DNA Interpretation Workshop 1

Data Interpretation Fundamentals

John M. Butler, PhD
U.S. National Institute of Standards and Technology (NIST)

john.butler@nist.gov
+1-301-975-4049
http://www.cstl.nist.gov/strbase/training.htm

ISFG Pre-Conference Workshop
Melbourne, Australia
September 2-3, 2013

NIST and NIJ Disclaimer

Funding: Interagency Agreement between the National Institute of Justice and NIST Office of Law Enforcement Standards

Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

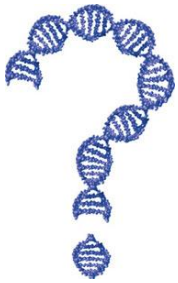
Our publications and presentations are made available at:
http://www.cstl.nist.gov/strbase/NISTpub.htm

Presentation Outline

- Introductions
- Data collection with ABI Genetic Analyzers
- Data interpretation overview
  - SWGDAM 2010 Interpretation Guidelines
- Stutter artifacts from PCR process
- Peak height ratios for heterozygous genotypes
- Number of contributors in mixed samples

Background of Participants…

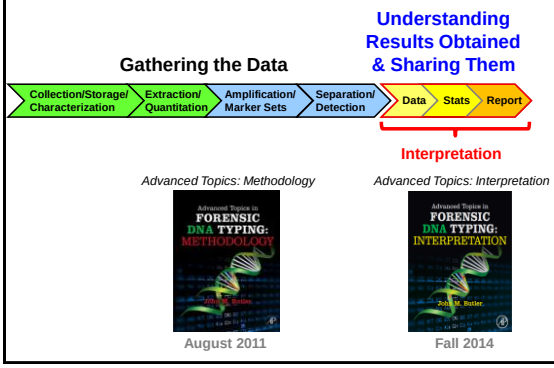
1) Your name
2) Where you are from (your organization)
3) **What you hope to learn from this workshop**



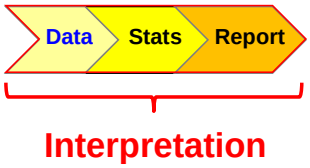
Importance of Improved Understanding Regarding DNA Mixture Interpretation

- Each DNA analyst may think his or her approach is correct – but misinterpretations have given rise to a variety of approaches being undertaken today, some of which are not correct…
- I believe that **a better understanding of general principles will aid consistency and quality of work being performed**

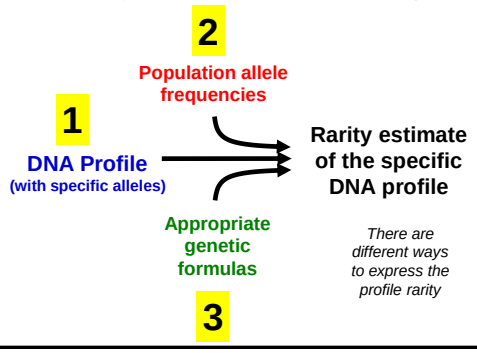
Steps in Forensic DNA Testing



Understanding Results Obtained & Sharing Them



Elements Going into the Calculation of a Rarity Estimate for a DNA Sample



Written summary of a recent interview…

The CAC News • 1st Quarter 2012 pp. 8-11

norah rudin & keith inman • the proceedings of lunch

The Discomfort of Thought
—a discussion with John Butler

"For the greatest enemy of truth is very often not the lie – deliberate, contrived and dishonest – but the myth – persistent, persuasive, and unrealistic. Too often we hold fast to the clichés of our forebears. We subject all facts to a prefabricated set of interpretations. We enjoy the comfort of opinion without the discomfort of thought."
—John F. Kennedy

Available at **http://www.cacnews.org/news/1stq12.pdf**

"…**we should spend as much time developing our interpretation skills as we do our methodological skills.** Technological progress (more sensitivity in detecting DNA, for example), can be a double–edged sword; without equivalent progress in interpretation skill, we are just as likely to cut ourselves as we are the target."

"Your interpretation and statistical methods should have consistent assumptions and go together for each assumption being made (e.g., you may interpret a mixture under alternative sets of assumptions)…"

President John F. Kennedy

Yale University commencement address (June 11, 1962)

"For the greatest enemy of truth is very often not the lie – deliberate, contrived and dishonest – but the myth – persistent, persuasive, and unrealistic. Too often we hold fast to the clichés of our forebears. We subject all facts to a prefabricated set of interpretations. **We enjoy the comfort of opinion without the discomfort of thought**."

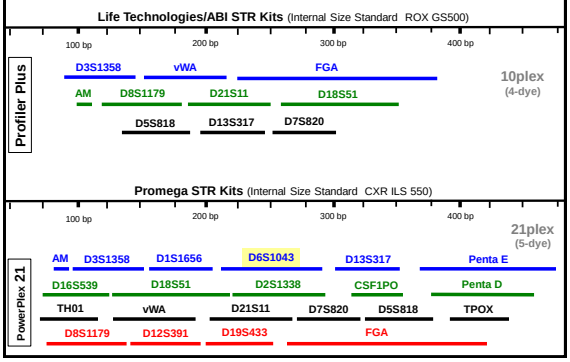
http://www.jfklibrary.org/Research/Ready-Reference/Kennedy-Library-Miscellaneous-Information/Yale-University-Commencement-Address.aspx

Different DNA Tests from Various STR Kits

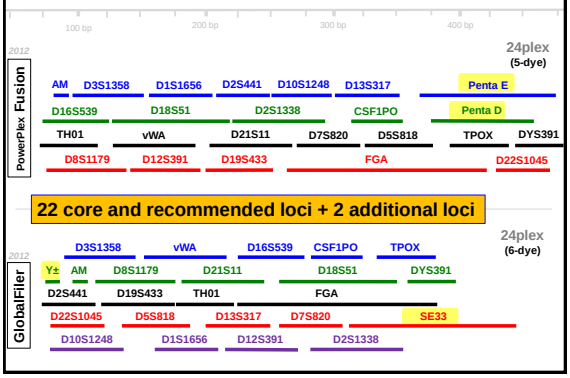
Kit Name	# STR Loci Tested	Manufacturer	Why Used?
Identifiler, **Identifiler Plus***	**15 autosomal STRs** (aSTRs) & amelogenin	Life Technologies (Applied Biosystems)	Covers the 13 core CODIS loci plus 2 extra
PowerPlex 16 **PowerPlex 16 HS***	**15 aSTRs** & amelogenin	Promega Corporation	Covers the 13 core CODIS loci plus 2 extra
Profiler Plus & COfiler (2 different kits)	**13 aSTRs** [9 + 6 with 2 overlapping] & amelogenin	Life Technologies (Applied Biosystems)	Original kits used to provide 13 CODIS STRs
Yfiler	**17 Y-chromosome STRs**	Life Technologies (Applied Biosystems)	Male-specific DNA test
MiniFiler	**8 aSTRs** & amelogenin	Life Technologies (Applied Biosystems)	Smaller regions examined; helps with degraded DNA samples
GlobalFiler*	**21 aSTRs**, DYS391, Y indel, & amelogenin	Life Technologies (Applied Biosystems)	Addresses future US core loci
PowerPlex Fusion*	**22 aSTRs**, DYS391, & amelogenin	Promega Corporation	Addresses future US core loci

***Newer kits** that contain improved PCR buffers and DNA polymerases to yield more sensitive results and recover data from difficult samples

Recent Australian STR Kit Change



STR Marker Layouts for New U.S. Kits

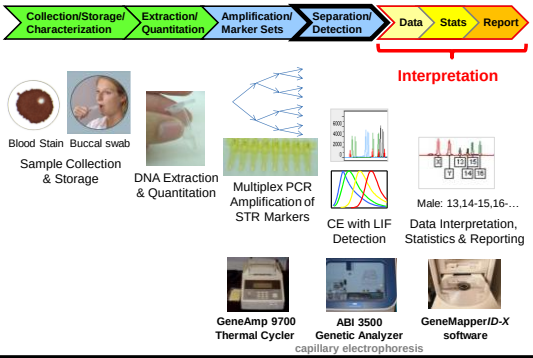


Questions for Workshop Participants

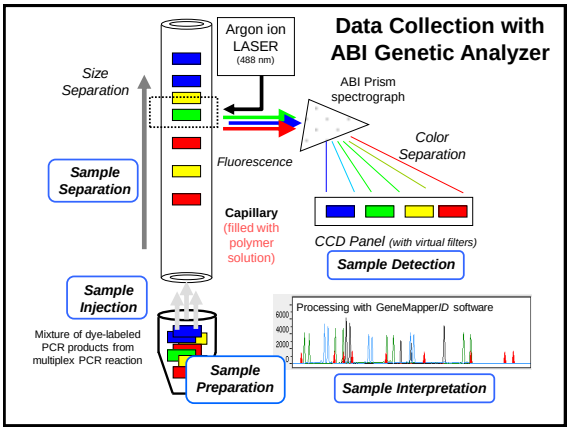
- **STR kits in your lab?**
  - Examples: Identifiler, NGM SElect, PP16, PP21
- **CE instrument(s)?**
  - Examples: ABI 310, ABI 3130xl, ABI 3500
- **Analysis software?**
  - Examples: GeneMapperID, GMID-X, GeneMarkerHID
- **Statistical analysis software?**
  - Examples: PopStats, in-house Excel program, LRmix, …

ABI Genetic Analyzer Data Collection

Steps in Forensic DNA Testing



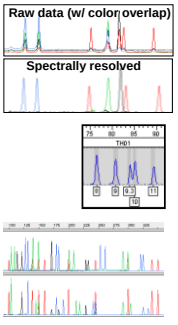
Data Collection with ABI Genetic Analyzer



Analytical Requirements for STR Typing

Butler *et al.* (2004) Electrophoresis 25: 1397-1412

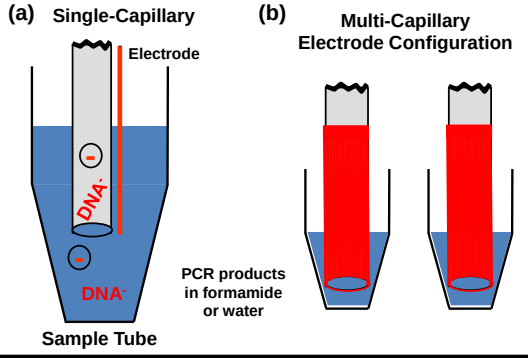
- Fluorescent dyes must be **spectrally resolved** in order to distinguish different dye labels on PCR products
- PCR products must be **spatially resolved** – desirable to have single base resolution out to >350 bp in order to distinguish variant alleles
- High **run-to-run precision** – an internal sizing standard is used to calibrate each run in order to compare data over time



Process Involved in Capillary Electrophoresis (ABI Genetic Analyzer) Data Collection

- **Injection**
  - Utilizes electrokinetic injection process (sample diluted in formamide)
  - Impacts sensitivity → peak signal height
- **Separation**
  - Capillary – 50um fused silica, 43 cm length (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

Capillary and Electrode Configurations



Sample Conductivity Impacts Amount Injected

$$[DNA_{inj}] = \frac{Et(\pi r^2)(\mu_{ep} + \mu_{eof})[DNA_{sample}](\lambda_{buffer})}{\lambda_{sample}}$$

$[DNA_{inj}]$ is the amount of sample injected

E is the electric field applied

t is the injection time

r is the radius of the capillary

μ_{ep} is the mobility of the sample molecules

μ_{eof} is the electroosmotic mobility

$[DNA_{sample}]$ is the concentration of DNA in the sample

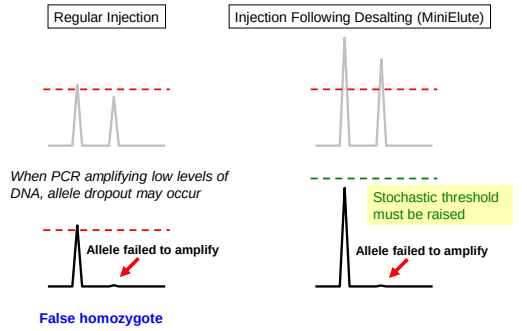
λ_{buffer} is the buffer conductivity

λ_{sample} is the sample conductivity

Cl- ions and other buffer ions present in PCR reaction contribute to the sample conductivity and thus will compete with DNA for injection onto the capillary

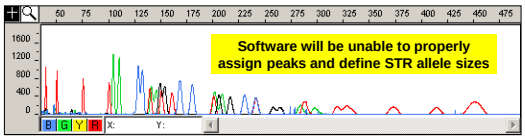
Butler *et al.* (2004) Electrophoresis 25: 1397-1412

Stochastic Effects and Thresholds

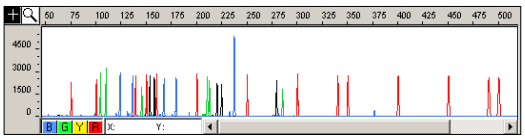


Important of Quality DNA Separations

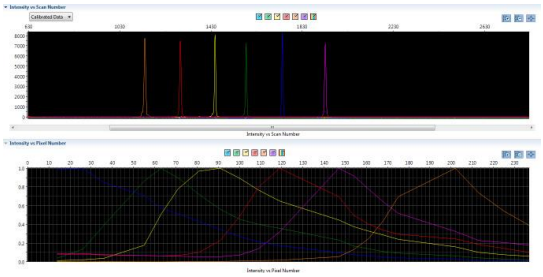
Poor resolution (wide peaks) impacts DNA typing



Sharp peaks are necessary to enable reliable typing

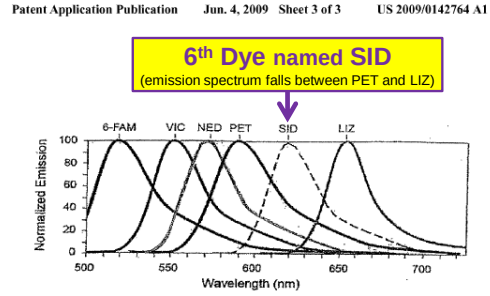


More Dye Channels Enables More Data to Be Collected in a Single Analysis



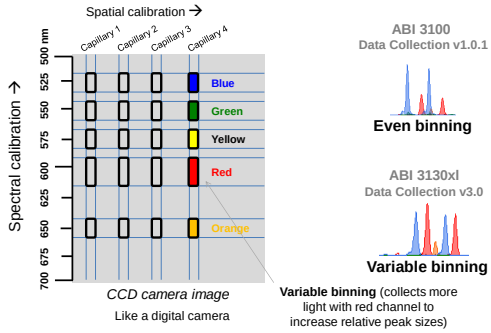
6-dye spectral (from GlobalFiler manual)

Applied Biosystems' Sixth Dye

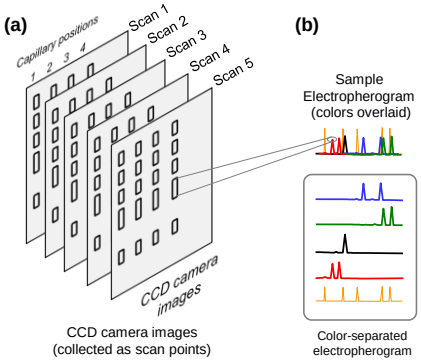


US Patent Application (filed October 30, 2008 by Lori Hennessy and Robert Green)
"Methods and Kits for Multiplex Amplification of Short Tandem Repeat Loci"

Depiction of CCD Camera Image from a Multi-Capillary Genetic Analyzer

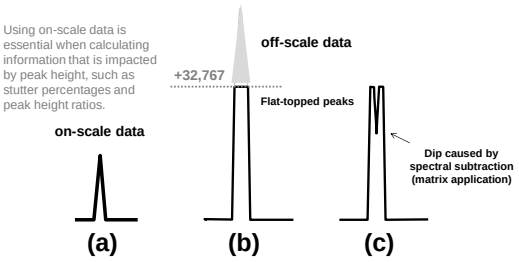


ABI Genetic Analyzer Data Collection



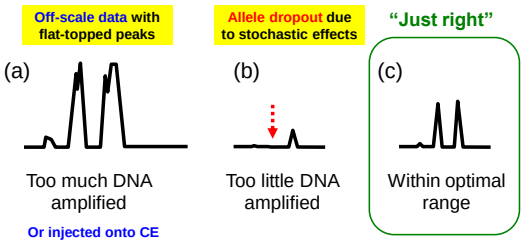
Upper Limit Exists for Data Storage

Using on-scale data is essential when calculating information that is impacted by peak height, such as stutter percentages and peak height ratios.



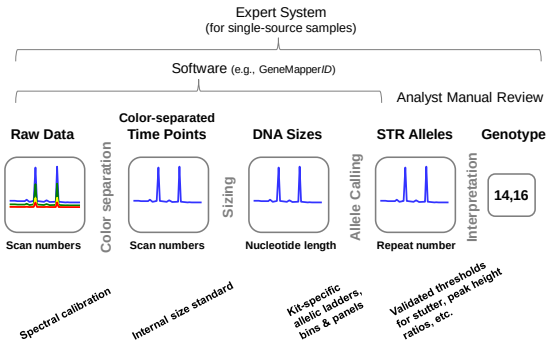
The CCD detector and computer software in the ABI Genetic Analyzers use a 2 byte system for data storage, which enables fluorescence values to be encoded with 16 bits (each bit holding a value of 0 or 1). This 2 byte storage format enables the data collection software to provide a range of 0 to 65,535 in decimal (base-10) digits, or 0000000000000000 to 1111111111111111 in binary (base-2) digits. If both positive and negative values are permitted, then the maxima of a 2 byte storage system are +32,767 and -32,767.

STR Typing Works Best in a Narrow Window of DNA Template Amounts

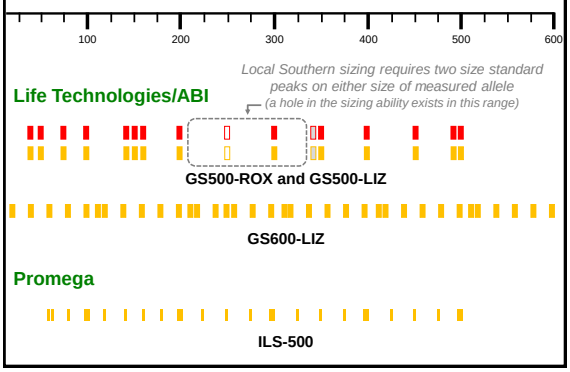


Typically best results are seen in the 0.5 ng to 1.5 ng range for most STR kits

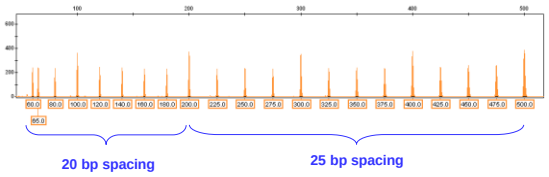
Transformation of Information at a Single STR Locus during Data Processing



Different Internal Size Standards



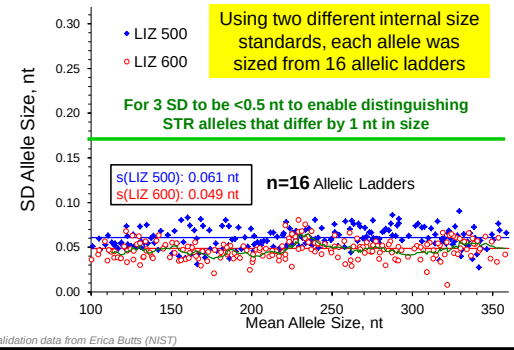
New Promega Size Standard



Labeled with 5-dye (orange)
Contains 21 fragments with even spacing
Low end: 60 & 65 bp
High end: 475 & 500 bp

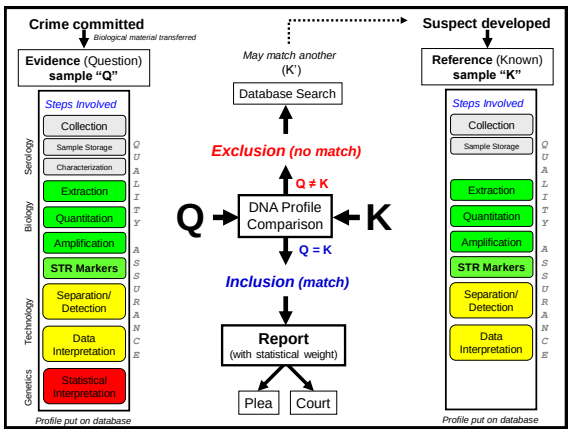
Local Southern sizing possible from 66 bp to 474 bp

Precision Data from ABI 3500

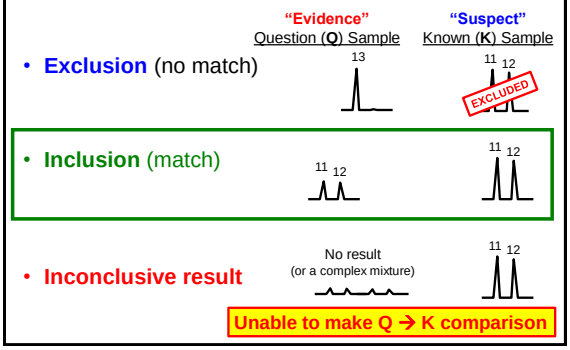


Validation data from Erica Butts (NIST)

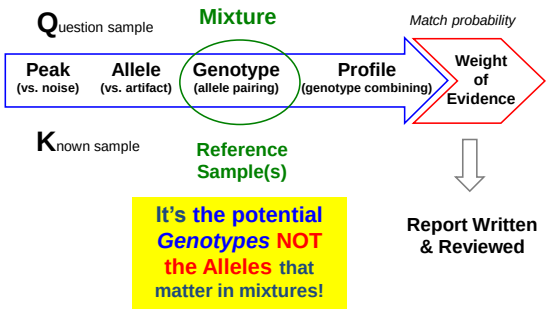
Data Interpretation Overview



Three Possible Outcomes of Evidence Examination



Steps in DNA Interpretation



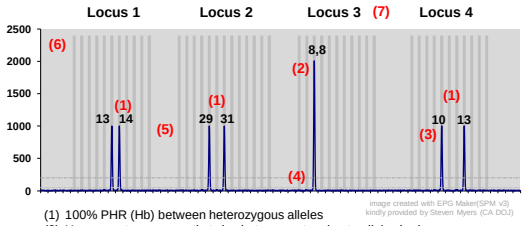
Overview of the SWGDAM 2010 Interp Guidelines

http://www.swgdam.org/Interpretation_Guidelines_January_2010.pdf

1. Preliminary evaluation of data – **is something a peak and is the analysis method working properly?**
2. Allele designation – **calling peaks as alleles**
3. Interpretation of DNA typing results – **using the allele information to make a determination about the sample**
   1. Non-allelic peaks
   2. Application of peak height thresholds to allelic peaks
   3. Peak height ratio
   4. Number of contributors to a DNA profile
   5. Interpretation of DNA typing results for mixed samples
   6. Comparison of DNA typing results
4. Statistical analysis of DNA typing results – **assessing the meaning (rarity) of a match**

Other supportive material: statistical formulae, references, and glossary

Using **Ideal Data** to Discuss Principles



(1) 100% PHR (Hb) between heterozygous alleles
(2) Homozygotes are exactly twice heterozygotes due to allele sharing
(3) No peak height differences exist due to size spread in alleles (any combination of resolvable alleles produces 100% PHR)
(4) No stutter artifacts enabling mixture detection at low contributor amounts
(5) Perfect inter-locus balance
(6) Completely repeatable peak heights from injection to injection on the same or other CE instruments in the lab or other labs
(7) Genetic markers that are so polymorphic all profiles are fully heterozygous with distinguishable alleles enabling better mixture detection and interpretation

Challenges in Real-World Data

- **Stochastic (random) variation** in sampling each allele during the PCR amplification process
  - This is highly affected by DNA quantity and quality
  - Imbalance in allele sampling gets worse with low amounts of DNA template and higher numbers of contributors
- **Degraded DNA** template may make some allele targets unavailable
- **PCR inhibitors** present in the sample may reduce PCR amplification efficiency for some alleles and/or loci
- **Overlap of alleles** from contributors **in DNA mixtures**
  - Stutter products can mask true alleles from a minor contributor
  - Allele stacking may not be fully proportional to contributor contribution

D.N.A. Approach to Understanding

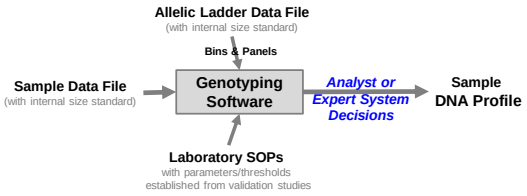
- **D**octrine or Dogma (why?)
  - A fundamental law of genetics, physics, or chemistry
    - Offspring receive one allele from each parent
    - Stochastic variation leads to uneven selection of alleles during PCR amplification from low amounts of DNA templates
    - Signal from fluorescent dyes is based on …
- **N**otable Principles (what?)
  - The amount of signal from heterozygous alleles in single-source samples should be similar
- **A**pplications (how?)
  - Peak height ratio measurements can associate alleles into possible genotypes

Greg Matheson on Forensic Science Philosophy

The CAC News – 2nd Quarter 2012 – p. 6
"Generalist vs. Specialist: a Philosophical Approach"
http://www.cacnews.org/news/2ndq12.pdf

- If you want to be a technician, performing tests on requests, then just focus on the policies and procedures of your laboratory. If you want to be a scientist and a professional, learn the policies and procedures, but go much further and learn the philosophy of your profession. **Understand the importance of why things are done** the way they are done, the scientific method, the viewpoint of the critiques, the issues of bias and the importance of ethics.

Overview of Data Interpretation Process



Decision during Data Interpretation

Input Information	Decision to be made	How decision is made
Data file	Peak or Noise	Analytical threshold
Peak	Allele or Artifact	Stutter threshold; precision sizing bin
Allele	Heterozygote or Homozygote or Allele(s) missing	Peak heights and peak height ratios; stochastic threshold
Genotype/ full profile	Single-source or Mixture	Numbers of peaks per locus
Mixture	Deconvolution or not	Major/minor mixture ratio
Low level DNA	Interpret or not	Complexity/uncertainty threshold
Poor quality data	Replace CE components (buffer, polymer, array) or call service engineer	Review size standard data quality with understanding of CE principles

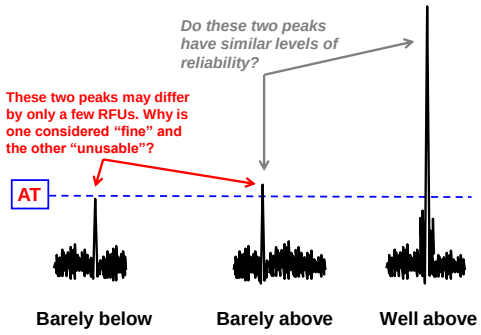
Analytical Threshold

Information on setting analytical thresholds:

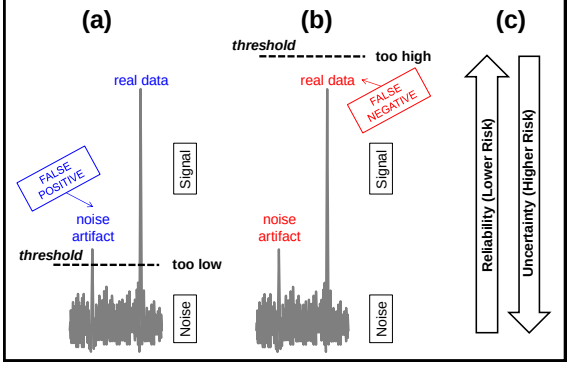
Rakay, C.A., et al. (2012). Maximizing allele detection: Effects of analytical threshold and DNA levels on rates of allele and locus drop-out. *Forensic Science International: Genetics, 6*, 723-728.

Bregu, J., et al. (2013). Analytical thresholds and sensitivity: establishing RFU thresholds for forensic DNA analysis. *Journal of Forensic Sciences, 58*, 120-129.

Analytical Threshold (AT)



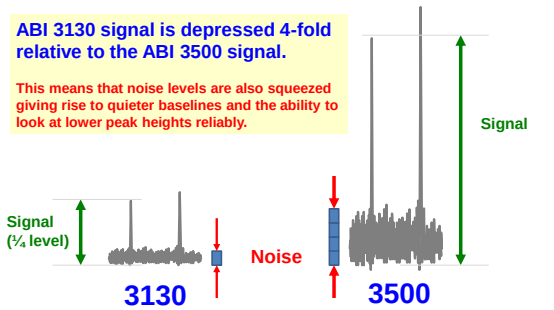
Setting Analytical (Detection) Threshold



Signal-to-Noise Ratio

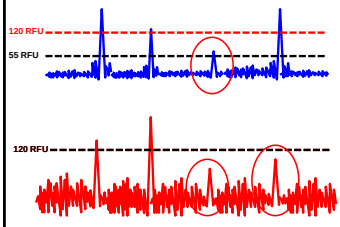
ABI 3130 signal is depressed 4-fold relative to the ABI 3500 signal.

This means that noise levels are also squeezed giving rise to quieter baselines and the ability to look at lower peak heights reliably.



On a relative basis in terms of the signal-to-noise ratio, these data are equivalent

Need for Dye-Specific Thresholds?



Single thresholds for all dye channels assumes all dye channels have the **same** amount of noise

Can cause data to fall below the analytical threshold and not be called

Dye-specific thresholds take into consideration that all dye channels do not have the same level of noise

Can increase the amount of data that is callable

Stutter Products

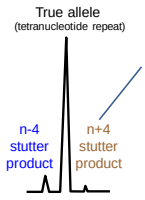
Stutter Products

- Peaks that show up primarily one repeat less than the true allele as a result of strand slippage during DNA synthesis

D18S51

Stutter Products

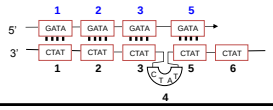
Typically **5-15%** of true allele in tetranucleotide repeats STR loci



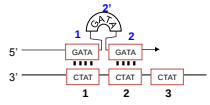
Occurs less frequently (typically <2%)

Walsh, P.S., et al. (1996). Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Research, 24*, 2807-2812.

Deletion caused by slippage on the copied (bottom) strand



Insertion caused by slippage of the copying (top) strand

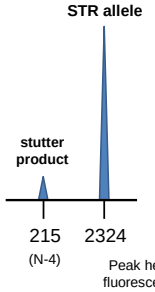


Interpretation of DNA Typing Results

SWGDAM Guideline 3.1.1.1.

In general, the empirical criteria are based on qualitative and/or quantitative characteristics of peaks. As an example, dye artifacts and spikes may be distinguished from allelic peaks based on morphology and/or reproducibility. ***Stutter and non-template dependent nucleotide addition peaks may be characterized based on size relative to an allelic peak and amplitude.***

Calculating Stutter



Peak heights (in relative fluorescence units, RFUs)

$$\text{Stutter \%} = \frac{\text{N-4 peak}}{\text{allele peak}}$$

$$= \frac{215}{2324}$$

$$= \mathbf{9.25\%}$$

STR_StutterFreq!

Welcome to STR_StutterFreq!
Version <04-Jan-10>

STR_StutterFreq is a specialty analysis tool for characterizing stutter frequency…
Development of STR_StutterFreq was funded in part by the National Institute of Justice.

- Program developed by Dave Duewer (NIST) to rapidly calculate stutter frequencies.

Available for free-download on NIST STRBase website:
http://www.cstl.nist.gov/strbase/tools/STR_AlleleFreq.xls

TPOX – $[AATG]_N$

Locus	Allele	Size	Stutter #	Stutter Median	Stutter MADe
TPOX	8	265.2	86	2.1	0.5
	9	269.2	21	2.9	0.4
	11	277.2	75	3.6	0.4
	12	281.2	14	4.3	0.4
		Avg	196	3.3	0.4
		SD		0.9	

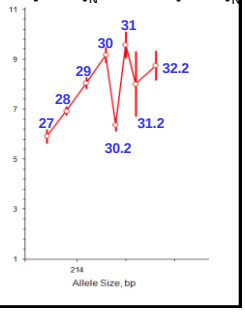
Average $Stutter_{Locus}$ + 3SD = 6%

6% >> 2%

D21S11 – a complex repeat

$[TCTA]_N$ $[TCTG]_N$ $[TCTA]_N$ TA $[TCTA]_N$ TCA $[TCTA]_N$ TCCATA $[TCTA]_N$

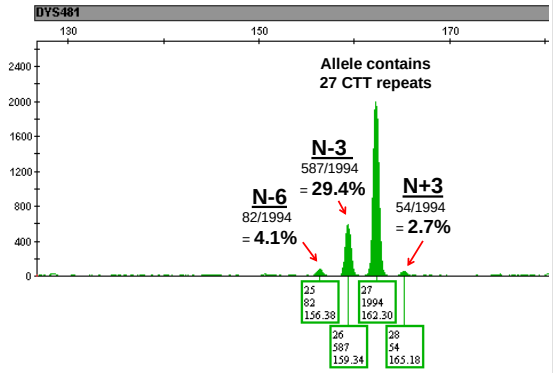
Locus	Allele	Size	Stutter #	Stutter Median	Stutter MADe
D21S11	27	207.8	20	5.9	0.6
	28	211.8	69	6.9	0.7
	29	215.8	59	8.0	0.8
	30	219.9	66	9.2	1.2
→	30.2	221.9	11	6.4	0.5
	31	223.9	21	9.6	1.2
	31.2	226.0	28	8.0	3.4
	32.2	230.0	33	8.7	1.7
		Avg	307	7.8	1.6
		SD		1.3	



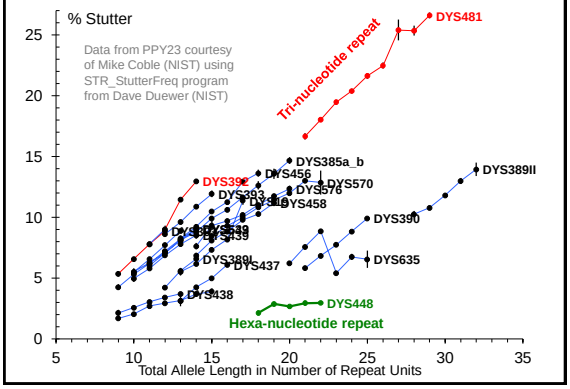
Average $Stutter_{Locus}$ + 3SD = 12%

12% >> 7%

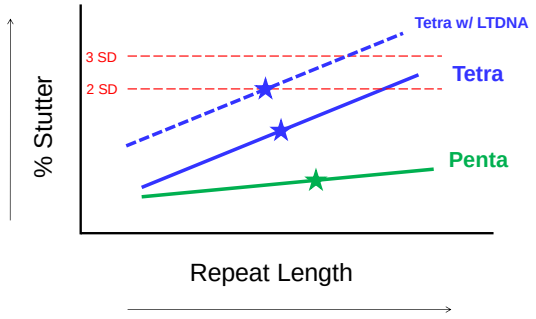
Stutter is Higher with a Tri-Nucleotide Repeat (DYS481)



Observed Stutter Trends



Simplified Illustration of Stutter Trends

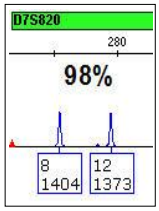


Assessing **Peak Height Ratios** to associate potential allele pairs into locus genotypes

Peak Height Ratio
What is PHR?

At each locus:

Peak A RFU (lower RFU peak) X 100 = PHR%
Peak B RFU (higher RFU peak)



1373 RFU X 100 = 98%
1404 RFU

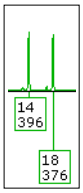
PHR also called **Hb**
(Heterozygote balance)

Slide from Charlotte Word (ISHI 2010 mixture workshop)

What is PHR?

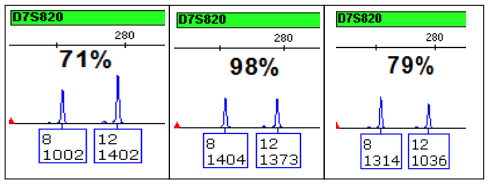
Theory: The two alleles for an individual who is heterozygous at a single locus should:

- **Have equal amounts in the genome**
- **Amplify equally**
- **Inject equally**
- **Have peak heights that are ~equal**
- **Value must be ≤100%**



Slide from Charlotte Word (ISHI 2010 mixture workshop)

Natural Variation in Peak Height Ratio During Replicate PCR Amplifications

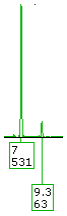


The heights of the peaks will vary from sample-to-sample, even for the same DNA sample amplified in parallel

Slide from Charlotte Word (ISHI 2010 mixture workshop)

Causes of Peak Height Imbalance

1. **Single-source samples**
   a. **Low Template DNA (LT DNA)**
   b. **Inhibited**
   c. **Degraded**
   d. **Preferential amplification**



2. **Mixture of DNA from 2 or more contributors is present**

Slide from Charlotte Word (ISHI 2010 mixture workshop)

How calculate Peak Height Ratios?

From Validation Studies

- **Sensitivity Study at different amounts of DNA**
- **Non-probative single-source samples with good quality profiles amplified with different amounts of DNA (or at least with different peak height ranges)**
- **Perform for each kit validated as PHRs may vary for the same locus amplified with different kits**

Slide from Charlotte Word (ISHI 2010 mixture workshop)

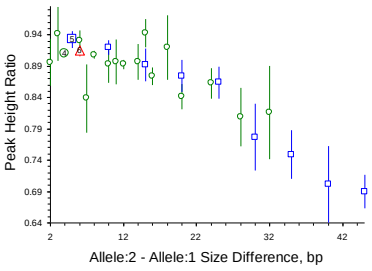
Peak Height Ratio **Data**

Locus	Δbp	#	Mean X	Mean s(X)	Median X	Median s(X)	Percentiles Min	Percentiles Max
D13S317	4	103	0.913	0.082	0.930	0.079	0.637	1.000
	8	49	0.879	0.083	0.900	0.091	0.652	0.998
	12	24	0.867	0.079	0.874	0.084	0.639	0.979
	16	20	0.855	0.080	0.847	0.070	0.696	0.997
	20	11	0.828	0.069	0.822	0.067	0.742	0.959
D18S51	4	63	0.878	0.097	0.900	0.100	0.554	0.998
	8	49	0.894	0.100	0.905	0.112	0.704	0.998
	12	44	0.866	0.104	0.876	0.116	0.583	0.997
	16	27	0.872	0.107	0.895	0.119	0.574	0.995
	20	22	0.807	0.100	0.796	0.112	0.644	0.963
	28	10	0.795	0.115	0.785	0.138	0.641	0.936
D8S1179	4	105	0.884	0.082	0.886	0.079	0.683	0.997
	8	61	0.895	0.090	0.908	0.085	0.714	0.990
	12	26	0.857	0.105	0.898	0.099	0.485	1.000
	16	14	0.886	0.088	0.891	0.094	0.620	0.999

PowerPlex 16 data from NIST

Slide from Charlotte Word (ISHI 2010 mixture workshop)

Peak Height Ratio **Data**

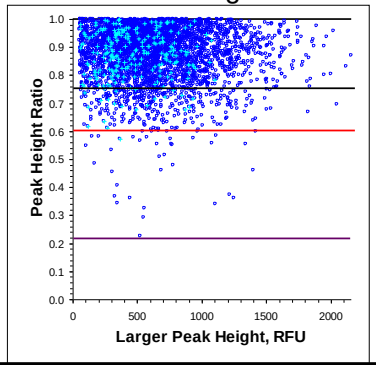


PowerPlex 16 data from NIST

>8000 alleles

Slide from Charlotte Word (ISHI 2010 mixture workshop)

Peak Height Ratio **Data**



Most peaks >1500 RFU

Most peaks

All peaks

Power Plex 16 data from NIST

>8000 alleles

Slide from Charlotte Word (ISHI 2010 mixture workshop)

Data Interpretation Overview

The Steps of Data Interpretation

Peak (vs. noise) → **Allele** (vs. artifact) → **Genotype** (allele pairing) → **Profile** (genotype combining)

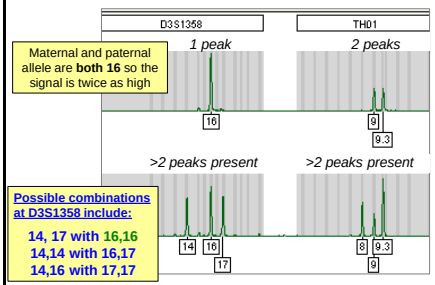
Analytical Threshold	Expected Stutter %	Stochastic Threshold	Peak Height Ratio (PHR)
	True allele; Stutter product	Allele 1; Dropout of Allele 2	Allele 1; Allele 2; genotype

Next step: Examine feasible genotypes to deduce possible contributor profiles

Moving from individual locus genotypes to profiles of potential contributors to the mixture is dependent on mixture ratios and numbers of contributors

Assessing **Number of Contributors** in a DNA Sample

Single-Source Sample vs Mixture Results

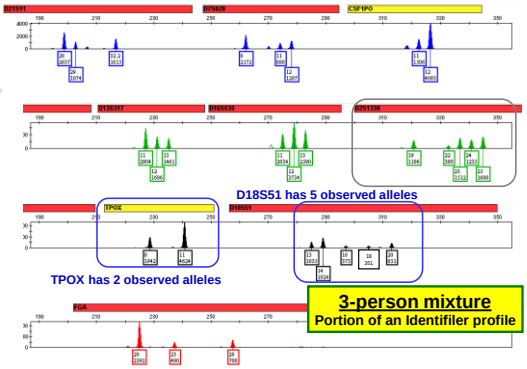


Single-Source

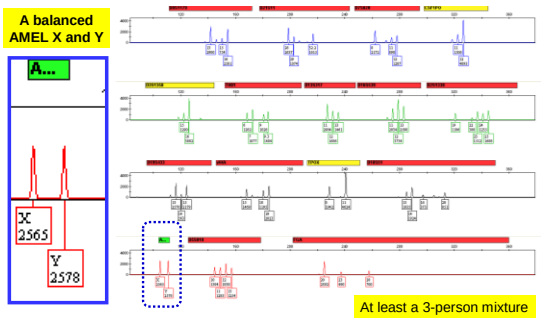
Mixture

Multiple possible combinations could have given rise to the mixture observed here

More Polymorphic Loci Tend to Exhibit a Greater Number of Alleles in Mixtures

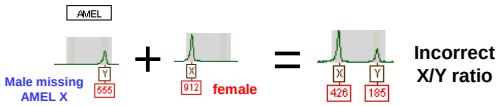


Amelogenin X/Y ratio can be a poor determinant of contributor number



Potential Problems with Amelogenin

- Works best with 2-person male/female mixtures (such as sexual assault cases)
  - Male/male mixture or multiple males with single female component limit usefulness
- Molecular reasons for alteration of expected ratio
  - Deletion of AMEL Y (or primer site mutation)
  - **Deletion of AMEL X** (or primer site mutation)

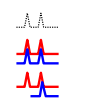


Possible genotype combinations in 2-person mixtures

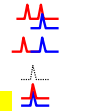
See Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, pp. 156-157

Four Peaks **(4 allele loci)**
- heterozygote + heterozygote, no overlapping alleles (genotypes are unique)

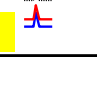
Three Peaks **(3 allele loci)**
- heterozygote + heterozygote, one overlapping allele
- heterozygote + homozygote, no overlapping alleles (genotypes are unique)



Two Peaks **(2 allele loci)**
- heterozygote + heterozygote, two overlapping alleles (genotypes are identical)
- heterozygote + homozygote, one overlapping allele
- homozygote + homozygote, no overlapping alleles (genotypes are unique)



Single Peak **(1 allele loci)**
- homozygote + homozygote, overlapping allele (genotypes are identical)



Must also consider the stutter position when the mixture ratio is large enough for the minor component(s) to be in PHR with stutter peaks

All possible 2-contributor combinations

14 total combinations

I	4 alleles		
	AB	CD	1
	AC	BD	2
	BC	AD	3

II	3 alleles		
	AA	BC	4
	BB	AC	5
	CC	AB	6

III			
	AB	AC	7
	AC	BC	8
	BC	AB	9

IV	2 alleles		
	AA	BB	10

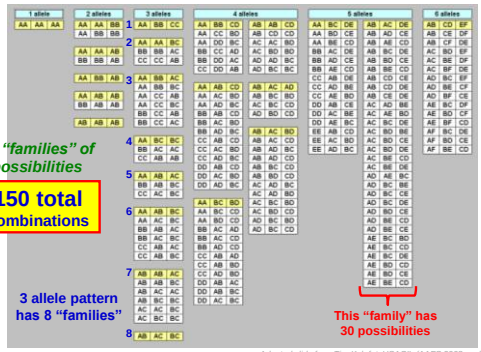
V			
	AA	AB	11
	BB	AB	12

VI			
	AB	AB	13

VII	1 allele		
	AA	AA	14

7 "families" (groups or classes) of possibilities

All possible 3-contributor combinations



Comparison of Expected and Simulated Mixture Results

Expected Results when estimating # of contributors:

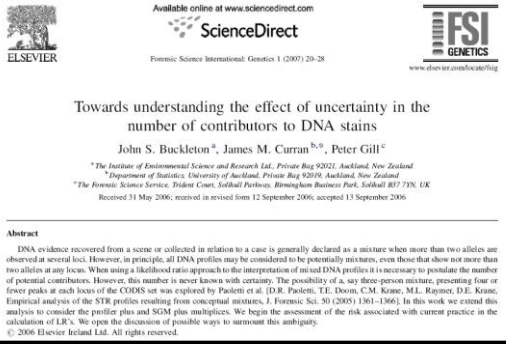
- If 2, 3, or 4 alleles are observed at every locus across a profile then 2 contributors are likely present
- If a maximum of 5 or 6 alleles at any locus, then 3 contributors are possible
- If >6 alleles in a single locus, then >3 contributors

Results from Simulation Studies:

- Buckleton *et al.* (2007) found with a simulation of four person mixtures that 0.02% would show four or fewer alleles and that 76.35% would show six or fewer alleles for the CODIS 13 STR loci.

Buckleton et al. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *FSI Genetics* 1:20-28

Simulation Study Regarding Detecting the Number of Contributors to a Mixture



ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 1 (2007) 20–28

FSI GENETICS
www.elsevier.com/locate/fsig

Towards understanding the effect of uncertainty in the number of contributors to DNA stains

John S. Buckleton [a], James M. Curran [b,*], Peter Gill [c]

[a] The Institute of Environmental Science and Research Ltd., Private Bag 92021, Auckland, New Zealand
[b] Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand
[c] The Forensic Science Service, Trident Court, Solihull Parkway, Birmingham Business Park, Solihull B37 7YN, UK

Received 31 May 2006; received in revised form 12 September 2006; accepted 13 September 2006

Abstract

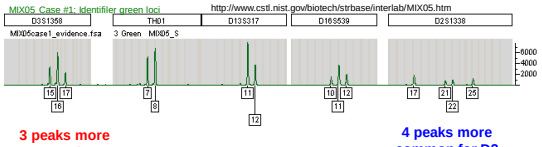
DNA evidence recovered from a scene or collected in relation to a case is generally declared as a mixture when more than two alleles are observed at several loci. However, in principle, all DNA profiles may be considered to be potentially mixtures, even those that show not more than two alleles at any locus. When using a likelihood ratio approach to the interpretation of mixed DNA profiles it is necessary to postulate the number of potential contributors. However, this number is never known with certainty. The possibility of a, say three-person mixture, presenting four or fewer peaks at each locus of the CODIS set was explored by Paoletti et al. [D.R. Paoletti, T.E. Doom, C.M. Krane, M.L. Raymer, D.E. Krane, Empirical analysis of the STR profiles resulting from conceptual mixtures, J. Forensic Sci. 50 (2005) 1361–1366]. In this work we extend this analysis to consider the profiler plus and SGM plus multiplices. We begin the assessment of the risk associated with current practice in the calculation of LR's. We open the discussion of possible ways to surmount this ambiguity.

Levels of Locus Heterozygosity Impact the Number of Alleles Observed in Mixtures

Buckleton et al. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *FSI Genetics* 1:20-28

Loci	No. of alleles	Simulated 2-Person Mixture		
	1	2	3	4
D3	0.011	0.240	0.559	0.190
vWA	0.008	0.194	0.548	0.250
D16	0.016	0.287	0.533	0.164
D2	0.003	0.094	0.462	0.441

Results from a 2-Person Mixture



3 peaks more common for D3

4 peaks more common for D2

Results Depend on Assumptions

- "Although courts expect one simple answer, statisticians know that **the result depends on how questions are framed and on assumptions tucked into the analysis**."
  – Mark Buchanan, Conviction by numbers. *Nature* (18 Jan 2007) 445: 254-255
- **We inform our assumptions with data from validation studies…**

Acknowledgments

$ National Institute of Justice and NIST OLES
John Buckleton and co-presenters

Slides and Discussions on DNA Mixtures
Mike Coble (NIST Applied Genetics Group)
Robin Cotton & Catherine Grgicak (Boston U.)
Bruce Heidebrecht (Maryland State Police)
Charlotte Word (consultant)

Contact info:
john.butler@nist.gov
+1-301-975-4049

Final version of this presentation will be available at:
http://www.cstl.nist.gov/strbase/NISTpub.htm