AMERICAN BAR ASSOCIATION
Criminal Justice Section

Fourth Annual
Prescription for Criminal Justice Forensics

New York City, NY
June 7, 2013

FORDHAM UNIVERSITY
THE JESUIT UNIVERSITY OF NEW YORK

DNA Mixture Interpretation:
History, Background, Thresholds, Statistical Methods, and SWGDAM

John M. Butler, Ph.D.
National Institute of Standards and Technology
Gaithersburg, Maryland

NIST

Disclaimers

Funding for research and training on forensic DNA performed by the NIST Applied Genetics Group has come from the National Institute of Justice and the NIST Law Enforcement Standards Office

Although I chaired the SWGDAM Mixture Committee that produced the 2010 STR Interpretation Guidelines**, I cannot speak for or on behalf of the Scientific Working Group on DNA Analysis Methods**

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.

Steps in Forensic DNA Testing

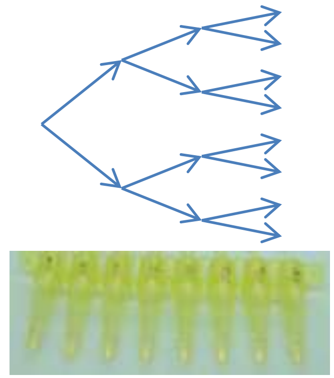
Collection/Storage/ Characterization → **Extraction/ Quantitation** → **Amplification/ Marker Sets** → **Separation/ Detection** → **Interpretation** → **Report**

Mixture interpretation

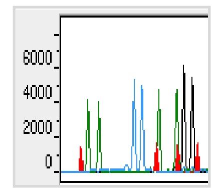
Blood Stain Buccal swab

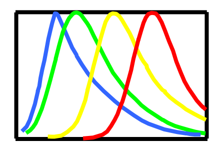
Sample Collection
& Storage

DNA Extraction
& Quantitation

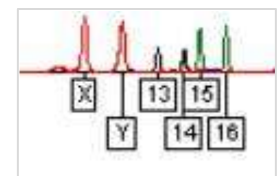


Multiplex PCR
Amplification of
STR Markers





CE with LIF
Detection



Male: 13,14-15,16-…

Data Interpretation ,
Review & Reporting

**GeneAmp 9700
Thermal Cycler**

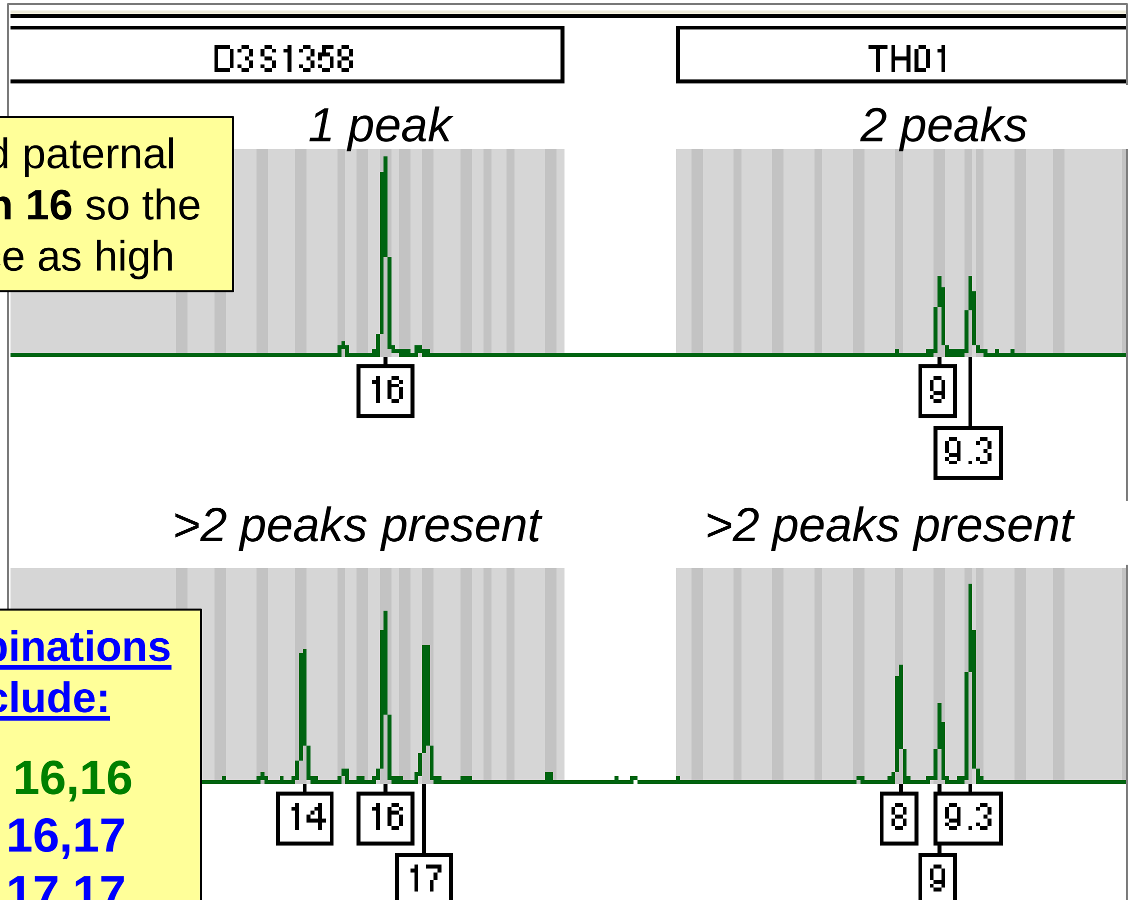
**ABI 3500
Genetic Analyzer**

capillary electrophoresis

**GeneMapper*ID-X*
software**

Single-Source Sample vs Mixture Results

Maternal and paternal allele are **both 16** so the signal is twice as high



Single-Source

Mixture

Possible combinations at D3S1358 include:

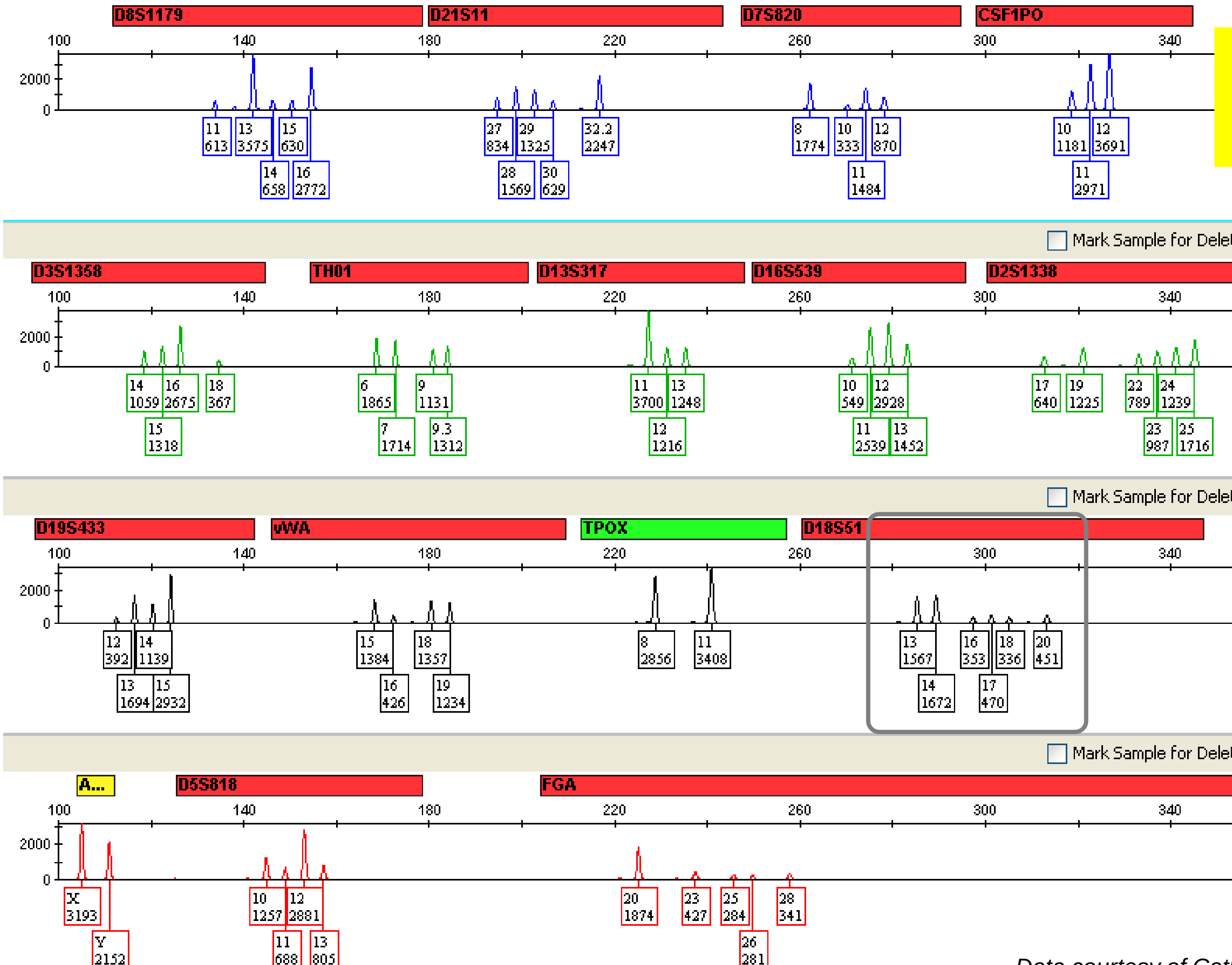
14, 17 with 16,16
14,14 with 16,17
14,16 with 17,17

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

Identifiler DNA test

Controlled mixture of **4 individuals**

DNA Mixture Result



More than two peaks per locus (DNA test site)

Data courtesy of Catherine Grgicak (Boston U.)

A Brief History of DNA Mixtures (1)

- **1995** – Mixtures presented in OJ Simpson trial
- **1996** – 9plex STR kits (Profiler Plus, PowerPlex 1.1)
- **1997** – Weir et al using Likelihood Ratios (LRs) for mixture statistics
- **1998** – Clayton et al (FSS) DNA mixture deconvolution
- **2000** – initial SWGDAM Interpretation Guidelines published
- **2000** – Combined Probability of Inclusion (CPI) statistic is allowed by DNA Advisory Board and pushed by the FBI
- **2000** – 16plex STR kits (PP16 and Identifiler)
- **2005** – NIST Interlaboratory Mixture Study **(MIX05) finds extensive variation in laboratory approaches**

A Brief History of DNA Mixtures (2)

- **2006** – ISFG Mixture Recommendations published emphasizing that LRs are a better method over CPI
- **2007** – informal SWGDAM study finds most labs doing 2-person mixtures (committee begins writing guidelines)
- **2008** – NIJ study shows value of DNA in burglary cases and more touch DNA samples with complex mixtures begin being processed
- **2010** – SWGDAM Interpretation Guidelines emphasize need for statistics and stochastic thresholds with CPI; probabilistic genotyping approach is mentioned
- **2012** – ISFG publishes LR with probability of dropout to cope with potential of allele dropout
- ***Present*** – a number of software programs exist to help with calculations but no universal approach exists

Statistical Approaches with Mixtures

See Ladd *et al.* (2001) *Croat Med J.* 42:244-246; SWGDAM (2010) section 5

1. **Random Match Probability (after inferring genotypes of contributors)** – Separate major and minor components into individual profiles and compute the random match probability estimate as if a component was from a single source

2. **Combined Probability of Exclusion/Inclusion** – CPE/CPI (RMNE) – Calculation of the probability that a random (unrelated) person would be excluded/included as a contributor to the observed DNA mixture

RMNE = Random Man Not Excluded (same as CPI)
CPE = Combined Probability of Exclusion (CPE = 1 – CPI)
CPI = Combined Probability of Inclusion (CPI = 1 – CPE)

3. **Likelihood Ratio (LR)** – Compares the probability of observing the mixture data under two alternative hypotheses; in its simplest form LR = 1/RMP

$$LR = \frac{\Pr(E \mid H_1)}{\Pr(E \mid H_2)}$$

DAB Recommendations on Statistics

February 23, 2000

Forensic Sci. Comm. 2(3); available on-line at

http://www.fbi.gov/hq/lab/fsc/backissu/july2000/dnastat.htm

"The DAB finds either one or both PE or LR calculations acceptable and strongly recommends that one or both calculations be carried out whenever feasible and a mixture is indicated"

- Probability of exclusion (PE)
  - Devlin, B. (1993) Forensic inference from genetic markers. *Statistical Methods in Medical Research*, 2, 241–262.
- Likelihood ratios (LR)
  - Evett, I. W. and Weir, B. S. (1998) *Interpreting DNA Evidence*. Sinauer, Sunderland, Massachusetts.

NIST Interlaboratory Studies on Mixtures

- 1997 - Mixed Stain Study 1 (MSS1)
- 1999 – MSS2
- 2001 – MSS3 (five 2-person and one 3-person mixture)

- **2005 – MIX05** (supplied data only with four 2-person mixtures)

- **2013 – study is planned to evaluate current variation in mixture interpretation**

SWGDAM = Scientific Working Group on DNA Analysis Methods
(http://www.swgdam.org/)

SWGDAM Mixture Interpretation Guidelines (2010)

- Provide guidance to labs for interpreting single-source and two-person mixtures
- **NOT** intended for Low Template DNA or >2 person mixtures
- Guidelines – NOT Standards
- Laboratories are not required to follow, but guidelines are STRONGLY RECOMMENDED

- **Require statistics when DNA inclusions are made** (SWGDAM 2010 section 4.1)

Stats Required for Inclusions

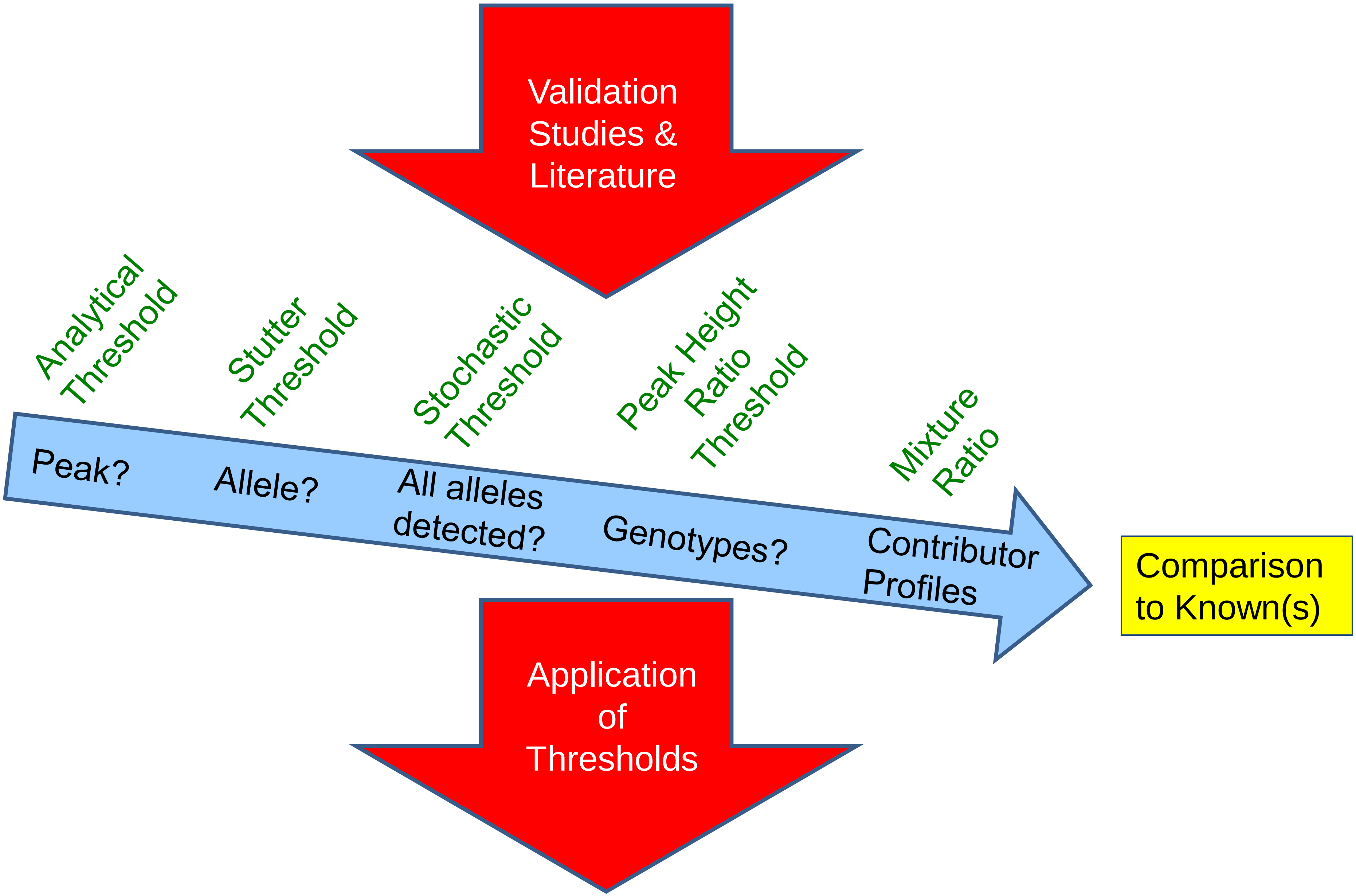
SWGDAM Interpretation Guideline 4.1:

“**The laboratory must perform statistical analysis in support of any inclusion** that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis.”

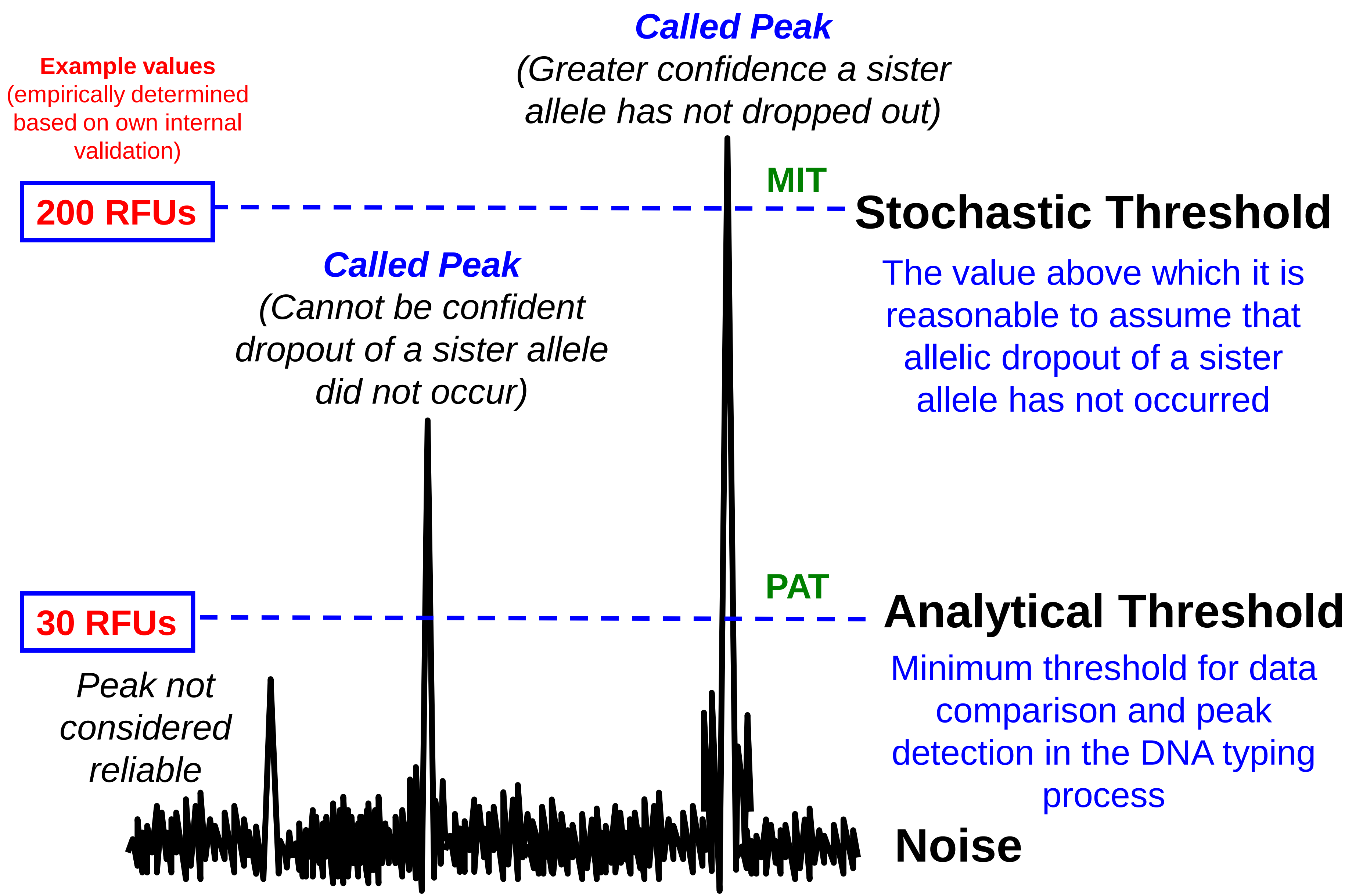
Buckleton & Curran (2008): “There is a considerable aura to DNA evidence. Because of this aura **it is vital that weak evidence is correctly represented as weak or not presented at all**.”

Buckleton, J. and Curran, J. (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Sci. Int. Genet.* 2: 343-348.

Steps in DNA Interpretation



Overview of Two Thresholds

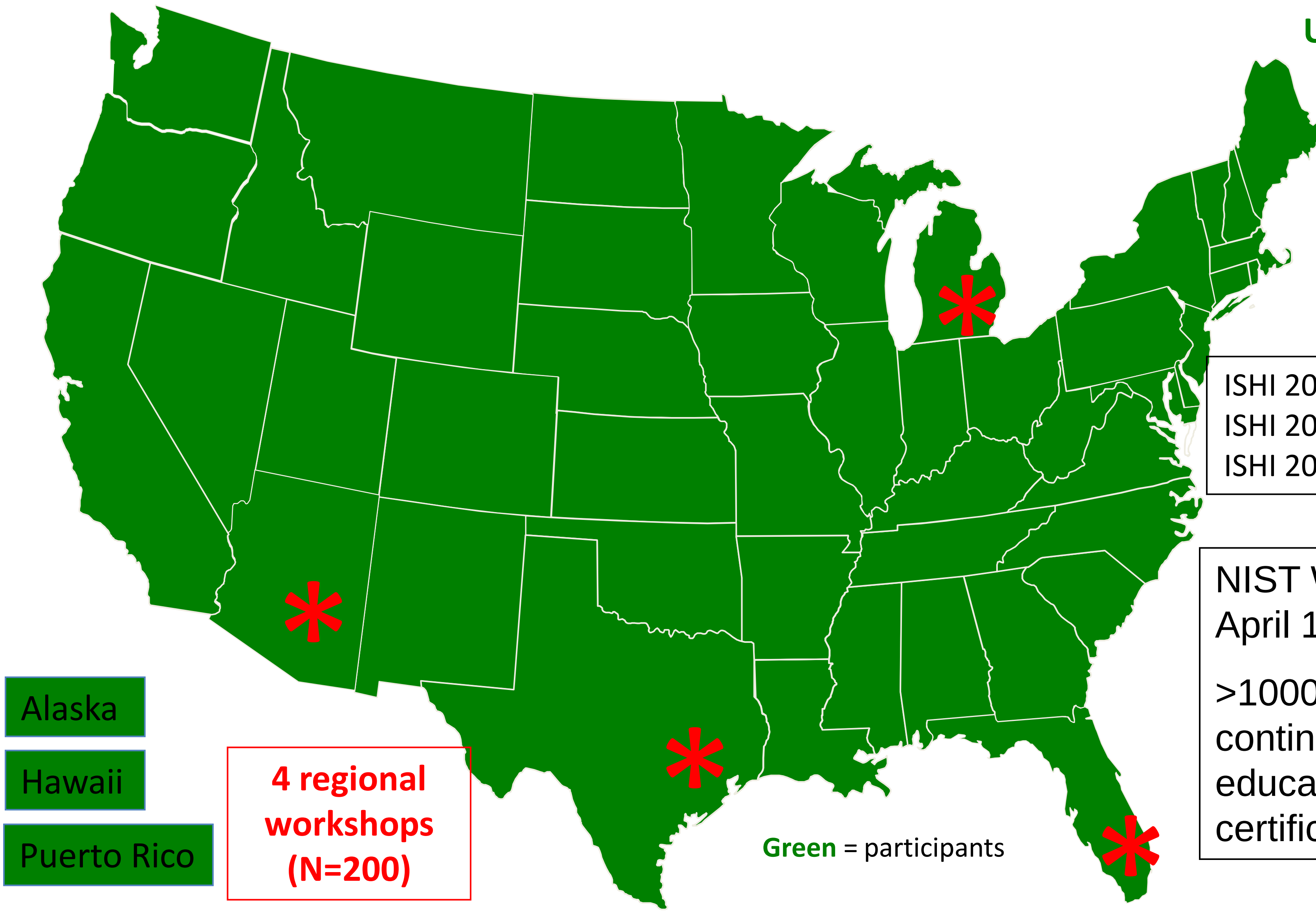


Butler, J.M. (2010) *Fundamentals of Forensic DNA Typing*. Elsevier Academic Press: San Diego.

Mixture Workshop Attendees

50 states and 25 other countries

Federal Labs
FBI
ATF
AFDIL
USACIL



ISHI 2010 (N=200)
ISHI 2011 (N=160)
ISHI 2012 (N=145)

NIST Webinar
April 12, 2013

>1000 continuing education certificates

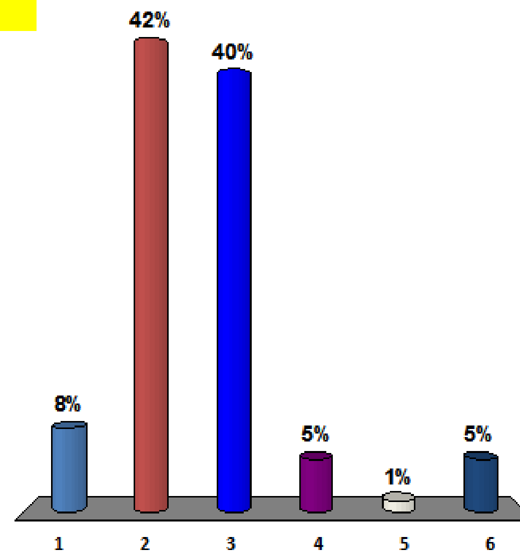
4 regional workshops (N=200)

Real-time interaction with the audience

How many DNA-related articles would you estimate that you read in a typical month?

Data from 2102 ISHI Mixture workshop (Oct 2012

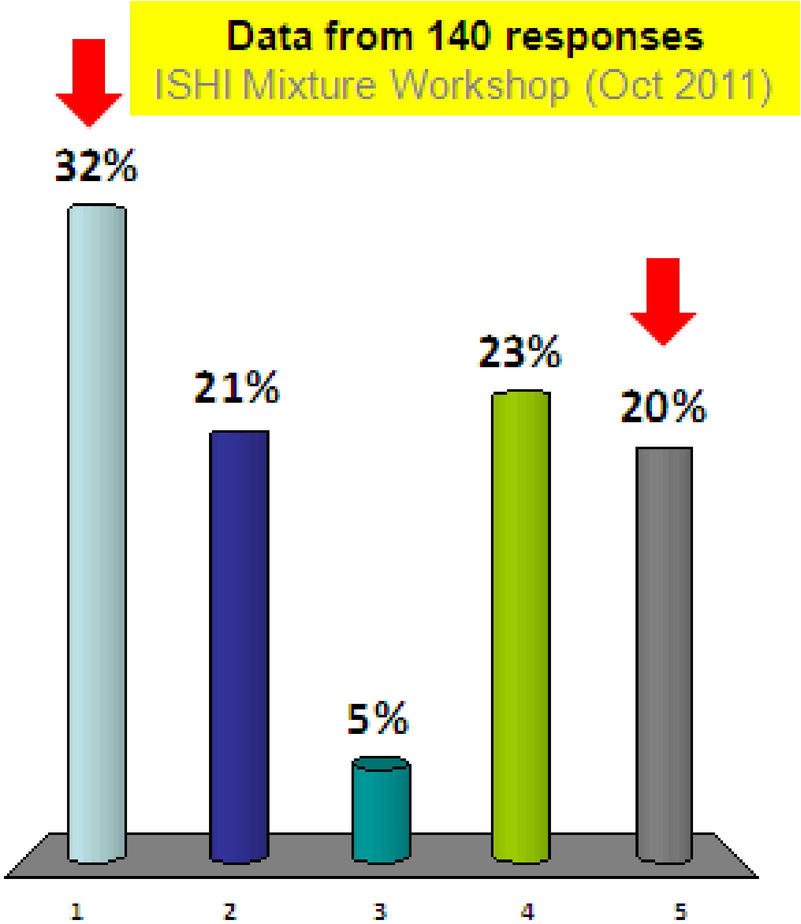
1. None
2. 1 article
3. 2 to 5 articles
4. More than 5 articles
5. None, I only read the abstracts
6. I don't make time to read!



2011 Response from ISHI Workshop

If your laboratory uses a stochastic threshold (ST), it is:

1. Same value as our analytical threshold (**we don’t use a ST**)
2. About twice as high as our AT (e.g., AT = 50 and ST = 100 RFU)
3. Less than twice as high as our AT
4. Greater than twice as high as our AT
5. I don’t know!

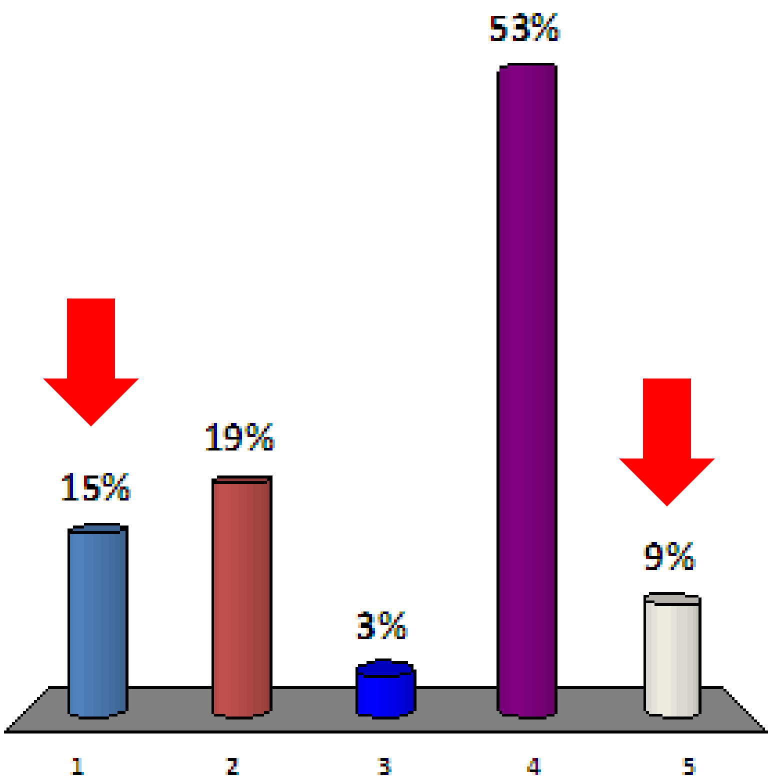


2012 Response from ISHI Workshop

If your laboratory uses a stochastic threshold (ST), it is:

1. Same value as our analytical threshold (**we don't use a ST**)
2. About twice as high as our AT (e.g., AT = 50 and ST = 100 RFU)
3. Less than twice as high as our AT
4. Greater than twice as high as our AT
5. I don't know!

Data from 120 responses
ISHI Mixture Workshop (Oct 2012)



Coupling of Statistics and Interpretation

- **The CPE/CPI approach** for reporting an inclusionary statistic **requires that all alleles be observed** in the evidence sample

- If allele drop-out is suspected at a locus, then any allele is possible and the probability of inclusion goes to 100% -- in other words, the locus is effectively dropped from consideration

- If alleles are seen below the established stochastic threshold, then the locus is typically eliminated (“INC” – declared inconclusive) in many current lab SOPs

Use of CPI is still widespread in U.S.

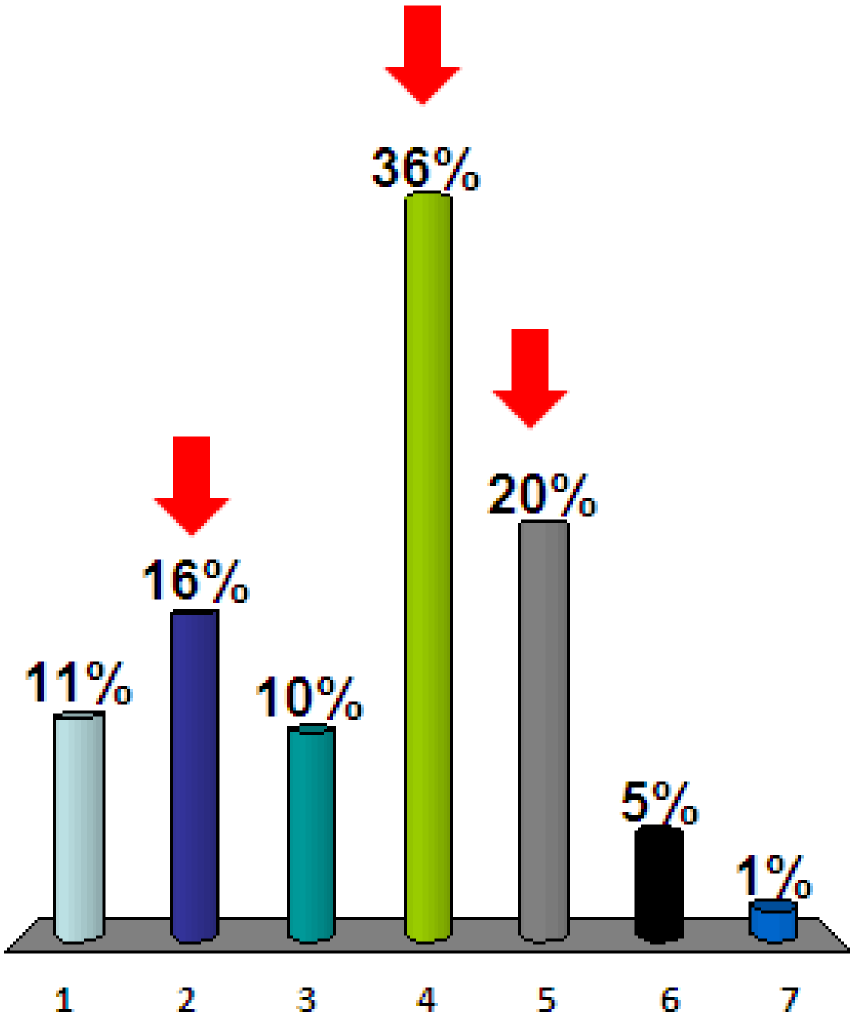
What kind of mixture statistic does your lab use?

2011 Response at Training Workshop

1. LR
2. CPE (RMNE, CPI)
3. RMP
4. CPE or RMP
5. Other combinations
6. Probabilistic modeling (e.g., TrueAllele)
7. We don't use stats (contradicting the new guidelines – section 4.1)

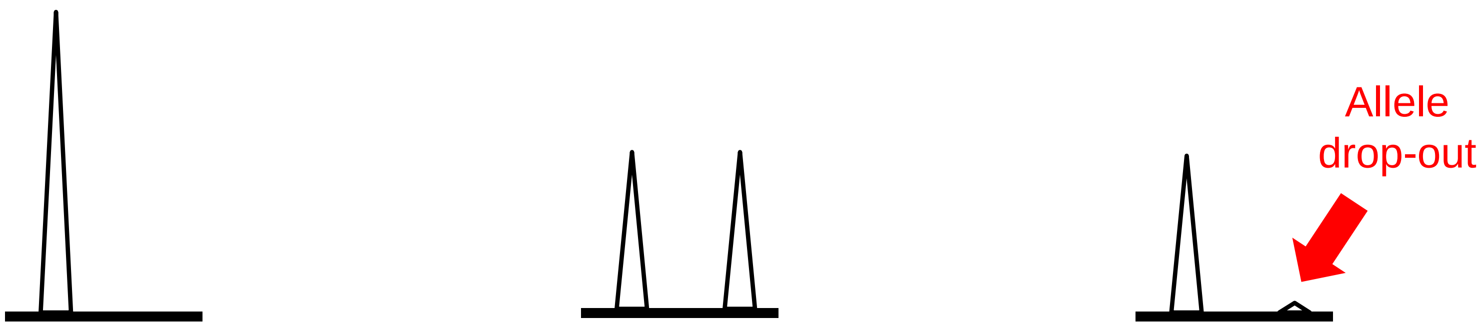
Data from 138 responses
ISHI Mixture Workshop (Oct 2011)

72% using CPI



Allele Drop-out

- If because of chemistry events sometimes associated with low levels of DNA (termed “stochastic effects”), one of the STR alleles “drop-out” and is not detected, then our sample at that locus looks like a homozygote instead of the heterozygote that it really is



True homozygote
(only a single peak)

True heterozygote
(both peaks detected)

False homozygote
(one peak has “dropped out” and fails to be detected)

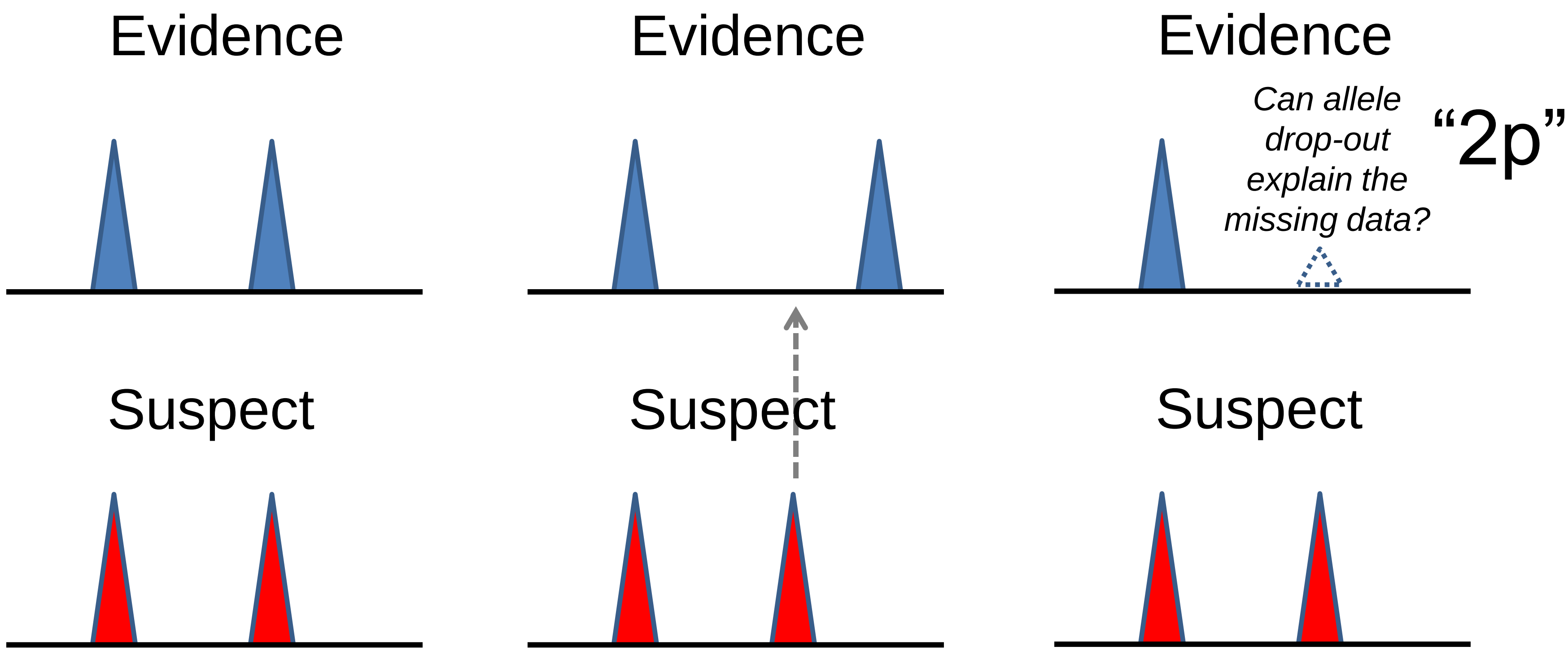
Statistical treatment

p^2

$2pq$

$2p$

Likelihood Ratios for Different Possibilities



$$LR = \frac{1}{2pq}$$

$$LR = \frac{0}{2pq}$$

$$LR = \frac{?}{2pq}$$

Binary LR approach (either 0 or 1)

Modified slide from Mike Coble (NIST)

New Statistical Tools/Software for Mixtures

- **Lab Retriever** (David Balding → Norah Rudin et al.)
  - Uses likelihood ratios (LRs) and probability of dropout [Pr(D) or P(Do)]
- **FST** – Forensic Statistical Tool (NYC OCME)
  - Uses LRs and empirically determined Pr(D) based on DNA quantity
- **Armed Xpert** (USACIL → Niche Vision)
  - Originally developed by US Army Crime Lab (USACIL)
  - Performs calculations typically manually done by analysts
- **TrueAllele** (Mark Perlin/Cybergenetics)
  - Uses probabilistic genotyping approach with LRs
- **STRmix** (John Buckleton/New Zealand ESR)
  - Like TrueAllele, uses LRs with computer simulations

New Efforts to Improve DNA Interpretation
(especially low-level DNA and mixtures)

Forensic Science International: Genetics 6 (2012) 677–678

Approaches to mixture data interpretation is in a state of change throughout the forensic DNA community

ELSEVIER

Forensic Science International: Genetics

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

Editorial

Focus issue—Analysis and biostatistical interpretation of complex and low template DNA samples

December 2012 – Forensic Science International: Genetics, volume 6, issue 6

Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in

Adele A. Mitchell *, Jeannie Tamariz, Kathleen O'Connell, Nubia Ducasse, Zoran Budimlija, Mechthild Prinz, Theresa Caragine

Department of Forensic Biology, Office of Chief Medical Examiner of The City of New York, 421 E 26th Street, New York, NY 10016, United States

NYC OCME Forensic Statistical Tool (FST) published

DNA Mixture Interpretation
April 12, 2013 Webcast

http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation.cfm

- **8-hours of DNA mixture interpretation training**
- **11 presentations** from **five different presenters**
  - John Butler, Mike Coble, Robin Cotton, Bruce Heidebrecht, Charlotte Word
- **20 poll questions** asked via SurveyMonkey (>600 participated)
  - Addressed additional questions sent via email or Twitter
- **>1000 participants** (almost entire U.S. represented and >10 countries)
- **Available for viewing or download** for at least six months (storage costs may limit longer-term storage)

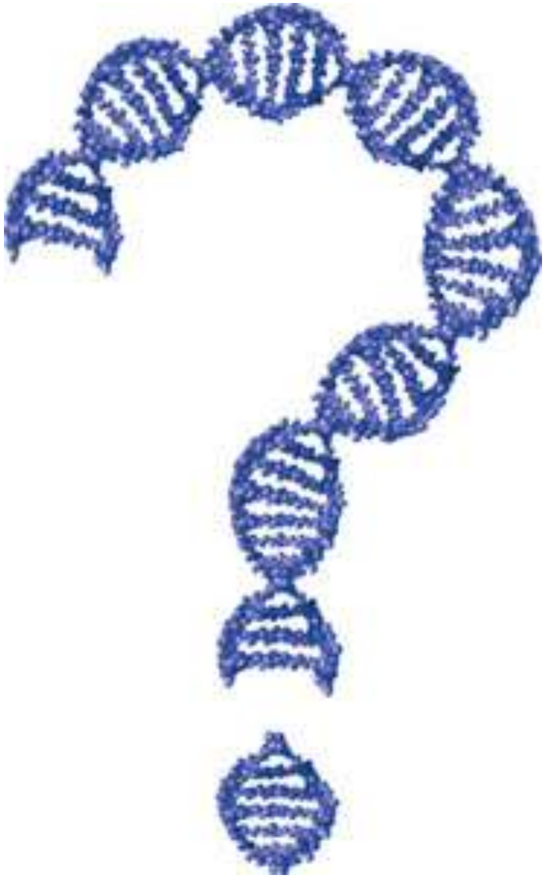
Left to right:
Gladys Arrisueno (NIST, Twitter feed monitor & poll questions)
John Paul Jones (NIST, webcast organizer)
Mike Coble (NIST, presenter)
John Butler (NIST, presenter & organizer)
Charlotte Word (Consultant, presenter)
Robin Cotton (Boston University, presenter)
Bruce Heidebrecht (Maryland State Police Lab, presenter)

Acknowledgments

National Institute of Justice funding to NIST and Boston University

<u>Slides and Discussions on DNA Mixtures</u>

- 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

301-975-4049