however, baseline noise is more affected by detector.
  - A good QC step is to monitor the ladder intensity

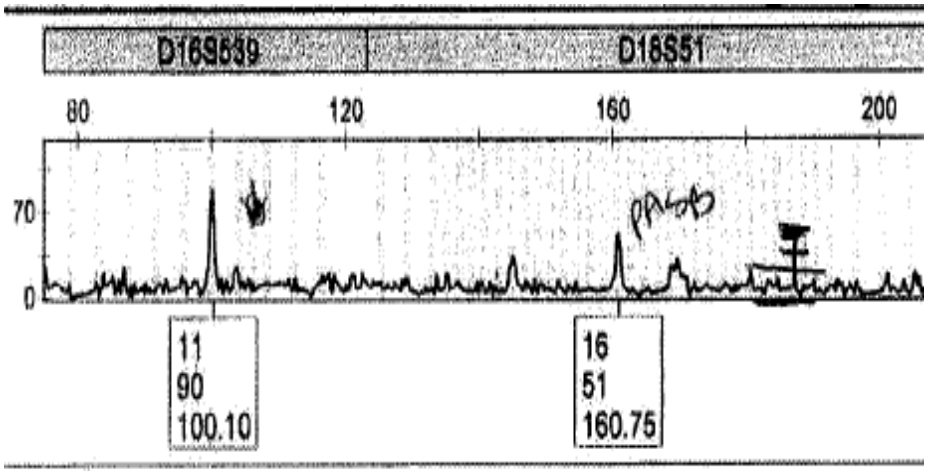


Part II Setting Thresholds

Now that we know how it works,

Where to go next?

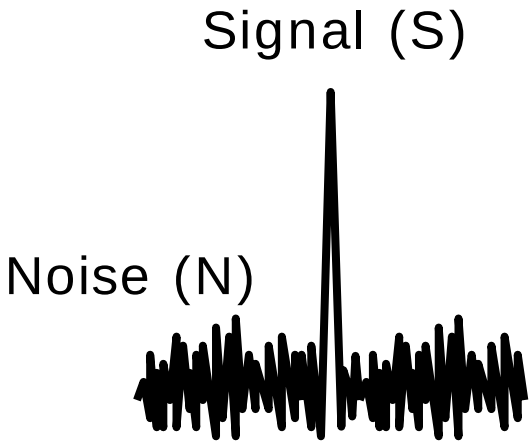
How to find the bottom?

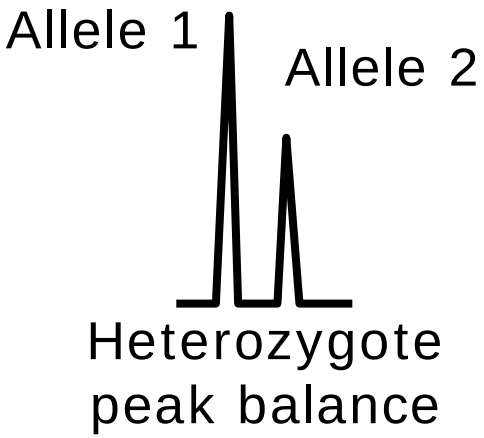


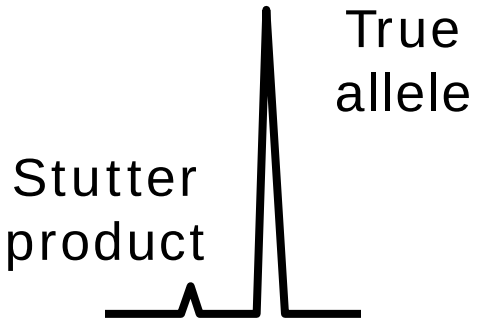
Setting thresholds for the ABI 310/3100

(Determining a true allele)

- Every laboratory needs to set an analytical threshold, a stochastic threshold, a limit of linearity and minimum peak height threshold.
- Can these values be set globally for the entire lab or are they instrument dependent?
- How do these values affect detection, stutter, pull-up, mixture interpretation, low copy DNA?







Fundamental parameters for allele detection by CE

Detection Limit (analytical threshold): 3x the standard deviation of the noise.

Estimated using 2x peak to peak noise. (approximately 35 - 50 RFUs)

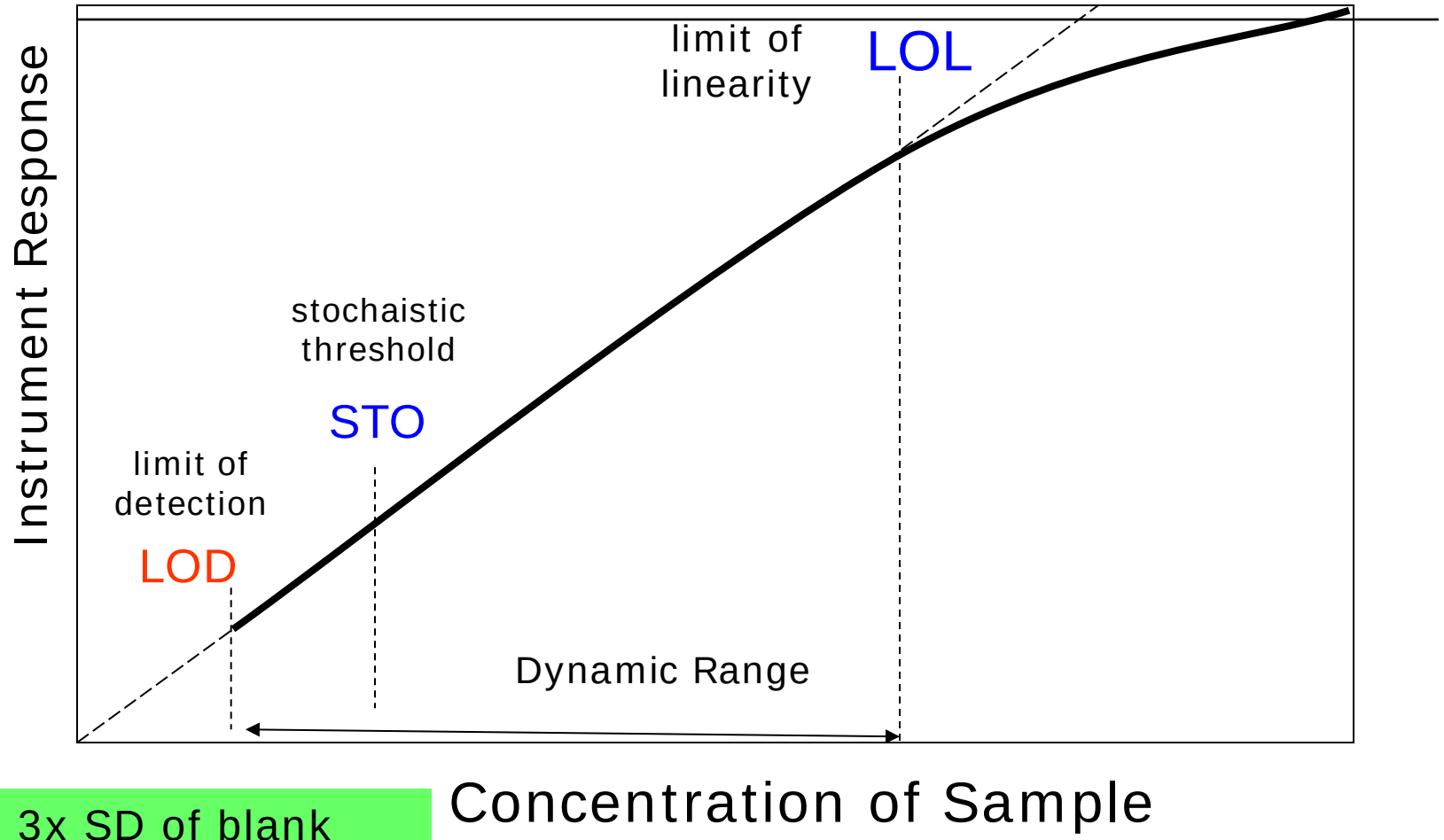
Peaks below this level may be random noise

Stochastic Threshold: Level of DNA below which a significant chance of allele dropout can occur. Set high enough that a heterozygous peak will produce its companion allele in the grey zone between stochastic and analytical threshold. (150-200 RFUs)

Limit of linearity: The level of DNA above which enhanced pull-up, flat top peaks and elevated stutter occurs. Determined by examining the relationship between input DNA and fluorescence signal varies (~4500 RFUs for ABI 310, ~3500 for a 3100, >20,000 for a 3500)

Heterozygous peak ratio: The minimum peak height ratio expected for a clean, single source DNA sample at a particular concentration (typically 60-70%)

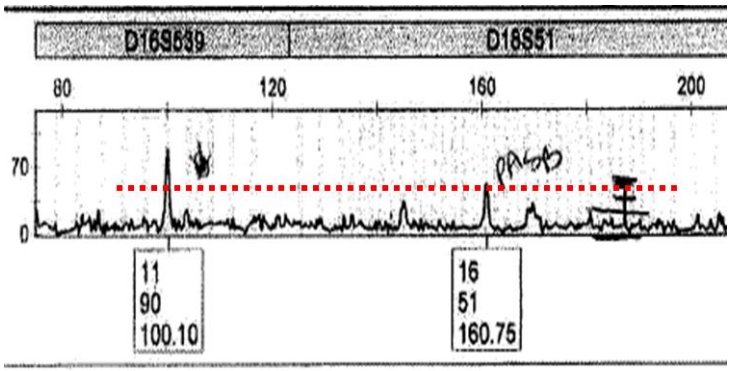
Useful Range of an Analytical Method



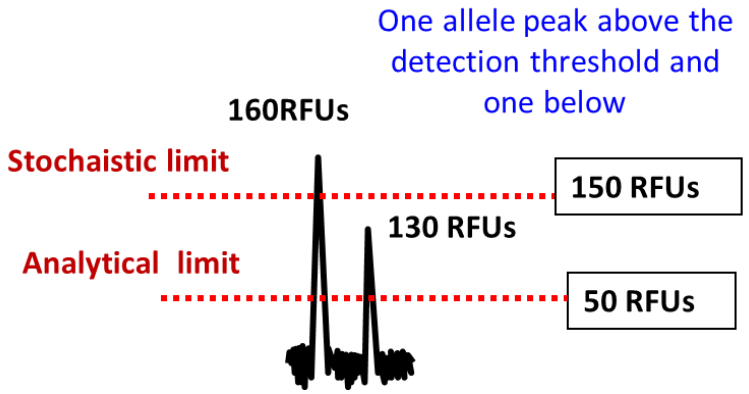
LOD = 3x SD of blank
STO = peak balance threshold
LOL = pull-up/mixture threshold (loss of linearity.)

Visual representations

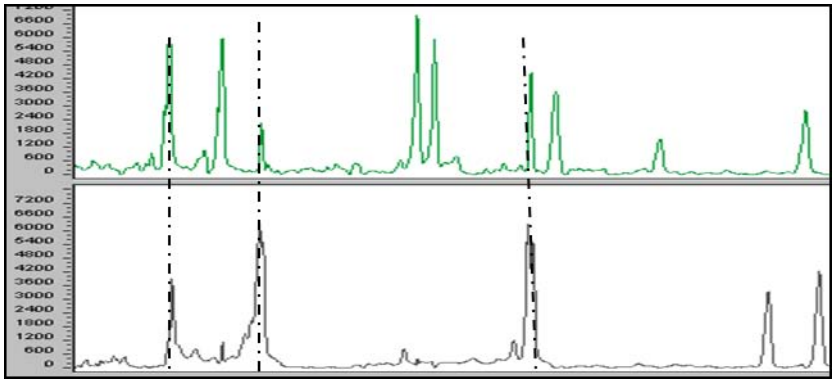
- **Analytical threshold** – peaks below 3x the baseline cant be distinguished from noise with scientific certainty.



- **Stochastic threshold** – peak height ratios below 60% may yield false homozygotes. Grey zone permits detection of 2nd homozygous peak



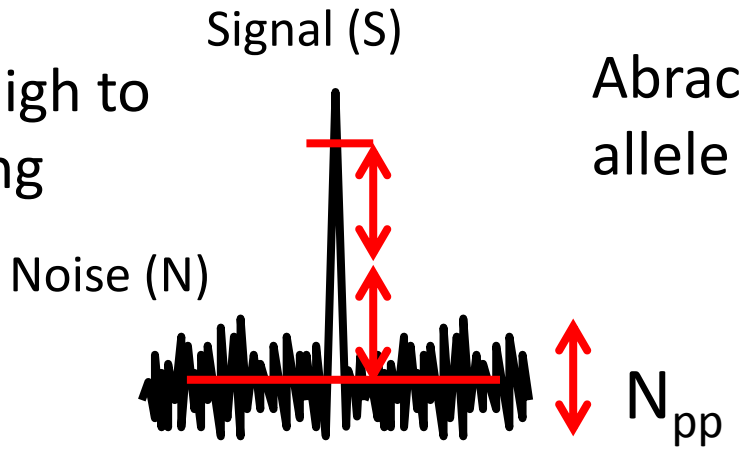
- **Limit of Linearity –** relationship between fluorescence and quantity breaks down. Pull-up appears and peaks become flat topped



The Scientific Reasoning behind the Concept of an Analytical Threshold/limit of detection

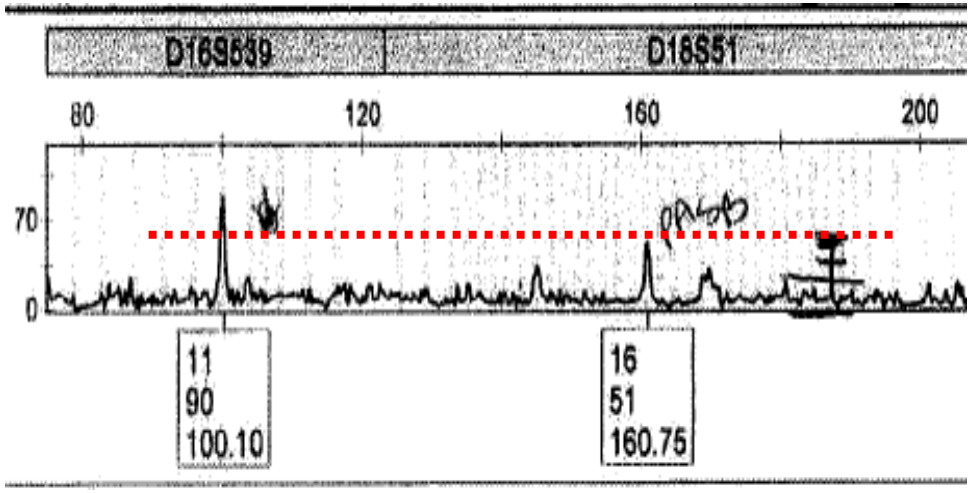
- This is fundamentally an issue of reliability
- **For a peak intensity below the LOD there is a very real chance that such a signal is the result of a random fluctuation**
- You want to be sure to avoid labeling noise!
- LOD = 2x N_{pp} or 3x SD_n
- Levels are typically set high to avoid constantly resetting thresholds

Abracadabra! It's an allele

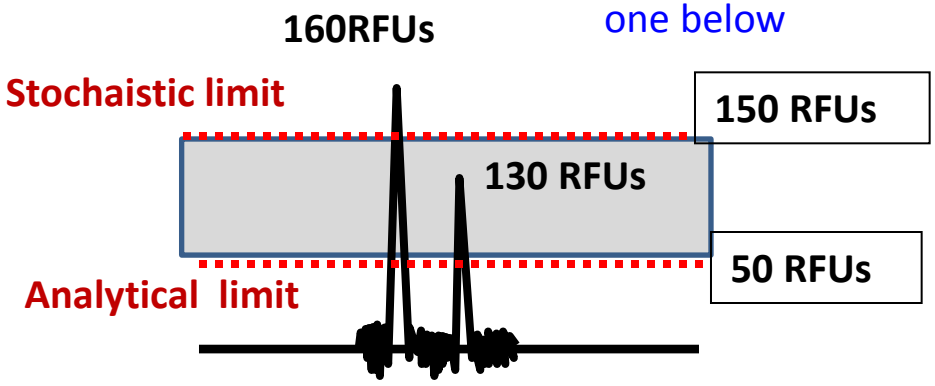


How thresholds are currently set

- **Analytical threshold** – Set to 3 times the baseline noise plus an additional amount to be conservative
- **Stochastic threshold** – Set to avoid the presence of peak height ratios below 60%
- **Grey zone** between two threshold permits detection of 2nd homozygous peak,



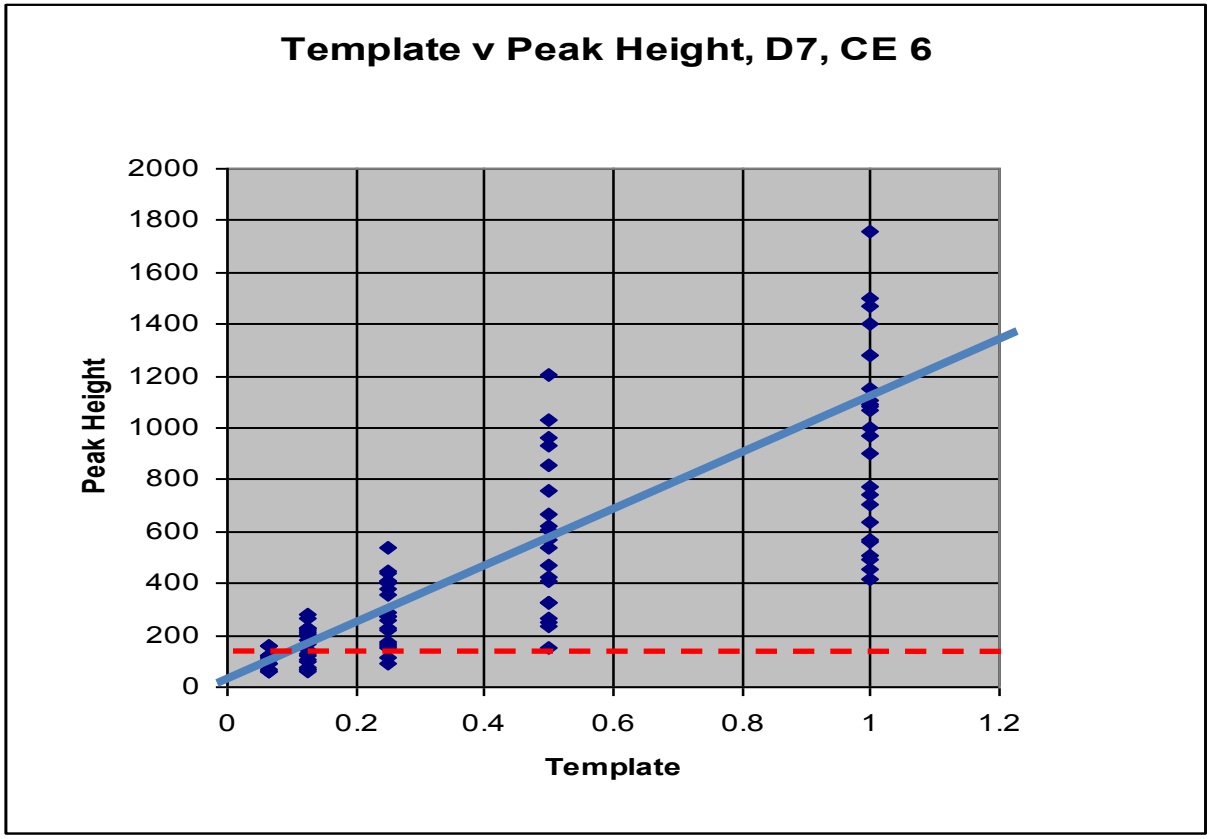
One allele peak above the detection threshold and one below



Sensitivity Study:

Profiler Plus

(Debbie Hobson-FBI)



Observation: Peak height variation increases with concentration

Therefore: its difficult to assess the quantity of DNA solely by peak height

The Scientific Reasoning behind the LOQ/Stochastic threshold

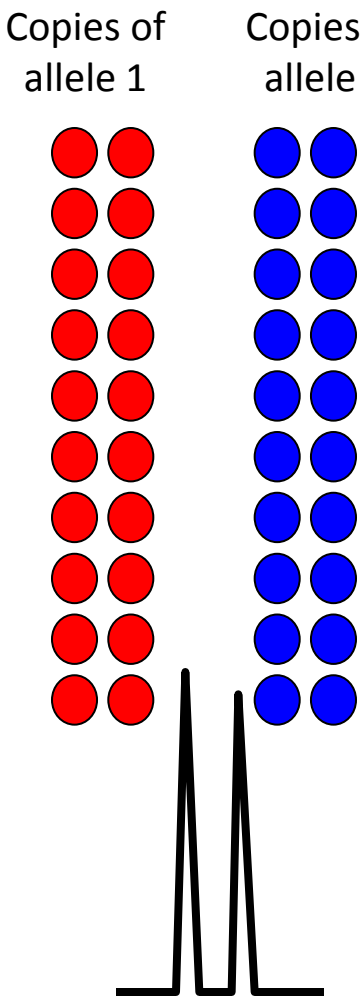
- With peak intensity below the LOQ, you have significant variation in height from one sample to the next.

- **Similarly due to stochastic fluctuation in peak height ratios, interpreting data below the stochastic threshold presents the real problem of allele dropout due to variation.**

- You rely on peak heights to detect major and minor profiles and **you need to be certain when calculating statistics that you do not have a dropped heterozygous allele**.

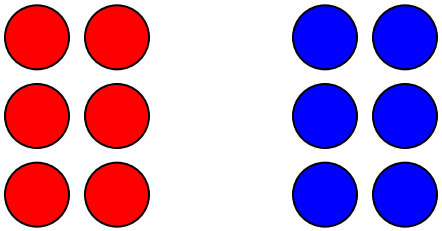
How low can you go?

Stochastic Statistical Sampling

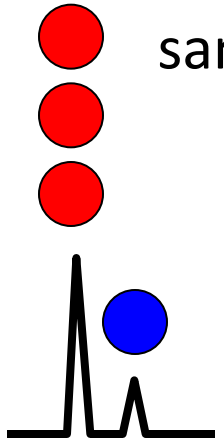


Resulting electropherogram

True amount

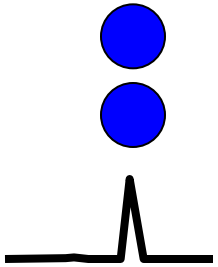


What might be sampled by the PCR reaction…



Allele imbalance

OR



Allele dropout

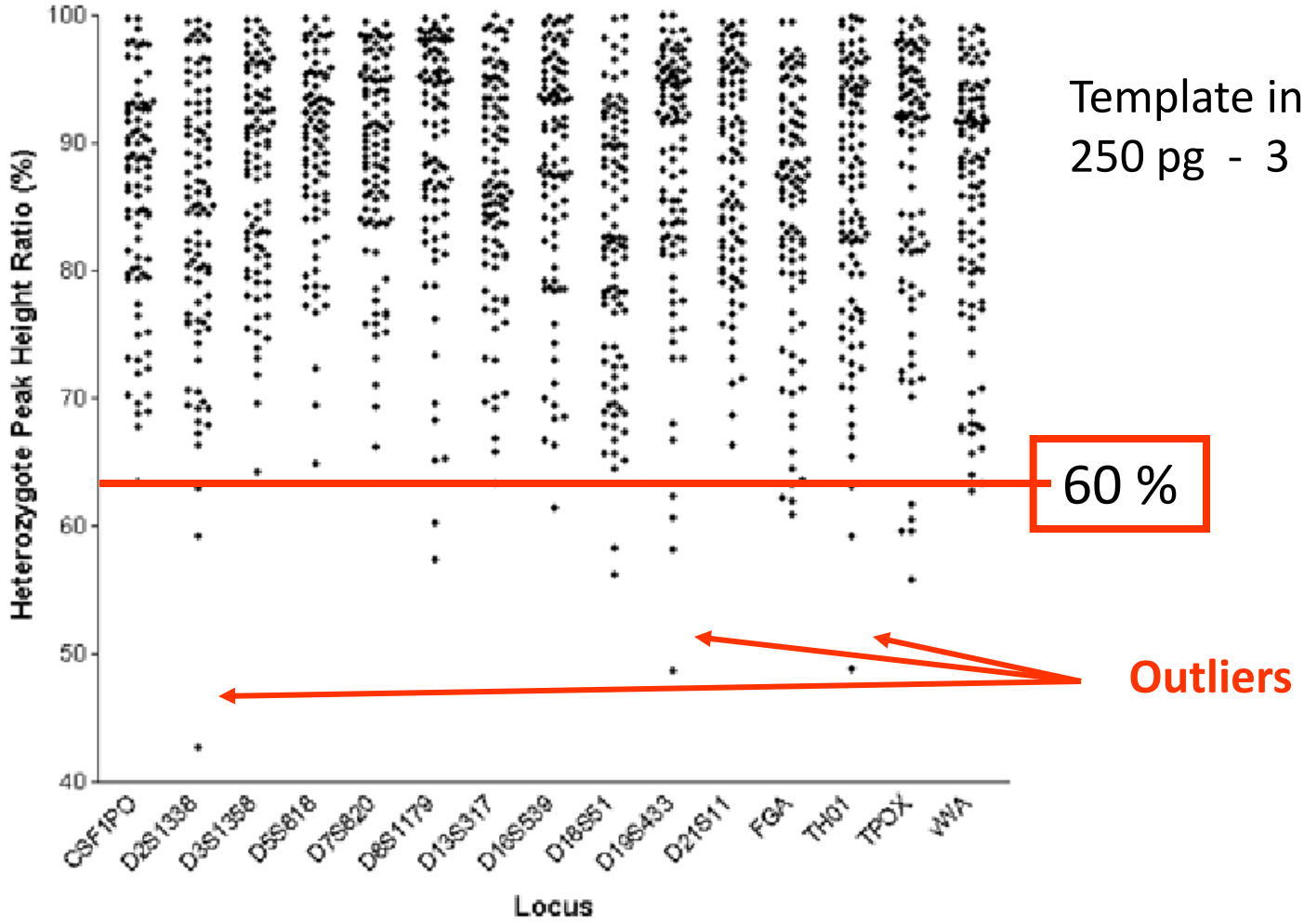
>20 copies per allele

6 copies copies per allele (LCN)

John Butler

Heterozygote Peak Height Ratios

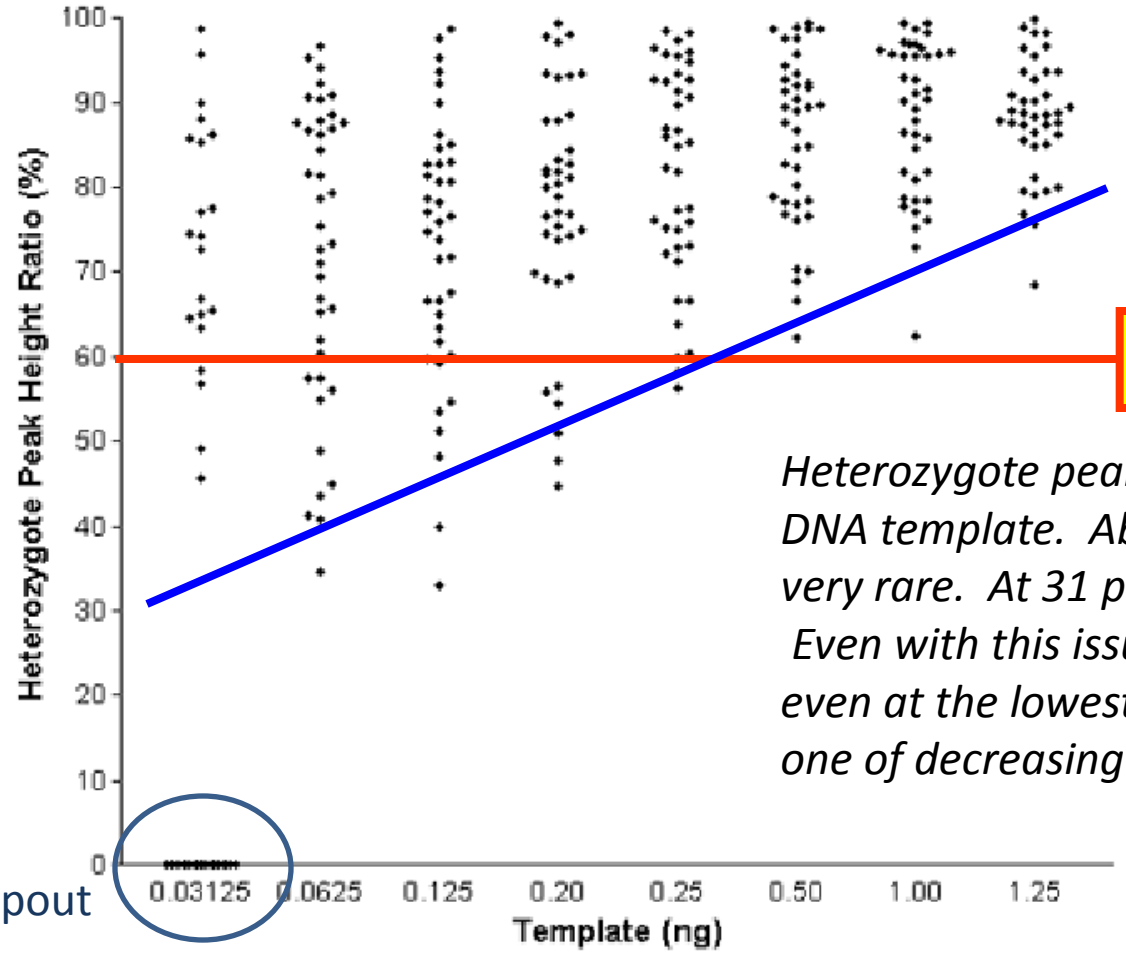
Identifiler STR Kit Developmental Validation



Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA. Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR amplification kit. *J. Forensic Sci.* 2004; 49(6): 1265-1277.

Heterozygote Peak Height Ratios

Identifiler STR Kit Developmental Validation



Allele dropout

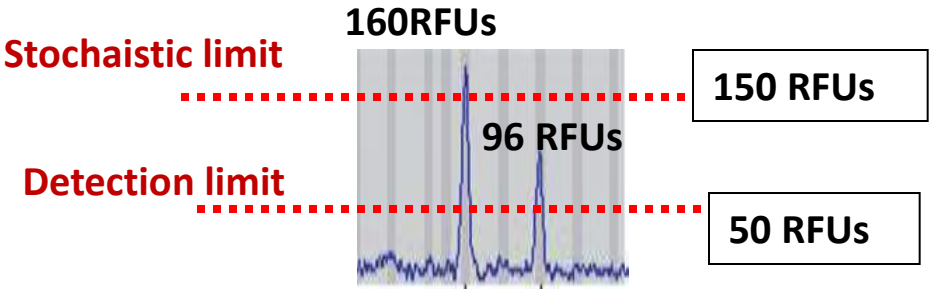
Heterozygote peak height ratios vary with DNA template. Above 250 pg, the effect is very rare. At 31 pg severe dropout occurs. Even with this issue, full profiles can be seen even at the lowest levels. The problem is one of decreasing reliability

Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA. Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR amplification kit. *J. Forensic Sci.* 2004; 49(6): 1265-1277.

The grey zone is predicated on a minimum peak height ratio (PHR)

Above the stochastic limit a peak at 60% RFU will show is partner allele in the grey zone.

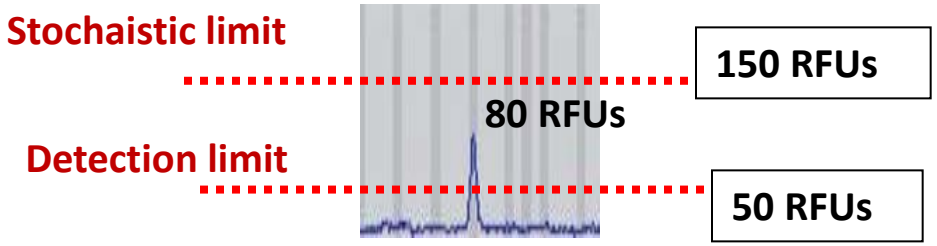
One allele peak above the detection threshold and one below



A single peak in the grey zone is considered unreliable as it may be heterozygous and its partner allele dropped in the noise.

Statistically how will you calculate the result?

Inconclusive result due to present of potential allele in between two thresholds.



Unusable result: (60% of 80 RFU is 48 and below threshold)

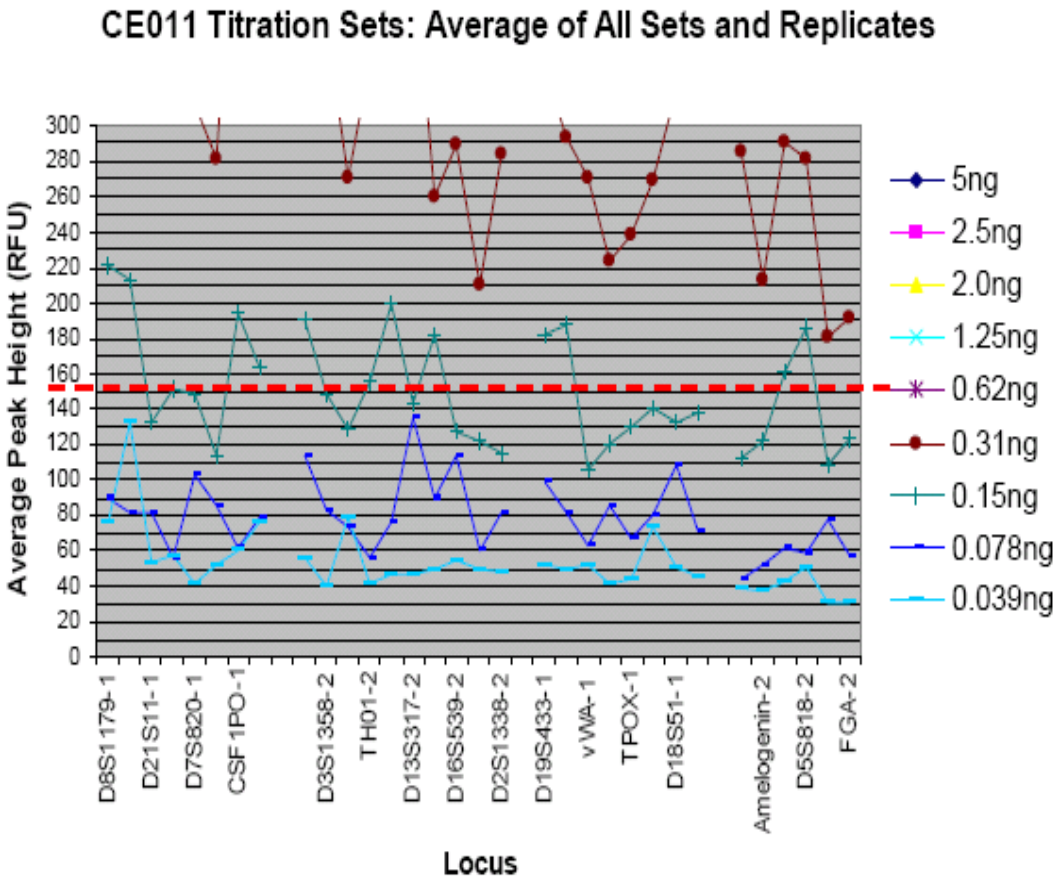
Alternative Procedure (Mass State Police)

1. Since most estimates for LCN show up from 100-250pg DNA, select a low level- say 150pg as your stochastic limit.

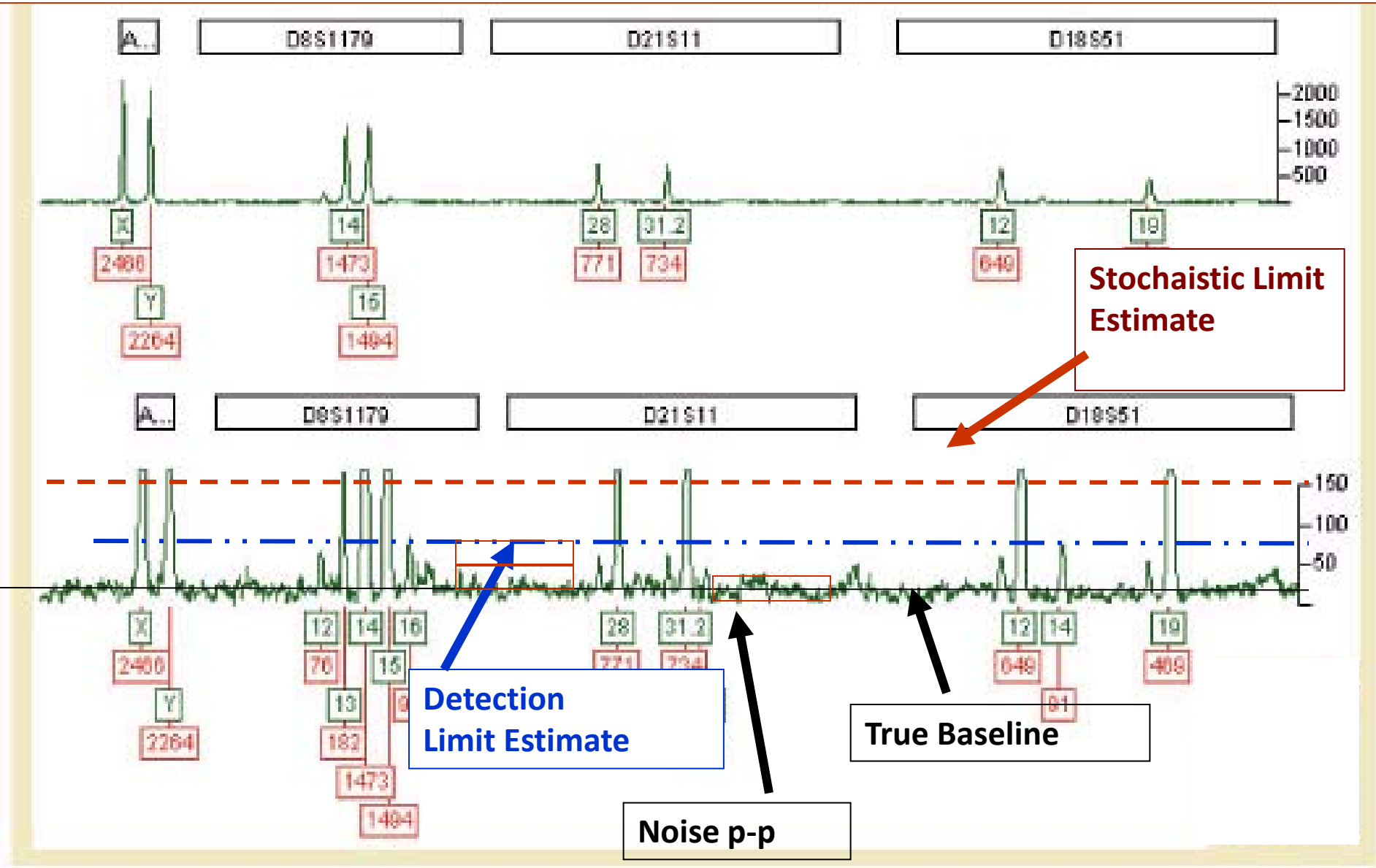
2. Amplify 2 or more samples at a range of concentrations (1.0-0.005) ng multiple times and score the intensity

3. The stochastic limit is the intensity (RFUs) at which half the alleles have intensity above this value and half are below

4. In this way you define straddle data as at the point 50% of your alleles will be above this mark



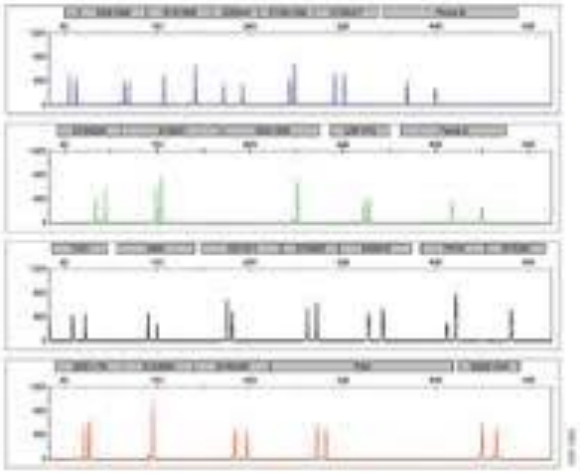
The issue with low level data interpretation



Thompson, W.C. et al. The Champion, April 2003, pp 16-25

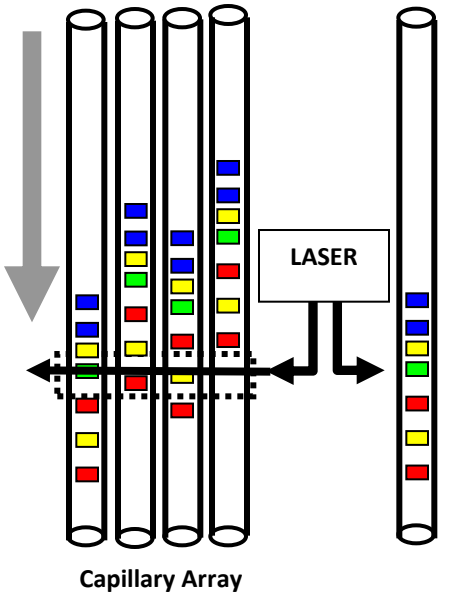
Sensitivity issues

- Improved STR multiplexes have better buffers, more mini STRs and increased sensitivity
- What is wrong with that?
- The fundamental problem with the PCR is stochastic amplification. It exists regardless of sensitivity of detection, and manifests itself as peak imbalance, enhanced stutter and peak dropout.
- These new kits and instruments detect better, but the PCR hasn't changed!!! New thresholds need to be set.



Powerplex Fusion-
24 loci multiplex

Previously suggested thresholds ABI310 vs 3100



Threshold (ABI)
310 50 RFUs
31xx 30 RFUs

Stochaistic
310 150 RFUs
31xx 90 RFU

Dynamic Range
310 4500
311 31xx 3500

Bottom line: 310 will appear more sensitive with a wider dynamic range. 3100 detects more samples with its array but has a more limited dynamic range.

However, new kits and new instruments may change these values!

The ABI 3500 a new paradigm?

This new instrument has 6 dye lanes and an improved laser, but the dynamic range of the system is much different-

RFU signals can range from 0 to over 20,000 RFU.
This compares to approximately 0-4500 RFU for a standard 310
What does this mean in practical terms?

1. The old threshold values for the 310 and 3100 can not be used, new ones must be calculated

2. 100pg of DNA produces heterozygous peaks of 750 or more indicating a stochastic threshold of 750 RFU – a very high value for labs used to 310 or 3100.

3. With the potential for pullup of up to 2 %, or more, laboratories may expect artifact peaks up to 400 RFU.

4. Signal to noise is quite low and caution must be taken when setting the analytical threshold particularly given the issue of artifacts– How to deal with this???

Peak Balance ABI-3500 ID+

Major problems with peak balance below 125pg This indicates strong potential for false homozygous peaks.

Input DNA	Average Peak Height (RFU)
0.03125 ng	305.83
0.0625 ng	476.21
0.125 ng	1213.63
0.1875 ng	1482.28
0.25 ng	2357.11
0.5 ng	5526.54
1 ng	7418.41
2 ng	16367.16

28 cycle/15 seconds

Input DNA	Replicate	D8S1179	D21S11	CSF1PO	D3S1358	TH01	D19S433	TPOX	D18S51	AMEL	D
5 ng	1	89.82%	98.81%	90.83%	99.33%	99.82%	90.65%	91.61%	92.09%	99.39%	9
	2	92.34%	91.22%	85.60%	96.74%	99.01%	88.18%	97.77%	85.61%	96.56%	9
	3	90.97%	93.56%	94.19%	92.84%	96.83%	89.85%	98.40%	92.09%	89.78%	9
2 ng	1	94.58%	89.40%	94.81%	98.13%	90.67%	87.63%	96.40%	85.03%	97.20%	8
	2	83.55%	86.88%	89.88%	96.09%	98.65%	90.37%	95.39%	92.68%	89.36%	9
	3	98.47%	96.31%	93.57%	94.27%	97.04%	90.81%	88.91%	90.55%	92.21%	9
1 ng	1	93.75%	83.79%	83.73%	93.96%	91.70%	90.34%	97.88%	90.61%	98.33%	9
	2	86.08%	99.69%	92.43%	85.63%	80.30%	96.74%	98.91%	95.11%	98.31%	9
	3	91.24%	98.18%	74.74%	94.98%	92.86%	95.44%	97.27%	90.49%	93.08%	8
0.5 ng	1	75.18%	[illegible]	75.89%	84.21%	91.88%	99.48%	93.53%	84.56%	99.58%	8
	2	[illegible]	[illegible]	94.40%	97.70%	92.27%	99.05%	98.61%	99.31%	74.02%	8
	3	[illegible]	[illegible]	89.25%	93.37%	99.84%	91.09%	77.25%	90.51%	91.25%	7
0.25 ng	[illegible]	[illegible]	65.16%	98.26%	53.13%	66.91%	91.08%	93.39%	87.48%	99.23%	85.17%
	2	74.91%	93.31%	95.95%	86.91%	77.42%	95.09%	69.47%	94.47%	66.58%	97.33%
	3	92.11%	96.49%	61.05%	86.28%	92.03%	94.49%	79.57%	96.84%	69.22%	61.30%
0.1875 ng	1	78.39%	83.20%	84.54%	74.75%	90.43%	77.88%	44.56%	41.61%	98.10%	96.89%
	2	83.15%	90.95%	62.90%	97.79%	94.73%	95.78%	90.35%	73.41%	90.19%	46.64%
	3	87.68%	74.84%	90.22%	98.43%	71.97%	83.44%	92.87%	98.91%	97.41%	77.74%
0.125 ng	1	98.11%	82.46%	70.98%	99.01%	84.72%	60.72%	37.24%	97.27%	93.26%	93.74%
	2	78.97%	85.81%	77.75%	61.81%	75.00%	69.97%	85.25%	83.91%	92.92%	96.68%
	3	[illegible]	66.61%	99.05%	54.27%	89.38%	83.33%	62.93%	84.38%	74.80%	70.73%
0.0625 ng	[illegible]	[illegible]	72.49%	97.96%	74.16%	71.63%	59.94%	61.48%	72.73%	89.91%	83.43%
	[illegible]	[illegible]	84.12%	41.48%	72.75%	75.39%	90.99%	75.97%	56.46%	46.61%	91.22%
	[illegible]	73.04%	27.59%	98.65%	60.84%	95.55%	91.03%	99.32%	93.26%	89.81%	66.67%
0.03125 ng	1	18.53%	55.62%	43.69%	83.98%	25.91%	99.10%	26.96%	51.28%	63.41%	40.63%
	2	59.58%	40.96%	39.94%	54.79%	64.73%	77.21%	97.94%	87.29%	54.87%	69.53%
	3	51.22%	84.62%	89.61%	66.98%	86.42%	70.40%			62.41%	

PHR starts to falter

Danger zone

Setting an analytical threshold based on method, not instrument

Many times there is more to the analytical method than just doing a reaction or submitting it to direct analysis. For example it might be necessary to heat a sample that is to be analyzed for a particular metal with the addition of acid first (this is called **digestion**). The sample may also be diluted or concentrated prior to analysis on an instrument. Additional steps in an analysis add additional opportunities for error.

Since detection limits are defined in terms of error, this will naturally increase the measured detection limit. This detection limit (with all steps of the analysis included) is called the MDL. The practical method for determining the MDL is to analyze 7 samples of concentration near the expected limit of detection. The standard deviation is then determined. The one-sided t-distribution is determined and multiplied versus the determined standard deviation. For seven samples (with six degrees of freedom) the t value for a 99% confidence interval is 3.14

(Wikipedia)

Using a Method Limit of Detection may be a more realistic way to set an Analytical Threshold than an LOD as it takes more into account.

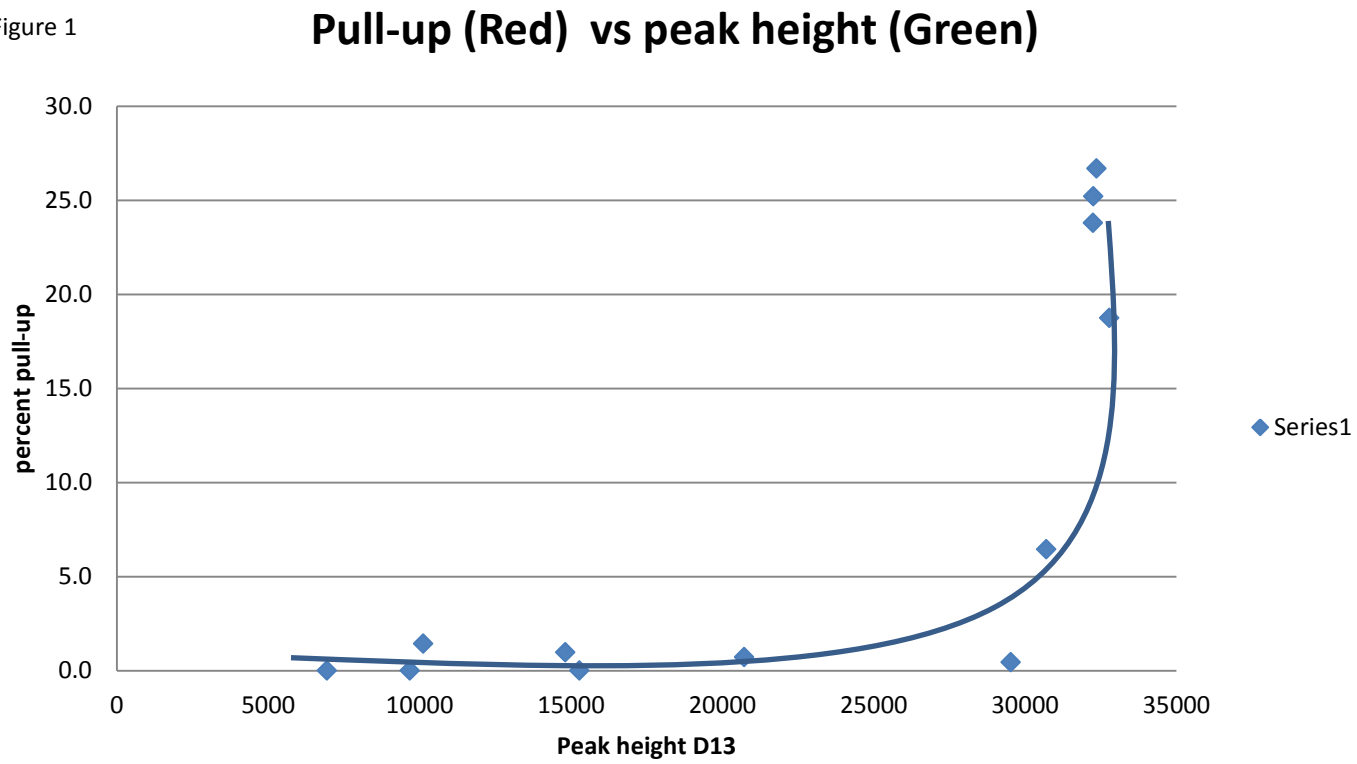
10 replicate amplifications of a low level sample indicating SD of 113 RFU.

10 Profiles from same extract-Male. Targeted 100 pg input DNA.						
Dye	FAM		Peak		Peak	
Locus	D8S1179	Allele 1	Height (RFU)	Allele 2	Height (RFU)	
Sample	JP20.1	11	464	15	227	
	JP20.2	11	453	15	494	
	JP20.3	11	276	15	120	
	JP20.4	11	412	15	338	
	JP20.5	11	433	15	450	
	JP20.6	11	173	15	332	
	JP20.7	11	324	15	389	
	JP20.8	11	214	15	208	
	JP20.9	11	215	15	321	
	JP20.10	11	202	15	353	
	Average Peak Height:		316.6		323.2	
	stdev		115.1271761	0	112.8	

Indicates a method limit of detection for an ABI 3500 of 3 x 113 ≈ 350 RFUs

3500 pull-up study (for determining limit of linearity)

Figure 1



Pull-up becomes excessive above 20,000 RFU

indicates limit of linearity for a 3500 around 20,000-25,000 RFU

Thresholds for the 3500

For 10 samples independently amplified and run on a 3500, the pooled standard deviation is approximately 113. Multipled by 3 this gives an analytical threshold of approximately 350.

At 100 pg the average peak intensity for a 3500 was measured at approximately 750 RFU this provides an estimate of the stochastic threshold.

Thus for the 3500 one could set a stochastic threshold at 750 RFU and an analytical threshold based on the MDL of 350.

The Limit of linearity is around 20,000 RFU and was set by examining the effect of pullup vs input DNA quantity.

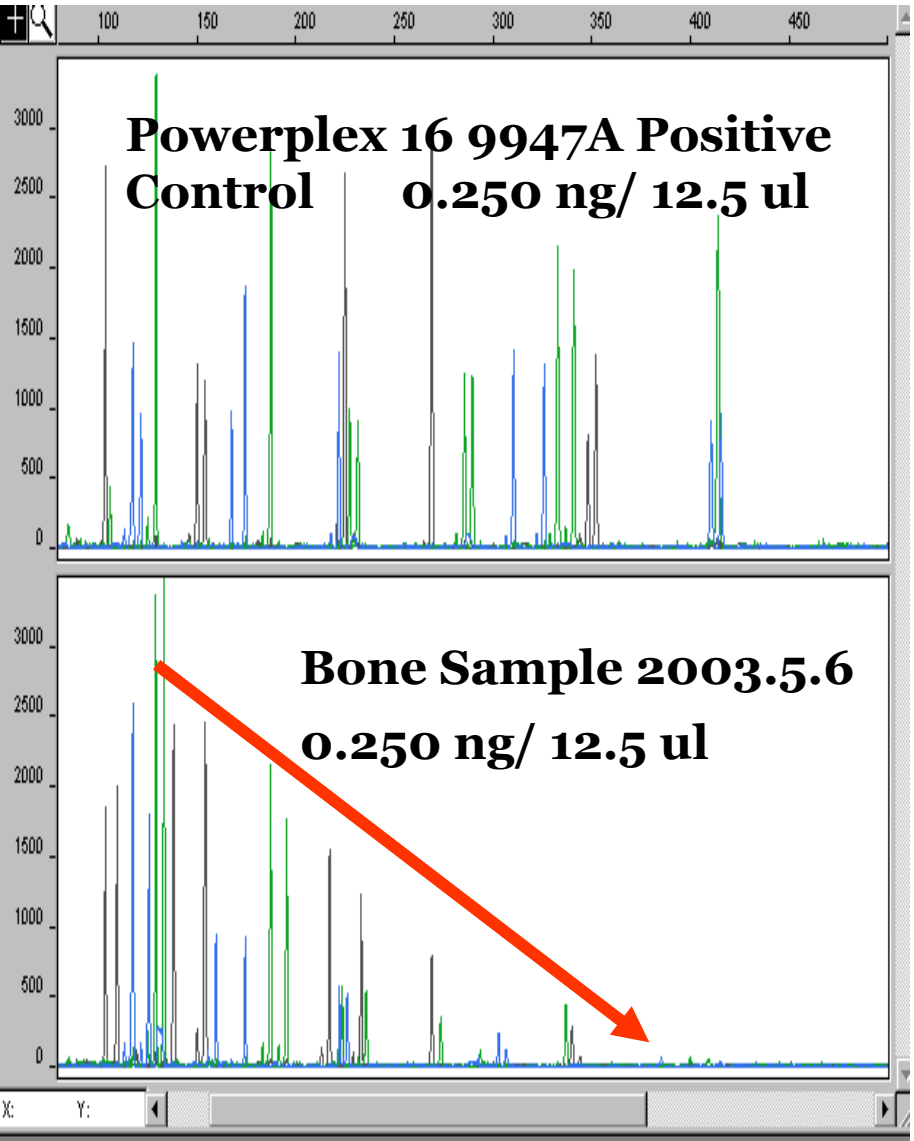
What else can go wrong?

Yarr,Take care mates!

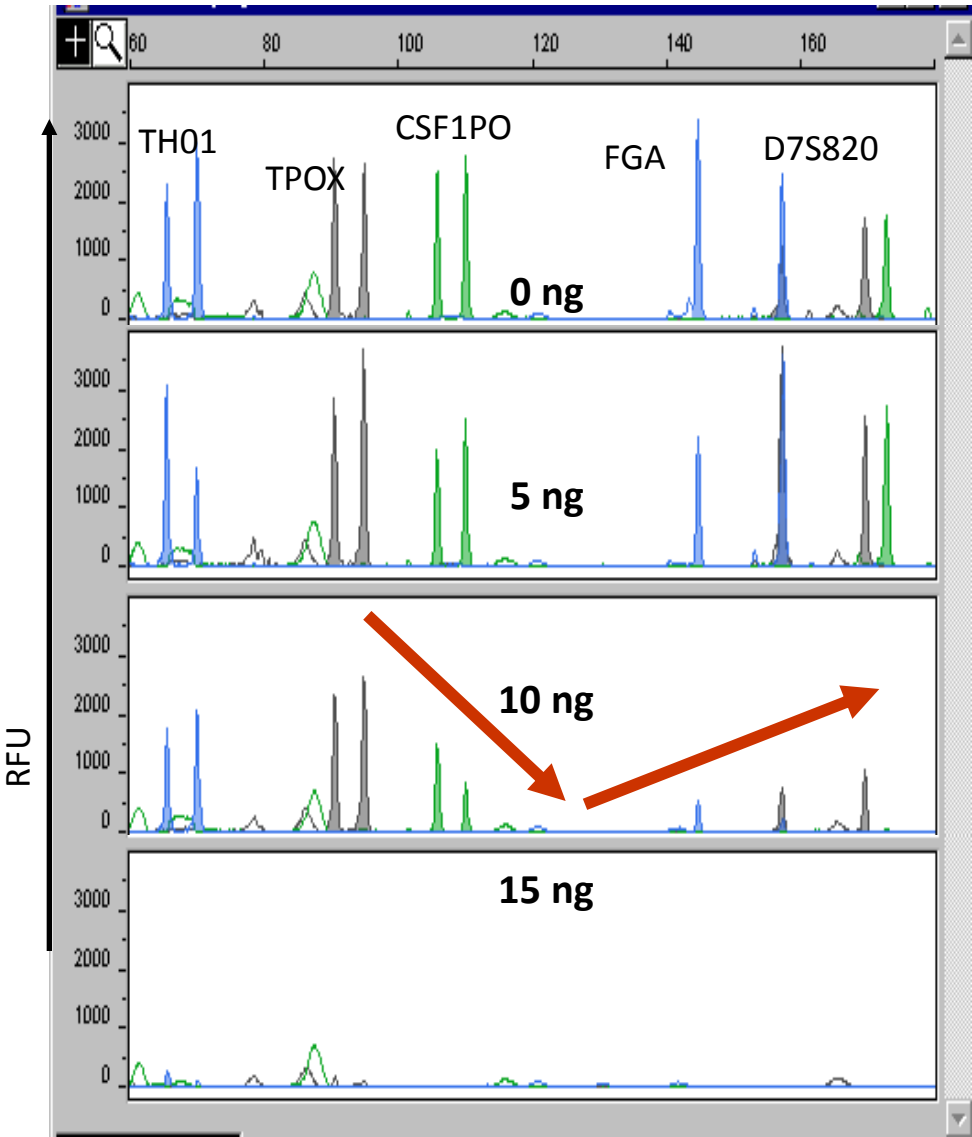
- Most validation studies are performed on pristine samples derived from clean sources.
- DNA degradation will result in dropped alleles from larger sized amplicons
- DNA inhibition will result in dropped alleles from any location and the effects are difficult to predict
- Inhibition and degradation can produce stochastic effects – peak balance issues and allele dropout.

Degradation vs Inhibition

Degraded DNA Sample
Ski slope effect



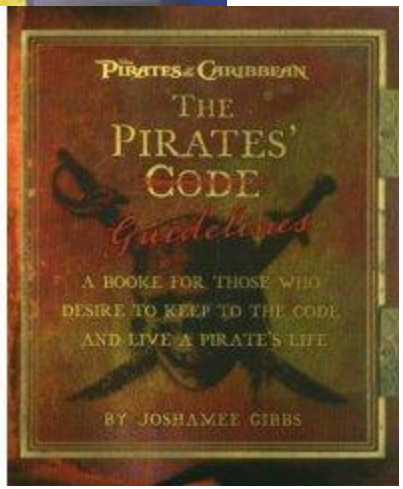
Humic Acid Inhibited DNA Sample
Less predictable effects



The bottom line:

1. Low signal levels are bad because:
   a. They may indicate low copy # DNA = inconsistent or confounding results
   b. They often coincide with peak imbalance
   c. PCR and instrumental artifacts appear at these levels
2. Relying on signal level to determine DNA quantity can be misleading
   a. There is wide variation in signal strength of amplified DNA
   b. Inhibitors and mixtures complicate interpretation
      1. peak imbalance can occur even in single source samples due to inhibition and degradation
      2. instruments can vary in sensitivity

Fuzzy Logic in Data Interpretation



- Capillary Electrophoresis is a dynamic processs

- Sensitivity varies with
  - Allele size
  - Injection solvent
  - Input DNA
  - Instrument factors
  - Presence of PCR inhibitors
  - Gel matrix

- Thus interpretation must be conservative and data from these studies yields guidelines, not rules.

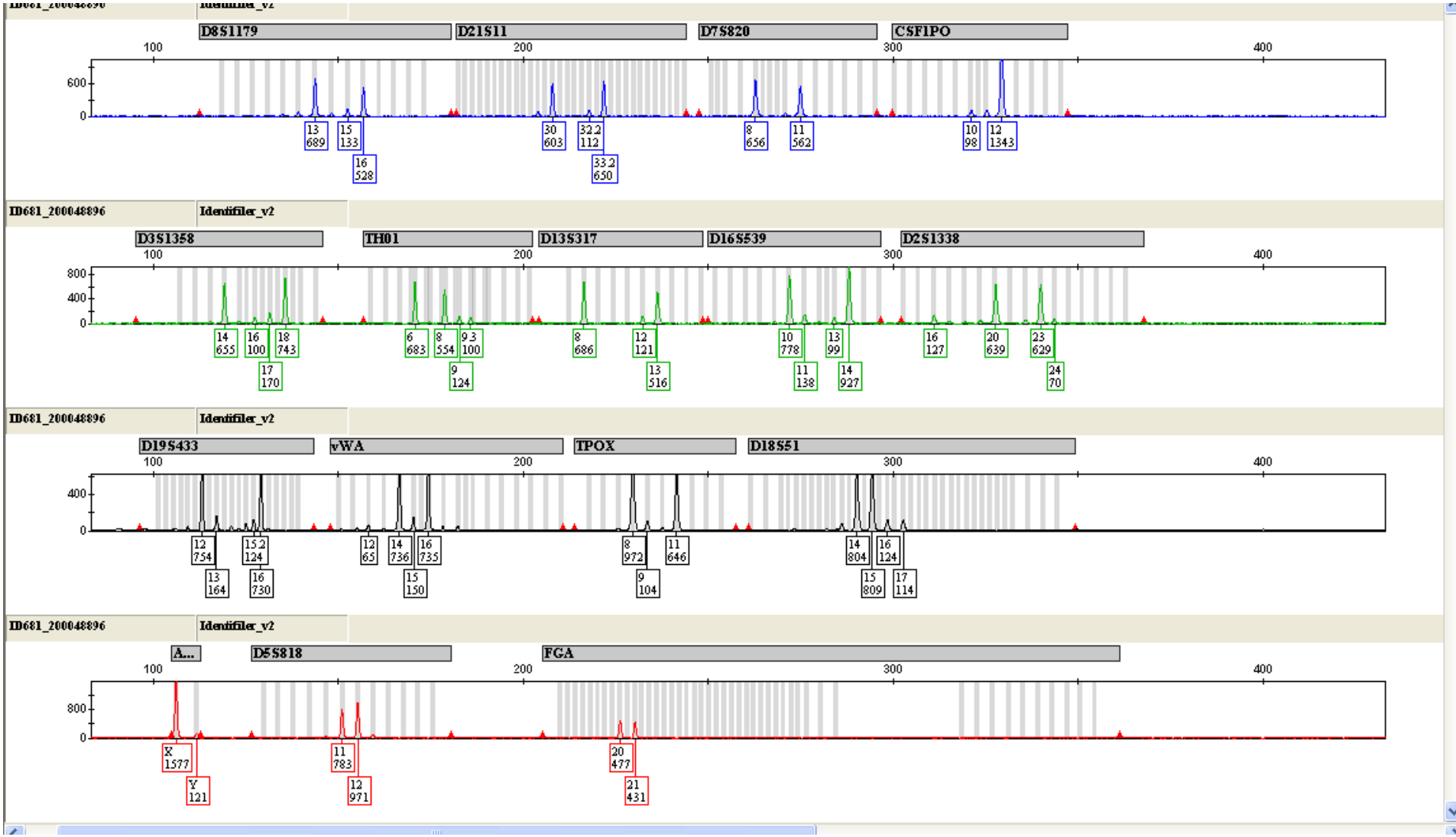
- In addition the interpretation and its significance cannot be dissociated from the overall facts of the case.

So why examine low level data at all?

- Touch DNA can be a powerful lead in a criminal investigation
- Detection of the presence of low level mixtures
- Clues to the presence of inhibited samples or poor injections
- Aids in determination if a suspect is excluded as a contributor
- **So How to Make it Reliable?**

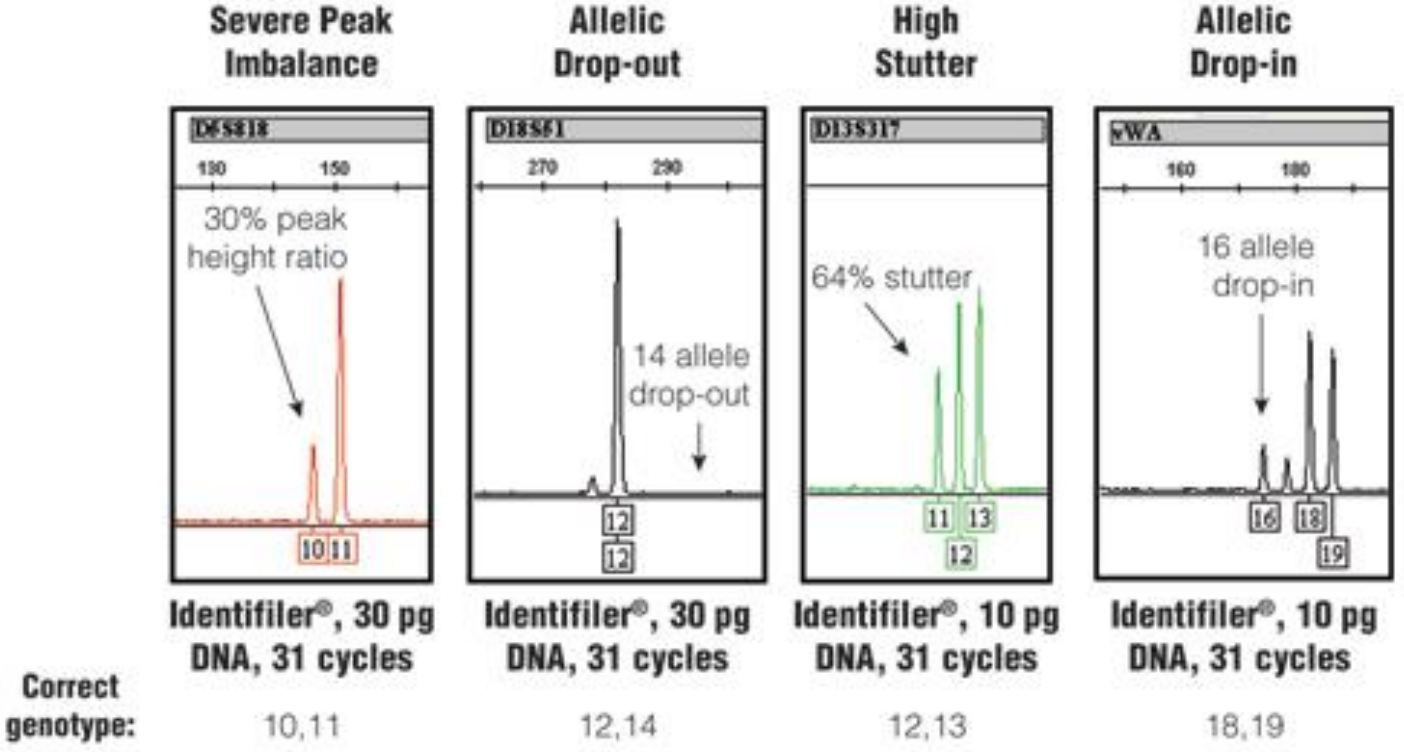
Extract of touch DNA

If the major profile is the victim, can we find the suspects alleles?



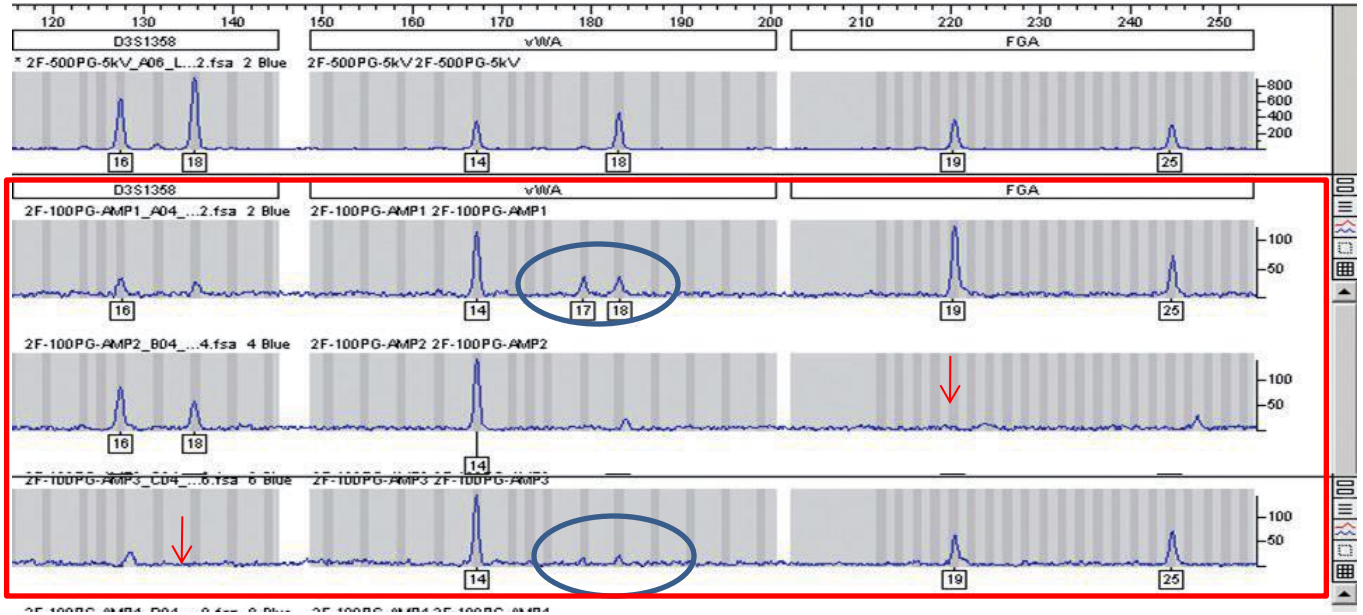
Analyst must consider allele dropout, drop-in, stutter and masking of peaks.
A very complex problem.

The problem occurs when you have to interpret a low level contributor or a single source profile along with these stochastic artifacts



Butler, J.M. and Hill, C.R. http://www.promega.com/resources/profiles-in-dna/2010/scientific-issues-with-analysis-of-low-amounts-of-dna/

Reproducibility over the long term



Bruce Budowle, Arthur J. Eisenberg, Angela van Daal, Validity of Low Copy Number Typing and Applications to Forensic Science, Croatian Medical Journal, 2009, 50, 207-216.

How to score this?

D3 16, Z ; vWA 14, Z ; FGA 19, 25 where Z indicates a potential partner allele

1. Allele dropouts everywhere but are not reproducible,
2. Allele dropins are mainly in the stutter position.
3. You need to run replicates to interpret this kind of data.

This kind of data is why we have thresholds and why these thresholds should be conservative.

Meatloaf Principle

- I want you
  I need you
  But -- there aint no way 'Im ever gonna love you
  Now don't be sad
  cause two out of three aint bad
  – Meatloaf

Catch 22 Principle

- The Catch 22 is that to do all these runs you end up wasting precious sample, further reducing the quality of results.

- It might be better to use more
  and amplify less – but is it scientific?

Interesting effects with low copy data

Table 1 Details of analysis

Number of single cells analyzed	226
Results obtained	206 (91%)
Amplification failure	20 (9%)
Full STR profile	114 (50%)
Acceptable profile (amelogenin,>4 STRs)	144 (64%)
Partial profile (1-4 STRs)	62 (27%)
Surplus alleles*	28 (12%)
False alleles**	11 (5%)
Allele dropout	88 (39%)

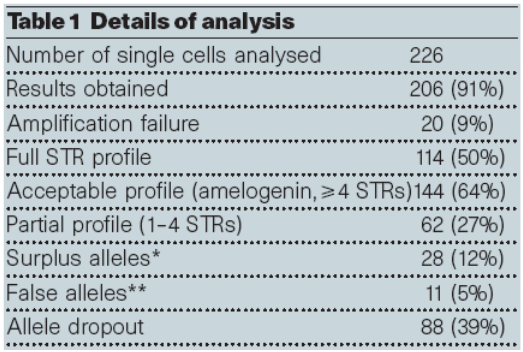


Table 1 Details of analysis

Number of single cells analysed	226
Results obtained	206 (91%)
Amplification failure	20 (9%)
Full STR profile	114 (50%)
Acceptable profile (amelogenin, ≥4 STRs)	144 (64%)
Partial profile (1–4 STRs)	62 (27%)
Surplus alleles*	28 (12%)
False alleles**	11 (5%)
Allele dropout	88 (39%)

Findley *et al.* (1997) *Nature* article

*Additional allele present in conjunction with true alleles.
**Additional allele in place of true allele.

So with one cell 64% of the time you get a usable profile and 27% a partial profile =91 % OK!

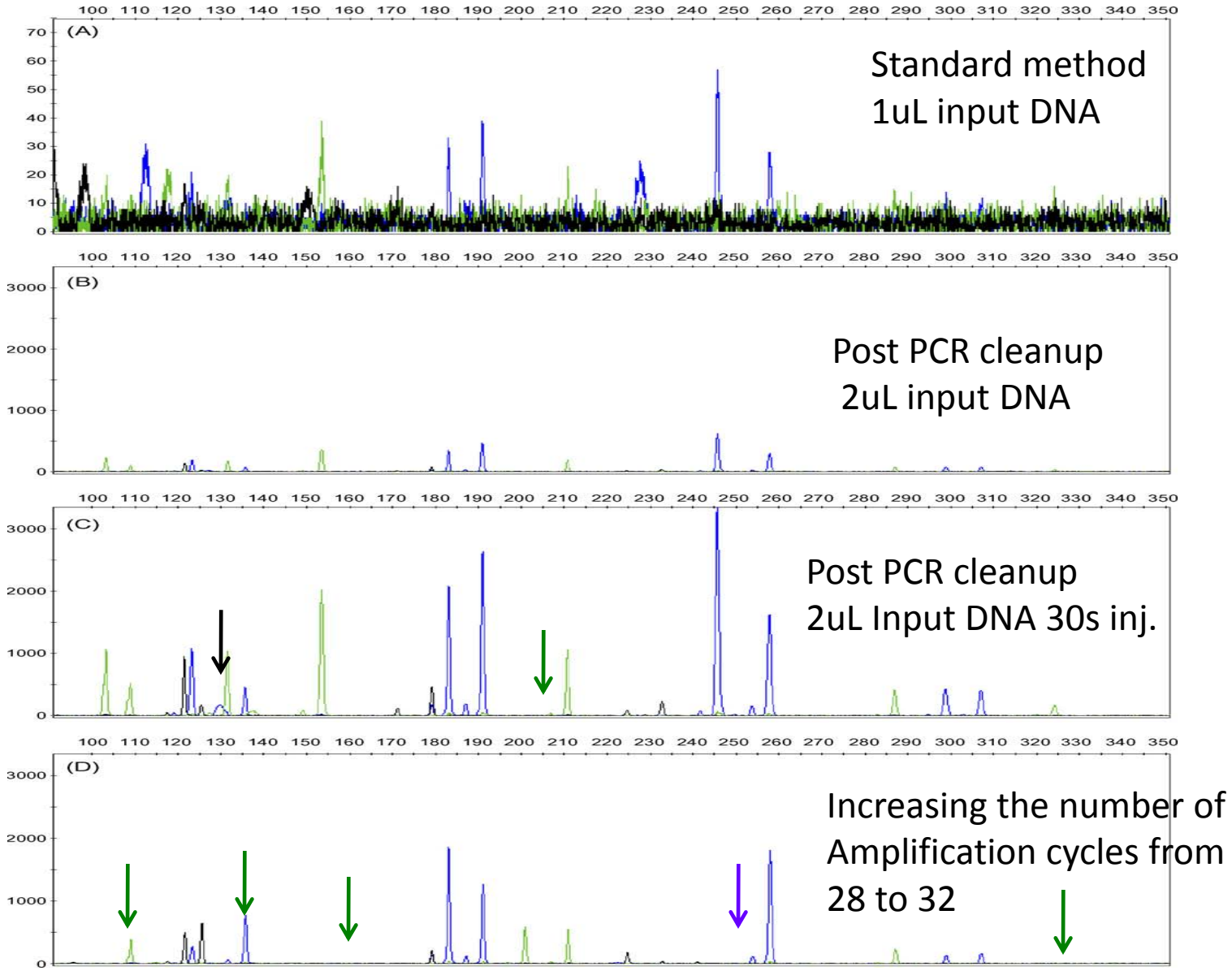
How would you know had a problem?

The danger is clearly false exclusion and the potential presence of low level exogenous DNA.

Post PCR manipulation

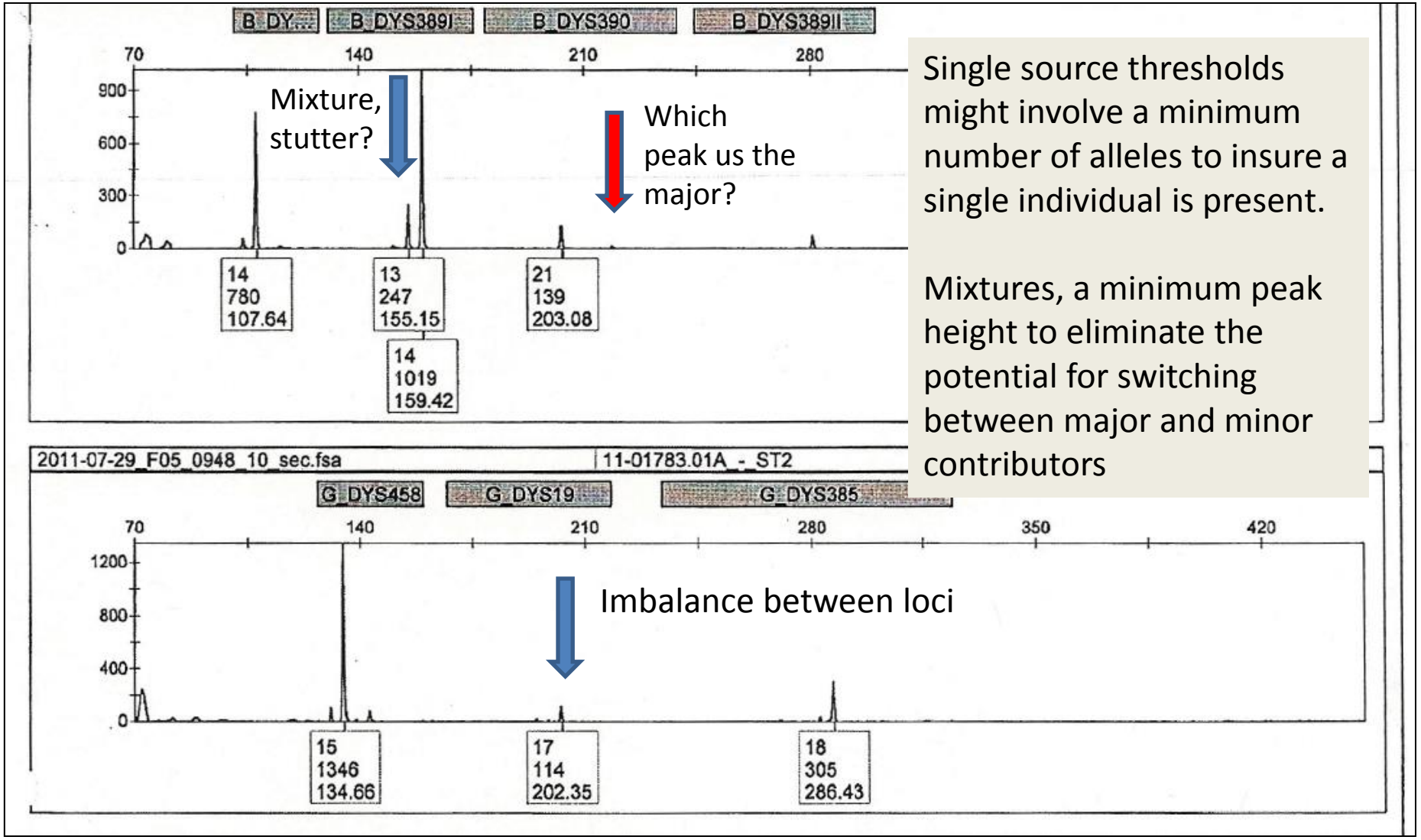
- Why not just cleanup after PCR?
  - PCR sample is concentrated but:
    - Spin filtration may result in removal of background salts,
    - This can greatly enhance sensitivity due to the stacking process
    - Best idea- remake sample up in buffer, not water to avoid massive issues with stochastic effects.

Various methods to increase the sensitivity of a low level sample.
Post PCR cleanup ≈ Extra Cycles (different peaks, same sample)



Forster, L. Thomson, J. and Kutranov, FSI genetics, 2 (2008) 318.

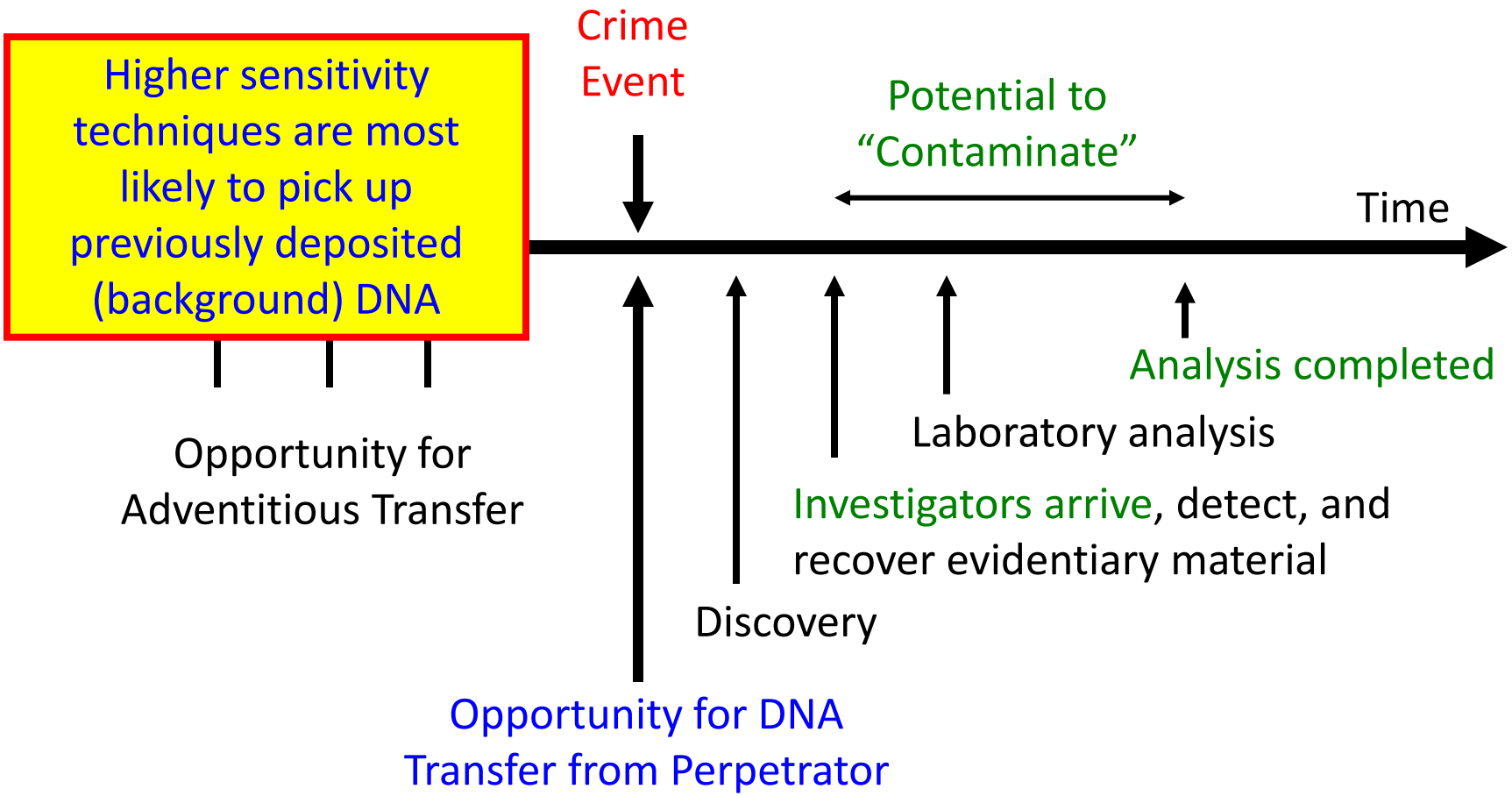
Does stochastic amplification occur with Y STRs? -YES!



Single source thresholds might involve a minimum number of alleles to insure a single individual is present.

Mixtures, a minimum peak height to eliminate the potential for switching between major and minor contributors

The problem with LCN DNA is that you cant be sure if it is from the case or if it is from some other event.



Adapted from Gill, P. (2002) *BioTechniques* 32(2): 366-385, Figure 5

John Butler

Guidelines for use with low template/low copy DNA

- Low copy DNA is not just more cycles, stochastic effects can occur anytime levels of input DNA are low.

  (effects begin below 250pg DNA, commonly set at 100pg)

- An allele should not be scored (considered real) unless it is present at least twice in replicate samples, usually 3 replicates are performed.

- Extremely sterile environments are required for PCR setup to avoid contamination from laboratory personnel or other sources. Personnel must be typed for contamination events.

- The potential for contamination from DNA not related to the events in the case must always be considered.

- Guidelines for minimum number of heterozygous alleles should be considered.

A Precautionary Tale

When we analyze lower amounts of DNA…

It is critical to understand the importance of setting thresholds and knowing input levels.

WHEN ARE SUCH SAMPLES SUBJECT TO LCN ISSUES?

Murder of a Witness by Rifle (recovered)

GENETIC MARKERS

ub #	ITEM	FGA	TPOX	D8S1179	vWA	Amelo	Penta E	D18S51	D21S11	TH01	D3S1358	Penta D	CSF1PO
2	Suspect	19 30.2	8 9	14	15 17	X Y	9 12	15 16	28	8 9	15 18	10 11	10 12
3	Accomplice	23.2 24	8	13	14 17	X Y	7 10	14 18	29 33.2	6 9	16 17	12 14	13 15

State Lab
Private Lab

GER GUARD-TRIGGER + HAMMER	19 (25) (26) (30.2)	8 9	14	15 17	X (Y)	9 (11) (12)	(14) (15) 16	28	8 9	15 18	(10) 11	12	
T. MINIFILER DATA:	21 PA = 25				X Y		PA	29*				10	?

State Lab
Private Lab

T SIDE OF BARREL STOCK -REAR	19 (23.2) 24 (30.2)	8 (9)	13 14	15 (17)	X Y	(7) 9 (10) 12	(14) 15 16 (18)	28 (29) 33.2	(6) (8) 9	15 (16) (17) 18	(10) 11 (12) (14)	10 (12) (13) (15)	
MINIFILER DATA:	21 25				X Y		17*	30*				14	

Recovered profile from laboratory matches suspect and accomplice

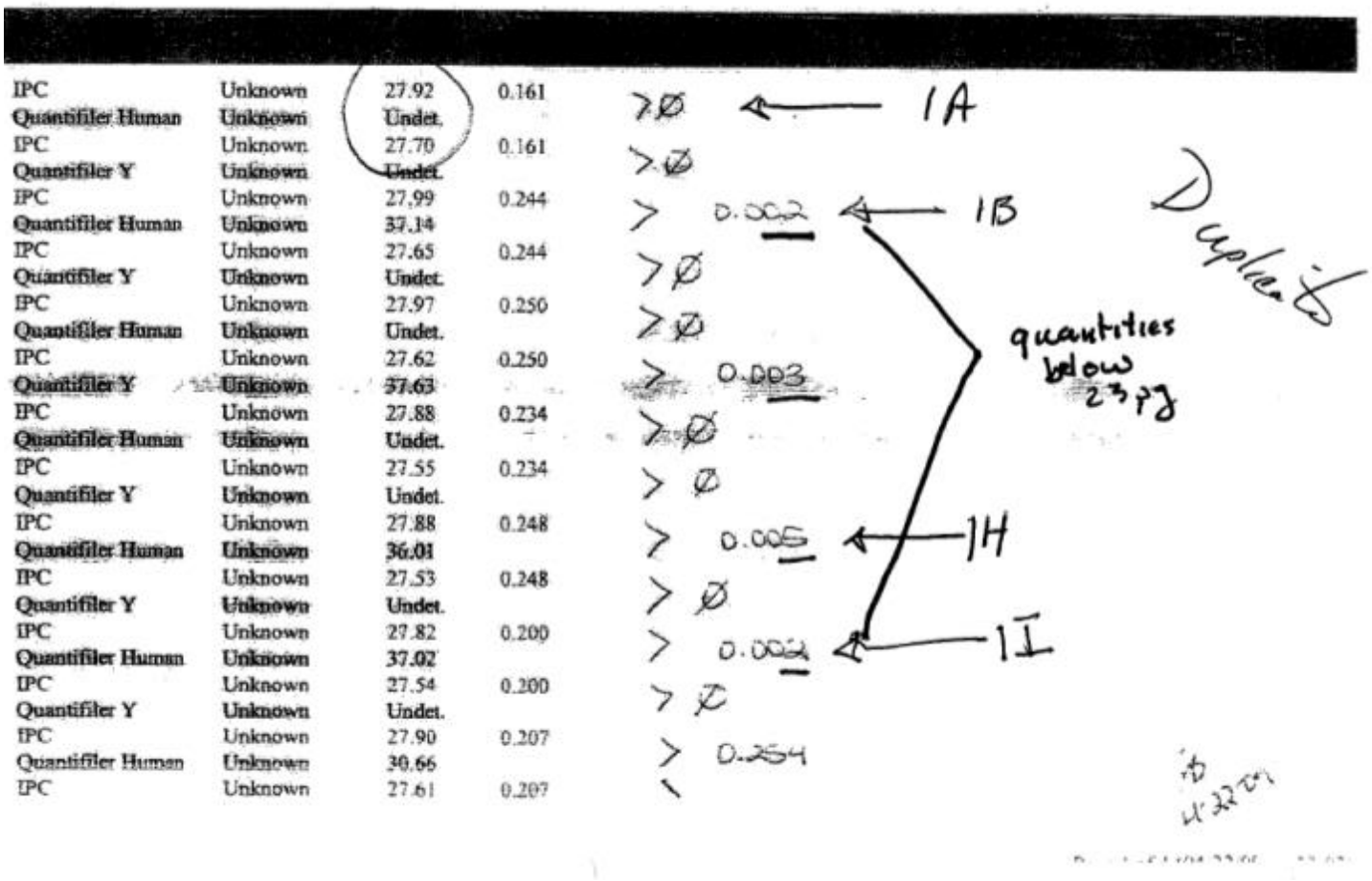
Second lab analyzes remainder of swab and gets a completely different result - minifiler

Would you exclude a suspect based on this data?

Look closer:

Quantitation indicated no DNA for the second sample…..

Even worse, the lab reported quantitative data below the lowest calibrator! (23pg/uL) A very bad practice….

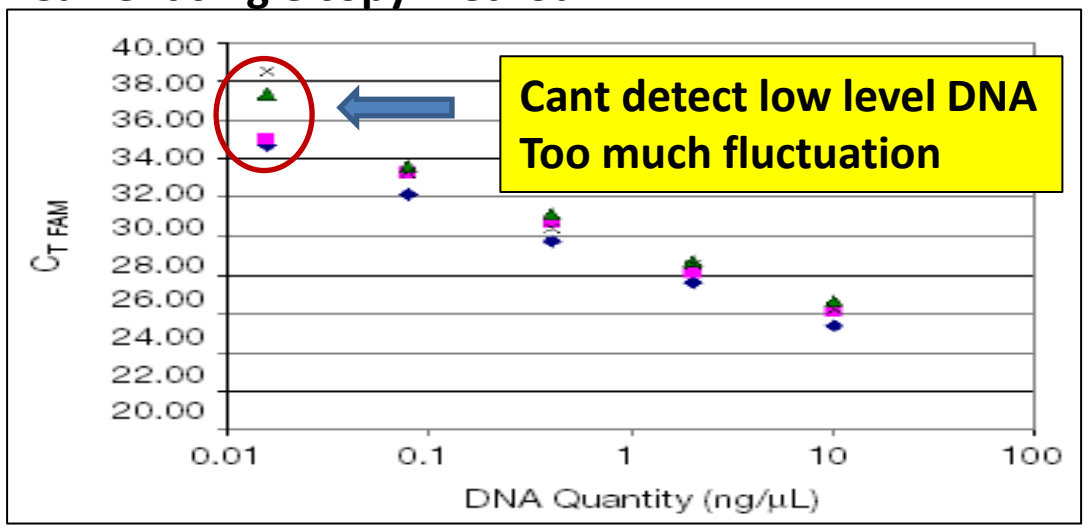


IPC	Unknown	27.92	0.161
Quantifiler Human	Unknown	Undet.	
IPC	Unknown	27.70	0.161
Quantifiler Y	Unknown	Undet.	
IPC	Unknown	27.99	0.244
Quantifiler Human	Unknown	37.14	
IPC	Unknown	27.65	0.244
Quantifiler Y	Unknown	Undet.	
IPC	Unknown	27.97	0.250
Quantifiler Human	Unknown	Undet.	
IPC	Unknown	27.62	0.250
Quantifiler Y	Unknown	37.63	
IPC	Unknown	27.88	0.234
Quantifiler Human	Unknown	Undet.	
IPC	Unknown	27.55	0.234
Quantifiler Y	Unknown	Undet.	
IPC	Unknown	27.88	0.248
Quantifiler Human	Unknown	36.01	
IPC	Unknown	27.53	0.248
Quantifiler Y	Unknown	Undet.	
IPC	Unknown	27.82	0.200
Quantifiler Human	Unknown	37.02	
IPC	Unknown	27.54	0.200
Quantifiler Y	Unknown	Undet.	
IPC	Unknown	27.90	0.207
Quantifiler Human	Unknown	30.66	
IPC	Unknown	27.61	0.207

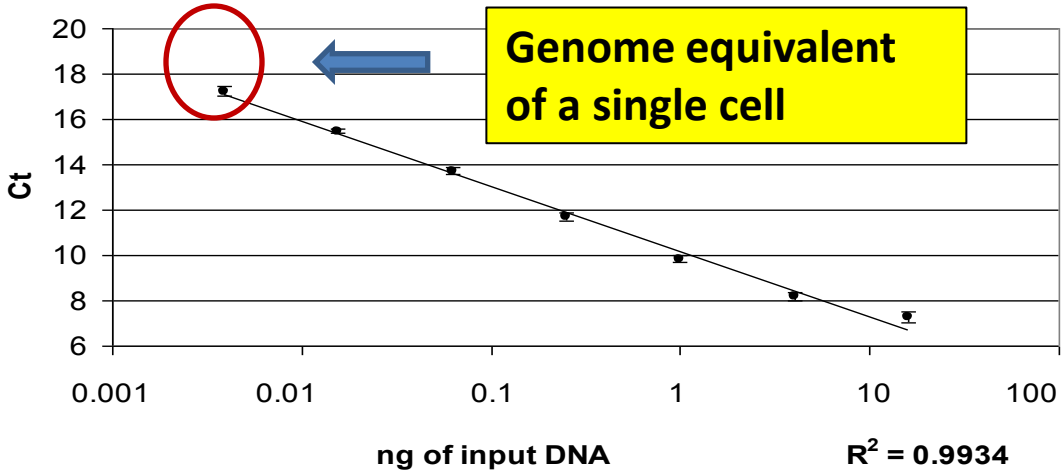
Defendant was declared guilty. It helped that there was a video of him buying the rifle….

If you are going to look at low level samples at least measure them by multicopy qPCR!

Current single copy method



MultiCopy Alu based Method



Why isnt everyone using multicopy techniques?

Nicklas, Buel, et. al. J Forensic Sci, 48(5), 2003.

1. Conclusions

1. CE based DNA analyses are complex

   a. Separations are affected by polymer length, concentration and field strength
   b. Injections vary greatly with salt content and PCR product quality
   c. Detectors require careful monitoring for pull-up and intensity

2. There are a multiplicity of instrument thresholds
   a. Analytical thresholds are based on standard deviation of noise.
   b. Stochastic thresholds are based on fundamental aspects of the PCR reaction which result in peak imbalances
   c. Limits of linearity affect ultimate peak height,stutter and pullup.
   d. Remember however, thresholds are guidelines- not rules be smart and keep things within the context of the case.

Acknowledgements

FIU

Maurice Aboud,

Kerry Opel

Jiri Drabek

Maximilien Blas

Robyn Thompson

Heather Eric

Jing Wang

Hye Hyun Oh

Martin Tracey

Maribel Funes

Ohio

Denise Chung

Brittany Bagley

Other Places

John Butler – NIST

George Duncan – Broward Sheriff's Office

Eric Buel, Jan Nicklas – Vermont Crime Lab

Debbie Fisher, Janet Doyle, Alice Isenberg - FBI

McCord Research Group

Major support provided by
The National Institute of Justice
as well as
Promega, Agilent, Qiagen, NSF, Streck,
Pressure Biosciences, FIU, OU.

Portions of this project were supported by the National Institute of Justice. Points of view in the document are those of the authors and do not necessarily represent the official view of the U.S. Department of Justice