American Academy of Forensic Sciences
Seattle, WA
February 21, 2014

SEATTLE 2014

DNA Mixture Interpretation:

History, Challenges, Statistical Approaches, and Solutions

John M. Butler, Ph.D.

National Institute of Standards and Technology

Gaithersburg, Maryland

Co-authors

Charlotte J. Word, Ph.D.

& Michael D. Coble, Ph.D.

Disclaimers

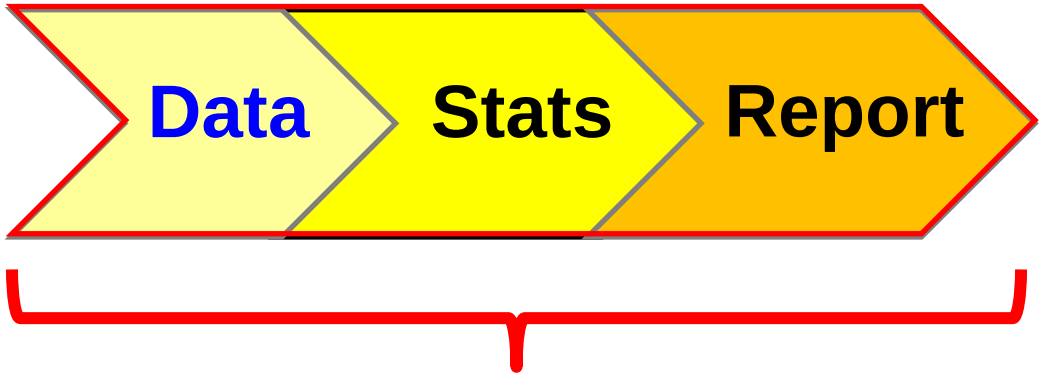
Funding for research and training on forensic DNA performed by the NIST Applied Genetics Group has come from the National Institute of Justice and the NIST Law Enforcement Standards Office

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.

Understanding Results Obtained & Sharing Them



Ian Evett on Interpretation

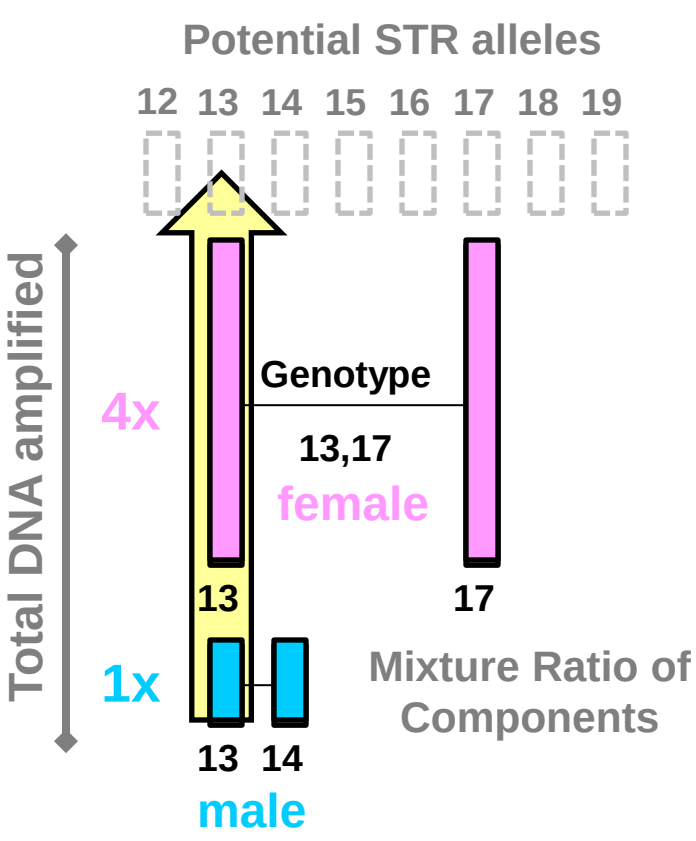
"The crucial element that the scientist brings to any case is the *interpretation* of those observations. This is the heart of forensic science: it is where the scientist adds value to the process."

Evett, I.W., et al. (2000). The impact of the principles of evidence interpretation on the structure and content of statements. *Science & Justice, 40,* 233-239.

True Sample Components

Sample Processing

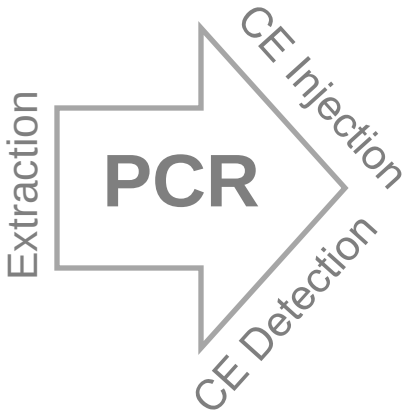
DNA Data Obtained



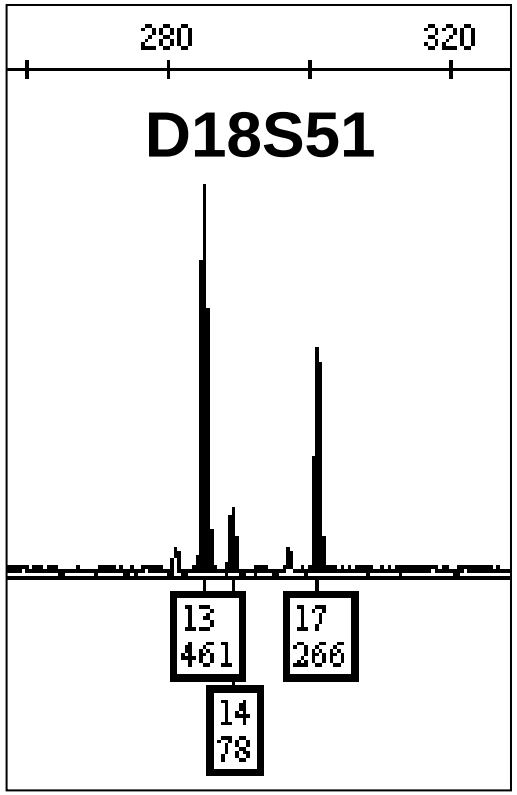
Number of Contributors
(sample components)

Validation

establishes variation and limits in the processes involved



portion of a CE electropherogram

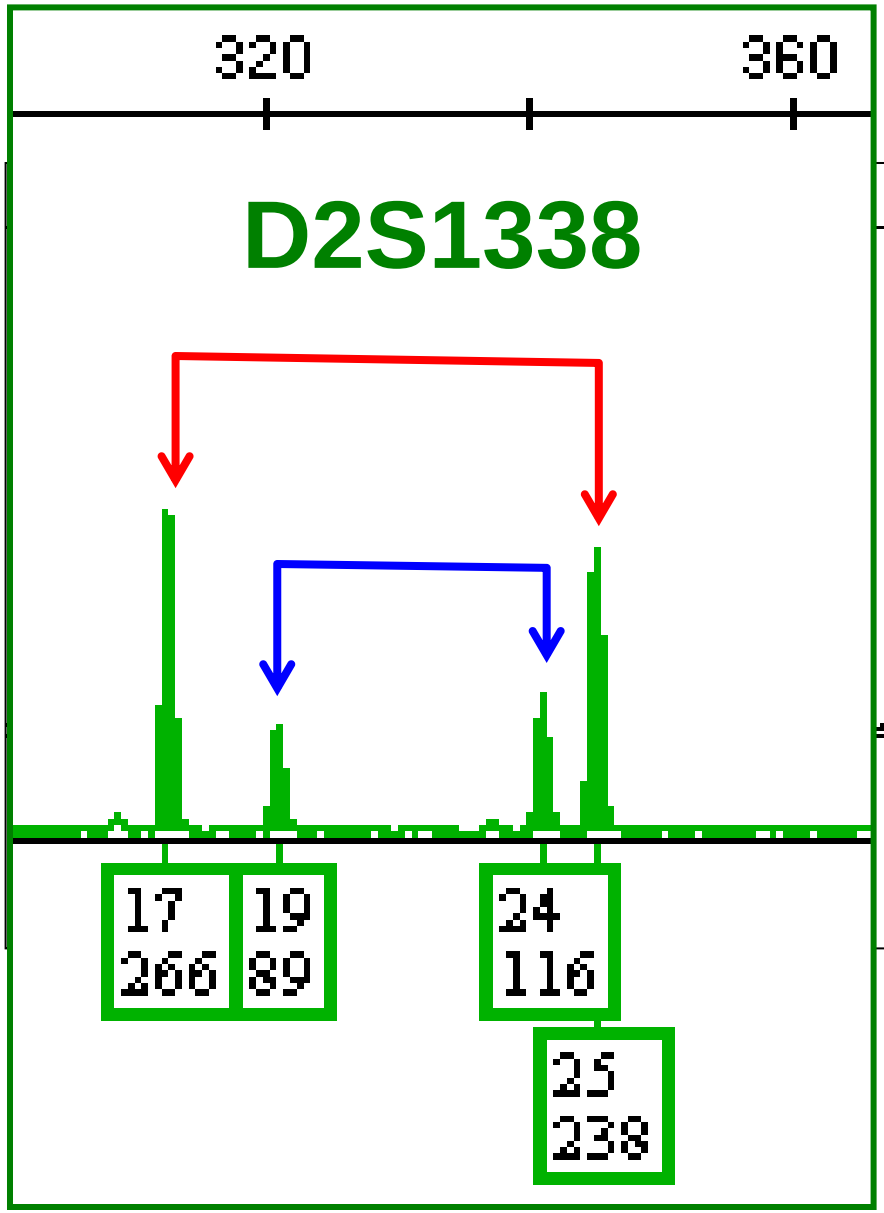


From available data

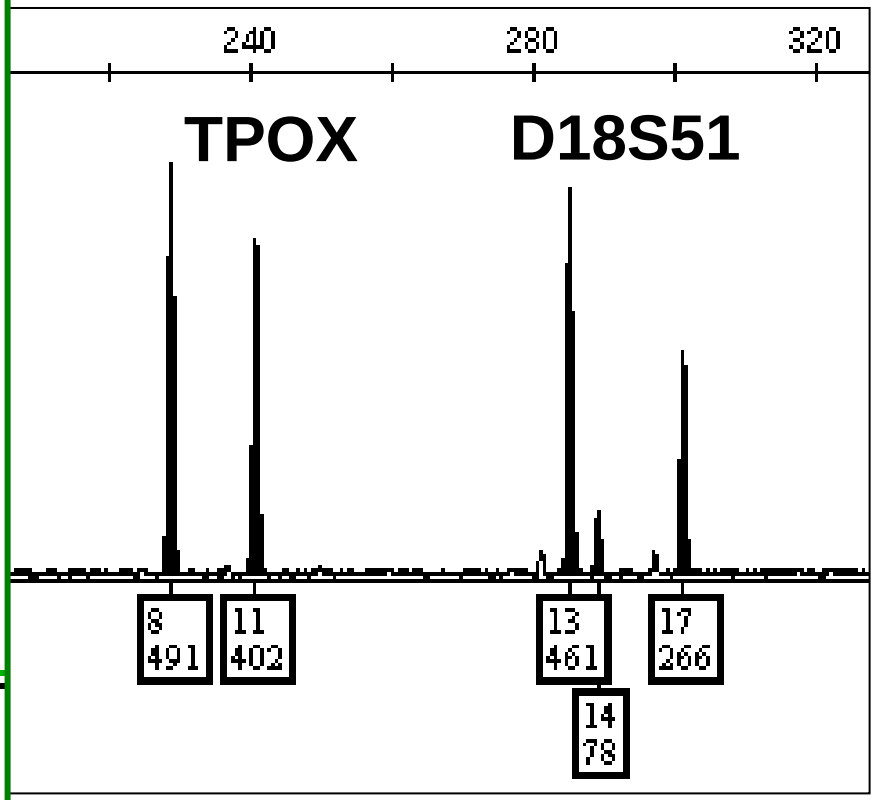
Infer possible genotypes & determine sample components

Goal of Interpretation

Other STR Loci in the Tested Help Inform the Interpretation



portion of a CE electropherogram



Non-overlapping alleles can help define mixture ratios

Complex Mixture Interpretation is **HARD**!!

Complex mixtures = mixtures of DNA with three or more contributors

Because:

1. Allele share is high

2. **Most complex mixtures have** at least one low template (LT) DNA contributor and therefore **stochastic effects**

3. The parameters used to interpret two person mixtures often may not be directly applied to complex mixtures

FAQ on SWGDAM.org website

Q: Are the 2010 SWGDAM Interpretation Guidelines applicable to all DNA mixtures?

SWGDAM Response: **These guidelines were written with single-source samples and two-person mixtures in mind**, …The *basic concepts* outlined in the 2010 SWGDAM Mixture Interpretation Guidelines hold true as they relate to DNA mixtures of three or more contributors, low-level DNA samples, and mixtures containing biologically related individuals. However, **there are nuances and limitations to the interpretation of these more complex mixtures, which are not fully explored in the 2010 guidelines**. …

http://www.swgdam.org/faq.html

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

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

A Brief History of DNA Mixtures (1)

- **1991** – Ian Evett (FSS) publishes Likelihood Ratios (LRs) as a method for DNA mixture interpretation using RFLP data
- **1992** – NRC I (p. 59) mentions Combined Probability of Inclusion (CPI) with a 2-person mixture
- **Early 1990s** – DQA1+PM not effective with mixtures
- **1995** – Mixtures presented in OJ Simpson trial
- **1996** – NRC II (p. 130) advocates LRs over CPI
- **1996** – 9plex STR kits (Profiler Plus, PowerPlex)
- **1997** – Weir et al LRs for mixture statistics

A Brief History of DNA Mixtures (2)

- **1998** – Clayton et al (FSS) DNA mixture deconvolution
- **2000** – initial SWGDAM Interpretation Guidelines published with very little on DNA mixtures
- **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**
- **2006** – ISFG Mixture Recommendations published emphasizing that LRs are a better method over CPI

A Brief History of DNA Mixtures (3)

- **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
- **2013** – NIST MIX13 study shows variation in approaches
- **2013** – 24plex STR kits (PowerPlex Fusion & GlobalFiler)
- ***Present*** – a number of software programs exist to help with calculations but no universal approach exists

Historical Perspective on DNA Mixture Approaches

Probabilistic genotyping software in development…

LR commonly used in Europe and other labs around the world

ISFG DNA Commission
LR with drop-out
2012

Today

2013 DNA TL Summit

SWGDAM guidelines
(RMP, CPI, LR)

ISFG DNA Commission
LR over CPI
2006

2010

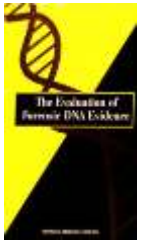
2008 NIJ burglary report increases touch evidence

Weir et al. describe LRs for mixtures

1997

NRC II report (p.130) supports LR

1996



Evett et al. describe LRs for mixtures

1991

Few mixtures

2-person mixtures predominate

complex mixtures increase

2000

DAB Stats
(Feb 2000)
CPI and LR okay

CPI becomes routine in U.S.

1992

NRC I report (p.59) supports CPI

1985

RMNE (CPI) used in paternity testing

LR = likelihood ratio
CPI = combined probability of inclusion
RMNE = random man not excluded

Published support for CPI approach? (1)

- One of the provided MIX13 laboratory reports cited the 1992 National Research Council report (NRC I) to support their use of the CPI mixture statistic

- NRC I (1992), p. 59:
  - "Typically, it will be impossible to distinguish the individual genotypes of each contributor. If a suspect's pattern is found within the mixed pattern, the appropriate frequency to assign such a 'match' is the sum of the frequencies of all genotypes that are contained within (i.e., that are a subset of) the mixed pattern." **[a description of the CPI statistic]**

Published support for CPI approach? (2)

- What does NRC II (1996) say on the topic?
  - Page 129: mentions the NRC I, p. 59 statement regarding CPI and illustrates an example calculation
  - Page 130: "**That calculation is hard to justify**, because it does not make use of some of the information available, namely, **the genotype of the suspect**. The correct procedure, we believe, was described by Evett et al. (1991)." [i.e., the likelihood ratio approach]

Evett, I.W., et al. (1991). A guide to interpreting single locus profiles of DNA mixtures in forensic cases. *Journal of Forensic Science Society, 31,* 41-47.

Why are we where we are today?

- The incredible success of DNA has lead to more sensitive methods and more samples being provided which has led to more complex mixtures (we are pushing the envelope)
  - Lower template DNA profiles have more uncertainty associated with them in terms of allele peak height variation

- Statistical interpretation techniques have not kept pace with the methodology improvements
  - Much of the forensic DNA community is effectively using a 1992 statistical tool on 21st century data

A Complexity/Uncertainty Threshold

New Scientist article (August 2010)

- **How DNA evidence creates victims of chance**
  - 18 August 2010 by Linda Geddes
- From the last paragraph:
  - **In really complex cases, analysts need to be able to draw a line** and say "This is just too complex, I can't make the call on it," says Butler. "Part of the challenge now, is that every lab has that line set at a different place. But the honest thing to do as a scientist is to say: **I'm not going to try to get something that won't be reliable**."

http://www.newscientist.com/article/mg20727743.300-how-dna-evidence-creates-victims-of-chance.html

Perhaps We Should Slow Down with Some of the DNA Mixtures That We (Scientists and Lawyers) Are Taking On…

Poor Quality Conditions

Large Numbers of Contributors

Resources to Learn More…

- Boston University DNA mixture training website
  - http://www.bu.edu/dnamixtures/ (2-, 3-, 4-person mixtures)

- NIST April 2013 webcast (8-hours of training)
  - http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation.cfm

- Listing of mixture literature & links to software tools
  - http://www.cstl.nist.gov/strbase/mixture.htm

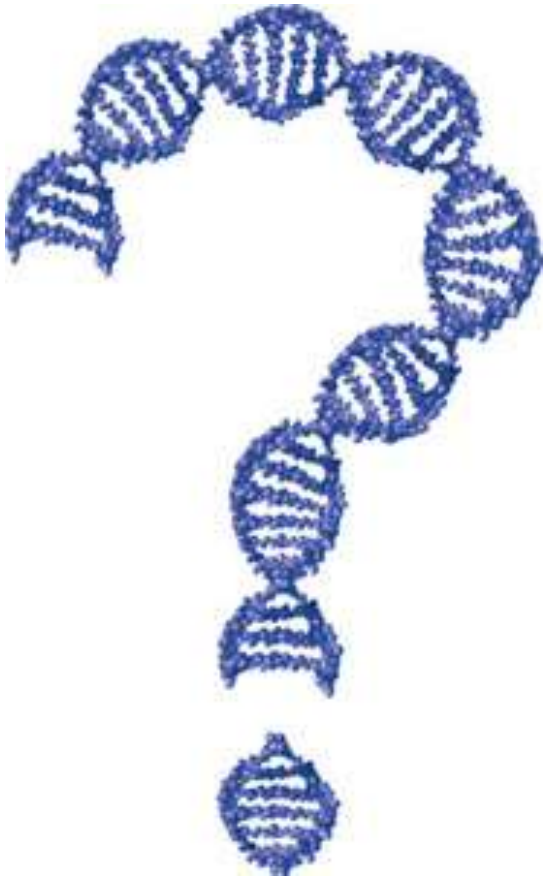
Word, C.J. (2011). Mixture interpretation: why is it sometimes so hard? *Profiles in DNA, 14(1)*. Available at http://www.promega.com/resources/profiles-in-dna/2011/mixture-interpretation-why-is-it-sometimes-so-hard/.

Acknowledgments

National Institute of Justice funding to NIST and Boston University

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

301-975-4049

NIST FORENSIC SCIENCES