MIXTURE INTERPRETATION:
Using Scientific Analysis

22nd International Symposium on Human Identification
October 3, 2011 (Washington, DC)

Presenters

John M. Butler, PhD	NIST, Applied Genetics Group
Michael D. Coble, PhD	NIST, Applied Genetics Group
Robin W. Cotton, PhD	Boston University, Biomedical Forensic Sciences
Catherine M. Grgicak, PhD	Boston University, Biomedical Forensic Sciences
Charlotte J. Word, PhD	Consultant

Supported by funding from the National Institute of Justice

NIJ Grant 2008-DN-BX-K158 to Boston University and Interagency Agreement 2008-DN-R-121 between NIJ and the NIST Office of Law Enforcement Standards funds the presenters. The Boston University grant also covers the registration for 175 US State and Local Crime Laboratory personnel.

ISHI 2011 Mixture Workshop - October 3, 2011

	Speaker	Topic
Introduction		
8:30 - 8:45 a.m.	Robin	Workshop introduction
8:45 - 9:00 a.m.	John	SWGDAM guidelines & mixture literature
Profile 1: Fundamental Principles of Mixture and Interpretation		
9:00 - 10:30 a.m.		
	Catherine	Profile overview & analytical threshold
	Mike	Stutter
	John	Stochastic threshold
	Charlotte	Profile analysis with assumption set 1
	Robin	Profile analysis with other assumptions
10:30 - 10:45 a.m.		BREAK
Profile 1: Weighting the Evidence (Statistics)		
10:45 a.m. - 12:15 p.m.		
	Mike	Statistics examples using specific loci
	Charlotte	Inclusion/exclusion comparison
		Report wording
12:15 - 1:00 p.m.		LUNCH
Profile 2: Complex mixtures		
1:00 - 2:45 p.m.		
	Robin	2-person indistinguishable mixture
	Mike	Statistical issues with 2-person indistinguishable mixture
	Charlotte	Complex mixtures (3 & 4 person, relatives)
2:45 - 3:00 p.m.		BREAK
Profile 3: Low Level DNA mixtures		
3:00 - 3:30 p.m.		
	John	Low level DNA mixture example: limitations of CPI
		Other confounding features of casework samples
Back to the Future		
3:30 - 4:45 p.m.		
	Robin	BU website introduction
	Mike	Are there software solutions?
	John	Literature review: value of emerging information
	Catherine	Validation U.S.A.
4:45 - 5:00 p.m.	Robin	Workshop evaluation

ISHI 2011
Mixture Interpretation Workshop: Using Scientific Analysis

October 3, 2011

NIJ National Institute of Justice

BOSTON UNIVERSITY

NIST

Welcome
to Washington DC, Maryland and Virginia

Thank you Promega for having us back this year!!

Promega 2011
Mixture Interpretation Workshop
Using **S**cientific **A**nalysis

8:30 am	Introduction of Presenters and Plan for Workshop
8:45 – 9:00 am	Profile 1: Fundamental Principles of Mixture Interpretation
9:00 -10:30 am	Profile 1: Fundamental Principles of Mixture Interpretation
10:30-10:45 am	BREAK
10:45 am – 12:15 pm	Profile 1: Weighting the Evidence
12:15 – 1:00 pm	LUNCH
1:00 – 2:45 pm	Profile 2: Complex Mixtures
2:45 – 3:00 pm	BREAK
3:00 – 3:30 pm	Profile 3: Low Level DNA Mixtures
3:30 - 4:45 pm	Back to the Future
4:45 – 5:00	Workshop Evaluation

Most participants are here through the sponsorship of NIJ

- NIJ Forensic Science Training Development and Delivery Program
- NIJ Grant # 2008-DN-BX-K158, awarded to Biomedical Forensic Science Program at Boston University School of Medicine
- Supporting registration for 175 participants from state and local laboratories

This grant also funds:

- Development of a web site for training in STR DNA profile mixture analysis
  - Lessons on specific topics related to mixtures
  - DNA profiles with 1, 2, 3, and 4 person mixtures
    - PowerPlex 16
    - Identifiler
    - Yfiler
    - Minifiler
- Over 2000 profiles in total
- Web site training material will be available for open use and .fsa files available for download

Mixture Workshop Presenters

Robin Cotton
Boston University

617-638-1952
rwcotton@bu.edu

John Butler
NIST

301-975-4049
john.butler@nist.gov

Catherine Grgicak
Boston University

617-638- 1968
cgrgicak@bu.edu

Mike Coble
NIST

301-975-4330
michael.coble@nist.gov

Charlotte Word
Consultant

301-527-1350
cjword@comcast.net

Presenters

- John Butler;
  - Ph.D. in Analytical Chemistry, University of Virginia
  - 20 years experience
  - Writes books on the side
  - The engine behind STRBase
- Mike Coble
  - Ph.D. in Genetics, George Washington University
  - 15 years DNA experience
  - Mitochondrial DNA and STRs at AFDIL
  - Now working even harder at NIST

Presenters

- Charlotte Word
  - Ph.D. in Microbiology, University of Virginia
  - 20 years casework and technical review experience for both public and private laboratories
  - More than 200 court testimonies in admissibility hearings and trials
  - Currently a private consultant in the Washington DC area

Presenters

- Robin Cotton
  - Ph.D. Molecular Biology and Biochemistry, University of California at Irvine
  - 18 years casework and testimony experience
  - Boston University School of Medicine since 2006
  - Program Director, Biomedical Forensic Sciences
- Catherine Grgicak
  - M.S. Forensic Science, University of Alabama
  - Ph.D. Chemistry, University of Ottawa
  - 3 years experience as DNA Analyst
  - Boston University School of Medicine since 2007

Clickers

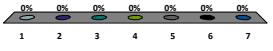
- Will allow real time audience participation
- We have practice slides to initiate audience participation.

Polling test for clickers; What time zone do you live in?

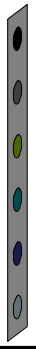
1. Eastern
2. Central
3. Mountain
4. Pacific
5. AK or HI
6. Other

What is your role in the laboratory?

1. DNA analyst
2. DNA technician
3. Database analyst
4. DNA technical leader
5. QA Manager
6. Attorney
7. Other



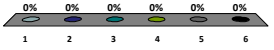
Your Experience Level as a DNA Analyst



1. Trainee — 0%
2. <2 years — 0%
3. 2-5 years — 0%
4. 5-10 years — 0%
5. 10+ years — 0%
6. I am not an analyst — 0%

How much DNA court testimony experience do you have?

1. Have not testified yet
2. 1 to 10 times
3. 11 to 25 times
4. 25 to 50 times
5. > 50 times
6. One more time and I will need a good shrink.



Goals of Workshop

Illustrate the process of mixture interpretation using a variety of mixture types.

Introduce relevant and current information in the literature.

Consider how the new information can inform and improve mixture and interpretation.

Expand Our Perspectives

We hope you have fun!!

Introduction to the SWGDAM Guidelines and the Mixture Literature

John M. Butler

NIJ National Institute of Justice

2011 Mixture Interpretation Workshop:
Using Scientific Analysis

BOSTON UNIVERSITY

NIST

Presentation Outline

- Highlights from the SWGDAM 2010 Guidelines
- Review steps (& purpose) in DNA interpretation
- Discuss reference list provided with handouts

Overview of the SWGDAM 2010 Interp Guidelines

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

Interpretation of Evidence Completed before Comparison to Known(s)

- "3.6.1. The laboratory **must establish guidelines** to ensure that, to the extent possible, **DNA typing results from evidentiary samples are interpreted before comparison with any known samples**, other than those of assumed contributors."

Q (question) before K (known)

  - While the FBI QAS do not address this issue, this is an example of an issue felt by the committee members to be of such importance that it warranted a "must."

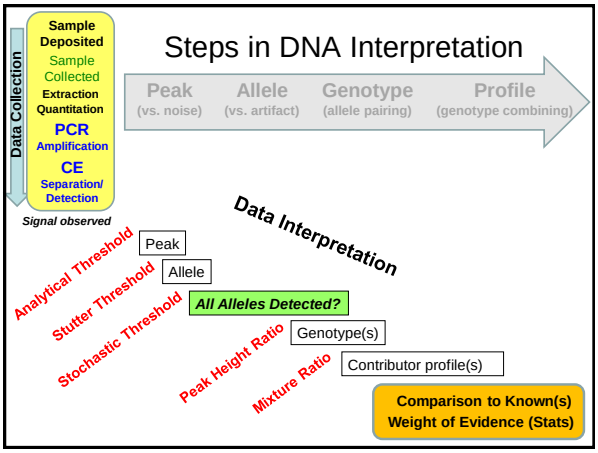
Stats Required for Inclusions

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.



Principles Behind Thresholds

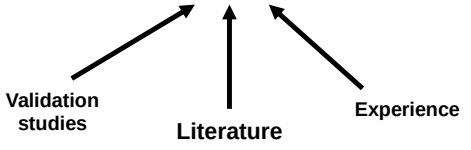
Thresholds (example values)	Principles Behind *(if properly set based on lab- & kit-specific empirical data)*
Analytical Threshold (e.g., 50 RFU)	Below this value, observed peaks cannot be reliably distinguished from instrument noise (baseline signal)
Limit of Linearity (e.g., 5000 RFU)	Above this value, the CCD camera can become saturated and peaks may not accurately reflect relative signal quantities (e.g., flat-topped peaks) and lead to pull-up/ bleed-through between dye color channels
Stochastic Threshold (e.g., 250 RFU)	Above this peak height value, it is reasonable to assume that allelic dropout of a sister allele of a heterozygote has not occurred at that locus; single alleles above this value in single-source samples are assumed to be homozygous
Stutter Threshold (e.g., 15%)	Below this value, a peak in the reverse (or forward) stutter position can be designated as a stutter artifact with single-source samples or some mixtures (often higher with lower DNA amounts)
Peak Height Ratio (e.g., 60%)	Above this value, two heterozygous alleles can be grouped as a possible genotype (often lower with lower DNA amounts)
Major/Minor Ratio (e.g., 4:1)	When the ratio of contributors is closer than this value in a two-person mixture, it becomes challenging and often impossible to correctly associate genotype combinations to either the major or minor contributor

Threshold Decisions

	Thresholds to Determine	Decisions to Make (lab & kit specific)	Useful Validation Data
Catherine	**Analytical = ____ RFU**	Single overall value or color specific	Noise levels in negative controls or non-peak areas of positive controls
John	**Stochastic = ____ RFU**	Minimum peak height RFU value or alternative criteria such as quantitation values or use of a probabilitistic genotype approach	Level where dropout occurs in low level single-source heterozygous samples under conditions used (e.g., different injection times, post-PCR cleanup)
Mike	**Stutter filter = ___%**	Profile, locus, or allele-specific	Stutter in single-source samples (helpful if examined at multiple DNA quantities)
Charlotte	**Peak Height Ratio = ___%**	Profile, locus, or signal height (quantity) specific	Heterozygote peak height ratios in single-source samples (helpful if examined at multiple DNA quantities)
Charlotte	**Major/Minor Ratio = ____**	When will you attempt to separate components of a mixture into major and minor contributors for profile deductions?	Defined mixture ratios (e.g., 1:1, 1:3, 1:9) with known samples to observe consistency across loci and to assess ability to deduce correct contributor profiles

Your Laboratory Interpretation Protocols

Standard Operating Procedures (SOPs)



SWGDAM Guidelines (2010) Introduction: *…the laboratory should utilize written procedures for interpretation of analytical results with the understanding that specificity in the standard operating protocols will enable greater consistency and accuracy among analysts within a laboratory. It is recommended that standard operating procedures for the interpretation of DNA typing results be sufficiently detailed that other forensic DNA analysts can review, understand in full, and assess the laboratory's policies and practices. The laboratory's interpretation guidelines should be based upon validation studies, scientific literature, and experience.*

The Mixture Literature

See provided reference list with **over 100 relevant references** for further information on each topic discussed in this workshop

Revised Quality Assurance Standard Requirement for Literature Review

Quality Assurance Standards for Forensic DNA Testing Laboratories
(effective September 1, 2011)

5.1.3.2. The laboratory shall have a program approved by the technical leader for the **annual review of scientific literature** that documents **the analysts' ongoing reading of scientific literature**. **The laboratory shall maintain or have physical or electronic access to a collection of current books, reviewed journals, or other literature applicable to DNA analysis**.

http://www.fbi.gov/hq/lab/fsc/backissu/oct2008/standards/2008_10_standards01b.htm

Useful Articles on DNA Mixture Interpretation

- Buckleton, J.S. and Curran, J.M. (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Sci. Int. Genet.* 2: 343-348.
- Budowle, B., *et al.* (2009) Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J. Forensic Sci.* 54: 810-821.
- Clayton, T.M., *et al.* (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci. Int.* 91: 55-70.
- Gill, P., *et al.* (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101.
- Gill, P., *et al.* (2008) National recommendations of the technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. *FSI Genetics* 2(1): 76–82.
- Schneider, P.M., *et al.* (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. *Int. J. Legal Med.* 123: 1-5.

German Mixture Classification Scheme

Schneider *et al.* (2009) *Int. J. Legal Med*. 123: 1-5

(German Stain Commission, 2006):

- **Type A**: no obvious major contributor, no evidence of stochastic effects
- **Type B**: clearly distinguishable major and minor contributors; consistent peak height ratios of ***approximately 4:1*** (major to minor component) for all heterozygous systems, no stochastic effects
- **Type C**: mixtures without major contributor(s), evidence for stochastic effects

SWGDAM

Type A — **"Indistinguishable"**

Type B — **"Distinguishable"**

Type C — **"Uninterpretable"**

Available for download from the ISFG Website:
http://www.isfg.org/Publication;Gill2006

ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Forensic Science International 160 (2006) 90–101

Forensic Science International
www.elsevier.com/locate/forsciint

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

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

[a] Forensic Science Service, Trident Court, 2960 Solihull Parkway, Birmingham, UK
[b] Forensic Science Group, School of Public Health, University of California, Berkeley, CA 510-339-1911, USA
[c] ESR, Private Bag 92021, Auckland, New Zealand
[j] University of Washington, Department of Biostatistics, Box 357232, Seattle, WA 98195, USA

Our discussions have highlighted a significant need for continuing education and research into this area.

Received 4 April 2006; accepted 10 April 2006
Available online 5 June 2006

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

ISFG Recommendations on Mixture Interpretation

http://www.isfg.org/Publication;Gill2006

1. The likelihood ratio (LR) is the preferred statistical method for mixtures over RMNE
2. Scientists should be trained in and use LRs
3. Methods to calculate LRs of mixtures are cited
4. Follow Clayton et al. (1998) guidelines when deducing component genotypes
5. Prosecution determines H_p and defense determines H_d and multiple propositions may be evaluated
6. When minor alleles are the same size as stutters of major alleles, then they are indistinguishable
7. Allele dropout to explain evidence can only be used with low signal data
8. No statistical interpretation should be performed on alleles below threshold
9. Stochastic effects limit usefulness of heterozygote balance and mixture proportion estimates with low level DNA

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

Budowle et al. (2009) Article from the FBI Mixture Committee

J Forensic Sci, May 2009, Vol. 54, No. 3
doi: 10.1111/j.1556-4029.2009.01046.x
Available online at: www.blackwell-synergy.com

Bruce Budowle,[1] Ph.D.; Anthony J. Onorato,[1] M.S.F.S., M.C.I.M.; Thomas F. Callaghan,[1] Ph.D.; Angelo Della Manna,[2] M.S.; Ann M. Gross,[3] M.S.; Richard A. Guerrieri,[1] M.S.; Jennifer C. Luttman,[1] M.F.S.; and David Lee McClure,[4] B.S.

Mixture Interpretation: Defining the Relevant Features for Guidelines for the Assessment of Mixed DNA Profiles in Forensic Casework*

In general we agree with the recommendations of Gill et al. that are: (i) when possible peak height/area should be included in mixture interpretation; (ii) stutter position peaks at similar peak height/area as that of obligate minor contributor alleles should be considered as potential alleles in the interpretation and statistics calculation; and (iii) a stochastic threshold (termed "dropout threshold") should be defined.

Budowle, B., *et al.* (2009) Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J. Forensic Sci.* 54: 810-821.

Responses to ISFG DNA Commission Mixture Recommendations

- UK Response
  - Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82
- **German Stain Commission**
  - Schneider *et al.* (2006) *Rechtsmedizin* 16:401-404 (German version)
  - Schneider *et al.* (2009) *Int. J. Legal Med.* 123: 1-5 (English version)
- ENFSI Policy Statement
  - Morling *et al.* (2007) *FSI Genetics* 1(3):291–292
- New Zealand/Australia Support Statement
  - Stringer *et al.* (2009) *FSI Genetics* 3(2):144-145
- **SWGDAM** – Interpretation Guidelines
  - Approved Jan 2010 and released April 2010 on FBI website

Elements of DNA Mixture Interpretation

Principles (theory)

↓

Protocols (validation)

↓

Practice (training & experience)

Consistency across analysts

ISFG Recommendations
SWGDAM Guidelines

⇣

Your Laboratory SOPs

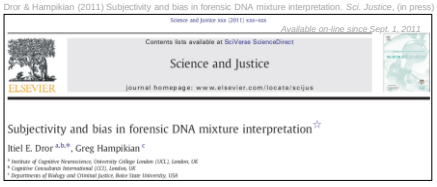
⇣⇡

Training within Your Laboratory

Periodic training will aid accuracy and efficiency within your laboratory.

Recent Article on "Problems" with Mixture Interpretation

Dror & Hampikian (2011) Subjectivity and bias in forensic DNA mixture interpretation. *Sci. Justice*, (in press)


Available on-line since Sept. 1, 2011

Contents lists available at SciVerse ScienceDirect

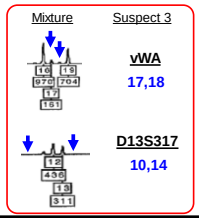
Science and Justice

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

Subjectivity and bias in forensic DNA mixture interpretation ☆

Itiel E. Dror [a,b,*], Greg Hampikian [c]

- DNA mixture from an adjudicated criminal case – and a "suspect 3" profile -- were provided to 17 DNA analysts without contextual case information
- **Results** (re: Suspect 3)**: 1 "cannot be excluded", 4 "inconclusive", and 12 "excluded"**
- Not consistent between analysts – authors suggest **subjectivity** in mixture interpretation
- Not consistent with original result – authors suggest **bias** due to availability of case context with original analysts



STRBase Mixture Section

http://www.cstl.nist.gov/biotech/strbase/mixture.htm

Section launched in October 2010 and will continue to develop over time

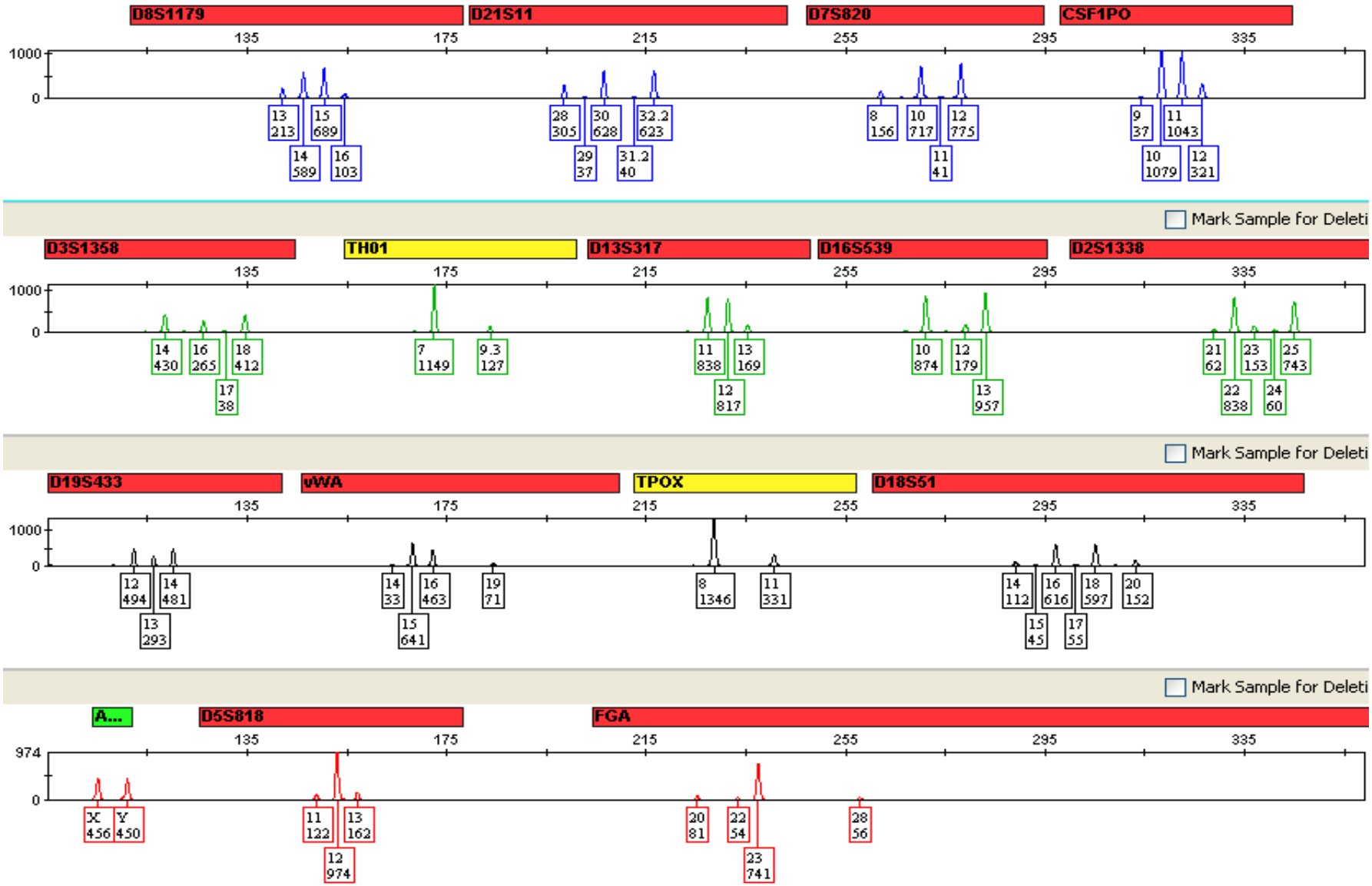
- Updated literature lists by topic
- Workshop slides and links to other info
- Useful freeware programs (e.g., Excel macros) will be available for download

Acknowledgments

- **SWGDAM Mixture Committee members** for their hard work through many long hours of discussing and writing these guidelines
- NIJ Funding to our NIST Group through NIST OLES interagency agreement 2008-DN-R-121

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

Profile 1 (stutter filter off)



Profile 1 Introduction
& Analytical Thresholds

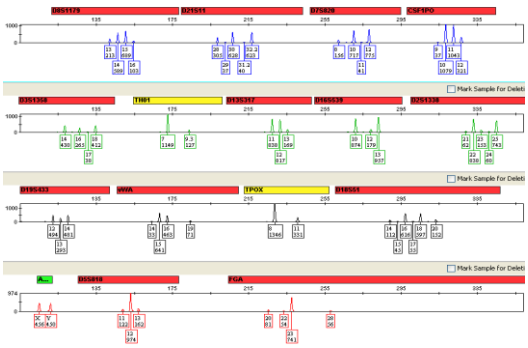
Catherine M. Grgicak

NIJ National Institute of Justice

ISHI 2011 Mixture Interpretation Workshop:
Using **S**cientific **A**nalysis

BOSTON UNIVERSITY
NIST

Profile 1 (stutter filter off)



SWGDAM Mixture Guidelines

- 1.1. "*Analytical threshold: The Laboratory should establish an analytical threshold based on signal-to-noise analyses of internally derived empirical data. As an example, an analytical threshold may be based on two times the intensity difference between the highest peak and lowest trough within the instrumental noise data. Other scientific methods may be used*"
  - What are these 'Other scientific methods'?
- 3.1.1.2. "*While the application of an analytical threshold may serve to filter out some non-allelic peaks, the analytical threshold should be established based on signal-to-noise considerations (i.e. distinguishing potential allelic peaks from background). The analytical threshold should not be established for purposes of avoiding artifact labeling as such may result in the potential loss of allelic data*"
  - How does one determine analytical threshold?

Methods Discussed in Literature

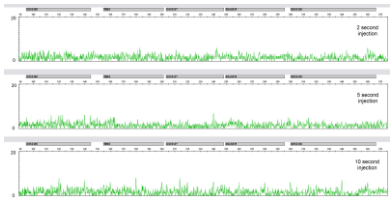
Use data from negatives (i.e. samples with no DNA)

- Method 1.
  - Kaiser (IUPAC 1976)
    - Winefordner 1983 and Krane 2007
- Method 2.
  - Currie (IUPAC 1995)
    - Winefordner 1983
- Method 3.
  - Example in SWGDAM Guidelines

Use data from DNA dilution series

- Method 4.
  - Miller & Miller. *Statistics for Analytical Chemistry (Ellis Horwood & Prentice Hall)*
    - IUPAC 1997 ElectroAnalytical Committee
- Method 5.
  - 1997 IUPAC ElectroAnalytical Committee Recommendations

Method 1, 2 and 3 - Negatives



- Run 33 negatives
- Suggest using amplification negatives (not just blanks) as they are more representative
- Analyze with GeneMapper ID at an RFU of 1 and remove labels within +/-2 bases of the Internal Size Standard

Method 1, 2 - Negatives

$$AT_{M1} = \bar{Y}_{bl} + k s_{bl}$$

$$AT_{M1} = 3.11 + (3 * 1.14) = 6.53$$

$$AT_{M1} = 7$$

Kaiser argued a value of $k = 3$ will result in an AT where we are at least 89% confident and at most 99.86% confident noise will be below this value.

$$AT_{M2} = \bar{Y}_{bl} + t_{1-\alpha,v} \frac{s_{bl}}{\sqrt{n}}$$

$$AT_{M2} = 3.11 + \left(2.46 * \frac{1.14}{\sqrt{30}} \right) = 3.68$$

$$AT_{M2} = 4$$

Both 95% and 99% confidence intervals have been suggested.

Method 3 and 4

$$AT_{M3} = 2(Y_{\max} - Y_{\min})$$

$$AT_{M3} = 2(9-0) = 18$$

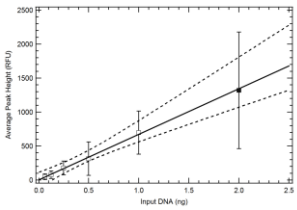
$$AT_{M3} = 18$$

NOTE: Because we are NOT using raw data (but analyzed GeneMapper data), data below 0 RFU is not 'observed' and therefore, the number calculated is smaller than expected!!!
HOWEVER, the calculated AT is still larger than either Method 1 or 2!

RFU Signal of Blank	No. of observations	Percent Rank	AT_{M4} at rank >=99%
1	206	3.87	
2	1481	31.73	
3	1884	67.16	
4	1161	89.00	
5	453	97.51	6
6	110	99.59	
7	18	99.92	
8	3	99.98	
9	1	100	

Methods 5 & 6 – Positives (Standard Curves)

- Regression of positive samples (i.e. single source samples)
- Amplified 0.0625-4ng dilution series, injected 5s using manufacturer's recommended protocol
- Plot of Input DNA (ng) versus average peak height (per color) – with error bars
  - If a peak was homozygous, the RFU was divided by 2



- The points at 2 and 4 ng fall off the line (PCR efficiency approaching a plateau)!
- The error bars become larger with increased DNA input!

- A weighted linear regression is within the linear range (i.e. 0.0625 – 1 ng) was used.

Method 5 & 6

- b (y-intercept) = -2.30
- S_y (standard error of regression) = 10.77

$$AT_{M4} = b + 3S_y$$

$$AT_{M4} = 31$$

- b (y-intercept) = -2.30
- S_y (standard error of regression) = 10.77
- t-stat (n-1=4) and alpha of 99% t=3.75

$$\mathrm{AT}_{\mathrm{M5}} = \mathrm{b} + t_{n-1,\alpha} S_y$$

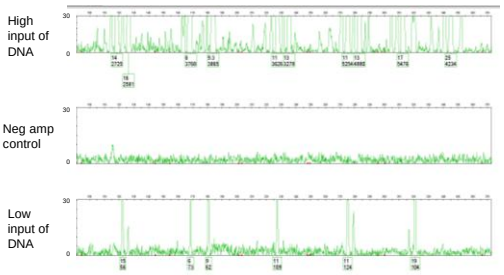
$$AT_{M5} = 39$$

Summary of Results

Method	Origin	Analytical Threshold for green 5s injection example
1	Negatives	7
2	Negatives	4
3	Negatives	18
4	Negatives	6
5	DNA Series	31
6	DNA Series	39

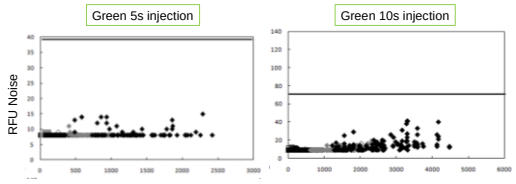
Before you choose, consider the following slide…

Baselines Positives ≠ Baselines Negatives



Type I Error – False Labeling of Noise Peaks

- This is not instrument baseline/noise
- Single source DNA data amplified from 0.0625 – 2 ng
  - Differentiated 'noise' from artifact
    - -A, pull-up, stutter (+ or -), spikes, dye artifacts
- Plotted RFU of the known/expected peak versus the highest 'noise' peak



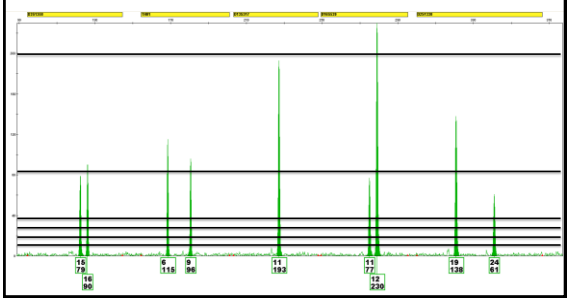
0.06 (□), 0.13 (■), 0.25 (▣), 0.5 (◊), 1 (♦) and 2 (♦) ng

Conclusions – Type 1 Error

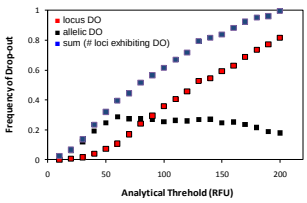
- Each of the methods can be used to describe the analytical threshold in SWGDAM Guidelines 1.1 and 3.1.1.2
  - -Data suggest 'noise' does not remain constant between negatives and samples with a significant amount of DNA (i.e. RFUs >1000).
- Methods 5 and 6 require that each run protocol (i.e. different injections times) be analyzed separately
- Different color channels behave differently – if possible, determine ATs for each color
- ATs derived from methods based on negative sample analysis (i.e. Method 1) may not be optimal for medium-high template samples – but reasonable for low-template ones.

Type II Error – False non-labeling of alleles (Drop-out)

Single source 0.125ng, 1ul 3130 prep volume and 10s injection



Type II Error – False non-labeling of alleles



$$freqDO(locus) = \frac{\#hetloci(2alleleDO)}{total\#hetloci}$$

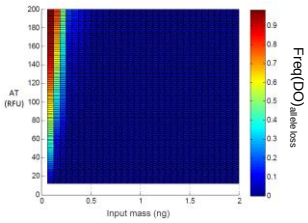
$$freqDO(allele) = \frac{\#hetloci(1alleleDO)}{total\#hetloci}$$

-As AT increases, locus DO increases, while allele DO stabilizes after 50 RFU then starts to decrease after AT of ~150 RFU.

-Although a higher AT (i.e. >150 RFU) begins to decrease the number of loci where allele DO occurs (less stochastic variation),

-Locus DO increases, resulting in an overall increase in DO with AT for Low-template samples

Type II Error – False non-labeling of alleles

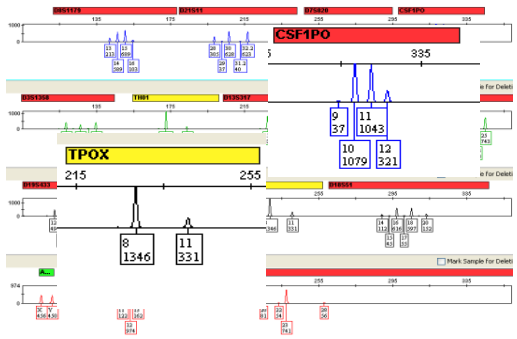


$$\text{Freq(DO)}_{\text{allele loss}} = \frac{\text{\# non-labeled alleles}_{\text{het loci}}}{\text{\# alleles expected}_{\text{het loci}}}$$

-AT's have a large effect on the ability to detect/label alleles.

- To take a 'conservative' approach and utilize high AT values leads to a substantial level of Type II errors for low-level samples (i.e. <1000RFU).

Profile (no stutter filter)



The AT is…..

#2.

- AT of 30 RFU (AT_{M5} 95% confidence) based on samples that contained DNA.

Color	AT_{M5} 95% confidence	RFU Threshold Applied for ISHI workshop
Blue	19	30
Green	24	
Yellow	16	
Red	13	

– NB: 30 RFU for all colors was used for simplicity and ATs applied on a per color basis is recommended

Stutter

Michael D. Coble

NIJ National Institute of Justice

2011 Mixture Interpretation Workshop:
Using Scientific Analysis

BOSTON UNIVERSITY
NIST

Review of the Literature

Study	Kit	Measured	TH01	vWA	D18S51
Greenspoon *et al.* (2004)	PP16 BIO	mean + 3SD	5	14	13
Krenke *et al.* (2002)	PP16	mean + 1SD	3	10	9
Moretti *et al.* (2001)	Pro+/CoFiler	mean + 3SD	15.9	11.7	13.9
Mulero *et al.* (2008)	MiniFiler	max %	-	-	**17.3**
Hill et al. (2010)	PP ESX	mean + 3SD	4.2	14.6	14.6
User Manual	Identifiler	max%	5.1	12.6	17
User Manual	IDfiler Direct	mean + 3SD	4.7	11.9	12.8
User Manual	IDfiler Plus	mean + 3SD	4	12.4	13.6

Many labs just use a flat 15%

Allele-Specific Stutter %

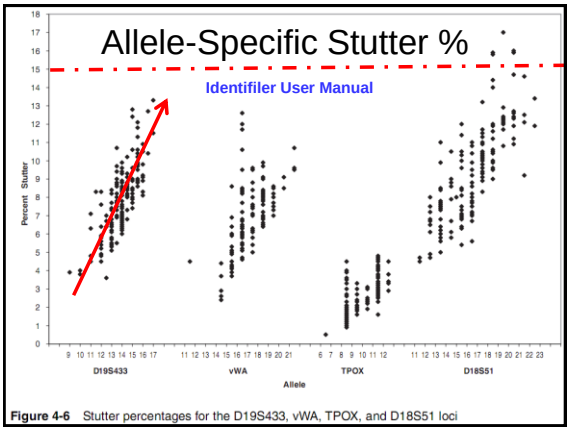
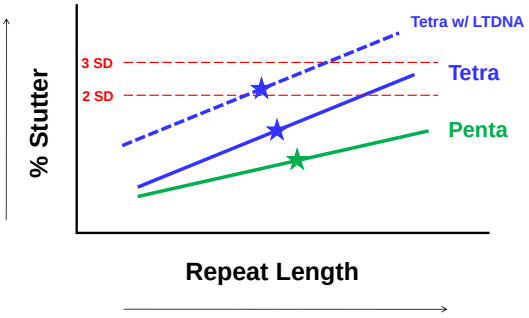


Figure 4-6 Stutter percentages for the D19S433, vWA, TPOX, and D18S51 loci

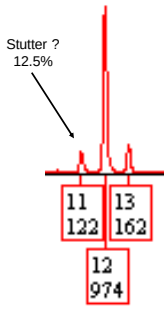
Stutter Trends



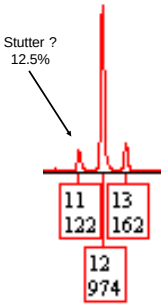
Interpretation of Potential Stutter Peaks in a Mixed Sample

- 3.5.8.1. For mixtures in which minor contributors are determined to be present, a peak in stutter position (generally n-4) may be determined to be 1) a stutter peak, 2) an allelic peak, or 3) indistinguishable as being either an allelic or stutter peak.

Profile 1 - D5S818



Profile 1 - D5S818



Possible Genotypes

13, 13
12, 13
11, 13

ISFG Recommendation #6 Example

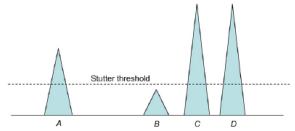


Fig. 2. A two person mixture with major peaks *C*, *D* and minor peaks *A*. There is an additional peak present in a stutter position (*B*).

Likely a AA

(homozygote)

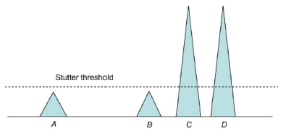


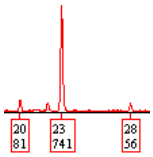
Fig. 3. A two person mixture with major peaks *C*, *D* and minor peaks *A*, *B*, where *B* is in a stutter position.

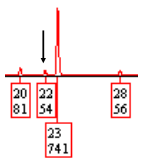
Possibly AB

(heterozygote)

Could also be AC, AD, AA, or A,? (dropout)

Profile 1 - FGA





Major	Minor
If Assume 2 Contributors….	
23,23	20,28
If Assume >2 Contributors…	
23,23	20,?; 28,?; 22,?; ?,?

Summary

- Stutter can vary across profiles, loci, or alleles.
- Stutter becomes especially problematic for mixtures when samples are at low [DNA] levels.
- Labs should decide when is it appropriate to turn off stutter filters, especially when the minor component alleles are nearly the same height as stutter peaks.

Setting and Applying
Stochastic Thresholds

John M. Butler

NIJ National Institute of Justice

ISHI 2011 Mixture Interpretation Workshop:
Using **S**cientific **A**nalysis

BOSTON UNIVERSITY
NIST

Presentation Outline

- What is a stochastic threshold and why is it important?
- How can you determine your stochastic threshold?
- How should you appropriately apply your stochastic threshold?
  - Remember that statistics and interpretation are coupled

General Definition of Stochastic

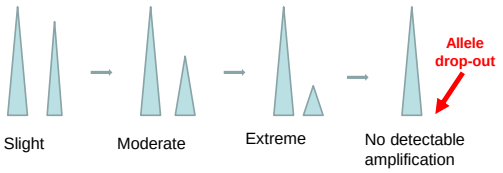
- Stochastic is synonymous with "random." The word is of Greek origin and means "pertaining to chance". … Stochastic is often used as counterpart of the word "deterministic," which means that random phenomena are not involved. Therefore, stochastic models are based on random trials, while deterministic models always produce the same output for a given starting condition.
- http://mathworld.wolfram.com/Stochastic.html

How can we characterize variation?

- Look at total amount of variation at end of process
  - Follow the positive control over time
- Experimentally break process into components and characterize using appropriate statistics
  - e.g., separate amplification variation from injection variation
- Analyze existing or new validation data, training sample data, SRM data, kit QC data
- Use casework data
  - e.g., variation between knowns (victim's DNA profile within an intimate sample) and matching single-source evidence profiles

Problem with Stochastic Effects

- **Allele drop-out** is an extension of the amplification disparity that is observed when heterozygous peaks heights are unequal
  - Occurs in single-source samples and mixtures
  - Analyst is unable to distinguish complete allele drop-out in a true heterozygote from a homozygous state



What is Allele Drop Out?

- Scientifically
  - **Failure to detect** an allele within a sample or failure to amplify an allele during PCR. *From SWGDAM Guidelines, 2010*
  - Note that: Failure to detect ≠ failure to amplify
- Operationally
  - Setting a threshold(s) or creating a process, based on validation data and information in the literature, which allows assessment of the likelihood of drop-out of an allele or a locus.

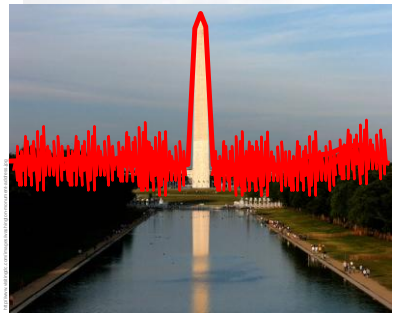
Stochastic Effects and Stochastic Threshold

SWGDAM 2010 Interpretation Guidelines glossary:

- **Stochastic effects**: the observation of intra-locus peak imbalance and/or allele drop-out resulting from random, disproportionate amplification of alleles in low-quantity template samples
- **Stochastic threshold**: the peak height value above which it is reasonable to assume that, at a given locus, allelic dropout of a sister allele has not occurred

http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines

Is this "peak" a homozygote or a heterozygote missing an allele?



Important Principle: With many casework sample, we cannot avoid stochastic effects and allele or locus drop-out.

Why ?

We do not know the number of contributors to a sample or the true contributor ratio in a mixture!

Sample Mixture Ratio Impacts Amount of DNA Available for PCR Amplification

Assume sample is a 1:3 mixture of two sources:

Amount of DNA	~ # of cells from major component	~ # of cells from minor component
1 ng	107	36
0.5 ng	53	18
0.25 ng	27	9
0.125 ng	12	4
0.063 ng	7	2

Stochastic effects expected with PCR amplification from <20 cells

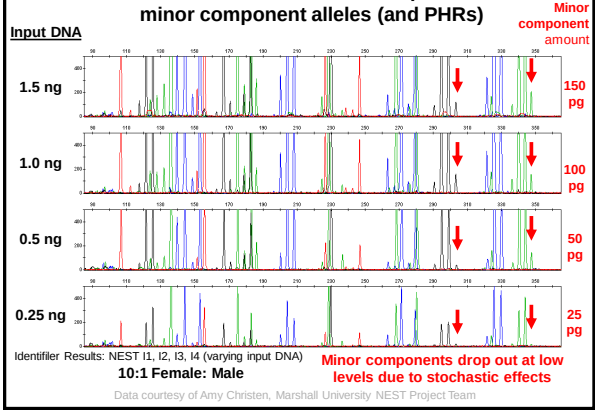
Stochastic Effects May Be Seen in Samples with Sufficient Total DNA

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

Amount of DNA	~ # of cells from major component	~ # of cells from minor component
1 ng	129	14
0.5 ng	64	7
0.25 ng	32	4
0.125 ng	16	2
0.063 ng	8	1

Stochastic effects expected with PCR amplification from <20 cells

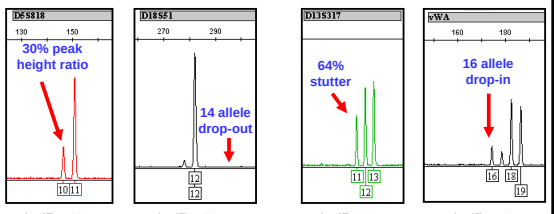
Stochastic effects can impact minor component alleles (and PHRs)



Stochastic Effects with Low Levels of DNA

When Combined with Higher Sensitivity Techniques

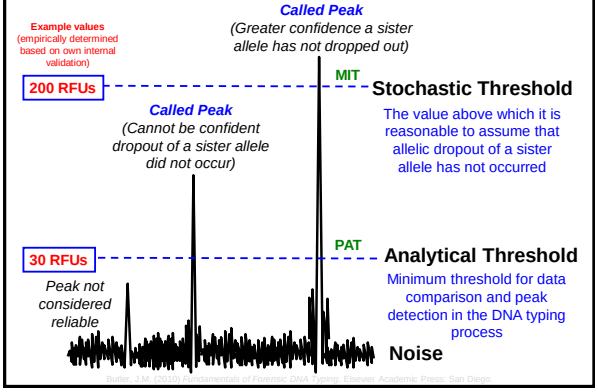
Loss of True Signal (False Negative)		*Gain of False Signal (False Positive)*	
Severe **Peak Imbalance**	Allelic **Drop-out**	**High Stutter**	Allelic **Drop-in**



Identifiler, 30 pg DNA, 31 cycles	Identifiler, 30 pg DNA, 31 cycles	Identifiler, 10 pg DNA, 31 cycles	Identifiler, 10 pg DNA, 31 cycles
Correct genotype: **10,11**	**12,14**	**12,13**	**18,19**

Approaches to Setting a Stochastic Threshold

Overview of Two Thresholds



Two Thresholds

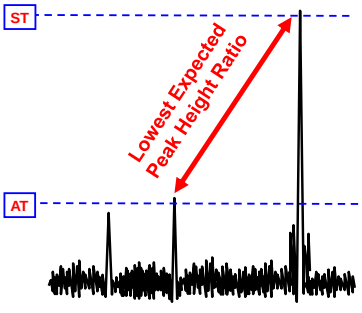
- Peak Amplitude Threshold (**PAT**)
- Match Interpretation Threshold (**MIT**)

Pat Buchanan

If between PAT and MIT, **can exclude** but not use statistics

Mitt Romney

Stochastic and Analytical Thresholds
Impact Lowest Expected Peak Height Ratio



Determining the Dropout (Stochastic) Threshold

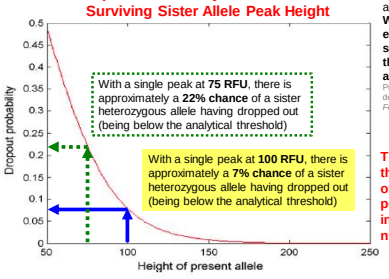
Gill *et al.* (2008) *FSI Genetics* 2(1): 76–82

- The dropout threshold can be determined experimentally for a given analytical technique from a series of pre-PCR dilutions of extracts of known genotype technique (it will probably vary between analytical methods). These samples can be used to determine the point where allelic dropout of a heterozygote is observed relative to the size of the survivor companion allele. The threshold is the maximum size of the companion allele observed. This is also the point where Pr(D) approaches zero (Fig. 4).

Dropout threshold will change depending on instrument and assay conditions (e.g., longer CE injection will raise dropout threshold)

Setting a Stochastic Threshold is Essentially Establishing a Risk Assessment

Drop Out Probability as a Function of Surviving Sister Allele Peak Height



"Currently, most laboratories use an arbitrary stochastic threshold. **When a protocol is changed, especially if it is made more sensitive to low-level DNA, then the stochastic threshold must also change.**"

Puch-Solis R, et al. (2011). Practical determination of the low template DNA threshold. *Forensic Sci. Int. Genet.* 5(5): 422-427.

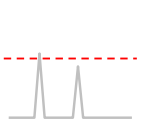
The position and shape of this curve may change based on anything that can impact peak detection (e.g., CE injection time, PCR cycle number, post-PCR cleanup).

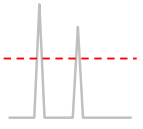
Gill, P., et al. (2009). The *low-template* (stochastic) threshold-Its determination relative to risk analysis for national DNA databases. *FSI Genetics*, 3, 104-111.

Stochastic Effects and Thresholds

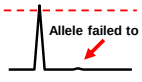
Regular Injection

Injection Following Desalting (MiniElute)

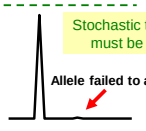




When PCR amplifying low levels of DNA, allele dropout may occur



False homozygote



Forensic Science International: Genetics 3 (2009) 222–226

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Estimating the probability of allelic drop-out of STR alleles in forensic genetics

Torben Tvedebrink [a,*], Poul Svante Eriksen [a,1], Helle Smidt Mogensen [b,2], Niels Morling [b,3]

[a] *Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark*
[b] *Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Fredrik V's Vej 11, DK-2100 Copenhagen East, Denmark*

Table 3
Mean peak heights (rfu) for various drop-out probabilities for 10 STR loci.

$P(\widehat{D}\vert\hat{H})$	D3	vWA	D16	D2	D8	D21	D18	D19	TH0	FGA	Overall
0.0001	556	577	622	562	558	461	531	722	723	692	648
0.0005	384	399	430	388	385	318	367	499	499	478	439
0.0010	327	340	366	331	328	271	313	425	426	407	371
0.0050	226	235	253	228	226	187	216	293	294	281	251
0.0100	192	200	215	194	193	159	184	250	250	239	212
0.0500	132	137	147	133	132	109	126	171	171	164	142
0.1000	111	115	124	112	111	92	106	144	144	138	119
0.2000	92	95	103	93	92	76	88	119	120	114	98
0.3000	81	84	91	82	81	67	78	105	106	101	86
0.4000	73	76	82	74	74	61	70	95	95	91	77
0.5000	67	69	75	68	67	55	64	87	87	83	70
0.6000	61	63	68	62	61	50	58	79	79	76	63
0.7000	55	57	62	56	55	46	53	71	71	68	57
0.8000	49	50	54	49	49	40	46	63	63	60	50
0.9000	40	42	45	41	40	33	39	52	52	50	41
0.9500	34	35	38	34	34	28	32	44	44	42	34
0.9900	23	24	26	23	23	19	22	30	30	29	23

Slide from Erica Butts (NIST) 3500 presentation in Innsbruck, Austria (Sept 5, 2011)

Setting Stochastic Methodology

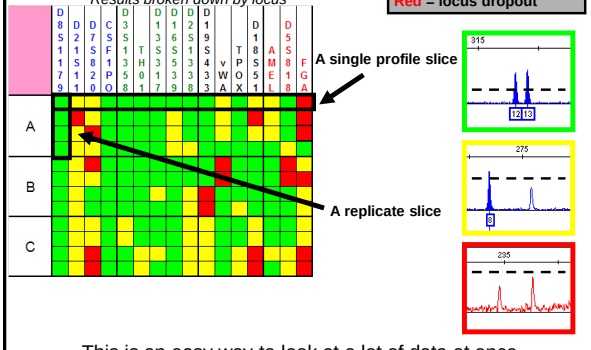
- Calculated with data from the sensitivity study (DNA dilution series) analyzed with dye specific analytical thresholds
- Examination of sample amounts where dropout is observed (50 pg, 30 pg, 10 pg for Identifiler and Identifiler Plus)
  - Focus on sample amounts with dropout present to examine stochastic effects including severe imbalance of heterozygous alleles and allele dropout
- Stochastic Threshold: The RFU value of highest surviving false homozygous peak per dye channel

Slide from Erica Butts (NIST) 3500 presentation in Innsbruck, Austria (Sept 5, 2011)

Heat Map Explanation

Green = full (correct) type
Yellow = allele dropout
Red = locus dropout

Results broken down by locus



This is an easy way to look at a lot of data at once

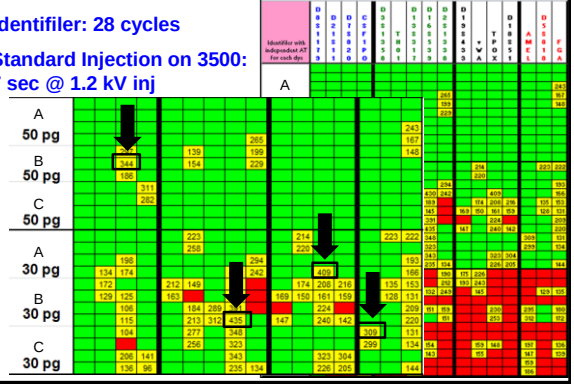
n=84 Samples

Slide from Erica Butts (NIST) 3500 presentation in Innsbruck, Austria (Sept 5, 2011)

Stochastic Threshold

Identifiler: 28 cycles

Standard Injection on 3500: 7 sec @ 1.2 kV inj



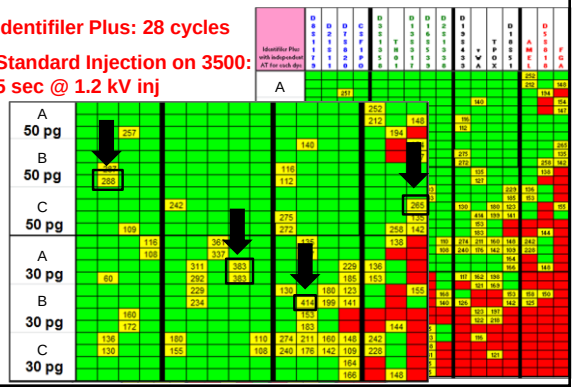
n=84 Samples

Slide from Erica Butts (NIST) 3500 presentation in Innsbruck, Austria (Sept 5, 2011)

Stochastic Threshold

Identifiler Plus: 28 cycles

Standard Injection on 3500: 5 sec @ 1.2 kV inj



n=84 samples

Slide from Erica Butts (NIST) 3500 presentation in Innsbruck, Austria (Sept 5, 2011)

Summary of Thresholds

Both AT and ST values rounded to the nearest 5 RFU value

Expected peak height ratio (PHR) is assuming the possibility of having one peak at the AT and one peak at the ST

Expected PHR = AT/ST

Identifiler: 7 sec @ 1.2 kV (28 cycles)

	AT (RFU)	Highest Surviving Peak (RFU)	ST (RFU)	Expected PHR
Blue	95	344	345	28%
Green	130	435	435	30%
Yellow	140	409	410	34%
Red	120	309	310	39%

Identifiler Plus: 7 sec @ 1.2 kV (28 cycles)

	AT (RFU)	Highest Surviving Peak (RFU)	ST (RFU)	Expected PHR
Blue	55	288	290	19%
Green	75	383	385	19%
Yellow	105	414	415	25%
Red	120	265	265	45%

Keep in Mind…

"The use of bounds **applied to data that show continuous variation** is common in forensic science and is often a pragmatic decision. However it should be borne in mind that applying such bounds has arbitrary elements to it and that **there will be cases where the data lie outside these bounds**."

Bright, J.A., et al. (2010). Examination of the variability in mixed DNA profile parameters for the Identifiler multiplex. *Forensic Science International: Genetics, 4,* 111-114.

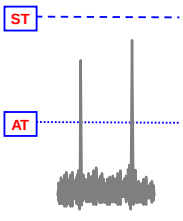
Appropriately Applying a Stochastic Threshold

ST = 200 RFU will be used for this workshop

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

Can This Locus Be Used for Statistical Calculations?



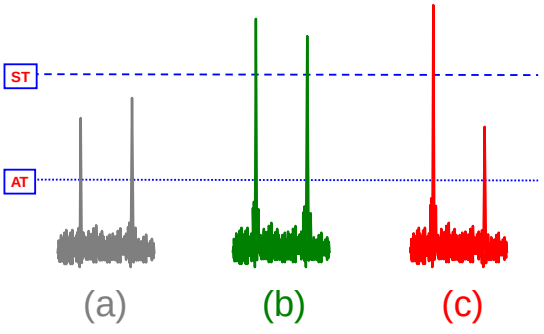
It depends on your assumption as to the number of contributors!

If you assume a single-source sample, then you can assume that the detection of two alleles fully represents the heterozygous genotype present at this locus.

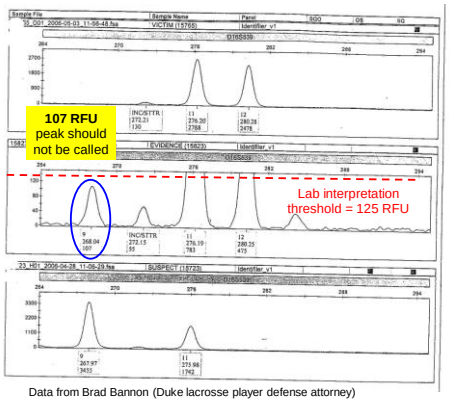
If you assume (from examining other loci in the profile as a whole) that the sample is a mixture of two or more contributors, then there may be allele drop-out and all alleles may not be fully represented.

Three Example Scenarios

Data from a Single Locus

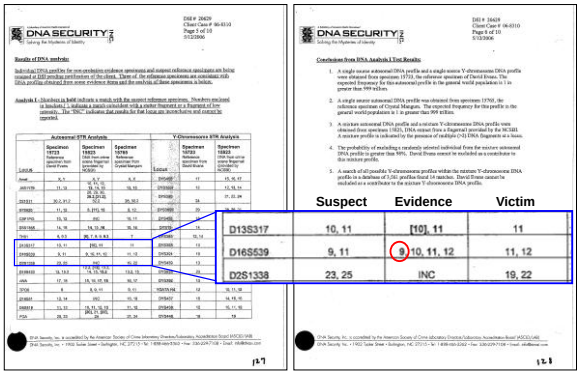


Don't Call Peaks Below Your Validated Threshold!



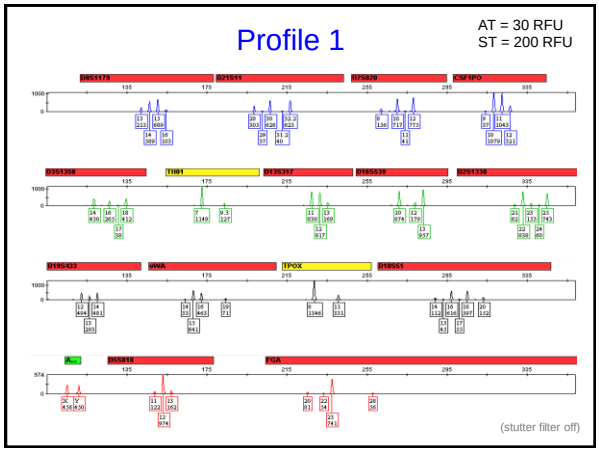
Data from Brad Bannon (Duke lacrosse player defense attorney)

May 12, 2006: DNA Security Report



	Suspect	Evidence	Victim
D13S317	10, 11	[10], 11	11
D16S539	9, 11	9, 10, 11, 12	11, 12
D2S1338	23, 25	INC	19, 22

Data from Brad Bannon (Duke lacrosse player defense attorney)



Limitations of Stochastic Thresholds

- The possibility of allele sharing with a complex mixture containing many contributors may make a stochastic threshold meaningless
- "Enhanced interrogation techniques" to increase sensitivity (e.g., increased PCR cycles) may yield false homozygotes with >1000 RFU
- **New turbo-charged kits with higher sensitivity will need to be carefully evaluated to avoid allele drop-out and false homozygotes**

Summary

- A stochastic threshold (ST) may be established for a specific set of conditions to reflect possibility of allele drop-out, which is essential for a CPE/CPI stats approach
- ST should be re-examined with different conditions (e.g., higher injection, sample desalting, increase in PCR cycles)
- ST will be dependent on the analytical threshold set with a method and impacts the lowest expected peak height ratio
- Assumptions of the number of contributors is key to correct application of ST

Profile 1:

Distinguishable Mixture

Number of Contributors, Assumptions & Genotypes

Charlotte J. Word

ISHI 2011 Mixture Interpretation Workshop:
Using Scientific Analysis

Mixture Interpretation

- Criteria for mixture
- Criteria for determining number of contributors
- Criteria for classifying mixture
  - Distinguishable vs. indistinguishable
- Calculating mixture ratio and use
- Criteria for major/minor contribtors
- Determining genotypes

Minimum Number of Contributors

- Can be determined based on the locus that exhibits the greatest number of allelic peaks
- 2 loci have 4 alleles – maximum number alleles observed
- 2 = minimum number of contributors
- What is the true number of contributors?
  - Must make assumptions

Impact of Assumptions on Interpretation and Statistical Calculations

With assumptions for # of contributor:

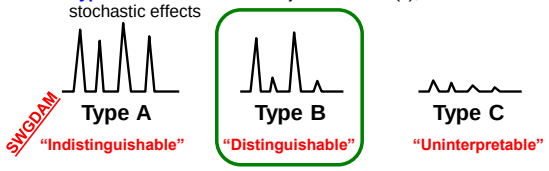
- **May be able to associate alleles into genotypes**
- **May be able to associate genotypes into single-source profiles**
- **Has an effect on the types of statistical calculations possible**

German Mixture Classification Scheme

Schneider *et al*. (2009) *Int. J. Legal Med*. 123: 1-5

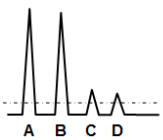
(German Stain Commission, 2006):

- **Type A**: no obvious major contributor, no evidence of stochastic effects
- **Type B**: **clearly distinguishable major and minor contributors**; consistent peak height ratios of ***approximately 4:1*** (major to minor component) for all heterozygous systems, no stochastic effects
- **Type C**: mixtures without major contributor(s), evidence for stochastic effects



Unrestricted vs. Restricted

Use of peak height information to select only certain genotype combinations based on assumptions used



Unrestricted

All combinations of alleles are deemed possible (relative peak height differences are not utilized)

AB + AC + AD + BC + BD + CD

Restricted

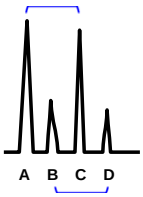
Based on relative peak heights, alleles are paired only where specific combinations of alleles are deemed possible

AB + ~~AC~~ + ~~AD~~ + ~~BC~~ + ~~BD~~ + CD

http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines

Calculation of Major/Minor Ratio

With 2-person mixtures, examine loci with four alleles



$$\frac{PH_A+PH_C}{PH_B+PH_D}$$

Sum of major allele peak heights

Sum of minor allele peak heights

$$\frac{PH_A+PH_C}{PH_A+PH_C+PH_B+PH_D}$$

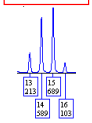
Sum of major allele peak heights

Sum of all allele peak heights at the locus

Major = A,C
Minor = B,D

Mixture Ratio Calculations

D8S1179

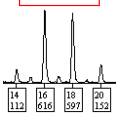


$$\frac{213 + 103}{589 + 689} = 0.25$$

$$\frac{213 + 103}{213 + 589 + 689 + 103} = 0.20$$

1 part + 4 parts

D18S51



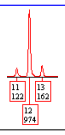
$$\frac{112 + 152}{616 + 597} = 0.22$$

$$\frac{112 + 152}{112 + 616 + 597 + 152} = 0.18$$

1 part + 4.5 parts

Mixture Ratio Calculations

D5S818

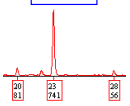


$$\frac{122 + 162}{974} = 0.29$$

$$\frac{122 + 162}{122 + 974 + 162} = 0.23$$

1 part + 3.3 parts

FGA



$$\frac{81 + 56}{741} = 0.19$$

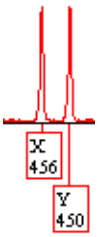
$$\frac{81 + 56}{81 + 741 + 56} = 0.16$$

1 part + 5 parts

Assumptions for Genotype Determination – Profile 1

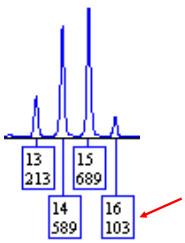
- Only 2 contributors
- 1 major contributor, 1 minor contributor
- Mixture ratio of ~4:1 (major:minor)
- No +4 or -4 stutter peaks, except where stated

Genotypes – Amelogenin



Major	Minor
X Y	?

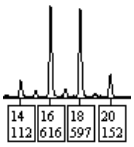
Genotypes – 4 Alleles – D8S1179



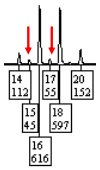
Major	Minor
14,15	13,16

Not concerned 16 is below stochastic threshold since have two minor peaks

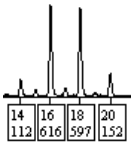
Genotypes – 4 Alleles – D18S51



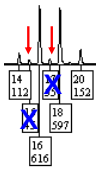
Stutter peaks irrelevant if assume only 2 contributors – can ignore since have 2 minor peaks



Genotypes – 4 Alleles – D18S51

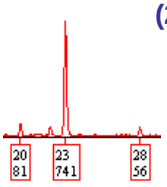


Stutter peaks irrelevant if assume only 2 contributors – can ignore since have 2 minor peaks

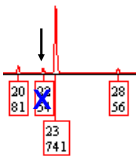


Major	Minor
16,18	14,20

Genotypes – 3 Alleles – FGA (2 minor peaks)

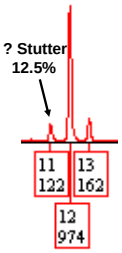


Stutter peak irrelevant if assume only 2 contributors – can ignore since have 2 minor peaks



Major	Minor
23,23	20,28

Genotypes – 3 Alleles – D5S818 (2 minor peaks; Stutter ?)

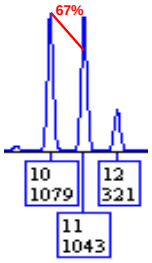


Major	Minor
12,12	11,13

stutter filter = 6.8%
it's an allele

If 11 is possibly stutter: 11, 13; 12, 13; or 13, ?

Genotypes – 3 Alleles – CSF1PO (1 minor allele above ST)

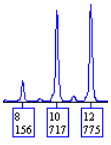


Major	Minor
10,11	12, 12 (10,12; 11,12)

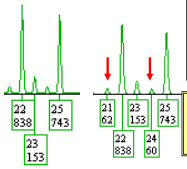
11,12 $\frac{1043-321}{1079}$ = 67%

10,12 $\frac{1079-321}{1043}$ = 73%

Genotypes – 3 Alleles – D7S820 & D2S1338 (1 minor peak, below ST)



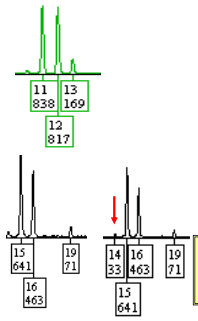
Major	Minor
10,12	8, ?



Major	Minor
22,25	23, ?

If 21 or 23 is possible stutter, accounted for with ?

Genotypes – 3 Alleles – D13S317 & vWA (1 minor peak, below ST)

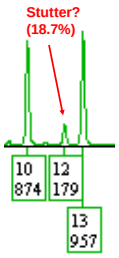


Major	Minor
11,12	13, ?

Major	Minor
15,16	19, ?

If 14 is possible stutter, accounted for with ?

Genotypes – 3 Alleles – D16S539 (1 minor peak, below ST, Stutter?)

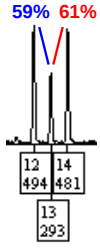


Major	Minor
10,13	12,?

Stutter filter = 10.4%
It's an allele

If stutter filter is 20% → stutter peak
Then no unique minor contributor peak is present
– inconclusive locus for minor contributor

Genotypes – 3 Alleles – D19S433 (high "minor allele")



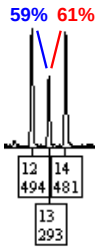
Using peak height ratio, all genotypes possible:

12,12	12,13
13,13	12,14
14,14	13,14

Is there a major:minor here?

Need Major:Minor Criteria to address this

Genotypes – 3 Alleles – D19S433 (high "minor allele")

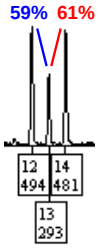


All possible genotype combinations:

12,12 + 13,14	1:1.6
13,13 + 12,14	1:3.3
14,14 + 12,13	1:1.6
12,13 + 12,14	1:1.4
~~12,13 + 13,14~~	~~1:1~~
12,14 + 13,14	1:1.4

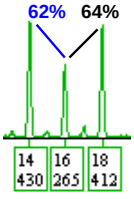
Using MIXTURE RATIO calculations, can eliminate genotype pairs

Genotypes – 3 Alleles – D19S433 (high "minor allele")



Major	Minor	
12,14	13,13	(3.3:1)

Genotypes – 3 Alleles – D3S1358 (high "minor allele")



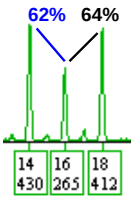
Using peak height ratio, all genotypes possible:

14,14	16,16
14,16	16,18
14,18	18,18

Is there a major:minor here?

Need Major:Minor Criteria to address this

Genotypes – 3 Alleles – D3S1358 (high "minor allele")

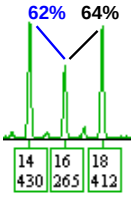


All possible genotype combinations:

14,14 + 16,18	1:1.5
16,16 + 14,18	1:3
18,18 + 14,16	1:1.7
14,16 + 14,18	1:1.3
~~14,16 + 16,18~~	~~1:1~~
14,18 + 16,18	1.3:1

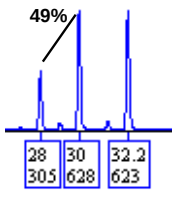
Using MIXTURE RATIO calculations, can eliminate genotype pairs

Genotypes – 3 Alleles – D3S1358 (high "minor allele")



Major	Minor
14,18	16,16 (3:1)

Genotypes – 3 Alleles – D21S11 (high "minor allele")



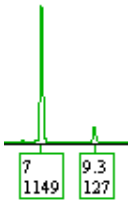
Major	Minor
30,32.2	28,28 (4:1)

Is there a major:minor here?

(30,30 28,32.2
32.2,32.2 28,30)

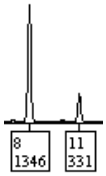
Need Major:Minor Criteria to address this

Genotypes – 2 Alleles – TH01 (unequal peak heights; minor <ST)



Major	Minor
7,7	9.3,?

Genotypes – 2 Alleles – TPOX (unequal peak heights; minor >ST)



Major	Minor
8,8	8,11 or 11,11

PROFILE INTERPRETATION WORKSHEET
IDENTIFILER

PROFILE NAME: Sample 1
ANALYST:
DATE: 10/3/2011
MIXTURE: ■ yes □ no □ unsure

Analytical threshold: 30 RFU
Stutter % used: ABI
Stochastic threshold: 200 RFU
Peak height ratio: 60% (major)
Comments: Mixture ratio calculated

Allele and Locus Assessments — Assuming 2 contributors, 4:1 ratio

ID LOCUS	Alleles above Analytical Threshold	Alleles above Stochastic Threshold	Other peaks to consider	Minimum # of Donors	All alleles likely present? Y/N	Stochastic issues? Elevated stutter? missing alleles?	Degradation Likely?poss. missing alleles?	Yes/ Min If mixture, distinq. profile? Y/N	Additional Comments
D8S1179	13, 14, 16, 16	13, 14, 16	-	2 (4:1)	Y, if 2*	16 +4 stutter?No*	No	14,15 13,16	* no, if >2 donors
D21S11	28,30, 32.2	28,30, 32.2	-	2	Y	No	No	30,32.2 28,28	
D7S820	8, 10, 12	10, 12	-	2	N?	Missing?	No	10,12 8,?	
CSF1PO	10, 11, 12	10, 11, 12	-	2	Y	No	No	10,11 12,12	10,12 11,12
D3S1358	14, 16, 18	14, 16, 18	-	2	Y	No	No	14,18 16,16	
TH01	7, 9.3	7	-	2 (PHR)	N?	Missing?	No	7,7 9.3,?	
D13S317	11, 12, 13	11, 12	-	2	N?	13 +4 stutter? Missing?	No	11,12 13, ?	
D16S539	10, 12, 13	10, 13	-	2	N?	12 stutter? Missing?	No	10,13 12,?	or inconcl
D2S1338	22, 23, 25	22, 25	21, 24?	2	N?	23 +4 stutter? Missing?	No	22,25 23,?	
D19S433	12, 13, 14	12, 13, 14	-	2	Y	13 +4 stutter?	No	12,14 13,13	
vWA	15, 16, 19	15, 16	14	2	N?	Missing?	No	15,16 19,?	
TPOX	8, 11	8, 11	-	2 (PHR)	Y	No*	No	8,8 8,11,11,11	*yes, if >2 donors

Summary

- **Do not need to interpret loci "in order"**
- **Treat loci with similar results the same way**
- **Have criteria in SOP to address possible data and interpretation options (e.g., # of contributors, mixture ratio, possible stutter peaks, major:minor)**
- **Be alert to loci that suggest alternative assumptions could/should be made**

GUIDELINES

Terminology – Mixture Ratio

SWGDAM Interpretation Guidelines glossary:

- **Mixture ratio:** the **relative ratio** of the DNA contributions of multiple individuals to a mixed DNA typing result, **as determined by the use of quantitative peak height information**; may also be expressed as a percentage.

GUIDELINES

Terminology

SWGDAM Guidelines glossary:

- **Major contributor(s):** an individual(s) who can account for the predominance of the DNA in a mixed profile.
- **Minor contributor(s):** an individual(s) who can account for the lesser portion of the DNA in a mixed profile.
- **Distinguishable mixture**: a DNA mixture in which **relative peak height ratios** allow deconvolution of the profiles of **major/minor contributor(s)**.
- **Indistinguishable mixture:** a DNA mixture in which **relative peak height ratios** are insufficient to attribute alleles to individual contributor(s).

GUIDELINES

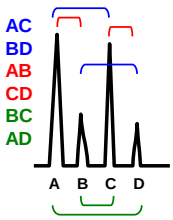
Terminology

SWGDAM Interpretation Guidelines glossary:

- **Conditional**: an interpretation category that **incorporates assumption(s) as to the number of contributors**.
- **Restricted**: referring to a statistical approach conditioned on the **number of contributors** and with consideration of quantitative peak height information and inference of contributor **mixture ratios**; used to limit the genotypic combinations of possible contributors.

Assessment of Major/Minor Contributor

With 2-person mixtures, examine loci with four alleles



Formation of possible genotypes depends on PHRs allowed and the mixture ratio

Major = A,C
Minor = B,D

PROFILE INTERPRETATION WORKSHEET

IDENTIFILER

PROFILE NAME: Sample 1

ANALYST: _____

DATE: 10/3/2011

MIXTURE: ■ yes ☐ no ☐ unsure

Analytical threshold: 30 RFU

Stutter % used: ABI

Stochastic threshold: 200 RFU

Peak height ratio: 60% (major)

Comments: Mixture ratio calculated

Allele and Locus Assessments

Assuming 2 contributors, 4:1 ratio

ID LOCUS	Alleles above Analytical Threshold	Alleles above Stochastic Threshold	Other peaks to consider	Minimum # of Donors	All alleles likely present? Y/N	Stochastic issues? Elevated stutter? missing alleles?	Degradation Likely? poss. missing alleles?	Yes M/m If mixture, distinguishable profile? Y/N	Additional Comments
D8S1179	13, 14, 15, 16	13, 14, 15	-	2 (4:1)	Y, if 2*	16 +4 stutter? No*	No	14,15 13,16	* no, if >2 donors
D21S11	28,30, 32.2	28,30, 32.2	-	2	Y	No	No	30,32.2 28,28	
D7S820	8, 10, 12	10, 12	-	2	N?	Missing?	No	10,12 8,?	
CSF1PO	10, 11, 12	10, 11, 12	-	2	Y	No	No	10,11 12,12	10,12 11,12
D3S1358	14, 16, 18	14, 16, 18	-	2	Y	No	No	14,18 16,16	
TH01	7, 9.3	7	-	2 (PHR)	N?	Missing?	No	7,7 9.3,?	
D13S317	11, 12, 13	11, 12	-	2	N?	13 +4 stutter? Missing?	No	11,12 13, ?	
D16S539	10, 12, 13	10, 13	-	2	N?	12 stutter? Missing?	No	10,13 12,?	or inconcl.
D2S1338	22, 23, 25	22, 25	21, 24?	2	N?	23 +4 stutter? Missing?	No	22,25 23,?	
D19S433	12, 13, 14	12, 13, 14	-	2	Y	13 +4 stutter?	No	12,14 13,13	
vWA	15, 16, 19	15, 16	14	2	N?	Missing?	No	15,16 19,?	
TPOX	8, 11	8, 11	-	2 (PHR)	Y	No*	No	8,8 8,11;11,11	*yes, if >2 donors
D18S51	14, 16, 18, 20	16, 18	15, 17?	3 (4.5:1)*	Y, if 2*	No*	No	16,18 14,20	*no, if >2 donors
Amel	X, Y	X, Y	-	-	Y	No	No	XY ?	
D5S818	11, 12, 13	12	-	2 (3:1)	Y, if 2*	No*	No	12,12 11,13;13,?	*no, if >2 donors
FGA	20, 23, 28	23	22	2 (5:1)	Y, if 2*	No*	No?	23,23 20,28	*no, if >2 donors

PROFILE INTERPRETATION WORKSHEET

PROFILE NAME: _____Sample 1_____

ANALYST: _____

DATE: _____

✓ = Included	✓A# = Included with assumption
X = Excluded	XA# = Excluded with assumption
? = Inconclusive	?A# = Inconclusive with assumption

Assumption 1: Number of contributors = ____2_____ Assumption 3: ___no stutter peaks_____
If distinguishable profiles, # of major contributors = ____1____
# of minor contributors = _____1___

Assumption 2: _____mixture ratio 4:1_____ Assumption 4: _____

Single-source, Deduced single-source, or Mixture with Distinguishable Major and/or Minor Profile Comparison

ID LOCUS	Alleles above Analytical Threshold	Alleles above Stochastic Threshold	Single Source, Major or Minor Contributor Alleles /Genotypes	Comparison Profiles						Additional Comments
				Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	
D8S1179			**14,15**	12,14	13,16	14,15	11,13	13,16		
D21S11			**30,32.2**	28,30	28,28	30,32.2	27,32.2	29,32.2		
D7S820			**10,12**	9,9	8,12	10,12	11,11	8,11		
CSF1PO			**10,11**	10,10*	12,12	10,11	10,11	11,12		*if only 1 major
D3S1358			**14,18**	16,17	16,16	14,18	14,16	15,16		
TH01			**7,7**	6,6	7,9.3	7,7	6,9.3	6,9		
D13S317			**11,12**	11,14	12,13	11,12	11,13	11,11		
D16S539			**10,13**	11,13	12,13	10,13	11,13	11,12		
D2S1338			**22,25**	22,23	23,25	22,25	17,25	19,24		
D19S433			**12,14**	12,14	13,13	12,14	14,15	15,15		
vWA			**15,16**	17,18	15,19	15,16	15,18	18,19		
TPOX			**8,8**	8,8	11,11	8,8	8,11	8,11		
D18S51			**16,18**	14,16	14,20	16,18	13,17	13,14		
Amel			**XY**	XY	XY	XY	XX	XY		
D5S818			**12,12**	12,13	11,13	12,12	12,12	10,12		
FGA			**23,23**	21,22	20,28	23,23	25,26	20,20		

PROFILE INTERPRETATION WORKSHEET

PROFILE NAME: Sample 1

ANALYST: _____

DATE: _____

✓ = Included	✓A# = Included with assumption
X = Excluded	XA# = Excluded with assumption
? = Inconclusive	?A# = Inconclusive with assumption

Assumption 1: Number of contributors = 2 Assumption 3: no stutter peaks

If distinguishable profiles, # of major contributors = 1

# of minor contributors = 1

Assumption 2: mixture ratio 4:1 Assumption 4: _____

Single-source, Deduced single-source, or Mixture with Distinguishable Major and/or Minor Profile Comparison

ID LOCUS	Alleles above Analytical Threshold	Alleles above Stochastic Threshold	Single Source, Major or Minor Contributor Alleles /Genotypes	Comparison Profiles						Additional Comments
				Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	
D8S1179			13,16	12,14	13,16	14,15	11,13	13,16		
D21S11			28,28	28,30	28,28	30,32.2	27,32.2	29,32.2		
D7S820			8,?	9,9	8,12	10,12	11,11	8,11		
CSF1PO			12,12; 10,12; 11,12	10,10	12,12	10,11	10,11	11,12		
D3S1358			16,16	16,17	16,16	14,18	14,16	15,16		
TH01			9.3,?	6,6	7,9.3	7,7	6,9.3	6,9		
D13S317			13,?	11,14	12,13	11,12	11,13	11,11		
D16S539			12,? or inconclusive	11,13	12,13	10,13	11,13	11,12		
D2S1338			23,?	22,23	23,25	22,25	17,25	19,24		
D19S433			13,13	12,14	13,13	12,14	14,15	15,15		
vWA			19,?	17,18	15,19	15,16	15,18	18,19		
TPOX			8,11 or 11,11	8,8*	11,11	8,8*	8,11	8,11		*if only 1 minor
D18S51			14,20	14,16	14,20	16,18	13,17	13,14		
Amel			?	XY	XY	XY	XX	XY		
D5S818			11,13 or 13,?	12,13	11,13	12,12	12,12	10,12		
FGA			20,28	21,22	20,28	23,23	25,26	20,20*		*if only 1 minor

Profile 1: Fundamental Principles of Mixture Interpretation

Distinguishable Mixture? Alternate Interpretation

Robin W. Cotton

NIJ National Institute of Justice

ISHI 2011 Mixture Interpretation Workshop:
Using **S**cientific **A**nalysis

BOSTON UNIVERSITY

NIST

Number of Contributors

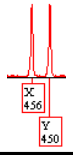
- How safe is the assumption of two contributors?
- Looks OK on the surface, but need to look at alternatives
- Allows analyst to consider alternate explanations and be prepared to discuss

Removing the assumption of 2 contributors:

- Unknown number of contributors
  - One major and one or more minor contributors?
  - OR
  - No assumption of major?
- No reason not to consider both
- How should we define major?
  - At what ratio of contributors is are the peaks sustainably higher?

Lets review the profile using the assumption of at least 3 contributors:

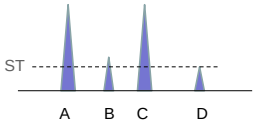
- Major contributor observed at how many loci?
- Minor contributor alleles are sometimes below stochastic threshold.
- If stutter peaks are included could there be more than one minor contributor?
- Is the assumption of only one minor is questionable?
- Amel and (and probably qPCR) are providing no information regarding number of minor contributors



Are we in-between with regard to the minor contributor(s)?

- **Distinguishable mixture**: a DNA mixture in which **relative peak height ratios** allow deconvolution of the profiles of **major/minor contributors**.

Without assumption of 2 contributors-



May be able to make assumption of major contributor and unknown number of minor contributors.

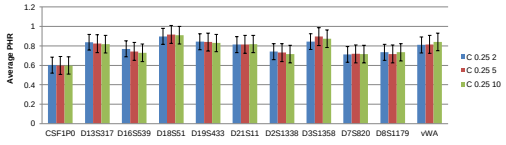
Going with that assumption our possible genotypes are: A,C + B,X + D, X

Conclusions regarding minor types will be limited.

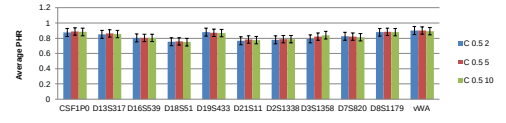
Considering peak height ratio data:

- We know there is deterioration in peak height ratio with decreasing amount of DNA
- This may impact the PHR of the minor contributor(s) in this profile

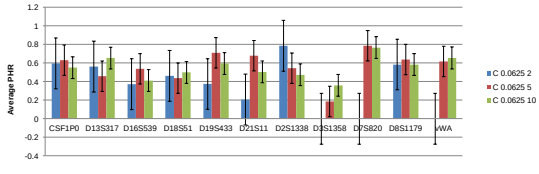
PHR for each locus, 0.25ng, 3 injection times



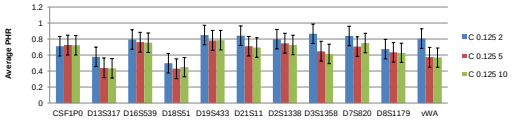
PHR for each locus, 0.5ng, 3 injection times



PHR for each locus, 0.0625ng, 3 injection times



PHR for each locus, 0.125ng, 3 injection times



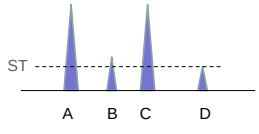
What does our peak height information tell us?

- What does the average tell us?
  - The average declines as the amount of target DNA is reduced.
- How frequently do we see numbers below the average?
  - The range of peak heights observed and the SD increase as the target DNA is reduced.

PROTOCOLS

Calculation of Major/Minor Ratio

Can still approximate proportion knowing that not all minor types are above analytical threshold.



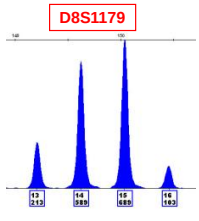
Major = A,C
Observed minor alleles: B & D

$$\frac{PH_A+PH_C}{PH_A+PH_C+PH_B+PH_D}$$

Sum of major allele peak heights

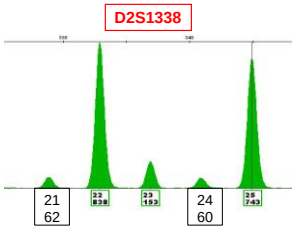
Sum of all allele peak heights at the locus

Estimate of proportion of minor based on observed peaks:



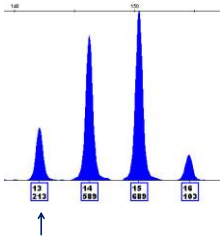
$$\frac{213 + 103}{213 + 589 + 689 + 103} = 0.20$$

Estimate of proportion of minor based on all observed peaks:



$$\frac{62 + 153 + 60}{62 + 838 + 153 + 60 + 743} = 0.15$$

Genotypes

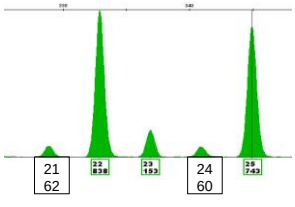


Possibilities:
14,15 Major

13, X
16, X
Minor cannot be defined

RFU = stutter + one or more contributors → Cannot assume that peak is above stochastic threshold

Genotypes

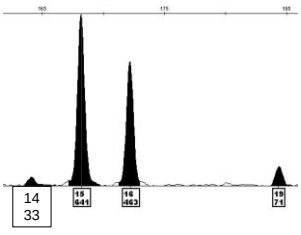


Possibilities:
22, 25 Major

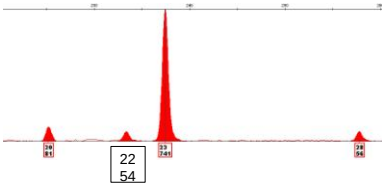
21, X
23, X
24, X
Minor cannot be defined

All three minor peaks are below stochastic threshold.

Genotype Possibilities



Genotype Possibilities



Summary

- Treat loci with similar results the same way
- SOP should address PHR variation with respect to mass of template
- Need basis for making assumptions regarding # of contributors
- Need methods for estimation of mixture ratio
- Must be alert to data that suggests alternative assumptions

Statistics

Michael D. Coble

NIJ National Institute of Justice

2011 Mixture Interpretation Workshop:
Using Scientific Analysis

BOSTON UNIVERSITY
NIST

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

Two Parts to Mixture Interpretation

- Determination of alleles present in the evidence and **deconvolution of mixture components** where possible
  - Many times through comparison to victim and suspect profiles
- **Providing some kind of statistical answer** regarding the weight of the evidence
  - There are multiple approaches and philosophies

Software tools can help with one or both of these…

4. Statistical Analysis of DNA Typing Results

- 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.

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.

Statistical Methods in Medical Research 1993; **2**: 241–262

Forensic inference from genetic markers

B Devlin Department of Epidemiology and Public Health, Yale University School of Medicine

Section 5.1 Exclusion probability

- **Discussion about exclusion probabilities in Paternity cases.**

Two types:

(1) Conditional Exclusion Probability - excluding a random man as a possible father, given the mother-child genotypes for a particular case.

(2) Average Exclusion Probability – excluding a random man as a possible father, given a randomly chosen mother-child pair.

Statistical Methods in Medical Research 1993; **2**: 241–262

Forensic inference from genetic markers

B Devlin Department of Epidemiology and Public Health, Yale University School of Medicine

Section 5.1 Exclusion probability

"The theoretical concept of exclusion probabilities, however, makes no sense within the framework of normal mixture models."

"The interpretation of conditional exclusion probability is obvious, which accounts for its value in the legal arena. Unlike [LR], however, it is not fully efficient."

Statistical Approaches with Mixtures

See Ladd *et al.* (2001) *Croat Med J.* 42:244-246

- **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
- **Calculation of Exclusion Probabilities** - CPE/CPI (RMNE) – The probability that a random person (unrelated individual) would be excluded as a contributor to the observed DNA mixture
- **Calculation of Likelihood Ratio Estimates** – Comparing the probability of observing the mixture data under two (or more) alternative hypotheses; in its simplest form LR = 1/RMP

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

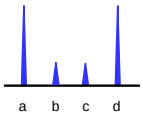
Statistical Approaches with Mixtures

See Ladd *et al.* (2001) *Croat Med J.* 42:244-246

"Exclusionary" Approach	**"Inferred Genotype" Approach**
Random Man Not Excluded (RMNE)	**Random Match Probability (RMP)**
Combined Prob. of Inclusion (CPI)	
Combined Prob. of Exclusion (CPE)	**Likelihood Ratio (LR)**

Statistical Approaches with Mixtures

- **Random Man Not Excluded (CPE/CPI)** - The probability that a random person (unrelated individual) would be included/excluded as a contributor to the observed DNA mixture.



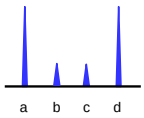
$$CPI = (f(a) + f(b) + f(c) + f(d))^2$$

$$CPI = PI_{M1} \text{ X } PI_{M2} \cdots$$

$$CPE = 1 - CPI$$

Statistical Approaches with Mixtures

- **Random Match Probability (RMP)** – The major and minor components can be successfully separated into individual profiles. A random match probability is calculated on the evidence as if the component was from a single source sample.



$$RMP_{major} = 2pq$$

$$= 2 \text{ x } f(a) \text{ x } f(d)$$

Statistical Approaches with Mixtures

- **Likelihood Ratio** - Comparing the probability of observing the mixture data under two (or more) alternative hypotheses

Basic Math Terms

- When '+' is used, this means 'OR'
- When 'x' is used, this means 'AND'
- Pr. is shorthand for probability

- Therefore…
  - the probability of *a* 'AND' *b* happening together is **Pr(*a* and *b*) = *a* x *b***
  - the probability of *a* 'OR' *b* happening together is **Pr(*a* or *b*) = *a* + *b***

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Conditioning

- **Probabilities are conditional**, which means that the probability of something is based on a hypothesis

- In math terms, conditioning is denoted by a vertical bar
  - Hence, **Pr(a|b)** means 'the probability of *a* ***given*** that *b* is true"

- The probability of an event *a* is dependent upon various assumptions—and these assumptions or hypotheses can change…

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Probability Example – Will It Rain? (1)

Defining the Event and Assumptions/Hypotheses

- Let's suppose that *a* is the probability of an event (e.g., **will it rain?**)
- What is the probability that it will rain in the afternoon – **Pr(*a*)**?

- This probability is dependent upon assumptions
  - We can look at the window in the morning and observe if it is sunny (s) or cloudy (c)
  - Pr(a) **if** it is sunny (s) is less than Pr(a) **if** it is cloudy (c)

- We can write this as **Pr(a|s)** and **Pr(a|c)**
  - Since sunny or cloudy are the only possibilities, Pr(s) + Pr(c) = 1
  - or **Pr(s) = 1 – Pr(c)**

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Probability Example – Will It Rain? (2)

Examining Available Data

- Pr(a|s) and Pr(a|c) can be calculated from data
- How often does it rain in the afternoon when its sunny in the morning?
  - 20 out of 100 observations so **Pr(a|s) = 0.2**
- How often does it rain in the afternoon when it is cloudy in the morning?
  - 80 out of 100 observations so **Pr(a|c) = 0.8**

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Probability Example – Will It Rain? (3)

Formation of the Likelihood Ratio (LR)

- The LR compares two probabilities to find out which of the two probabilities is the most likely

The probability that it will rain in the afternoon when it is cloudy in the morning or **Pr(a|c)** is divided by the probability that it will rain in the afternoon when it is sunny in the morning or **Pr(a|s)**

$$LR = \frac{\Pr(a \mid c)}{\Pr(a \mid s)} = \frac{0.8}{0.2} = 4$$

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Probability Example – Will It Rain? (4)

Explanation of the Likelihood Ratio

$$LR = \frac{\Pr(a \mid c)}{\Pr(a \mid s)} = \frac{0.8}{0.2} = 4$$

- The probability that it will rain is 4 times more likely **if** it is cloudy in the morning than **if** it is sunny in the morning.
- The word **if** is very important here. It must always be used when explaining a likelihood ratio otherwise the explanation could be misleading.

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Likelihood Ratios in Forensic DNA Work

- We evaluate the evidence (*E*) relative to alternative pairs of hypotheses
- Usually these hypotheses are formulated as follows:
  - The probability of the evidence if the crime stain originated with the suspect or Pr(*E*|*S*)
  - The probability of the evidence if the crime stain originated from an unknown, unrelated individual or Pr(*E*|*U*)

$$LR = \frac{\Pr(E \mid S)}{\Pr(E \mid U)}$$

The numerator

The denominator

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

The Likelihood Ratio Must Be Stated Carefully

- The probability of the evidence is *x* times more likely **if** the stain came from the suspect Mr. Smith than **if** it came from an unknown, unrelated individual.
- It is not appropriate to say: "The probability that the stain came from Mr. Smith." because we must always include the conditioning statement – i.e., **always make the hypothesis clear in the statement.**
- Always use the word '**if**' when using a likelihood ratio to avoid this trap

Slide information from Peter Gill (ISFG 2007 workshop, Copenhagen, August 20-21, 2007)

Likelihood Ratio (LR)

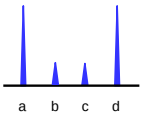
- Provides ability to express and evaluate both the prosecution hypothesis, H_p (the suspect is the perpetrator) and the defense hypothesis, H_d (an unknown individual with a matching profile is the perpetrator)

$$LR = \frac{H_p}{H_d}$$

- **The numerator, H_p, is usually 1** – since in theory the prosecution would only prosecute the suspect if they are 100% certain he/she is the perpetrator
- The denominator, $\mathbf{H_d}$, is typically the profile frequency in a particular population (based on individual allele frequencies and assuming HWE) – i.e., **the random match probability**

Statistical Approaches with Mixtures

- **Likelihood Ratio** - Comparing the probability of observing the mixture data under two (or more) alternative hypotheses; in its simplest form LR = 1/RMP



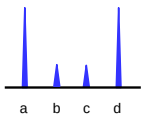
$$\frac{P(E|H_1)}{P(E|H_2)}$$

E = Evidence
H_1 = Prosecutor's Hypothesis (the suspect did it) = 1
H_2 = Defense Hypothesis (the suspect is an unknown, random person)

$$= \frac{1}{P(E|H_2)} = \frac{1}{2pq} = 1/RMP$$

Statistical Approaches with Mixtures

- **Unrestricted Likelihood Ratio** - All combinations of alleles are deemed possible (relative peak height differences are not utilized).



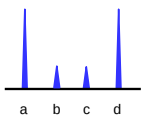
Possible Combinations

AB BC
AC BD
AD CD

= (AB + AC + AD + BC + BD + CD)

Statistical Approaches with Mixtures

- **Restricted Likelihood Ratio** - Based on relative peak heights, alleles are paired only where specific combinations of alleles are deemed possible



(without victim subtraction)

Possible Combinations

~~AB~~ BC
~~AC~~ ~~BD~~
AD ~~CD~~

= (AD + BC)
2pq + 2pq

Advantages and Disadvantages RMNE and LR

RMNE (CPE/CPI)

Advantages
- Does not require an assumption of the number of contributors to a mixture
- Easier to explain in court

Disadvantages
- Weaker use of the available information (robs the evidence of its true probative power because this approach does not consider the suspect's genotype)
- LR approaches are developed within a consistent logical framework

Likelihood Ratios (LR)

Advantages
- Enables full use of the data including different suspects

Disadvantages
- More difficult to calculate (software programs can assist)
- More difficult to present in court

Summarized from John Buckleton, *Forensic DNA Evidence Interpretation*, p. 223 Buckleton and Curran (2008) *FSI-G* 343-348.

FSI GENETICS

Forensic Science International: Genetics 2 (2008) 343–348

A discussion of the merits of random man not excluded and likelihood ratios

John Buckleton [a,*], James Curran [b]

[a] *ESR, PB 92021, Auckland, New Zealand*
[b] *Department of Statistics, University of Auckland, PB 92019, Auckland, New Zealand*

Received 15 January 2008; received in revised form 29 April 2008; accepted 1 May 2008

We conclude that the two matters that appear to have real force are:
(1) LRs are more difficult to present in court and
(2) the RMNE statistic wastes information that should be utilised.

Curran and Buckleton (2010)

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01446.x
Available online at: interscience.wiley.com

PAPER
CRIMINALISTICS; GENERAL

James M. Curran,[1] M.Sc.(Hons.), Ph.D. and John Buckleton,[2] Ph.D.

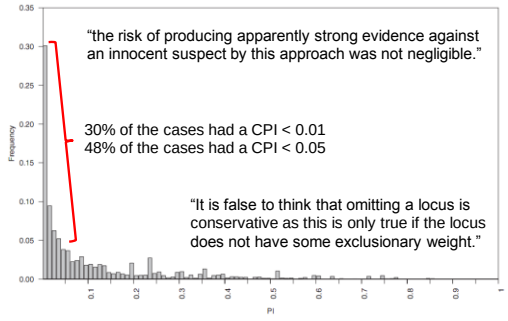
Inclusion Probabilities and Dropout

Created 1000 Two-person Mixtures (Budowle *et al.*1999 AfAm freq.).

Created 10,000 "third person" genotypes.

Compared "third person" to mixture data, calculated PI for included loci, ignored discordant alleles.

Curran and Buckleton (2010)

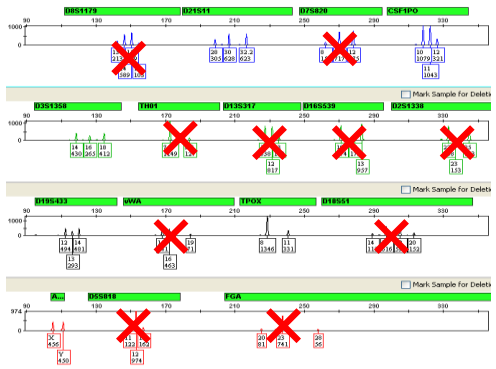


If CPI/CPE Stats are Used

Since exclusionary statistics cannot adjust for the possibility of dropout, and does not take the number of contributors into account, any loci where alleles are below stochastic levels cannot be used in the CPI statistic.

CPI/CPE

Profile 1: ID_2_SCD_NG0.5_R4,1_A1_V1.2



If CPI/CPE Stats are Used

Can use	Cannot use	
D21	D8	D2
CSF	D7	vWA
D3	TH01	D18
D19	D13	D5
TPOX	D16	FGA

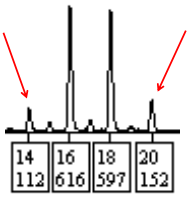
If CPI/CPE Stats are Used

- CPI statistics using FBI Caucasian Frequencies
- 1 in 71 Caucasians included
- 98.59% Caucasians excluded

If RMP Stats are Used

- Since there is an assumption to the number of contributors, it is possible to use data that falls below the ST.

RMP - D18S51



If Assume 2 Contributors….

Major	Minor
16,18	14,20

RMP_{minor} = 2pq

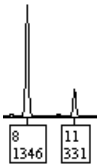
= 2 x f(14) x f(20)

= 2 x (0.1735) x (0.0255)

= 0.00884 or 1 in 113

(LR = 113)

RMP - TPOX



If Assume 2 Contributors….

Major	Minor
8,8	11,8 ***OR*** 11,11

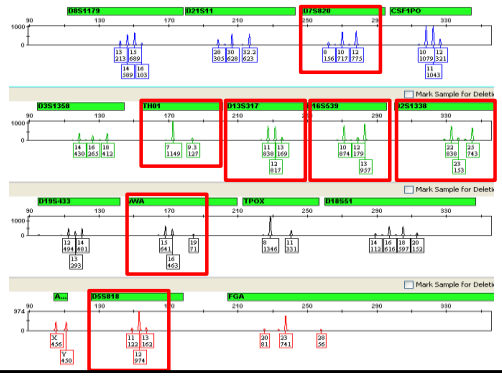
RMP = 8,11 + 11,11

RMP = 2pq + $(q^2 + q(1-q)\theta)$

RMP = 2(0.5443)(0.2537) +
$(0.2537)^2$ + (0.2537)(0.7463)(0.01)
= 0.3424 or 1 in 2.9

RMP/LR

Profile 1: ID_2_SCD_NG0.5_R4,1_A1_V1.2

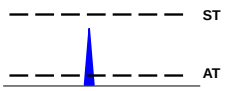


If RMP/LR Stats are Used

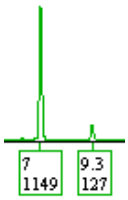
Can use	Loci with potential D-out	
D8	D7	D2
D21	TH01	vWA
D18	D13	D5
D3	D16	
D19		
TPOX		
FGA		
CSF		

The "2p" Rule

- The "2p" rule can be used to statistically account for zygosity ambiguity – i.e. is this single peak below the stochastic threshold the result of a homozygous genotype or the result of a heterozygous genotype with allele drop-out of the sister allele?



Profile 1 - TH01



Major – 7, 7

Possible Minor Contributors

7, 9.3 (2pq)

9.3, 9.3 p^2

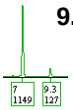
9.3, ? 2p (or $p^2 + 2p(1 - p)$)

Profile 1 - TH01 (LR)

$$\frac{P(E|H_1)}{P(E|H_2)} = \frac{V+S}{V+U} = \frac{f_7^2 + f_7(1-f_7)\theta + 1}{f_7^2 + f_7(1-f_7)\theta + p^2 + 2p(1-p)}$$

V = 7, 7
U = 7, 9.3
9.3, 9.3
9.3, ?

$$= \frac{1}{f_{9.3}^2 + 2f_{9.3}(1-f_{9.3})}$$



$f_{9.3} = 0.3054$

$= 1 / 0.5175 = \boxed{1.93}$

Profile 1 - TH01 (LR)

$$\frac{P(E|H_1)}{P(E|H_2)} = \frac{V+S}{V+U} = \frac{1}{p^2 + p(1-p)\theta + 2pq}$$

V = 7, 7
U = 7, 9.3
9.3, 9.3

$$= \frac{1}{f_{9.3}^2 + f_{9.3}(1-f_{9.3})\theta + 2f_{9.3}f_7}$$

Let ST = 125 RFU

$f_{9.3} = 0.3054$
$f_7 = 0.1724$

$= 1 / 0.2007 = \boxed{4.98}$

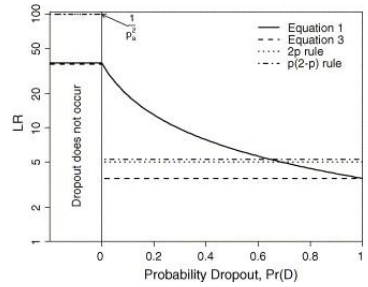
The "2p" Rule

- "This rule arose during the VNTR era. At that time many smaller alleles "ran off the end of the gel" and were not visualised."

  - Buckleton and Triggs (2006)

"Is the 2*p* rule always conservative?"

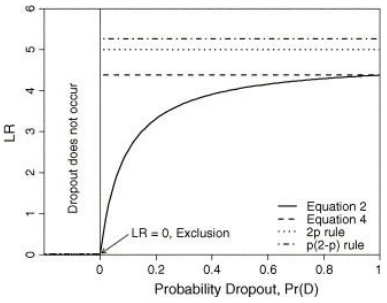
The “2p” Rule



Stain = aa

Suspect = aa

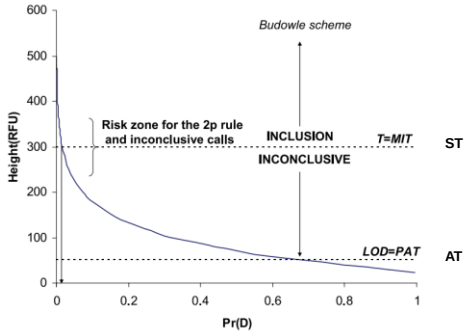
The “2p” Rule



Stain = aa

Suspect = ab

Gill and Buckleton (2010)



How to Handle Low Level Peaks Where Dropout May be Present

- Ignore the locus
- Re-amplify – then what?
  - Choose the least or most informative profile **(replicate shopper)**
  - Use the consensus approach (alleles are authentic if observed in both replicates)
  - Use a composite approach
  - Use a Bayesian model
  - Use a continuous model (such as TrueAllele)

Bright, Gill and Buckleton
Composite profiles in DNA analysis
FSI-G in press

Strategies to statistically evaluate low-level DNA evidence

- ~~The Binary LR Method (Alleles are present or not).~~
- The semi-Continuous Method (Gill *et al.* 2000 – Interpretation of LCN data).
- The fully-Continuous Method (Perlin *et al.* 2011 – True Allele).

The Drop-out Model

The interpretation of low level DNA mixtures

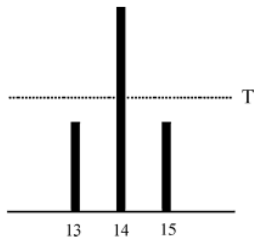
Hannah Kelly [a,*], Jo-Anne Bright [a], James Curran [b], John Buckleton [a]

[a] ESR, PB 92021 Auckland, New Zealand
[b] Department of Statistics, University of Auckland, PB 92019 Auckland, New Zealand

FSI-Genetics, in press

Article included in workshop handout

First – Convert Peaks to Alleles

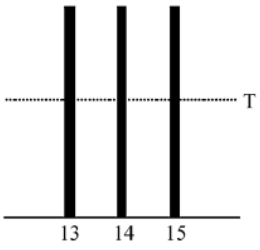


Assume 2 Contributors
3 peaks – 4 alleles

Allelic Vector
13
14
14
15

13,14,14,15

Ambiguity in Determining Vectors



Assume 2 Contributors

Allelic Vectors
13, 13, 14, 15
13, 14, 14, 15
13, 14, 15, 15

3 possibilities

Permutations

- The number of permutations is the number of ways that the alleles can be arranged as pairs.

Permutations

- An easier way to compute using factorials.

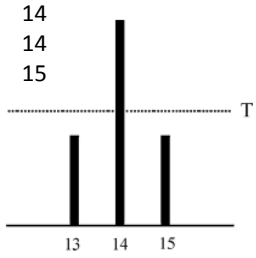
$$\binom{n}{m_1, m_2, \ldots m_l} = \frac{n!}{m_1!m_2!, \ldots, m_l!}$$

n = total number of alleles at the locus.
m = number of times each allele is seen.

Determine the Permutations for this example

Allelic Vectors

13
14
14
15

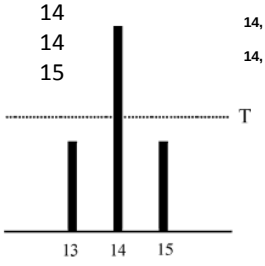


$$= \frac{4!}{1!2!1!} = \frac{4x3x\cancel{2}x\cancel{1}}{1x\cancel{2}x\cancel{1}} = 12$$

Let's Prove It!

Allelic Vectors

13
14
14
15



13, 14 and 14, 15 = 2ab x 2bc = $4ab^2c$

13, 15 and 14, 14 = 2ac x b^2 = $2ab^2c$

14, 15 and 13, 14 = 2bc x 2ab = $4ab^2c$

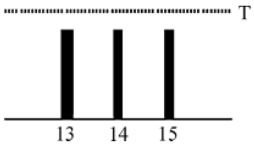
14, 14 and 13, 15 = b^2 x 2bc = $2ab^2c$

= $12ab^2c$

= 12

Assign Allele Designations

- Use "F" as a placeholder to consider alleles that may have dropout.

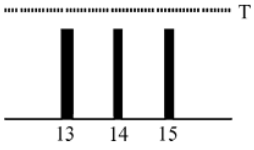


Assume 2 Contributors
3 peaks – 3 alleles

Allelic Vector
13,14,15,F

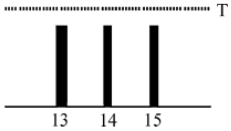
Assign Probability using the F-model

- Calculate the number of permutations using "F" as a placeholder and then drop it from the equation.



Assign Probability using the F-model

$$Pr(13,14,15,F|X) = \frac{4!}{1!1!1!1!} Pr(13,14,15,F|X)$$



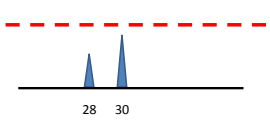
$$= 24Pr(13,14,15|X)$$

Apply the Sampling Formula (Balding and Nichols 1994)

$$\frac{x\,\theta + (1-\theta)pA}{1 + (n-1)\,\theta}$$

x = value calculated from the F-model.
p_a = frequency of the "a" allele.
Θ = coancestry coefficient (F_{ST}).
n = number of alleles.

A Worked Example



POI = 28, 30

2 peaks – 4 alleles

D21
Assume 2 contributors
Allele 28 = 107 RFU
Allele 30 = 198 RFU
ST = 200 RFU

Allelic Vector
28,30,F,F

Permutations and Probability

$$\Pr(28,30,F,F \mid 28,30) = \frac{4!}{1!1!2!}\ \Pr(28,30,F,F \mid 28,20)$$

$$= 12\Pr(28,30 \mid 28,30)$$

Apply the Sampling Formula (Balding and Nichols 1994)

$$\Pr(A|X) = \frac{x\theta + (1-\theta)\,p_a}{1+(n-1)\theta}$$

Pr(E|Hp) =1
Pr(E|Hd) =12Pr(28,30|28,30)

$$\frac{12(\theta(1-\theta)\,p_{28})(\theta+(1-\theta)\,p_{30})}{(1+\theta)(1+2\theta)}$$

LR = 1.86

Kelly *et al. (in press)*

- Other models including the "Q" method and the Unconstrained Combinatorial "UC" method (no peak height info).
- The UC method overestimates the LR and is not appropriate. The "Q" model performs better than the "F" model, but is more mathematically intense…

The "Q" Model for D21 (28,30)

Allelic vector (28,30)

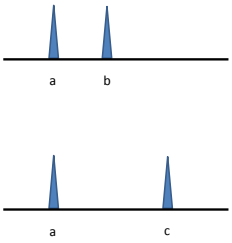
$\Pr(E|Hp) = 1$

$$4\Pr(28,28,28,30|28,30) + 6\Pr(28,28,30,30|28,30) + 4\Pr(28,30,30,30|28,30) + 12\Pr(28,28,30,Q|28,30) + 12\Pr(28,30,Q,Q|28,30)$$

$$\Pr(E|Hd) = 2\Pr(28,30|28,30) \times \begin{bmatrix} 6 - 6\Pr(28|28,28,30,30) - 6\Pr(30|28,28,30,30) + 2\Pr(28,28|28,28,30,30) \\ +2\Pr(30,30|28,28,30,30) \\ +3\Pr(28,30|28,28,30,30) \end{bmatrix}$$

$$\frac{2(\theta(1-\theta)p_{28})(\theta+(1-\theta)p_{30})}{(1+\theta)(1+2\theta)} \times \left[6 - \frac{6(2\theta+(1-\theta)p_{28})}{(1+3\theta)} - \frac{6(2\theta+(1-\theta)p_{30})}{(1+3\theta)} + \frac{2(2\theta+(1-\theta)p_{28})(3\theta+(1-\theta)p_{28})}{(1+3\theta)(1+4\theta)} + \frac{2(2\theta+(1-\theta)p_{30})(3\theta(1-\theta)p_{30})}{(1+3\theta)(1+4\theta)} + \frac{3(2\theta+(1-\theta)p_{28})(2\theta+(1-\theta)p_{30})}{(1+3\theta)(1+4\theta)} \right]$$

Including the Pr(Dropout) in the LR



Suspect

Stain

H_P

Allele b has dropped out

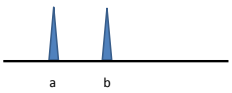
Allele c is a contaminant

H_D

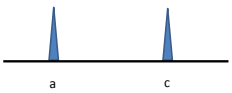
The stain is from an Unknown person

Some Assumptions

Suspect



Stain



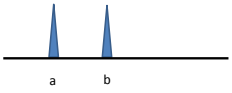
D = Pr(dropout) = 0.5

C_c = Contamination of allele c = 0.03

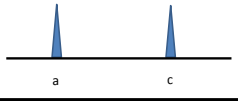
Let $f_a = f_b = f_c = 0.1$

Possible Random Men	Pr(genotypes)	Pr(E | genotypes)	Multiply Columns	Denominator (H_D)	Numerator (H_P)
a, b	$2f_af_b$	$\bar{D}\ D\ C_c$	$2f_af_b\bar{D}DC_c$	0.00015	0.00075
a, c	$2f_af_c$	$\bar{D}\ \bar{D}\ \bar{C}_c = \bar{D}^2\bar{C}_c$	$2f_af_c\bar{D}^2\bar{C}_c$	0.0049	
				0.00505	

Suspect



Stain



$$LR = \frac{0.00075}{0.00505}$$

$$= 0.1485$$

This favors the H_D

Summary

- The laboratory must perform statistical analysis in support of any inclusion (Guidelines 4.1).
- There are advantages and disadvantages to both RMNE and LR stats. As a general rule, RMNE does not take full advantage of all the data. LRs are more difficult to explain, but can incorporate drop out, drop in, contamination, stutter, etc…

Summary

- The issue of how to handle low level data is not going away – samples examined ("touch" DNA) and more sensitive kits from the manufacturers.
- We've spent a lot of time and money on validating new extraction methods, kits, instruments – there is a need to improve interpretation and statistical analyses of the data.

Profile 1:

Inclusion/Exclusion
Comparison & Report Wording

Charlotte J. Word

NIJ National Institute of Justice

ISHI 2011 Mixture Interpretation Workshop:
Using Scientific Analysis

BOSTON UNIVERSITY
NIST

Report Wording – Conclusions

Before Comparison, we know:

1. Mixture obtained with STR kit
2. Know loci with results/partial/no results
3. At least two contributors
4. At least one contributor is male

Conclusion:
The profile contains a mixture of DNA from at least two contributors at least one of whom is a male.

Starting the Report

- Enough information to start conclusions of report and calculate statistics even before doing the comparison
- Can generate Table of alleles and genotypes included (worksheet)

Report Wording – Assumptions

Before Comparison, we know:

5a. If no assumptions are made regarding the number of contributors and we are only interpreting the loci with all peaks above stochastic threshold (CSF1PO, D19, D3, D21, TPOX, D18), need to state that only those loci are being interpreted and why the others are inconclusive. (i.e., calculating CPE/CPI)

Report Wording – Assumptions

Before Comparison, we know:

5b. If assuming there is a single major contributor (based on lab SOP criteria) and using no information from secondary contributors, need to state that is what is being done and what loci are being used.

Report Wording - Assumptions

Before Comparison, we know:

5c. If assume only two contributors:

1. Assume mixture ratio of 4:1
2. Assume one major contributor
3. Assume one minor contributor
4. No filtered stutter peaks considered

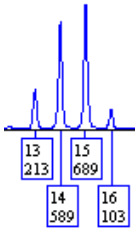
Report Wording – Conclusions

If the data are interpreted with the assumption that there are only two contributors, then the mixture contains DNA from one major contributor who is a male and from one minor contributor whose gender is unknown. The results were interpreted using a mixture ratio of 4:1 for the major:minor contributors.

Comparisons

- Compare profiles from known standards to genotypes determined previously using the assumptions and recorded on the worksheet (i.e., my previous talk)

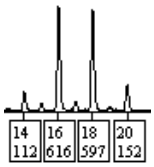
Genotypes – 4 Alleles – D8S1179



Major	Minor
14,15	13,16

1	12,14	Exc	Exc
2	13,16	Exc	Included
3	14,15	Included	Exc
4	11,13	Exc	Exc
5	13,16	Exc	Included

Genotypes – 4 Alleles – D18S51

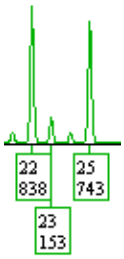


Major	Minor
16,18	14,20

1	14,16	Exc	Exc*
2	14,20	Exc	Included
3	16,18	Included	Exc
4	13,17	Exc	Exc
5	13,14	Exc	Exc

*if only two contributors

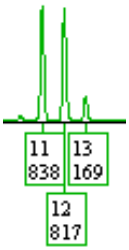
Genotypes – 3 Alleles – D2S1338 (1 minor peak, below ST)



Major	Minor
22,25	23, ?

1	22,23	Exc	Included
2	23,25	Exc	Included
3	22,25	Included	Exc
4	17,25	Exc	Exc
5	19,24	Exc	Exc

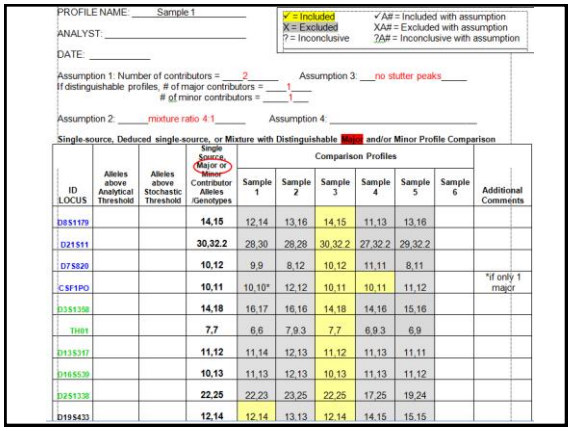
Genotypes – 3 Alleles – D13S317 (1 minor peak, below ST)



Major	Minor
11,12	13, ?

1	11,14	Exc	Exc
2	12,13	Exc	Included
3	11,12	Included	Exc
4	11,13	Exc	Included
5	11,11	Exc*	Exc*

* If only two contributors



PROFILE NAME: Sample 1
ANALYST:
DATE:

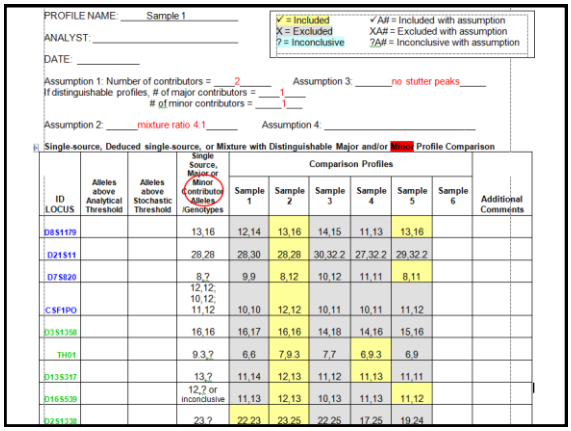
✓ = Included
X = Excluded
? = Inconclusive
✓A# = Included with assumption
XA# = Excluded with assumption
?A# = Inconclusive with assumption

Assumption 1: Number of contributors = 2 Assumption 3: no stutter peaks
If distinguishable profiles, # of major contributors = 1
# of minor contributors = 1

Assumption 2: mixture ratio 4:1 Assumption 4:

Single-source, Deduced single-source, or Mixture with Distinguishable Major and/or Minor Profile Comparison

ID LOCUS	Alleles above Analytical Threshold	Alleles above Stochastic Threshold	Single Source, Major or Minor Contributor Alleles /Genotypes	Comparison Profiles: Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Additional Comments
D8S1179			14,15	12,14	13,16	14,15	11,13	13,16		
D21S11			30,32.2	28,30	28,28	30,32.2	27,32.2	29,32.2		
D7S820			10,12	9,9	8,12	10,12	11,11	8,11		
CSF1PO			10,11	10,10*	12,12	10,11	10,11	11,12		*if only 1 major
D3S1358			14,18	16,17	16,16	14,18	14,16	15,16		
TH01			7,7	6,6	7,9.3	7,7	6,9.3	6,9		
D13S317			11,12	11,14	12,13	11,12	11,13	11,11		
D16S539			10,13	11,13	12,13	10,13	11,13	11,12		
D2S1338			22,25	22,23	23,25	22,25	17,25	19,24		
D19S433			12,14	12,14	13,13	12,14	14,15	15,15		



PROFILE NAME: Sample 1
ANALYST:
DATE:

✓ = Included
X = Excluded
? = Inconclusive
✓A# = Included with assumption
XA# = Excluded with assumption
?A# = Inconclusive with assumption

Assumption 1: Number of contributors = 2 Assumption 3: no stutter peaks
If distinguishable profiles, # of major contributors = 1
# of minor contributors = 1

Assumption 2: mixture ratio 4:1 Assumption 4:

Single-source, Deduced single-source, or Mixture with Distinguishable Major and/or Minor Profile Comparison

ID LOCUS	Alleles above Analytical Threshold	Alleles above Stochastic Threshold	Single Source, Major or Minor Contributor Alleles /Genotypes	Comparison Profiles: Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Additional Comments
D8S1179			13,16	12,14	13,16	14,15	11,13	13,16		
D21S11			28,28	28,30	28,28	30,32.2	27,32.2	29,32.2		
D7S820			8,?2	9,9	8,12	10,12	11,11	8,11		
CSF1PO			12,12; 10,12; 11,12	10,10	12,12	10,11	10,11	11,12		
D3S1358			16,16	16,17	16,16	14,18	14,16	15,16		
TH01			9.3,?2	6,6	7,9.3	7,7	6,9.3	6,9		
D13S317			13,?2	11,14	12,13	11,12	11,13	11,11		
D16S539			12,?2 or inconclusive	11,13	12,13	10,13	11,13	11,12		
D2S1338			23,?2	22,23	23,25	22,25	17,25	19,24		

Report Wording – Conclusions

Using the stated assumuptions (in 5c):

Person 1, person 4, and person 5, are excluded as (the major and minor) contributors to the mixture.

Person 3 is included as the major contributor. Statistical frequencies….

Person 4 is included as the minor contributor. Statistical frequencies…

Inclusions

Remember:

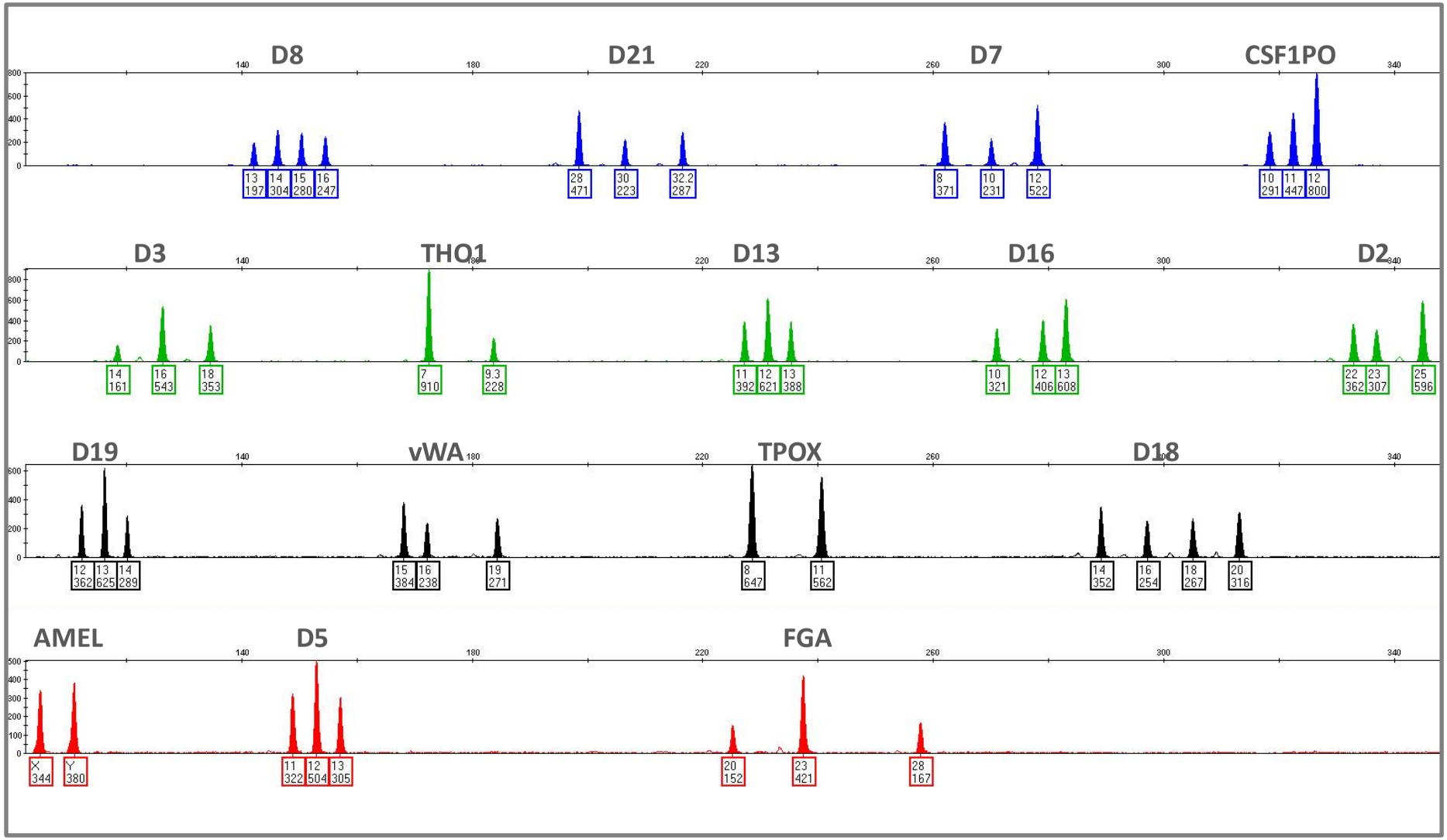
Inclusion
Included
Cannot be excluded

All have the same meaning → need statistical frequencies reported with the statement

Additional Report Wording

- Additional assumptions that should be considered and reported
  - More than 2 contributors (>1 major? >1 minor?)
  - Filtered stutter peaks that should be included
  - Deducing/Subtracting out a Profile: Assuming one contributor was Person X, then….
- Any locus that was inconclusive and why
  - Partial profile, locus drop-out

Profile 2: Identifiler, 30 rfu, w/stutter filter



Profile 2: Complex Mixtures

2-Person Indistinguishable Mixture

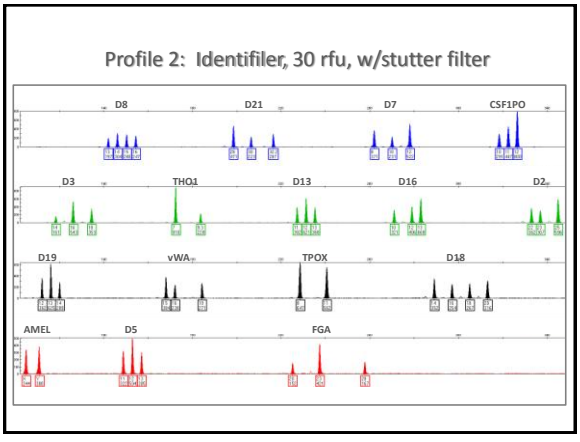
Robin W. Cotton

NIJ National Institute of Justice

2011 Mixture Interpretation Workshop:
Using Scientific Analysis

BOSTON UNIVERSITY

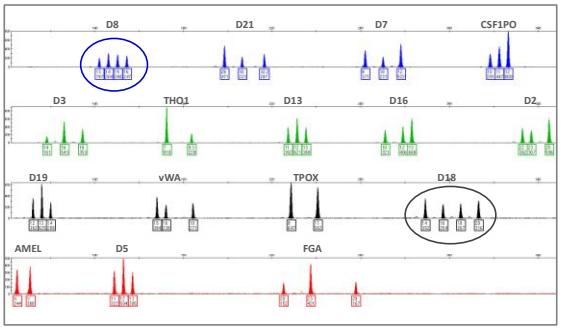
NIST



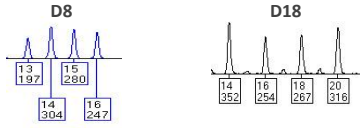
General Assessment

- Maximum # of alleles at a locus = 4
- No evidence of degradation in the profile
- A few peaks are above analytical threshold, but below stochastic threshold
- There are no relatively small (minor) peaks except those in stutter positions
- No indication of a third contributor
  - (consider the points shown in grey above)

Profile 2: Four Allele Loci



Beginning with the four allele loci and assuming two contributors can we estimate the ratio of the two contributors?



Possible genotype combinations

13,14 + 15,16
13,15 + 14,16
13,16 + 14,15

&

14,16 + 18,20
14,18 + 16,20
14,20 + 16,18

Example: (rfu 13+ rfu 14) ÷ total rfu = proportion of contributor 1
(197 + 304) ÷ 1028 = 0.49

Proportion of contributor 1 to total rfu

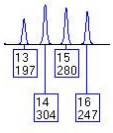
0.49
0.46
or
0.43

≈

0.49
0.48
or
0.44

Consider peak height ratio at the 4 allele loci:

D8



Possible genotype pairs

1) 13,14 + 15,16
2) 13,15 + 14,16
3) 13,16 + 14,15

Peak height ratios

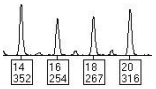
1) 0.65 & 0.88
2) 0.70 & 0.81
3) 0.80 & 0.92

Based on the allowed peak height ratio, all three pairs of genotypes are possible.

Cannot restrict the possible combinations of genotypes

Consider peak height ratio at the 4 allele loci:

D18



possible pairs

1) 14,16 + 18,20
2) 14,18 + 16,20
3) 14,20 + 16,18

Peak height ratios

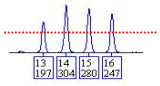
1) 0.72 & 0.84
2) 0.76 & 0.80
3) 0.90 & 0.95

Based on the allowed peak height ratio, all three pairs of genotypes are possible.

Cannot restrict the possible combinations of genotypes

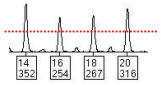
Combined Probability of Inclusion

D8



ST = 150 RFU

D18



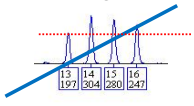
$PI = (p + q + r + s)^2$
$PI = (\text{freq } 13 + \text{freq } 14 + \text{freq } 15 + \text{freq } 16)$
$PI = (0.305 + 0.166 + 0.114 + 0.031)^2$
$PI = (0.616)^2$
PI = 0.379

$PI = (p + q + r + s)^2$
$PI = (\text{freq } 14 + \text{freq } 16 + \text{freq } 18 + \text{freq } 20)$
$PI = (0.137 + 0.139 + 0.076 + 0.022)^2$
$PI = (0.374)^2$
PI = 0.140

CPI = (0.379)(0.140) = 0.053
CPE = 1 – CPI = 0.947

Combined Probability of Inclusion

D8

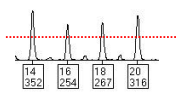


ST = 200 RFU

OR

ST = 250 RFU

D18



PI = 1.0

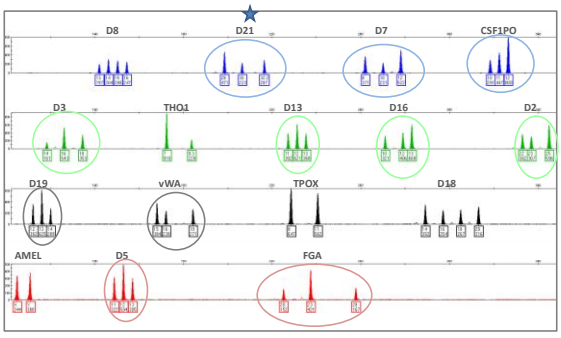
$PI = (p + q + r + s)^2$
$PI = (\text{freq } 14 + \text{freq } 16 + \text{freq } 18 + \text{freq } 20)$
$PI = (0.137 + 0.139 + 0.076 + 0.022)^2$
$PI = (0.374)^2$
PI = 0.140

CPI = (1.0)(0.140) = 0.140
CPE = 1 – CPI = 0.86

What do we know so far?

- If two contributors, the proportion of contributor 1 ≈ 0.46 and contributor 2 ≈ 0.54
- Using a stochastic threshold of __ for the profile:
  - 150 rfu, there are **no** peaks below the threshold
  - 200 rfu, there are 4 peaks below the threshold
  - 250 rfu, there are 10 peaks below the threshold

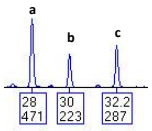
Profile 2: Three Allele Loci



How will we analyze?

- Assume two contributors at a ratio of ≈ 1:1.
- List possible contributing genotypes.
- List possible pairs of contributing genotypes.
- Calculate the resulting peak height ratios.
- Use ½ rfu in this calculation when a peak would be shared between the two contributing genotypes.
  - (use ½ for ease of calculation today, could use range of proportions based on PHR data)
  - For this exercise we are rounding the proportion of 0.46 and 0.54 to 0.5 and 0.5 ≈ 1:1.

D21



Possible Genotypes

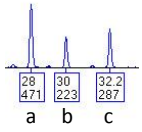
a, a
a, b
a, c
b, b
b, c
c, c

Possible Genotype Combinations

a, a + b, c	28, 28 + 30, 32.2
a, b + a, c	28, 30 + 28, 32.2
b, b + a, c	30, 30 + 28, 32.2
a, b + b, c	28, 30 + 30, 32.2
c, c + a, b	32.2, 32.2 + 28, 30
a, c + b, c	28, 32.2 + 30, 32.2

Can we rule out any of these combinations by looking at peak height ratio and incorporating the estimated ratio of contributors using the two contributor assumption?

D21



Possible Genotype Combinations

28, 28 + 30, 32.2 •
28, 30 + 28, 32.2 •
30, 30 + 28, 32.2 •
28, 30 + 30, 32.2 •
32.2, 32.2 + 28, 30 •
28, 32.2 + 30, 32.2 •

Peak Height Ratios ….

30, 32.2 = 0.78
½(28), 30 = 0.95 & ½(28), 32.2 = 0.82

28, 32.2 = 0.61
28, ½(30) = 0.47 & ½(30), 32.2 = 0.39

28, 30 = 0.47
28, ½(32.2) = 0.31 & 30, ½(32.2) = 0.65

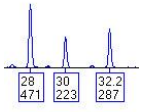
28, 28 + 30, 32.2

28, 30 + 28, 32.2

~~30, 30 + 28, 32.2~~

Possible Genotype Combinations

Remembering the assumption of two people:



28, 28 + 30, 32.2

28, 30 + 28, 32.2

30, 30 + 28, 32.2 ?

Possible Genotype Combinations

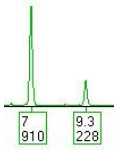
Included or excluded???

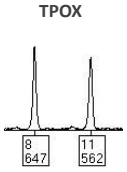
28, 28	28, other*
28,30	
28, 32.2	
30,30	30, other*
30,32.2	
32.2,32.2	32.2, other*

Under the assumption of two contributors and using the calculated contributor ratio, the 32.2, 32.2 genotype is excluded.

*Other = not: 28, 30, or 32.2

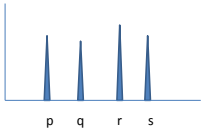
Two allele loci:





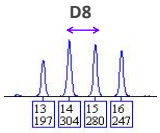
Deducing a Second Contributor

Four allele locus, assume two contributors:

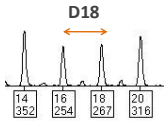


4 alleles are observed:
Known is heterozygous ⟶ deduce 2nd person
Known is: p, q ⟶ 2nd person is: r, s

Four Allele Loci



Known : 14, 15



Known : 16, 18

Can calculate the proportion of this contributor

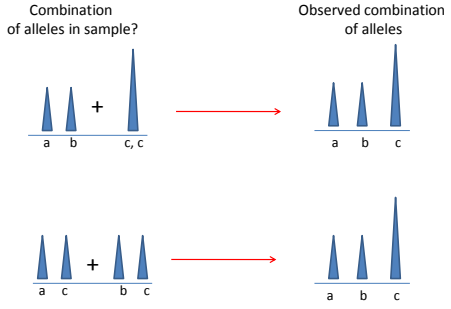
Known proportion = 0.57

Deduced 2nd contributor: 13, 16

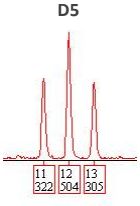
Known proportion = 0.44

Deduced 2nd contributor: 14, 20

Three allele loci, assume 2 contributors and approximately 1:1 mixture



Three allele loci: known homozygous



Deduced 2nd contributor = 11, 13

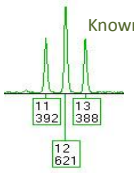
Peak height ratio (11, 13) = 0.94 &
Contributor ratio = 0.45

Contributor ratio ~ 1:1

Known : 12, 12

Easy!!

D 13 Three allele loci: Known heterozygous



Known : 11, 12

Contributor ratio ~ 1:1

Obligate allele of 2nd contributor = 13

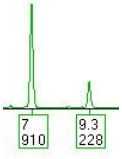
Second contributor = 13, 13 / 11, 13 / 12, 13

known	2nd	PkR (K)	PkR(2nd)	Proportion K & 2nd
11, 12	13, 13	0.63	NA	0.72 &0.28
11, 12	11, 13	0.32	0.51	0.58 & 0.42
11, 12	12, 13	0.79	0.80	0.50 & 0.50

Used ½(rfu) as estimate for shared allele rfu in this example.

Two allele loci:

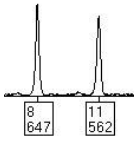
THO1



Known = 7, 7

Second contributor = ?

TPOX



Known = 8, 8

Deducing a Second Contributor

- Need a specified process:
  - Takes into account the mixture proportion data from the profile
  - and the range of allowable peak height ratio
- Computer processes are needed and available to test

Profile 2: Complex Mixtures

Statistical Issues with 2-Person Indistinguishable Mixtures

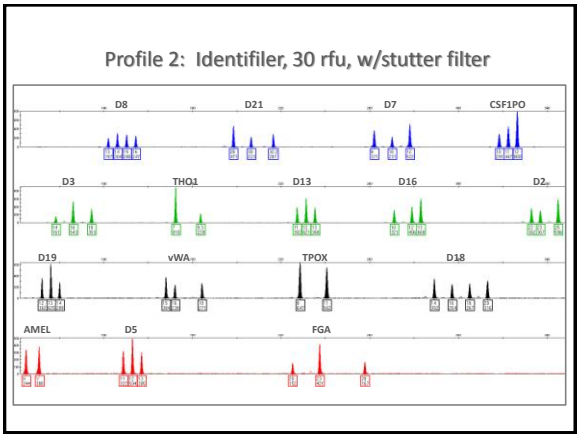
Michael D. Coble

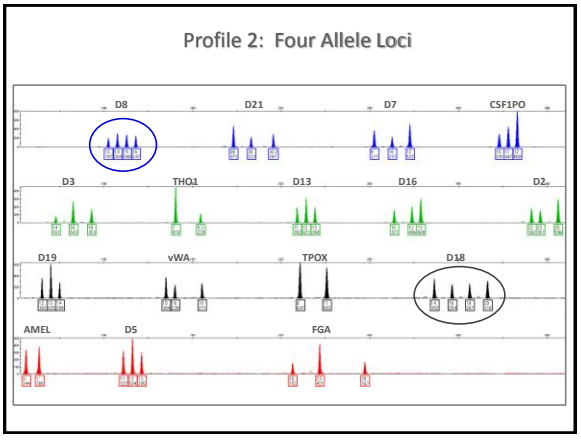
NIJ National Institute of Justice

BOSTON UNIVERSITY

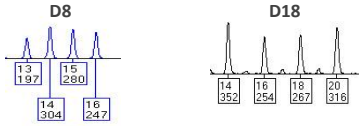
NIST

2011 Mixture Interpretation Workshop:
Using Scientific Analysis





Beginning with the four allele loci and assuming two contributors can we estimate the ratio of the two contributors?



Possible genotype combinations

13,14 + 15,16
13,15 + 14,16
13,16 + 14,15

&

14,16 + 18,20
14,18 + 16,20
14,20 + 16,18

Example: (rfu 13+ rfu 14) ÷ total rfu = proportion of contributor 1
(197 + 304) ÷ 1028 = 0.49

Proportion of contributor 1 to total rfu

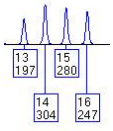
0.49
0.46
or
0.43

≈

0.49
0.48
or
0.44

Consider peak height ratio at the 4 allele loci:

D8



Possible genotype pairs

1) 13,14 + 15,16
2) 13,15 + 14,16
3) 13,16 + 14,15

Peak height ratios

1) 0.65 & 0.88
2) 0.70 & 0.81
3) 0.80 & 0.92

Based on the allowed peak height ratio, all three pairs of genotypes are possible.

Cannot restrict the possible combinations of genotypes

Combined Probability of Inclusion

$PI = (p + q + r + s)^2$
PI = (freq 13 + freq 14 + freq 15 + freq 16)
$PI = (0.305 + 0.166 + 0.114 + 0.031)^2$
$PI = (0.616)^2$
PI = 0.379

$PI = (p + q + r + s)^2$
PI = (freq 14 + freq 16 + freq 18 + freq 20)
$PI = (0.137 + 0.139 + 0.076 + 0.022)^2$
$PI = (0.374)^2$
PI = 0.140

CPI = (0.379)(0.140) = 0.053
CPE = 1 – CPI = 0.947

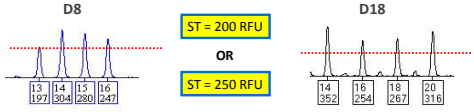
Combined Probability of Inclusion

PI = 1.0

$PI = (p + q + r + s)^2$
$PI = (\text{freq } 14 + \text{freq } 16 + \text{freq } 18 + \text{freq } 20)$
$PI = (0.137 + 0.139 + 0.076 + 0.022)^2$
$PI = (0.374)^2$
PI = 0.140

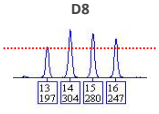
CPI = (1.0)(0.140) = 0.140
CPE = 1 – CPI = 0.86

Likelihood Ratio



We can use the information from D8!

Conditioning on the Victim



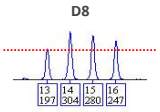
Victim = 14, 15

Perp = 13, 16 (obligate alleles)

$$\frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{V + S}{V + U} = \frac{1}{2pq} = \frac{1}{2f_{13}f_{16}}$$

$$= \frac{1}{2(0.3393)(0.013)} = 113$$

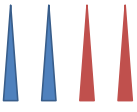
No Conditioning



Suspect = 13, 16

Unknown = 14, 15 (obligate alleles)

$$\frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{U + S}{U + U}$$



for the H_P

[ab] [cd]

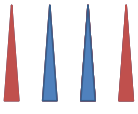
Suspect & Unk

[A,B] = Suspect

[C,D] = Unknown

[A,B & C,D] = [1 x 2cd]

= 2cd



for the H_D

[A,B + C,D] = [2ab x 2cd]

[A,C + B,D] = [2ac x 2bd]

[A,D + B,C] = [2ad x 2bc]

[A,B + C,D] = [2ab x 2cd]

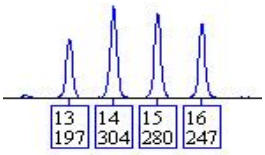
[A,C + B,D] = [2ac x 2bd]

[A,D + B,C] = [2ad x 2bc]

[2ab x 2cd] = 4abcd

4abcd x 6 = 24abcd

Let's do that with Permutations!!



Allelic Vector
13
14
15
16

$$= \frac{4!}{1!1!1!1!}$$

$$= \frac{4x3x2x1}{1x1x1x1}$$

$$= 24 \checkmark$$

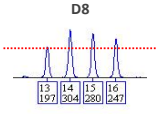
$$LR = \frac{Pr(E|H_P)}{Pr(E|H_D)} = \frac{[ab]\quad[cd]}{\text{Suspect + Unk}}$$

2 Unknowns

$$= \frac{2cd}{24abcd} = \frac{1}{12ab}$$

12 ab

No Conditioning



Suspect = 13, 16

Unknown = 14, 15 (obligate alleles)

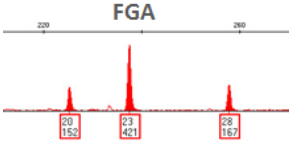
$$\frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{U + S}{U + U} = \frac{1}{12f_{13}f_{16}}$$

$$= \frac{1}{12(0.3393)(0.013)} = 18.9$$

Comparison of the Results (D8)

CPI	LR (victim)	LR (no conditioning)
0.379	133	18.9

Consider FGA…



Complex Mixtures

Charlotte J. Word

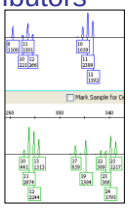
NIJ National Institute of Justice

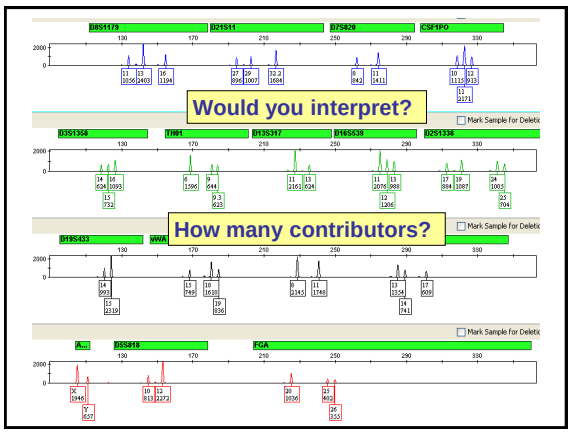
ISHI 2011 Mixture Interpretation Workshop:
Using **S**cientific **A**nalysis

BOSTON UNIVERSITY
NIST

Complex Mixtures

- Multiple contributors
  - –3- & 4- Person (or more!) vs. 2-Person Mixtures
- Determining Number of Contributors
- Relatives in Mixtures
- Allele Sharing in Mixtures
- Major/minor contributors

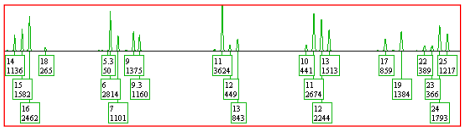


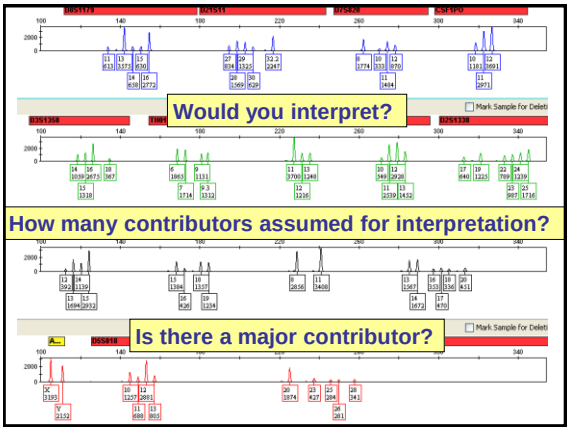


Complex Mixtures

Multiple Contributors

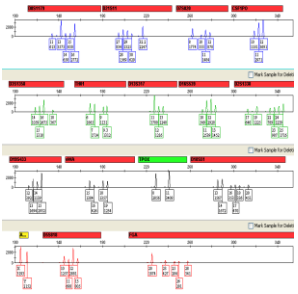
Determining the Number of Contributors





Complex Mixture – Allele Summary

- 6 alleles at 2 loci
- 5 alleles at 3 loci
- 4 alleles at 7 loci
- 3 alleles at 2 loci
- 2 alleles at 1 locus
- 1 allele at 0 loci
- 63 total alleles



Two Person Mixtures

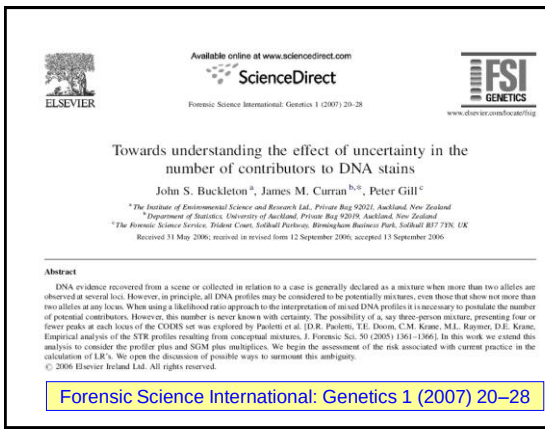
14 total combinations

Observed profile	A	B	
			**4 alleles** All heterozygotes and non-overlapping alleles
			**3 alleles** Heterozygote + heterozygote, one overlapping allele Heterozygote + homozygote, no overlapping alleles
			**2 alleles** Heterozygote + heterozygote, two overlapping alleles Heterozygote + homozygote, one overlapping allele Homozygote + homozygote, no overlapping alleles
			**1 allele** Homozygote + homozygote, overlapping allele

3-Person Mixtures

150 total combinations

Observed profile	
	**6 alleles** All heterozygotes and non-overlapping alleles
	**5 alleles** Two heterozygotes and one homozygote Three heterozygotes, one overlapping alleles
	**4 alleles** Six combinations of heterozygotes, homozygotes and overlapping alleles
	**3 alleles** Eight combinations of heterozygotes, homozygotes, and overlapping alleles
	**2 alleles** Five combinations of heterozygotes, homozygotes, and overlapping alleles
	**1 allele** All homozygotes, overlapping allele



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.

Forensic Science International: Genetics 1 (2007) 20–28

Two-Person Simulated Mixtures – SGM+ Number of Alleles at each Locus

Table 1
The probability of observing a given number of alleles in a two-person mixtures for simulated profiles at the SGM+™ loci

Loci	No. of alleles			
	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
D8	0.011	0.194	0.521	0.274
D21	0.007	0.147	0.505	0.341
D18	0.003	0.095	0.472	0.430
D19	0.020	0.261	0.516	0.203
THO	0.016	0.271	0.547	0.166
FGA	0.003	0.116	0.500	0.381

Buckleton et al. Forensic Science International: Genetics 1 (2007) 20–28

Three-Person Simulated Mixtures – SGM+ Number of Alleles at each Locus

Table 2
The probability of observing a given number of alleles in a three-person mixtures for simulated profiles at the SGM+™ loci

Loci	No. of alleles showing					
	1	2	3	4	5	6
D3	0.000	0.053	0.366	0.463	0.115	0.002
vWA	0.000	0.037	0.285	0.468	0.194	0.016
D16	0.001	0.086	0.397	0.411	0.100	0.005
D2	0.000	0.008	0.104	0.385	0.393	0.110
D8	0.001	0.041	0.258	0.436	0.236	0.029
D21	0.000	0.023	0.192	0.428	0.302	0.055
D18	0.000	0.007	0.109	0.392	0.396	0.096
D19	0.003	0.078	0.352	0.401	0.152	0.014
THO	0.001	0.074	0.395	0.439	0.088	0.002
FGA	0.000	0.012	0.144	0.424	0.346	0.074

Buckleton et al. Forensic Science International: Genetics 1 (2007) 20–28

2, 3, 4-Person Simulated Mixtures – CODIS Loci Number of Alleles at each Locus

J Forensic Sci, Nov. 2005, Vol. 50, No. 6
Paper ID JFS2004475
Available online at: www.astm.org

David R. Paoletti,[1] M.S.; Travis E. Doom,[1,2] Ph.D.; Carissa M. Krane,[3] Ph.D.; Michael L. Raymer,[1,2] Ph.D.; and Dan E. Krane,[4] Ph.D.

Empirical Analysis of the STR Profiles Resulting from Conceptual Mixtures

ABSTRACT: Samples containing DNA from two or more individuals can be difficult to interpret. Even ascertaining the number of contributors can be challenging and associated uncertainties can have dramatic effects on the interpretation of testing results. Using an FBI genotypes dataset, containing complete genotype information from the 13 Combined DNA Index System (CODIS) loci for 959 individuals, all possible mixtures of three individuals were exhaustively and empirically computed. Allele sharing between pairs of individuals in the original dataset, a randomized dataset and datasets of generated cousins and siblings was evaluated as were the number of loci that were necessary to reliably deduce the number of contributors present in simulated mixtures of four or less contributors. The relatively small number of alleles detectable at most CODIS loci and the fact that some alleles are likely to be shared between individuals within a population can make the maximum number of different alleles observed at any tested loci an unreliable indicator of the maximum number of contributors to a mixed DNA sample. This analysis does not use other data available from the electropherograms (such as peak height or peak area) to estimate the number of contributors to each mixture. As a result, the study represents a worst case analysis of mixture characterization. Within this dataset, approximately 3% of three-person mixtures would be mischaracterized as two-person mixtures and more than 70% of four-person mixtures would be mischaracterized as two- or three-person mixtures using only the maximum number of alleles observed at any tested locus.

Paoletti et al. J Forensic Sci, Nov. 2005, Vol. 50, No. 6

2-5-Person Simulated Mixtures – Identifiler Number of Alleles vs. Likelihood Estimator

J Forensic Sci, January 2011, Vol. 56, No. 1
doi: 10.1111/j.1556-4029.2010.01550.x
Available online at: interscience.wiley.com

PAPER
CRIMINALISTICS

Hinda Haned,[1] M.S.; Laurent Pène,[2] M.S.; Jean R. Lobry,[1] Ph.D.; Anne B. Dufour,[1] Ph.D.; and Dominique Pontier,[1] Ph.D.

Estimating the Number of Contributors to Forensic DNA Mixtures: Does Maximum Likelihood Perform Better Than Maximum Allele Count?

Haned et al. J Forensic Sci, January 2011, Vol. 56, No. 1

Number of Contributors – Total Number of Alleles

314 FORENSIC SCIENCE CMJ

doi: 10.3325/cmj.2011.52.314

Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts

Jaheida Perez, Adele A. Mitchell, Nubia Ducasse, Jeannie Tamariz, Theresa Caragine

Office of Chief Medical Examiner of the City of New York, The Department of Forensic Biology, New York, NY, USA

Perez et al., Croat Med J. 2011; 52:314-26

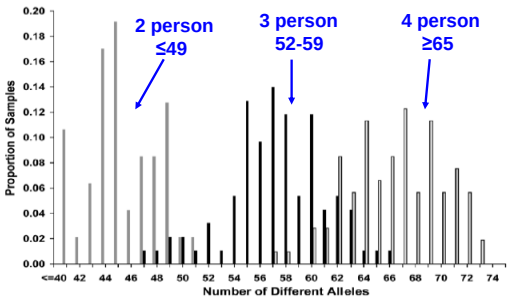


Figure 1. Expected # of different alleles from mixtures.

Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts
Perez et al., Croat Med J. 2011; 52:314-26

Two Person Mixture Studies Summary

- **Always** recognized as a mixture – no risk of confusing as a single-source
  - Loci with 3 or 4 alleles
  - Peak height ratio imbalance at loci with 2 alleles
- Observe more loci with 2 or 3 alleles than 4 alleles – even when profiles from two heterozygous profiles mixed
- 49 or fewer total alleles

Buckleton et al. Forensic Science International: Genetics 1 (2007) 20–28; Paoletti et al. J Forensic Sci, Nov. 2005, Vol. 50, No. 6; Haned et al. J Forensic Sci, January 2011, Vol. 56, No. 1; Perez et al., Croat Med J. 2011; 52:314-26

Three Person Mixture Studies Summary

- No risk of confusing as a single-source
- Small risk of confusing with two-person mixture
  - Observe at least one locus with 5 or 6 alleles in ~97% of profiles (3% have ≤4 alleles)
  - 3% profiles look like 2-person mixture
  - Risk if LT-DNA, degradation, inhibition, primer mutation to look like 2-person mixture
- Most loci have 3 or 4 alleles
- 52-59 total alleles

Buckleton et al. Forensic Science International: Genetics 1 (2007) 20–28; Paoletti et al. J Forensic Sci, Nov. 2005, Vol. 50, No. 6; Haned et al. J Forensic Sci, January 2011, Vol. 56, No. 1; Perez et al., Croat Med J. 2011; 52:314-26

Four Person Mixture Studies Summary

- No risk of confusing as a single-source
- Very small risk of confusing with two-person mixture
  - Likely to have peak height imbalance
- Very small number of loci with 8 alleles and very few with 7 alleles
  - High risk of confusing with three-person mixture
  - Risk if LT-DNA, degradation, inhibition, primer mutation
- ≥65 total alleles

Buckleton et al. Forensic Science International: Genetics 1 (2007) 20–28; Paoletti et al. J Forensic Sci, Nov. 2005, Vol. 50, No. 6; Haned et al. J Forensic Sci, January 2011, Vol. 56, No. 1; Perez et al., Croat Med J. 2011; 52:314-26

Four Person Mixture Studies Summary

>70% of 4-person mixtures would NOT be recognized as 4-person mixtures based on allele count

Buckleton et al. Forensic Science International: Genetics 1 (2007) 20–28; Paoletti et al. J Forensic Sci, Nov. 2005, Vol. 50, No. 6; Haned et al. J Forensic Sci, January 2011, Vol. 56, No. 1; Perez et al., Croat Med J. 2011; 52:314-26

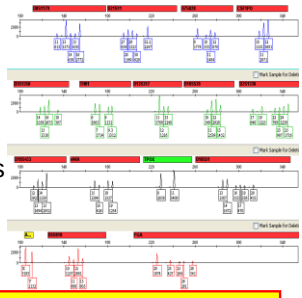
Five, Six Person Mixture Studies Summary

- >99% of 5 person mixtures would look like 4 person mixtures (~60%) or 3-person mixtures (~40%)
- Most 6 person mixtures would look like 5 person mixture (6%), 4-person mixtures (80%) or 3-person mixtures (14%)

Wang, T.W., Kalet, P., Pendleton, J., Gilbert, K., Lucas, L. and Birdwell, J.D. 2005 The probable number of contributors to a STR DNA mixture. http://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/16th-ishi-poster-abstracts/; Haned et al. J Forensic Sci, January 2011, Vol. 56,(1), 23-28

Complex Mixture – Allele Summary

- 6 alleles at 2 loci
- 5 alleles at 3 loci
- 4 alleles at 7 loci
- 3 alleles at 2 loci
- 2 alleles at 1 locus
- 1 allele at 0 loci
- 63 total alleles

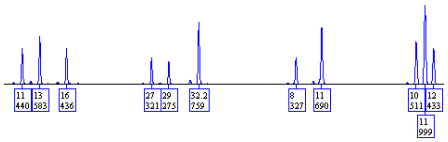


A 4-person mixture @ 1:1:1:2 ratio!!

Complex Mixtures

Mixtures with Relatives

Parent-Child
Sibling-Sibling



Parent + Child	Mixture DNA Profile Pattern
A B + B C	**Maximum: 3 alleles** Both heterozygote, one shared allele → A B C
A B + A B	**2 alleles** Heterozygote + heterozygote, two shared alleles → A B
A B + A	Heterozygote + homozygote, one shared allele → A B
A + A B	Homozygote + heterozygote, one shared allele → A B
A + A	**1 allele** Homozygote + homozygote, one shared allele → A

ALLELE SHARE AT EACH LOCUS

P1 + P2	Genotypes of Children	% Sibling Allele Sharing
A B + C D	AC or AD or BC or BD	0%, 50% or 100%
A B + B C	AB or AC or BB or BC	0%, 50% or 100%
A B + A B	AB/BA or AA or BB	0%, 50% or 100%
A B + C	AC or BC	50% or 100%
A B + A	AA or BA	50% or 100%
A + B	AB	100%
A + A	AA	100%

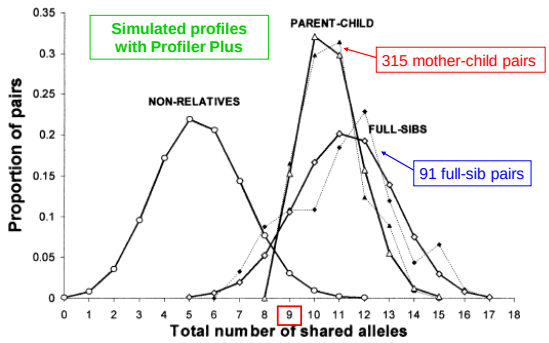
P1 = Parent 1; P2 = Parent 2

Allele Sharing in Relatives

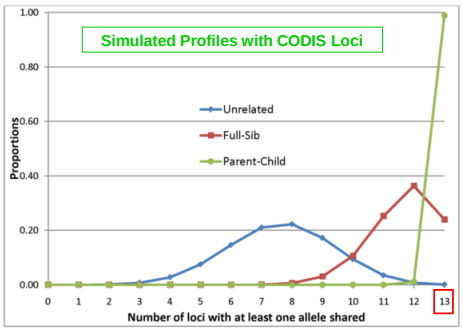
Allele sharing in first-degree and unrelated pairs of individuals in the Ge.F.I. AmpFlSTR® Profiler Plus™ database

Silvano Presciuttini[a,1,*], Francesca Ciampini[a], Milena Alù[b], Nicoletta Cerri[c], Marina Dobosz[d], Ranieri Domenici[e], Gabriella Peloso[f], Susi Pelotti[g], Andrea Piccinini[h], Elena Ponzano[f], Ugo Ricci[i], Adriano Tagliabracci[g], J.E. Baley-Wilson[j], Francesco De Stefano[m], Vincenzo Pascali[g,1]

Presciuttini et al. Forensic Science International 131 (2003) 85-89



Presciuttini et al. Forensic Science International 131 (2003) 85-89



Ge et al. Comparisons of the familial DNA databases searching policies. J. For. Sci. 2011, in press.

Mixtures with Relatives – Summary

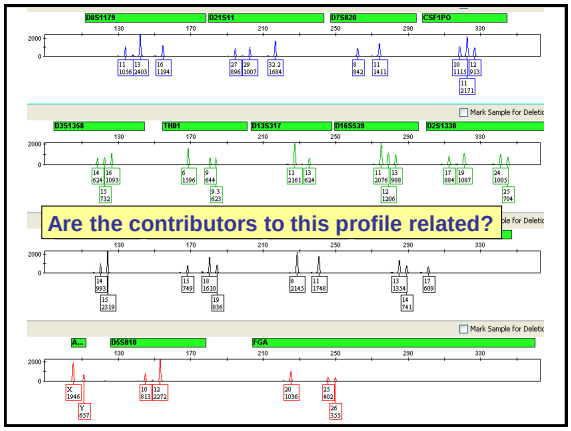
Parent-Child

- Expect at least 50% allele share
- Expect at least one shared allele at each locus
- Maximum 3 alleles per locus (in absence of mutation)
- If test X loci, expect >X allele shares (9-14 Profiler Plus; 13-20 CODIS)

Mixtures with Relatives – Summary

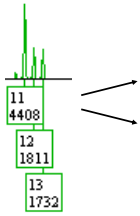
Sibling-Sibling

- Expect at least 50% allele share overall, but variable: 7-16 Profiler Plus; 12-22 CODIS (≥X-1)
- Expect 0, 50 or 100% allele share at each locus
- Expect at least one allele share at 9-13 loci (CODIS data)



Mixtures with Relatives – Working Backwards from Mixed DNA Profile

- With mixed DNA profile from unknowns, may not know if alleles are shared
- Data in the graphs are not helpful



11,12 + 11,13

Relative?
Parent-Child?
Sibs?

or

11,11 + 12,13

Relative?
Sibs?

Unrelated?

True Known Contributors to Previous Profile

- Share 14 alleles over 15 loci
  - 8 alleles at 9 Profiler Plus loci
  - 13 alleles at 13 CODIS loci
  - 15 alleles at 17 loci (Identifiler + PowerPlex 16 HS)
- One allele in common at each locus, except D2, FGA and Penta E
- Likely not parent
- Sibs? Inconclusive from allele #; Ge locus data suggests sibs
- Provided as DNA from non-relatives

Issues with Complex Mixtures

- ***Minimum number*** of contributors and the ***true number*** of contributors may be underestimated, especially when there are **three or more contributors** to the DNA mixture
  - May need to interpret data under several assumptions
- Number of allele shares increases when **relatives** are in the mixture as compared to unrelated contributors

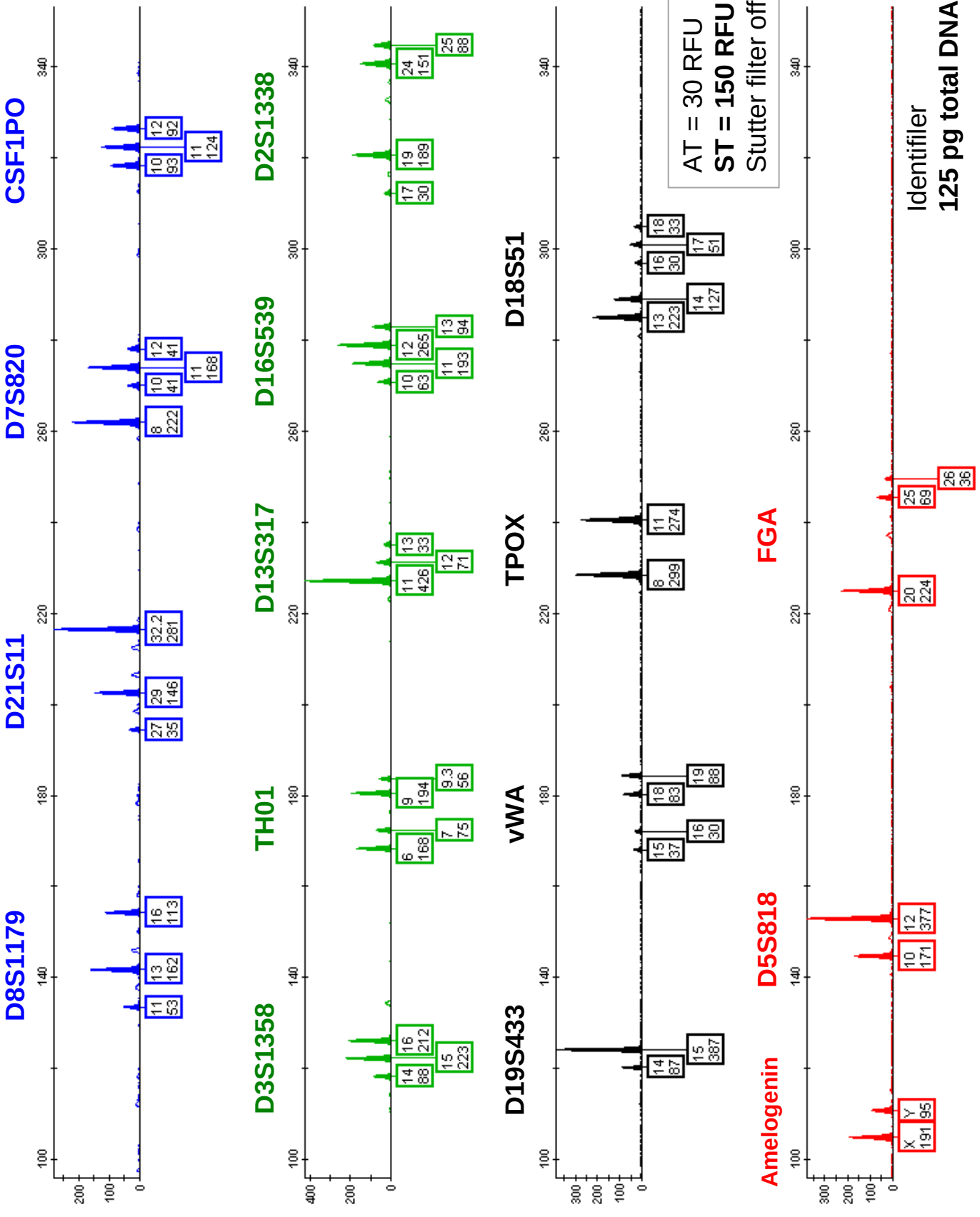
Issues with Complex Mixtures

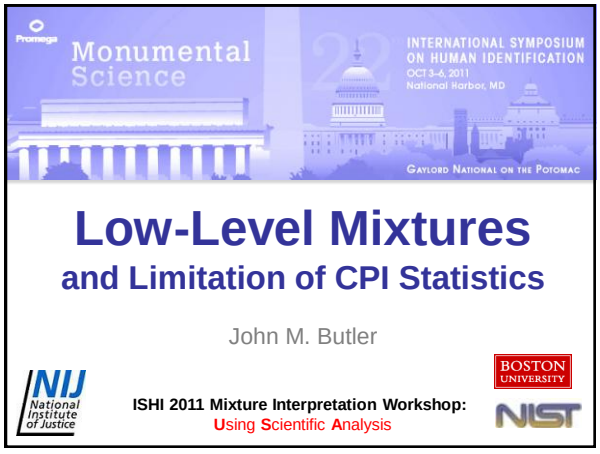
- **Peak height ratios** and **mixture ratios** may not be helpful
- Decreased ability to associate alleles and **determine genotypes** for contributors
- Increased likelihood of **falsely including** a non-contributor
- Increased likelihood of being **inconclusive**

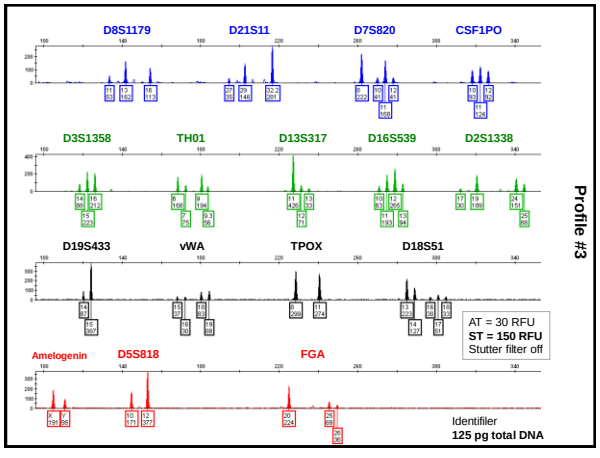
Issues with Complex Mixtures

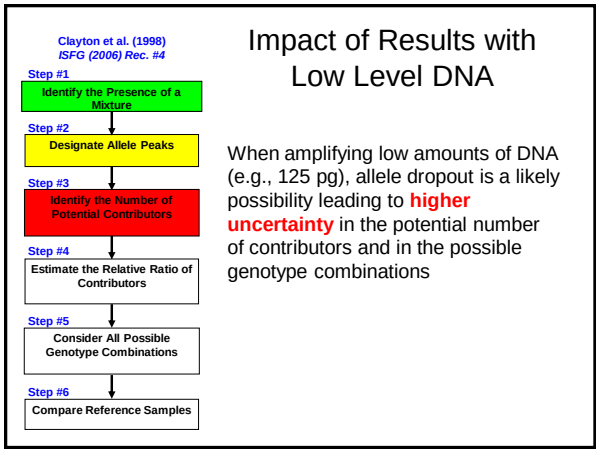
- **High number of allele shares** (homozygous, related & unrelated) cause high peaks from multiple contributors
  - May be **falsely interpreted** to be from a **single major** DNA contributor
  - **Stochastic threshold** becomes useless
  - May lead to conclusion all alleles are present when, in fact, **allelic drop-out** is likely
    - May be **heterozygous** allele (and not homozygous) for one or more contributors
  - Increased likelihood of **missing or off-scale data** even when amounts of DNA are similar

Case Example #3




Promega
Monumental Science
INTERNATIONAL SYMPOSIUM ON HUMAN IDENTIFICATION
OCT 3-6, 2011
National Harbor, MD
GAYLORD NATIONAL ON THE POTOMAC
Low-Level Mixtures
and Limitation of CPI Statistics
John M. Butler
NIJ National Institute of Justice
ISHI 2011 Mixture Interpretation Workshop:
Using Scientific Analysis
BOSTON UNIVERSITY
NIST

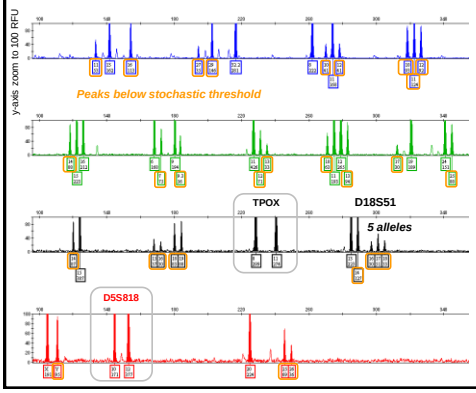

D8S1179
D21S11
D7S820
CSF1PO
D3S1358
TH01
D13S317
D16S539
D2S1338
D19S433
vWA
TPOX
D18S51
AT = 30 RFU
ST = 150 RFU
Stutter filter off
Amelogenin
D5S818
FGA
Identifiler
125 pg total DNA
Profile #3


Clayton et al. (1998)
ISFG (2006) Rec. #4
Step #1
Identify the Presence of a Mixture
Step #2
Designate Allele Peaks
Step #3
Identify the Number of Potential Contributors
Step #4
Estimate the Relative Ratio of Contributors
Step #5
Consider All Possible Genotype Combinations
Step #6
Compare Reference Samples
Impact of Results with Low Level DNA
When amplifying low amounts of DNA (e.g., 125 pg), allele dropout is a likely possibility leading to higher uncertainty in the potential number of contributors and in the possible genotype combinations

Identifiler
125 pg total DNA

Case Example #3

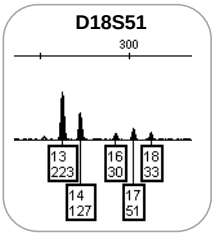
AT = 30 RFU
ST = 150 RFU
Stutter filter off



What Can We Say about this Result?

- Low level DNA (only amplified 125 pg total DNA)
  - likely to exhibit stochastic effects and have allele dropout
- Mixture of at least 3 contributors
  - Based on detection of 5 alleles at D18S51
  - If at equal amounts, ~40 pg of each contributor (if not equal, then less for the minor contributors); **we expect allele dropout**
- At least one of the contributors is male
  - Based on presence of Y allele at amelogenin
- Statistics if using CPI/CPE
  - Would appear that we can only use TPOX and D5S818 results with a stochastic threshold of 150 RFU *(will explore this further)*
- **Due to potential of excessive allele dropout, we are unable to perform any meaningful Q-K comparisons**

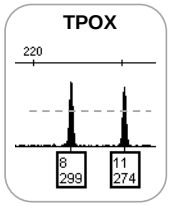
Uncertainty in the Potential Number of Contributors with this Result



5 alleles observed

- Several of the peaks are barely above the analytical threshold of 30 RFU
  - In fact, with an analytical threshold of 50 RFU or even 35 RFU, there would only be three detected alleles at D18S51
- Stochastic effects could result in a high degree of stutter off of the 17 allele making alleles 16 and 18 potential stutter products
- No other loci have >4 alleles detected

All Detected Alleles Are Above the Stochastic Threshold – **Or Are They?**



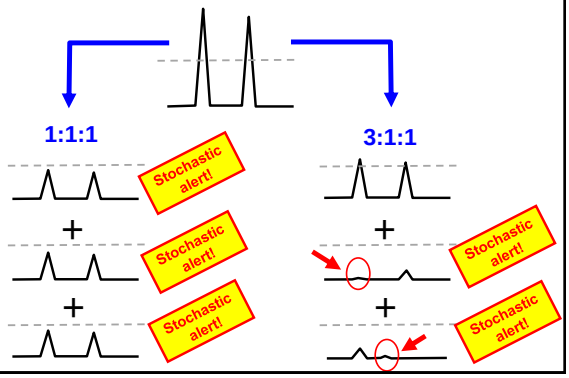
Stochastic threshold = 150 RFU

Does this result guarantee no allele drop-out?

We have assumed three contributors. If result is from an equal contribution of 3 individuals…

Then some alleles from individual contributors would be below the stochastic threshold and we could not assume that all alleles are being observed!

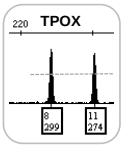
Assuming Three Contributors…
Some Possible Contributions to This Result



All Loci Are Not Created Equal
when it comes to mixture interpretation

- In the case of less polymorphic loci, such as TPOX, there are fewer alleles and these occur at higher frequency. Thus, there is a greater chance of allele sharing (peak height stacking) in mixtures.
- **Higher locus heterozygosity is advantageous for mixture interpretation** – we would expect to see more alleles (within and between contributors) and thus have a better chance of estimating the true number of contributors to the mixture

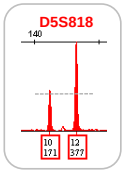
Even if you did attempt to calculate a CPI/CPE statistic using loci with all observed alleles above the stochastic threshold on this result…



TPOX Allele Frequencies (NIST Caucasian, Butler et al. 2003)
8 = 0.53
11 = 0.24
CPI = $(0.53 + 0.24)^2$ = 0.59 or **59%**

Combine loci = 0.59 x 0.18 = 0.11 or **11%**

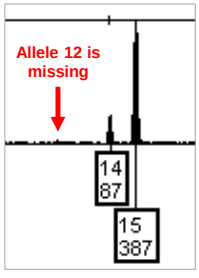
Approximately 1 in every 9 Caucasians could be included in this mixture



D5S818 Allele Frequencies (NIST Caucasian, Butler et al. 2003)
10 = 0.05
12 = 0.38
CPI = $(0.05 + 0.38)^2$ = 0.18 or **18%**

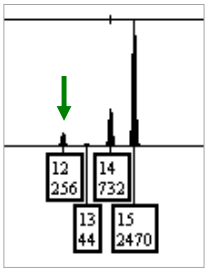
Impact of Amplifying More DNA

D19S433



125 pg total DNA amplified

D19S433



500 pg total DNA amplified

True Contributors
3 contributors with a 2:1:1 mixture

15,15 (2x)
14,15 (1x)
12,14 (1x)

How should you handle the suspect comparison(s) with this case result?

- **No suspect comparisons should be made** as **the mixture result has too much uncertainty** with stochastic effects that may not account for all alleles being detected
- **Declare the result "inconclusive"**

How not to handle this result

- "To heck with the analytical and stochastic thresholds", **I am just going to see if the suspect profile(s) can fit into the mixture allele pattern observed** – and then if an allele is not present in the evidentiary sample try to explain it with possible allele dropout due to stochastic effects
- This is what Bill Thompson calls "painting the target around the arrow (matching profile)…"

Thompson, W.C. (2009) Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. *Law, Probability and Risk* 8: 257-276

Value of Using a Profile Interpretation Worksheet

PROFILE INTERPRETATION WORKSHEET
IDENTIFILER

PROFILE NAME: *Case Example #3*

ANALYST: *John Butler*

DATE: *11 October 2010*

MIXTURE: ■ yes □ no □ unsure

Analytical threshold: *30 RFU*

Stutter % used: *0% (filter turned-off)*

Stochastic threshold: *150 RFU*

Peak height ratio: *60%*

Comments: *low level DNA (125 pg)*

Allele and Locus Assessments

ID LOCUS	Alleles called	Alleles above Stochastic Threshold	Stutter or other peaks to consider	Possible allele dropout ? Y/N	Stochastic issues? (e.g., elevated stutter, PHR imbalance, drop-in, etc.) Y/N	Degradation / inhibition (obvious)? Y/N	If mixture, restricted genotypes can be used? Y/N	Can this locus be interpreted ? Y/N	Additional Comments
D8S1179	11,13,16	13	Maybe	Y	Y	N	N	N	

Make decisions on the evidentiary sample and document them prior to looking at the known(s) for comparison purposes

What to do with low level DNA mixtures?

- **German Stain Commission "Category C"**
  (Schneider et al. 2006, 2009)
  - Cannot perform stats because stochastic effects make it uncertain that all alleles are accounted for
- **ISFG Recommendations #8 & #9** (Gill et al. 2006)
  - Stochastic effects limit usefulness
- *Fundamentals of Forensic DNA Typing* (2010)
  Butler 3rd edition (volume 1), chapter 18
  - Don't go "outside the box" without supporting validation

ISFG Recommendations on Mixture Interpretation

http://www.isfg.org/Publication;Gill2006

1. The likelihood ratio (LR) is the preferred statistical method for mixtures over RMNE
2. Scientists should be trained in and use LRs
3. Methods to calculate LRs of mixtures are cited
4. Follow Clayton et al. (1998) guidelines when deducing component genotypes
5. Prosecution determines H_p and defense determines H_d and multiple propositions may be evaluated
6. When minor alleles are the same size as stutters of major alleles, then they are indistinguishable
7. Allele dropout to explain evidence can only be used with low signal data
8. No statistical interpretation should be performed on alleles below threshold
9. Stochastic effects limit usefulness of heterozygote balance and mixture proportion estimates with low level DNA

Gill *et al.* (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

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

Summary

- Do not blindly use a stochastic threshold with complex mixtures as assumptions regarding the number of contributors can impact interpretation
- Going back to try and get a better sample from the evidence (if available) is wiser than spending a lot of time trying to work with a poor quality DNA result

Future of Complex, Low-level Mixtures

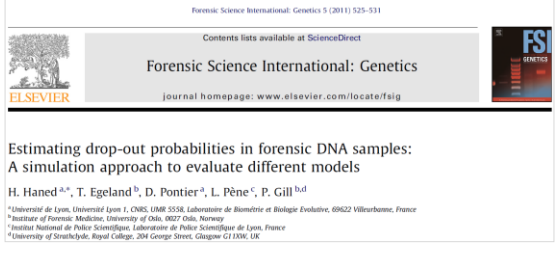
- **If you want to work in this area, you need supporting validation data** (collecting a few results at high DNA levels and extrapolating to greater complexity and smaller amounts of DNA will not be sufficient)
- Recent efforts in Europe are focused on **modeling uncertainty through probabilistic genotype approaches**
- Will require software to perform all of the calculations
- See articles included in the workshop reference list to learn more…

A Statistical Modeling Approach


G Model
FSIGEN-740; No. of Pages 7
ARTICLE IN PRESS
Forensic Science International: Genetics xxx (2011) xxx–xxx
Contents lists available at ScienceDirect
Forensic Science International: Genetics
journal homepage: www.elsevier.com/locate/fsig
ELSEVIER
The interpretation of low level DNA mixtures
Hannah Kelly [a,*], Jo-Anne Bright [a], James Curran [b], John Buckleton [a]
[a] ESR, PB 92021 Auckland, New Zealand
[b] Department of Statistics, University of Auckland, PB 92019 Auckland, New Zealand

Development of statistical models that account for the possibility of allele drop-out

A Simulation Approach


Forensic Science International: Genetics 5 (2011) 525–531
Contents lists available at ScienceDirect
Forensic Science International: Genetics
journal homepage: www.elsevier.com/locate/fsig
ELSEVIER
Estimating drop-out probabilities in forensic DNA samples:
A simulation approach to evaluate different models
H. Haned [a,*], T. Egeland [b], D. Pontier [a], L. Pène [c], P. Gill [b,d]
[a] Université de Lyon, Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, 69622 Villeurbanne, France
[b] Institute of Forensic Medicine, University of Oslo, 0027 Oslo, Norway
[c] Institut National de Police Scientifique, Laboratoire de Police Scientifique de Lyon, France
[d] University of Strathclyde, Royal College, 204 George Street, Glasgow G1 1XW, UK

A Logistic Regression Model

Forensic Science International: Genetics 3 (2009) 222–226

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Estimating the probability of allelic drop-out of STR alleles in forensic genetics

Torben Tvedebrink [a,*], Poul Svante Eriksen [a,1], Helle Smidt Mogensen [b,2], Niels Morling [b,3]

[a] Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark
[b] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Frederik V's Vej 11, DK-2100 Copenhagen East, Denmark

Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out

Torben Tvedebrink [a,*], Poul Svante Eriksen [a,1], Helle Smidt Mogensen [b,2], Niels Morling [b,3]

[a] Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark
[b] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Frederik V's Vej 11, DK-2100 Copenhagen East, Denmark

Allelic drop-out probabilities estimated by logistic regression—Further considerations and practical implementation

Torben Tvedebrink [a,*], Poul Svante Eriksen [a,1], Maria Asplund [b,2], Helle Smidt Mogensen [b,3], Niels Morling [b,4]

[a] Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark
[b] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Frederik V's Vej 11, DK-2100 Copenhagen East, Denmark

A Logistic Regression Model

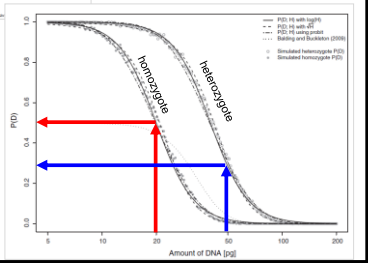
ARTICLE IN PRESS

Forensic Science International: Genetics

Allelic drop-out probabilities estimated by logistic regression—Further considerations and practical implementation

At **20 pg**, approximately **50% of homozygote** alleles will have dropped out

At **50 pg**, approximately **30% of heterozygote** alleles will have dropped out



CPE/CPI (RMNE) Limitations

- A CPE/CPI approach assumes that all alleles are present (i.e., cannot handle allele drop-out)
- Thus, statistical analysis of low-level DNA CANNOT be correctly performed with a CPE/CPI approach because some alleles may be missing
- Charles Brenner in his AAFS 2011 talk addressed this issue
- Research is on-going to develop allele drop-out models and software to enable appropriate calculations

BOSTON UNIVERSITY

Boston University Biomedical Forensic Sciences

DNA Mixture Analysis Training Tool

DNA Mixture Analysis Training Tool-Profiles

- Profiles
  - Single source
  - Two person
  - Three person
  - Four person
- Varying amounts
  - 0.625 to 4ng
- Ratio of contributors varies in both directions
- Amplification kits
  - PowerPlex 16
  - Identifiler
  - Yfiler
  - Minifiler
- Three injection times per profile

BOSTON UNIVERSITY

DNA Mixture Analysis Training Tool-Website

- Lessons on basic feature of DNA profile
- Lessons topics similar to those covered today
- Profiles in lessons can be viewed and compared
- All website profiles with corresponding ladders and controls can be downloaded as .fsa files for use in training or other purposes

BOSTON UNIVERSITY

DNA Mixture Analysis Training Tool-Website

- Scheduled to go live November 30, 2011
  - Website built by BU Information Services and Technology
  - Content Authors: Robin Cotton, Charlotte Word, Margaret Terrill and Catherine Grgicak
  - Profiles by Catherine Grgicak
- Funded by:
  - NIJ Forensic Science Training Development and Delivery Program
  - NIJ Grant # 2008-DN-BX-K158, awarded to Biomedical Forensic Science Program at Boston University School of Medicine

BOSTON UNIVERSITY

Software Solutions?

Michael D. Coble

NIJ National Institute of Justice

2011 Mixture Interpretation Workshop:
Using **S**cientific **A**nalysis

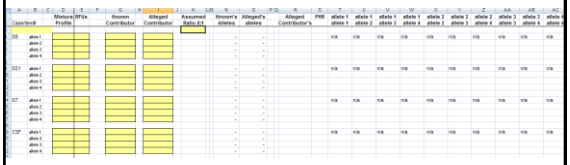
BOSTON UNIVERSITY
NIST

Official Disclaimer

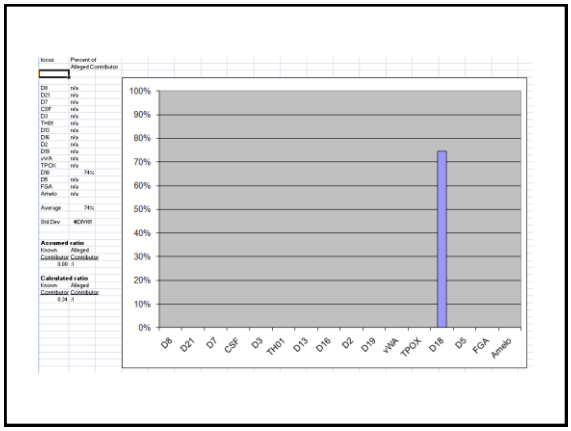
The opinions and assertions contained herein are solely those of the author and are not to be construed as official or as views of the U.S. Department of Commerce, the U.S. Department of Justice, or the National Institute of Justice.

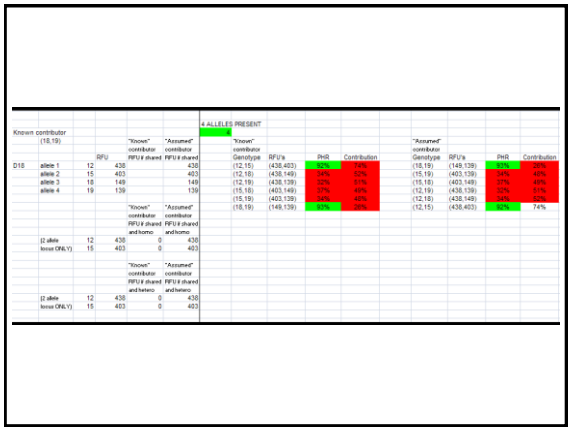
Commercial software, equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the U.S. Department of Commerce, the U.S. Department of Justice, or the National Institute of Justice nor does it imply that any of the software, materials, instruments or equipment identified are necessarily the best available for the purpose.

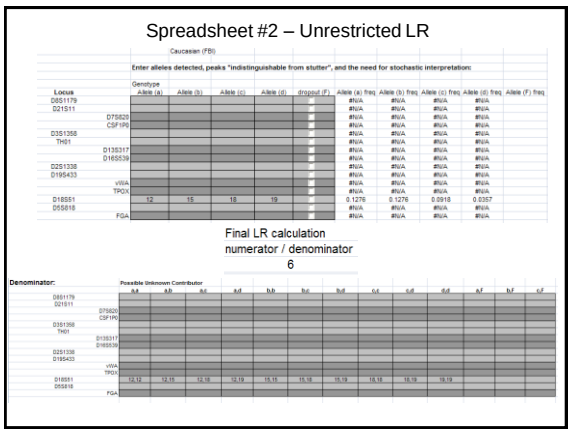
Deconvolutinator 2.0.5
(Bruce Heidebrecht, MDSP)



bheidebrecht@mdsp.org


locus
Percent of Alleged Contributor
D8
D21
D7
CSF
D3
TH01
D13
D16
D2
D19
vWA
TPOX
D18
D5
FGA
Amelo
Average
Std Dev
Assumed ratio
Known Contributor
Alleged Contributor
Calculated ratio
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%


4 ALLELES PRESENT
Known contributor (18,19)
"Known" contributor RFU # shared
"Assumed" contributor RFU # shared
"Known" contributor Genotype
RFU's
PHR
Contribution
"Assumed" contributor Genotype
RFU's
PHR
Contribution
D18
allele 1
allele 2
allele 3
allele 4
(12,15)
(12,18)
(12,19)
(15,18)
(15,19)
(18,19)
(438,403)
(438,149)
(438,139)
(403,149)
(403,139)
(149,139)
(15,19)
(15,18)
(12,18)
(12,15)
(403,139)
(403,149)
(438,139)
(438,149)
(438,403)
"Known" contributor RFU # shared and homo-
"Assumed" contributor RFU # shared and homo-
(2 allele locus ONLY)
"Known" contributor RFU # shared and hetero-
"Assumed" contributor RFU # shared and hetero-


Spreadsheet #2 – Unrestricted LR
Caucasian (FB)
Enter alleles detected, peaks "indistinguishable from stutter", and the need for stochastic interpretation:
Locus
Genotype
Allele (a)
Allele (b)
Allele (c)
Allele (d)
dropout (F)
Allele (a) freq
Allele (b) freq
Allele (c) freq
Allele (d) freq
Allele (F) freq
D8S1179
D21S11
D7S820
CSF1PO
D3S1358
TH01
D13S317
D16S539
D2S1338
D19S433
vWA
TPOX
D18S51
D5S818
FGA
12
15
18
19
0.1276
0.1276
0.0918
0.0357
#N/A
Final LR calculation
numerator / denominator
6
Denominator:
Possible Unknown Contributor
a,a
a,b
a,c
a,d
b,b
b,c
b,d
c,c
c,d
d,d
a,F
b,F
c,F

Software for Mixture Analysis

GeneMapper ID-X Software — Now Featuring an Integrated Mixture Analysis Tool

Cybergenetics TrueAllele®

FSS-i3

GeneMarker® HID

NicheVision Inc.

Automated Mixture Software

Advantages

- Can calculate parameters for mixture deconvolution: PHR, mixture ratio, etc…
- Speed of Analysis
- Statistical Analyses – RMNE, RMP, LR

Limitations

- The ultimate decision is up to the DNA analyst.
- 3+ person mixtures are limited to RMNE (CPE) Statistics

"On the Threshold of a Dilemma"

- Gill and Buckleton (2010)
- Although most labs use thresholds of some description, this philosophy has always been problematic because there is an inherent illogicality which we call the falling off the cliff effect.

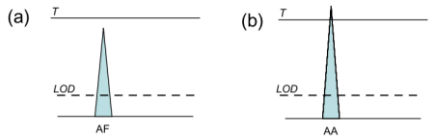
JOURNAL OF FORENSIC SCIENCES

Commentary on: Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forensic Sci 2009;54(4):810–21.

J Forensic Sci, January 2010, Vol. 55, No. 1
doi: 10.1111/j.1556-4029.2009.01257.x
Available online at: interscience.wiley.com

"Falling off the Cliff Effect"

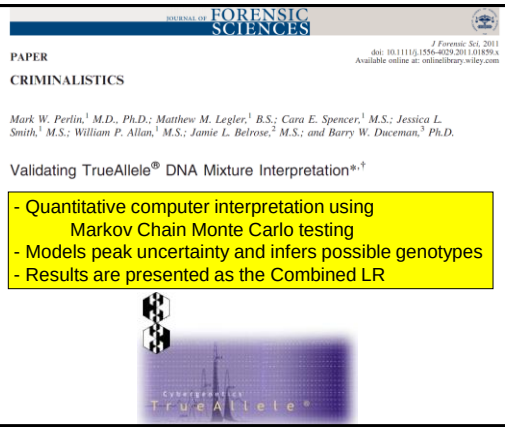
- If T = an arbitrary level (e.g., 150 rfu), an allele of 149 rfu is subject to a different set of guidelines compared with one that is 150 rfu even though they differ by just 1 rfu (Fig. 1).



Gill and Buckleton *JFS* 55: 265-268 (2010)

Gill and Buckleton *JFS* **55:** 265-268 (2010)

- "The purpose of the ISFG DNA commission document was to provide a way forward to demonstrate the use of ***probabilistic models to circumvent the requirement for a threshold*** and to safeguard the legitimate interests of defendants."


JOURNAL OF FORENSIC SCIENCES

PAPER

CRIMINALISTICS

J Forensic Sci, 2011
doi: 10.1111/j.1556-4029.2011.01859.x
Available online at: onlinelibrary.wiley.com

Mark W. Perlin,[1] M.D., Ph.D.; Matthew M. Legler,[1] B.S.; Cara E. Spencer,[1] M.S.; Jessica L. Smith,[1] M.S.; William P. Allan,[1] M.S.; Jamie L. Belrose,[2] M.S.; and Barry W. Duceman,[3] Ph.D.

Validating TrueAllele® DNA Mixture Interpretation*,†

- Quantitative computer interpretation using Markov Chain Monte Carlo testing
- Models peak uncertainty and infers possible genotypes
- Results are presented as the Combined LR

True Allele Software (Cybergenetics)

- We purchased the software in September 2010.
- Three day training at Cybergenetics (Pittsburgh, PA) in October.
- Software runs on a Linux Server with a Mac interface.

True Allele Casework Workflow
5 Modules

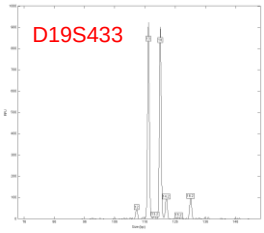
Analyze

.fsa files imported
Size Standard check
Allelic Ladder check
Alleles are called

True Allele Casework Workflow
5 Modules

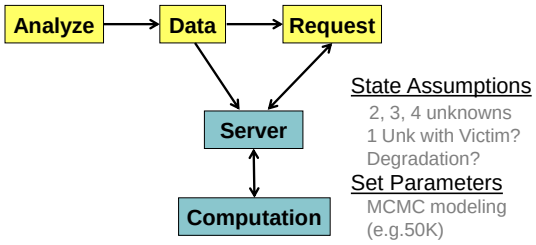
Analyze → **Data**

Server



All Peaks above 10 RFU are considered

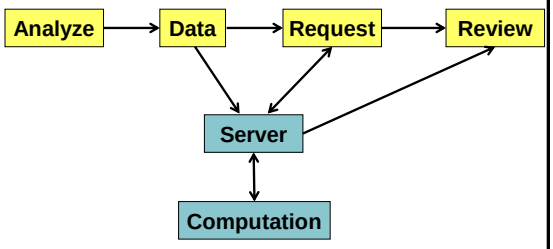
True Allele Casework Workflow
5 Modules



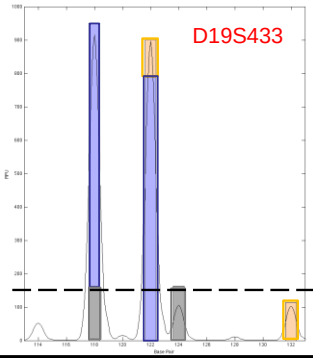
State Assumptions
- 2, 3, 4 unknowns
- 1 Unk with Victim?
- Degradation?

Set Parameters
- MCMC modeling (e.g.50K)

True Allele Casework Workflow
5 Modules



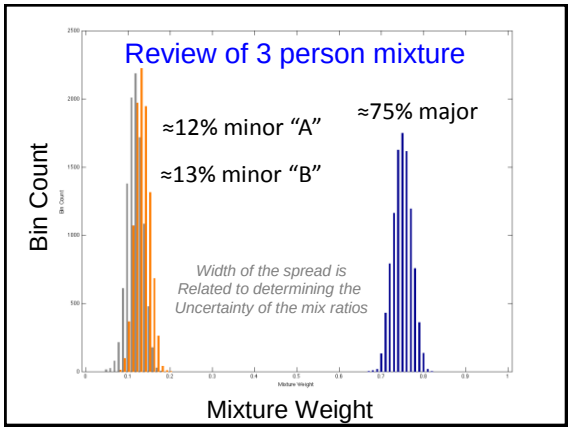
Review of One Replicate (of 50K)

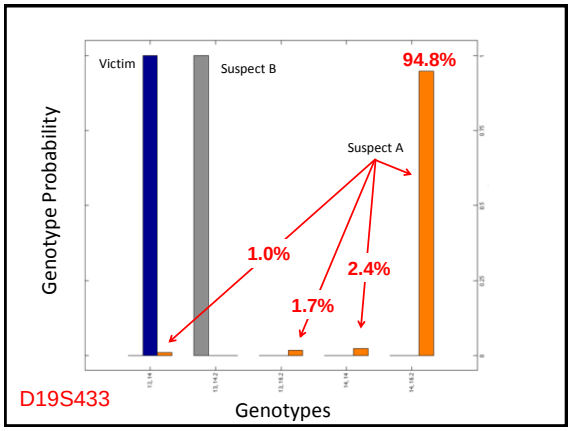


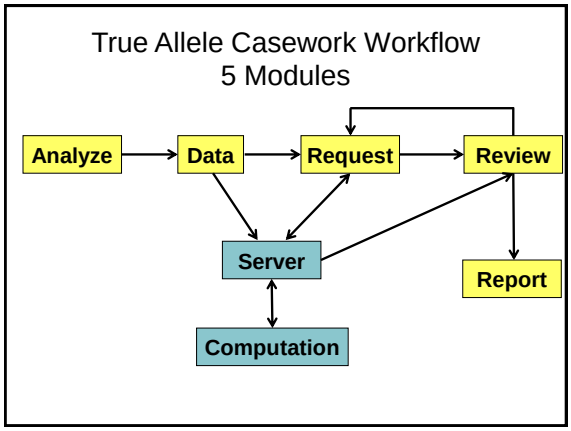
3P mixture,
2 Unknowns,

Conditioned on the Victim (major)

Good fit of the data to the model


Review of 3 person mixture
≈12% minor "A"
≈13% minor "B"
≈75% major
Width of the spread is
Related to determining the
Uncertainty of the mix ratios
Bin Count
Mixture Weight


Victim
Suspect B
Suspect A
94.8%
1.0%
1.7%
2.4%
Genotype Probability
D19S433
Genotypes


True Allele Casework Workflow
5 Modules
Analyze
Data
Request
Review
Server
Report
Computation

Determining the LR for D19S433

Suspect A = 14, 16.2

$H_P = 1 * 0.967$

Allele Pair	Probability Before Conditioning
14, 16.2	0.967
14, 14	0.003
13, 16.2	0.026
13, 14	0.001

$$LR = \frac{0.967}{\quad}$$

Determining the LR for D19S433

Suspect A = 14, 16.2

$H_P = 1 * 0.967$

Allele Pair	Probability Before Conditioning	Genotype Frequency	Probability * Genotype Freq
14, 16.2	0.967	0.0120	0.01164
14, 14	0.003	0.0498	0.00013
13, 16.2	0.026	0.0131	0.00034
13, 14	0.001	0.1082	0.00009
		sum	0.0122

$$LR = \frac{0.967}{0.0122} = 79.26$$

H_D

Combined LR = 5.6 Quintillion

			Genotype Probability Distribution			Weighted Likelihood		Likelihood Ratio	
	allele pair	Likelihood	Questioned	Reference	Suspect	Numerator	Denominator	LR	log(LR)
locus	x	l(x)	q(x)	r(x)	s(x)	l(x)*s(x)	l(x)*r(x)		
CSF1PO	11, 12	0.686	0.778	0.1448	1	0.68615	0.1292	5.31	0.725
D13S317	9, 12	1	1	0.0291	1	0.99952	0.02913	34.301	1.535
D16S539	9, 11	0.985	0.995	0.1238	1	0.98451	0.12188	8.036	0.905
D18S51	13, 17	0.999	1	0.0154	1	0.99915	0.01543	64.677	1.811
D19S433	14, 16.2	0.967	0.948	0.012	1	0.96715	0.01222	79.143	1.898
D21S11	28, 30	0.968	0.98	0.0872	1	0.96809	0.08648	11.194	1.049
D2S1338	23, 24	0.998	1	0.0179	1	0.99831	0.01787	55.866	1.747
D3S1358	15, 17	0.988	0.994	0.1224	1	0.98759	0.12084	8.14	0.911
D5S818	11, 11	0.451	0.394	0.0537	1	0.45103	0.07309	6.17	0.79
D7S820	11, 12	0.984	0.978	0.0356	1	0.98383	0.03617	27.198	1.435
D8S1179	13, 14	0.203	0.9	0.1293	1	0.20267	0.02993	6.771	0.831
FGA	21, 25	0.32	0.356	0.028	1	0.31986	0.01906	16.783	1.225
TH01	7, 7	0.887	0.985	0.1739	1	0.88661	0.15588	5.687	0.755
TPOX	8, 8	1	1	0.1375	1	1	0.13746	7.275	0.862
vWA	15, 20	0.998	0.996	0.0057	1	0.99808	0.00569	174.834	2.243

Results

- Results are expressed as logLR values

$$LR = 1{,}000{,}000 = 10^6$$

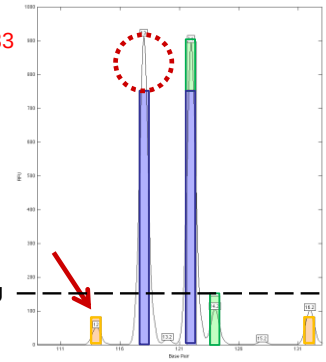
$$\log(LR) = \log 10^6$$

$$\log(LR) = 6 * \log 10 \ (1)$$

$$\log(LR) = 6$$

Review of One Replicate (of 50K)

D19S433

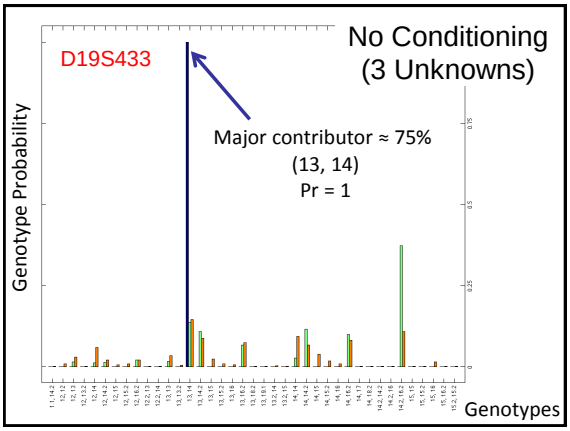


3P mixture,

3 Unknowns

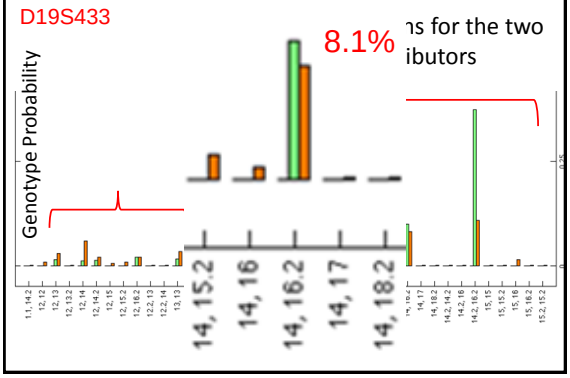
Poor fit of the data to the model

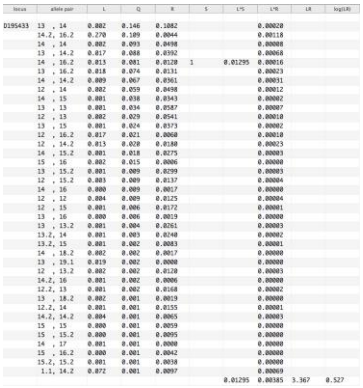
150 RFU



No Conditioning (3 Unknowns)

D19S433





Suspect "A" Genotype

39 probable genotypes

D19S433

Suspect A = 14, 16.2

$H_P = 1 * 0.013$

Allele Pair	Probability	Genotype Frequency	Prob * GenFreq
13,14	0.002	0.1082	0.00020
14.2, 16.2	0.270	0.0044	0.00118
14, 14	0.002	0.0498	0.00008
13, 14.2	0.017	0.0392	0.00068
14, 16.2	0.013	0.0120	0.00016
13, 16.2	0.018	0.0131	0.00023
etc…	etc…	etc…	etc…
		Sum	0.00385

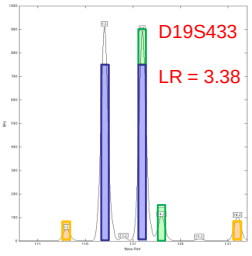
H_D

$$LR = \frac{0.013}{0.00385} = 3.38$$

D19S433

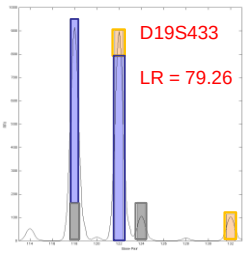
No Conditioning (3 Unknowns)

No Conditioning



Profile - Combined log(LR)
Suspect A log(LR) = 8.03
Suspect B log(LR) = 7.84

Conditioned on Victim



Profile - Combined log(LR)
Suspect A log(LR) = 18.72
Suspect B log(LR) = 19.45

Exploring the Capabilities

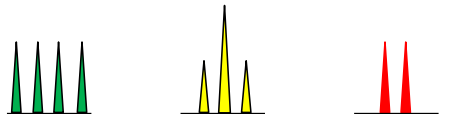
- **Degree of Allele Sharing**
- **Mixture Ratios**
- DNA Quantity

Mixture Data Set

- Mixtures of pristine male and female DNA amplified at a total concentration of 1.0 ng/μL using Identifiler (standard conditions).
- Mixture ratios ranged from 90:10, 80:20, 70:30 60:40, 50:50, 40:60, 30:70, 20:80, and 10:90
- Each sample was amplified twice.

Mixture Data Set

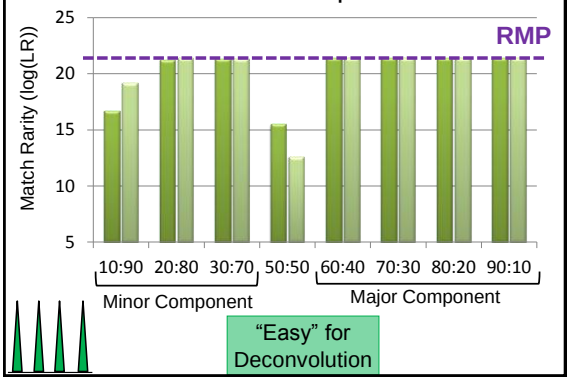
- Three different combinations:



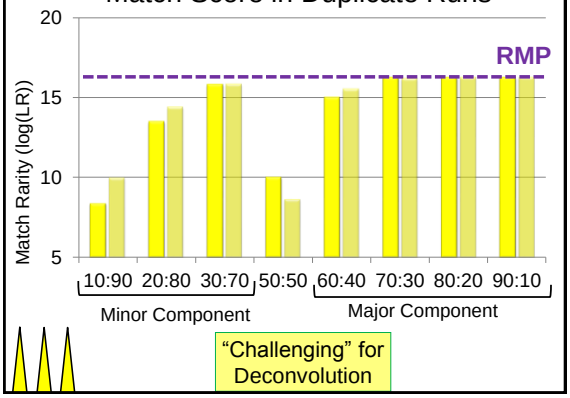
"Low" Sharing	"Medium" Sharing	"High" Sharing
4 alleles – 10 loci	4 alleles – 3 loci	4 alleles – 0 loci
3 alleles – 5 loci	3 alleles – 8 loci	3 alleles – 6 loci
2 alleles – 0 loci	2 alleles – 4 loci	2 alleles – 8 loci
1 allele – 0 loci	1 allele – 0 loci	1 allele – 1 loci

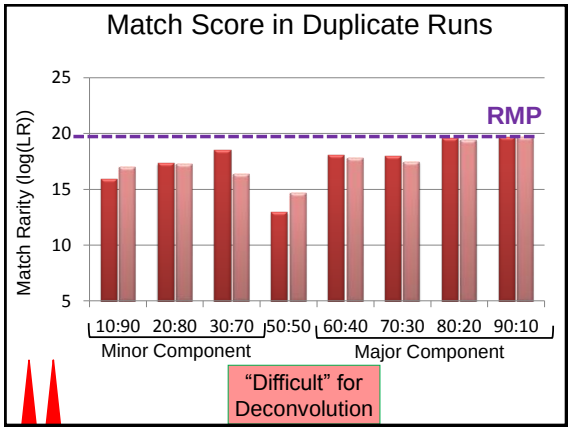
Virtual MixtureMaker - http://www.cstl.nist.gov/strbase/software.htm

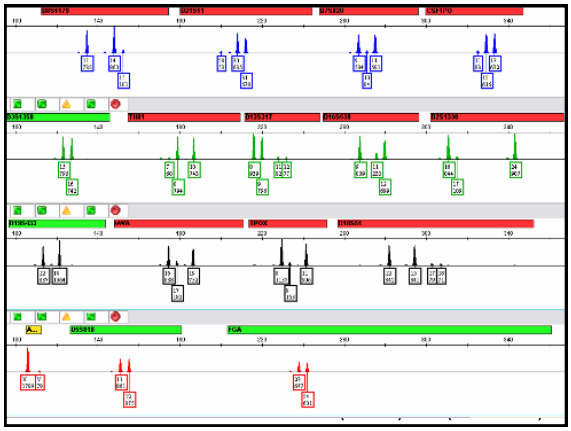
Match Score in Duplicate Runs



Match Score in Duplicate Runs

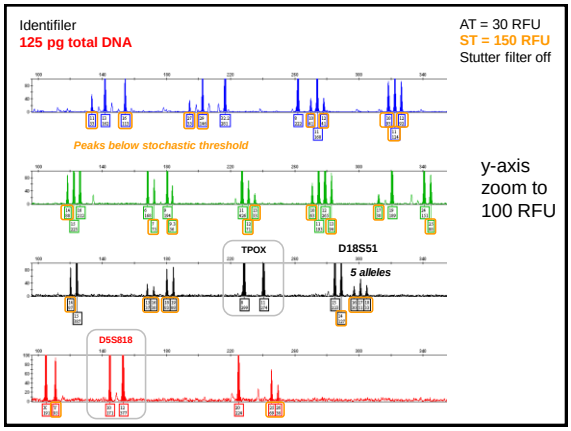


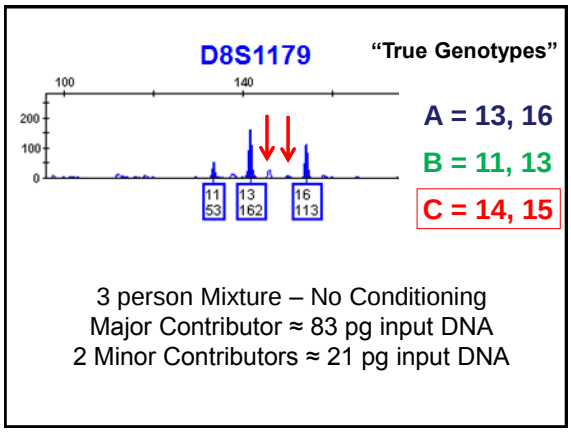


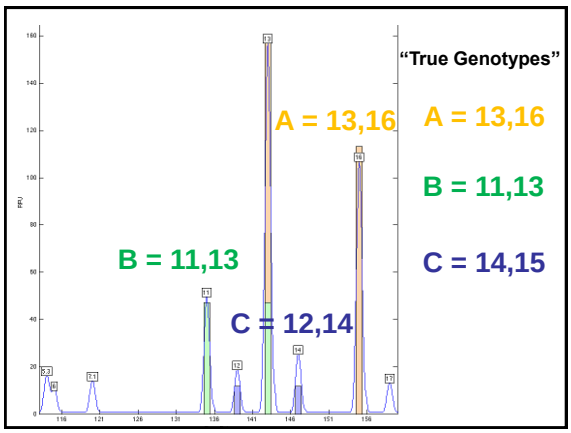


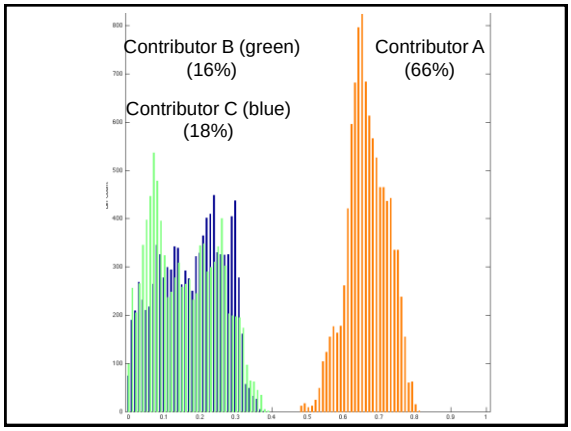
Exploring the Capabilities

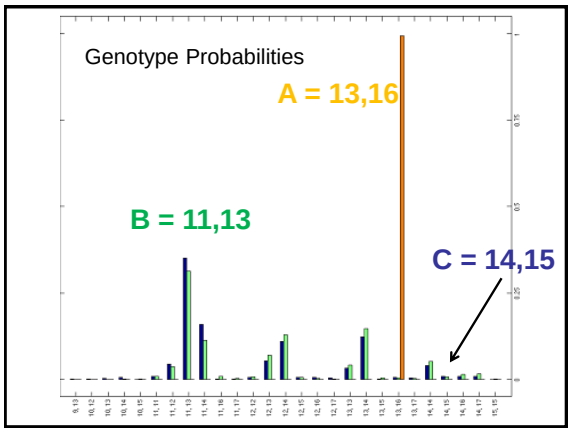
- Degree of Allele Sharing
- Mixture Ratios
- **DNA Quantity**


Identifiler
125 pg total DNA
AT = 30 RFU
ST = 150 RFU
Stutter filter off
Peaks below stochastic threshold
y-axis zoom to 100 RFU
TPOX
D18S51
5 alleles
D5S818


D8S1179
"True Genotypes"
100
140
200
100
0
11
53
13
162
16
113
A = 13, 16
B = 11, 13
C = 14, 15
3 person Mixture – No Conditioning
Major Contributor ≈ 83 pg input DNA
2 Minor Contributors ≈ 21 pg input DNA


"True Genotypes"
A = 13,16
B = 11,13
C = 14,15
A = 13,16
B = 11,13
C = 12,14





Results for Contributor A (male)

Locus	Allele Pair	Probability Likelihood	Genotype Frequency	Suspect	H_p Numerator	H_d Denominator	LR
CSF1PO	10, 11	0.572	0.1292			0.07395	
	11, 12	0.306	0.2133	1	0.30563	0.0652	
	10, 12	0.12	0.1547			0.01861	
					0.30563	0.15791	1.935
D13S317	11, 11	1	0.1149	1	1	0.11488	8.704
D8S1179	13, 16	0.998	0.0199	1	0.99786	0.0199	49.668

The match rarity between the evidence and suspect is 1.21 quintillion

Results for Contributor B (female)

Locus	Allele Pair	Probability Likelihood	Genotype Frequency	Suspect	H_p Numerator	H_d Denominator	LR
D8S1179	11, 13	0.073	0.0498	1	0.07338	0.00366	
	11, 14	0.034	0.0271			0.00092	
	13, 14	0.006	0.0996			0.00065	
	12, 14	0.011	0.0606			0.00068	
	12, 13	0.005	0.1115			0.0006	
	11, 12	0.018	0.0303			0.00054	
	14, 14	0.004	0.0271			0.00012	
	13, 13	0.003	0.0916			0.00031	
	14, 16	0.003	0.0108			0.00003	
	14, 15	0.001	0.0379			0.00003	
	etc…						**9.197**

The match rarity between the evidence and suspect is 1.43 million

Results for Contributor C (male)

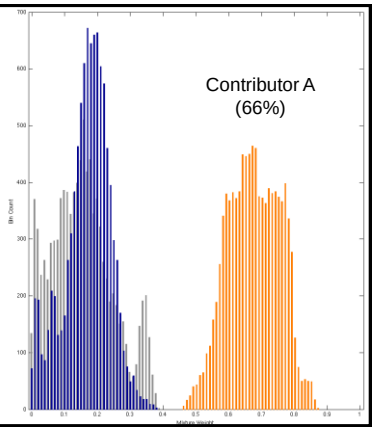
Locus	Allele Pair	Probability Likelihood	Genotype Frequency	Suspect	H_p Numerator	H_d Denominator	LR
D8S1179	11, 13	0.056	0.0498			0.00279	
	13, 14	0.007	0.0996			0.00066	
	12, 14	0.011	0.0606			0.00068	
	11, 14	0.021	0.0271			0.00056	
	12, 13	0.006	0.1115			0.00066	
	14, 14	0.005	0.0271			0.00013	
	etc…	etc…	etc…			etc…	
	14, 15	0.001	0.0379	1	0.00056	0.00002	
	12, 15	0.001	0.0424			0.00003	
	etc…	etc…	etc…			etc…	
	10, 15	0	0.0227			0.00001	
					0.00056	0.00665	**0.084**

The match rarity between the evidence and suspect is 9.16 thousand

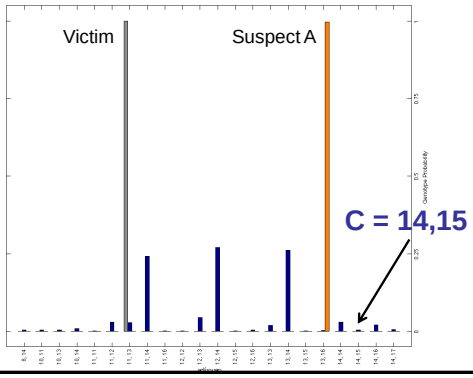
Contributor B (gray) (16%)

Contributor C (blue) (18%)

Conditioned on the Victim



The Power of Conditioning



The Power of Conditioning

	LR (no conditioning, 3unk)
Contributor A	1.21 Quintillion
Contributor B (victim)	1.43 Million
Contributor C	9.16 Thousand

	LR (conditioned on victim + 2unk)
Contributor A	1.32 Quintillion
Contributor B (victim)	2.19 Million
Contributor C	59.8 Thousand

Ranged from 1.13 to 800K

Summary

- True Allele utilizes probabilistic genotyping and makes better use of the data than the RMNE approach.
- However, the software is computer intensive. On our 4 processor system, it can take 12-16 hours to run up to four mixture samples.

Summary

- **Allele Sharing:** Stacking of alleles due to sharing creates more uncertainty.
- **Mixture Ratio:** With “distance” between the two contributors, there is greater certainty. Generally, True Allele performs better than RMNE and the classic LR with low level contributors.

Summary

- **DNA Quantity:** Generally, with high DNA signal, replicates runs on True Allele are very reproducible.
- However, with low DNA signal, higher levels of uncertainty are observed (as expected).
- There is a need to determine an appropriate threshold for an inclusion log(LR).

Future Work

- More work will be performed with low level, complex (3 and 4 person) mixtures.

Literature Review:
Value of Emerging Information

John M. Butler

NIJ National Institute of Justice

ISHI 2011 Mixture Interpretation Workshop:
Using **S**cientific **A**nalysis

BOSTON UNIVERSITY
NIST

Read to Maintain a Big Picture View!

If you are not following the recent literature, you would have missed:

- Software applications & implementation
- Impact of allele dropout on stats
- Studies on number of contributors

- The literature is changing very fast
  - Read more than *Journal of Forensic Sciences* to stay caught up
- Make time in your schedule to read and ask critical questions

Useful Articles on DNA Mixture Interpretation

- **Buckleton, J.S. and Curran, J.M. (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Sci. Int. Genet.* 2: 343-348.**
- Budowle, B., *et al.* (2009) Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J. Forensic Sci.* 54: 810-821.
- Clayton, T.M., *et al.* (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Sci. Int.* 91: 55-70.
- **Gill, P., *et al.* (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101.**
- Gill, P., *et al.* (2008) National recommendations of the technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. *FSI Genetics* 2(1): 76–82.
- Schneider, P.M., *et al.* (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. *Int. J. Legal Med.* 123: 1-5.

Several articles have been included in the workshop handouts

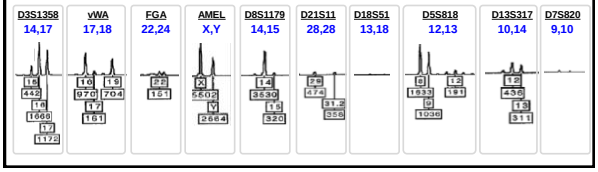
Would you include, exclude, or call inconclusive the profile shown above the mixture?

COfiler data

D3S1358	D16S539	AMEL	TH01	TPOX	CSF1PO	D7S820
14,17	9,13	X,Y	7,8	9,9	11,11	9,10

Profiler Plus data

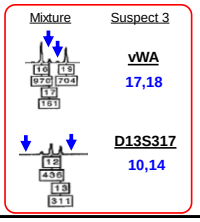
D3S1358	vWA	FGA	AMEL	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820
14,17	17,18	22,24	X,Y	14,15	28,28	13,18	12,13	10,14	9,10



Recent Article on "Problems" with Mixture Interpretation

Dror & Hampikian (2011) Subjectivity and bias in forensic DNA mixture interpretation. *Sci. Justice*, (in press)

- DNA mixture from an adjudicated criminal case – and a "suspect 3" profile -- were provided to 17 DNA analysts without contextual case information
- **Results with Suspect 3: 1 "cannot be excluded", 4 "inconclusive", and 12 "excluded"**
- Not consistent between analysts – authors suggests **subjectivity** in mixture interpretation
- Not consistent with original result – authors suggest **bias** due to availability of case context with original analysts



Know the Literature

- Sometimes articles may not be all that they claim to be – evaluate them critically
- Stay informed in order to be a good scientist
- **U**sing **S**cientific **A**nalysis involves knowing the literature (past and present)

Validation and SOPs:

Re-evaluating new kits, instruments and methods

Catherine M. Grgicak

NIJ National Institute of Justice

ISHI 2011 Mixture Interpretation Workshop:
Using **S**cientific **A**nalysis

BOSTON UNIVERSITY

NIST

Validation

- Validation is necessary to characterize and help predict and/or understand future events
  - Baseline noise - Analytical thresholds
  - Frequencies of DO - Stochastic thresholds and/or probabilities of DO
  - Stutter – stutter filters (if appropriate)
  - PHRs – compositional information related to the mixture (i.e. major/minor)

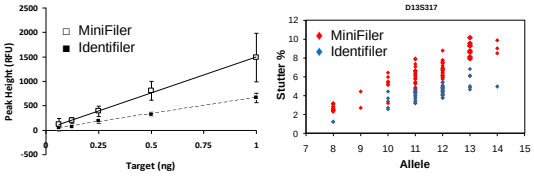
Q: Do you have to do this for every kit, instrument, method?

Can you apply parameters derived from prior studies?

- "Relevance of prior studies:
  - √ Demonstration that a comparable precision to that obtained previously can be achieved
  - √ Demonstration that use of bias results obtained previously is justified
  - √ Continued performance within statistical control as shown by regular QC sample results
- When the conditions above are met and the method is operated within its scope and field of application, it is normally acceptable to apply the data from prior studies directly to uncertainty estimates"

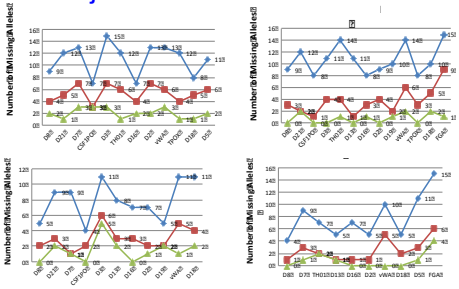
Eurachem/CITAC Guide CG 4 "Quantifying Uncertainty in Analytical Measurement" 2nd ed.

Different Kits – Do you need to re-evaluate all filters and thresholds for each kit?



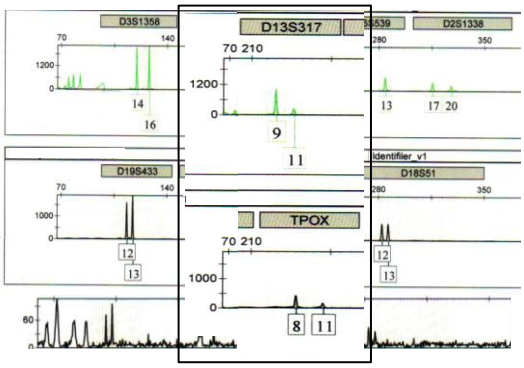
Target (ng)	Identifiler* Average Peak Height Ratio Range	Minifiler* Average Peak Height Ratio Range
0.0625	0.52-0.88	0.38-0.64
0.125	0.54-0.91	0.43-0.78
0.25	0.60-0.86	0.63-0.87
0.5	0.76-0.95	0.63-0.86
1	0.79-0.94	0.72-0.85
2	0.86-0.95	0.84-0.90
4	0.85-0.96	N/A

Different Injection Times?

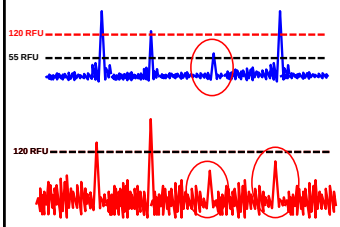


Number of missing alleles per locus per injection time for 0.0625 and 0.125 ng. (▲) 2, (■) 5 and (♦) 10 seconds injections on 3130. Number of missing allele for each locus represents the data fr om each 2 alleles in 2 different amounts with 4 replicate samples.

Post PCR Clean-up



Within one instrument



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

Courtesy of: Erica L.R. Butts National Institute of Standards and Technology, Innsbruck, Austria, September 5th, 2011

Within one instrument – over time

Thresholds – Analytical and Stochastic - are signal thresholds determined during validation

Therefore, after major instrumentation changes (i.e. laser alignment or CCD replacement), a change in sensitivity (i.e. signal to input) may result in different ST or ATs.

Within Instrument Types (i.e. 5 different 3130's)

Instrument – to – Instrument variability/sensitivities

Results in different ATs and STs for different instruments

Between Instrument Types (i.e. 3130 v. 3500)

	31xx Platforms	3500 Platforms
Laser	Argon ion (AR+) with 488/514 nm wavelength	Single-line 505 nm, solid-state, long-life laser
Power Requirement	220V	110V
File Generated	.fsa files	.hid files
Normalization	None	Instrument-to-instrument; only with AB kits
Optimal Signal Intensity	1500-3000 RFU	4x greater than 31xx platforms

Courtesy of: Erica L.R. Butts National Institute of Standards and Technology, Innsbruck, Austria, September 5th, 2011

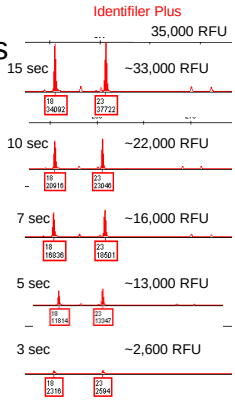
n=15: Identifiler Plus

Injection Parameters

- Injection voltage/time:
  - 1.2 kV for 15 sec
  - 1.2 kV for 10 sec
  - 1.2 kV for 7 sec
  - 1.2 kV for 5 sec
  - 1.2 kV for 3 sec

Standard injection parameters set based on samples with:

1. No pull-up present
2. No drop out present



Courtesy of: Erica L.R. Butts National Institute of Standards and Technology, Innsbruck, Austria, September 5th, 2011

Conclusion

- Stutter ratios, PH ratios, sensitivities, amp noise, LOD's change between methods
- Because 'thresholds' are in the 'signal space' each analysis parameter (i.e. AT, ST, etc) must be determined for each METHOD in the lab
  - Kit -Chemistries, injection time, post-pcr enhancements, instrument-type
- Choose an optimized method (i.e. 10 s injection, no post-PCR clean-up, choose your amp target, etc) as a default method.
- If enhanced methods are introduced new analysis parameters specific to that method are recommended.

- How far do you go?

  "What about allele size variation that could impact ATs and/or STs?. If you measure the same allele a few hundred times across the detection lanes of an array (spatial location in the CCD) or even the lifespan of the array itself, this has noticeable variations."

- "The precision should be estimated as far as possible over an extended time period, and chosen to allow natural variation of all factors affecting the result. This can be obtained from
  - The standard deviation of results for a typical sample analyzed several times over a period of time, using different analysts and equipment where possible (the results of measurements on QC check samples can provide this information).
  - The standard deviation obtained from replicate analyses performed on each of several samples. NOTE: Replicates should be performed at materially different times to obtain estimates of intermediate precision; within-batch replication provides estimates of repeatability only."

  Eurachem/CITAC Guide CG 4 "Quantifying Uncertainty in Analytical Measurement" 2nd ed.

DNA Mixtures Reference List

Background on Elements of Mixture Interpretation and Resources for Further Learning

Mixture Principles & Recommendations

Buckleton, J.S., & Curran, J.M. (2008). A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Science International: Genetics, 2*, 343-348.

Budowle, B., et al. (2009). Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *Journal of Forensic Sciences, 54*, 810-821.

DNA Advisory Board (2000) Statistical and population genetic issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s) (approved 23 February 2000). Forensic Science Communications, July 2000. Available at: http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2000/dnastat.htm.

Gill, P., et al. (2006). DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International, 160*, 90-101.

Gill, P., et al. (2008). National recommendations of the technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. *Forensic Science International: Genetics, 2*, 76-82.

Morling, N., et al. (2007). Interpretation of DNA mixtures – European consensus on principles. *Forensic Science International: Genetics, 1,* 291-292.

Schneider, P.M., et al. (2006). Editorial on the recommendations of the DNA commission of the ISFG on the interpretation of mixtures. *Forensic Science International, 160*, 89-89.

Schneider, P.M., et al. (2009). The German Stain Commission: recommendations for the interpretation of mixed stains. *International Journal of Legal Medicine, 123*, 1-5. (originally published in German in 2006 -- *Rechtsmedizin* 16:401-404).

Stringer, P., et al. (2009). Interpretation of DNA mixtures—Australian and New Zealand consensus on principles. *Forensic Science International: Genetics, 3*, 144-145.

SWGDAM (2010). SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. Available at http://www.fbi.gov/about-us/lab/codis/swgdam.pdf.

Wickenheiser, R.A. (2006). General guidelines for categorization and interpretation of mixed STR DNA profiles. *Canadian Society of Forensic Science Journal, 39*, 179-216.

Setting Thresholds

Currie, L. (1999). Detection and quantification limits: origin and historical overview. *Analytica Chimica Acta*, 391, 127–134.

Gilder, J.R., et al. (2007). Run-specific limits of detection and quantitation for STR-based DNA testing. *Journal of Forensic Sciences, 52*, 97-101.

Gill, P., et al. (2009). The *low-template-DNA* (stochastic) threshold -- its determination relative to risk analysis for national DNA databases. *Forensic Science International: Genetics, 3*, 104-111.

Gill, P. and Buckleton, J. (2010). A universal strategy to interpret DNA profiles that does not require a definition of *low-copy-number. Forensic Science International: Genetics, 4,* 221-227.

Kaiser, H. (1970). Report for analytical chemists: part II. Quantitation in elemental analysis. *Analytical Chemistry, 42,* 26A-59A.

Long, G.L., & Winefordner, J.D. (1983). Limit of detection: a closer look at the IUPAC definition. *Analytical Chemistry, 55,* 712A-724A.

Miller J.C., & Miller J.N. (2005). Errors in instrumental analysis; regression and correlation in *Statistics for Analytical Chemistry*, Ellis Horwood and Prentice Hall, pp. 101-137.

Mocak, J., Bond, A.M., Mitchell, S., & Scollary, G. (1997). A statistical overview of standard (IUPAC and ACS) and new procedures for determining the limits of detection and quantification: application to voltammetric and stripping techniques. *Pure and Applied Chemistry, 69,* 297-328.

Puch-Solis, R., et al. (2011). Practical determination of the low template DNA threshold. *Forensic Science International: Genetics, 5(5),* 422-427.

Rubinson, K.A,, & Rubinson, J.F. (2000). Sample size and major, minor, trace, and ultratrace components. *Contemporary Instrumental Analysis*. Upper Saddle River: Prentice Hall, pp. 150–158.

DNA Mixtures Reference List

Background on Elements of Mixture Interpretation and Resources for Further Learning

Stutter Products & Peak Height Ratios

Blackmore, V.L., et al. (2000). Preferential amplification and stutter observed in population database samples using the AmpFlSTR Profiler multiplex system. *Canadian Society of Forensic Sciences Journal, 33*, 23-32.

Bright, J.-A., et al. (2010). Examination of the variability in mixed DNA profile parameters for the Identifiler multiplex. *Forensic Science International: Genetics, 4,* 111-114.

Bright, J.-A., et al. (2011). Determination of the variables affecting mixed MiniFiler™ DNA profiles. *Forensic Science International: Genetics, 5(5)*, 381-385.

Brookes, C., Bright, J.A., Harbison, S., Buckleton, J. (2011). Characterising stutter in forensic STR multiplexes. *Forensic Sci Int Genet.* (in press).

Buckleton, J. (2009). Validation issues around DNA typing of low level DNA. *Forensic Science International: Genetics*, *3*, 255-260.

Buse, E.L., et al. (2003). Performance evaluation of two multiplexes used in fluorescent short tandem repeat DNA analysis. *Journal of Forensic Sciences, 48,* 348-357.

Debernardi, A., et al. (2011). One year variability of peak heights, heterozygous balance and inter-locus balance for the DNA positive control of AmpFlSTR Identifiler STR kit. *Forensic Science International: Genetics, 5(1),* 43-49.

Gibb, A.J., et al. (2009). Characterisation of forward stutter in the AmpFlSTR SGM Plus PCR. *Science & Justice, 49*, 24-31.

Gilder, J.R., et al. (2011). Magnitude-dependent variation in peak height balance at heterozygous STR loci. *International Journal of Legal Medicine, 125,* 87-94.

Gill, P., et al. (1997). Development of guidelines to designate alleles using an STR multiplex system. *Forensic Science International, 89*, 185-197.

Gill, P., et al. (1998). Interpretation of simple mixtures when artifacts such as stutters are present—with special reference to multiplex STRs used by the Forensic Science Service. *Forensic Science International, 95*, 213-224.

Hill, C.R., et al. (2011). Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems. *Forensic Science International: Genetics, 5,* 269-275.

Leclair, B., et al. (2004). Systematic analysis of stutter percentages and allele peak height and peak area ratios at heterozygous STR loci for forensic casework and database samples. *Journal of Forensic Sciences, 49*, 968-980.

Moretti, T.R., et al. (2001). Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *Journal of Forensic Sciences, 46*, 647-660.

Moretti, T.R., et al. (2001). Validation of STR typing by capillary electrophoresis. *Journal of Forensic Sciences, 46*, 661-676.

Mulero, J.J., et al. (2006). Characterization of the N+3 stutter product in the trinucleotide repeat locus DYS392. *Journal of Forensic Sciences, 51,* 1069-1073.

Wallin, J.M., et al. (1998). TWGDAM validation of the AmpFlSTR Blue PCR amplification kit for forensic casework analysis. *Journal of Forensic Sciences, 43*, 854-870.

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.

Stochastic Effects & Allele Dropout

Balding, D.J., & Buckleton, J. (2009). Interpreting low template DNA profiles. *Forensic Science International: Genetics,* 4: 1-10.

Bright, J.-A., et al. (2011). A comparison of stochastic variation in mixed and unmixed casework and synthetic samples. *Forensic Science International: Genetics,* (in press).

Gill, P., et al. (2005). A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. *Nucleic Acids Research*, *33*, 632-643.

Haned, H., et al. (2011). Estimating drop-out probabilities in forensic DNA samples: a simulation approach to evaluate different models. *Forensic Science International: Genetics, 5,* 525-531.

Kelly, H., Bright, J.A., Curran, J., Buckleton, J. (2011). The interpretation of low level DNA mixtures. *Forensic Science International: Genetics,* (in press).

Puch-Solis, R., et al. (2009). Assigning weight of DNA evidence using a continuous model that takes into account stutter and dropout. *Forensic Science International: Genetics Supplement Series*, *2*, 460-461.

Stenman, J., & Orpana, A. (2001). Accuracy in amplification. *Nature Biotechnology, 19,* 1011-1012.

Taberlet, P., et al. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research, 24*, 3189-3194.

Tvedebrink, T., et al. (2008). Amplification of DNA mixtures—missing data approach. *Forensic Science International: Genetics Supplement Series, 1,* 664-666.

Tvedebrink, T., et al. (2009). Estimating the probability of allelic drop-out of STR alleles in forensic genetics. *Forensic Science International: Genetics, 3,* 222-226.

Tvedebrink, T., et al. (2011). Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. *Forensic Science International: Genetics, (in press).*

Tvedebrink, T., et al. (2011). Allelic drop-out probabilities estimated by logistic regression – further considerations and practical implementation. *Forensic Science International: Genetics,* (in press).

Walsh, P.S., et al. (1992). Preferential PCR amplification of alleles: Mechanisms and solutions. *PCR Methods and Applications, 1*, 241-250.

Weiler, N.E.C., et al. (2011). Extending PCR conditions to reduce drop-out frequencies in low template STR typing including unequal mixtures. *Forensic Science International: Genetics, (in press).*

Estimating the Number of Contributors

Brenner, C.H., et al. (1996). Likelihood ratios for mixed stains when the number of donors cannot be agreed. *International Journal of Legal Medicine 109*, 218-219.

Buckleton, J.S., et al. (1998). Setting bounds for the likelihood ratio when multiple hypotheses are postulated. *Science & Justice 38*, 23-26.

Buckleton, J.S., et al. (2007). Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *Forensic Science International: Genetics, 1*, 20-28.

Clayton, T.M., et al. (2004). A genetic basis for anomalous band patterns encountered during DNA STR profiling. *Journal of Forensic Sciences, 49*, 1207-1214.

Egeland, T., et al. (2003). Estimating the number of contributors to a DNA profile. *International Journal of Legal Medicine, 117*, 271-275.

Ge, J., et al. (2011). Comparisons of familial DNA database searching strategies. *Journal of Forensic Sciences, (in press).*

Haned, H., et al. (2011). The predictive value of the maximum likelihood estimator of the number of contributors to a DNA mixture. *Forensic Science International: Genetics, 5(5),* 281-284.

Haned, H., et al. (2011). Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? *Journal of Forensic Sciences, 56(1),* 23-28.

Lauritzen, S.L., & Mortera, J. (2002). Bounding the number of contributors to mixed DNA stains. *Forensic Science International 130*, 125-126.

Paoletti, D.R., et al. (2005). Empirical analysis of the STR profiles resulting from conceptual mixtures. *Journal of Forensic Sciences, 50*, 1361-1366.

Paoletti, D.R., et al. (2011). Inferring the number of contributors to mixed DNA profiles. *IEEE/ACM Trans Comput Biol Bioinform.* (in press).

Perez, J., et al. (2011). Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts. *Croatian Medical Journal, 52(3)*, 314-326.

Presciuttini, S., et al. (2003) Allele sharing in first-degree and unrelated pairs of individuals in the Ge. F.I. AmpFlSTR Profiler Plus database. *Forensic Science International*, *131*, 85-89.

Tvedebrink, T., et al. (2011). Identifying contributors of DNA mixtures by means of quantitative information of STR typing. *Journal of Computative Biology, (in press)*.

Mixture Ratios

Clayton, T.M., et al. (1998). Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International*, *91*, 55-70.

Cowell, R.G., et al. (2007). Identification and separation of DNA mixtures using peak area information. *Forensic Science International*, *166*, 28-34.

Evett, I.W., et al. (1998). Taking account of peak areas when interpreting mixed DNA profiles. *Journal of Forensic Sciences*, *43*, 62-69.

Frégeau, C.J., et al. (2003). AmpFlSTR Profiler Plus short tandem repeat DNA analysis of casework samples, mixture samples, and nonhuman DNA samples amplified under reduced PCR volume conditions (25 microL). *Journal of Forensic Sciences, 48,* 1014-1034.

Gill, P., et al. (1998). Interpreting simple STR mixtures using allelic peak areas. *Forensic Science International*, 91, 41-53.

Perlin, M.W., & Szabady, B. (2001). Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *Journal of Forensic Sciences, 46*, 1372-1378.

Wang, T., et al. (2006). Least-squares deconvolution: a framework for interpreting short tandem repeat mixtures. *Journal of Forensic Sciences, 51*, 1284-1297.

Statistical Approaches

Balding, D.J. (2005) *Weight-of-evidence for Forensic DNA Profiles*. John Wiley & Sons; see mixture section on pp. 101-110.

Chung, Y.K., et al. (2010). Evaluation of DNA mixtures from database search. *Biometrics, 66,* 233-238.

Chung, Y.K., & Fung, W.K. (2011). The evidentiary values of “cold hits” in a DNA database search on two-person mixture. *Science & Justice, 51(1),* 10-15.

Curran, J.M., et al. (1999). Interpreting DNA mixtures in structured populations. *Journal of Forensic Sciences, 44*, 987-995.

Curran, J.M., & Buckleton, J. (2010). Inclusion probabilities and dropout. *Journal of Forensic Science, 55, 1171-1173.*

Devlin, B. (1993). Forensic inference from genetic markers. *Statistical Methods in Medical Research*, 2, 241–262.

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.

Evett, I.W., & Weir, B.S. (1998). *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists*. Sunderland, MA: Sinauer Associates.

Fung, W.K., & Hu, Y.-Q. (2001). The evaluation of mixed stains from different ethnic origins: general result and common cases. *International Journal of Legal Medicine, 115*, 48-53.

Fung, W.K., & Hu, Y.-Q. (2002). The statistical evaluation of DNA mixtures with contributors from different ethnic groups. *International Journal of Legal Medicine, 116*, 79-86.

Fung, W.K., & Hu, Y.-Q. (2002). Evaluating mixed stains with contributors of different ethnic groups under the NRC-II Recommendation 4.1. *Statistics in Medicine, 21,* 3583-3593.

Fung, W.K., & Hu, Y.-Q. (2008). *Statistical DNA Forensics: Theory, Methods and Computation*. Wiley: Hoboken, NJ.

Hu, Y.-Q., & Fung, W.K. (2003). Interpreting DNA mixtures with the presence of relatives. *International Journal of Legal Medicine, 117*, 39-45.

Hu, Y.-Q., & Fung, W.K. (2003). Evaluating forensic DNA mixtures with contributors of different structured ethnic origins: a computer software. *International Journal of Legal Medicine, 117*, 248-249.

Hu, Y.-Q., & Fung, W.K. (2005). Evaluation of DNA mixtures involving two pairs of relatives. *International Journal of Legal Medicine, 119(5),* 251-259.

Ladd, C., et al. (2001). Interpretation of complex forensic DNA mixtures. *Croatian Medical Journal, 42*, 244-246.

Puch-Solis, R., et al. (2010). Calculating likelihood ratios for a mixed DNA profile when a contribution from a genetic relative of a suspect is proposed. *Science & Justice, 50(4),* 205-209.

van Niewerburgh, F., et al. (2009). Impact of allelic dropout on evidential value of forensic DNA profiles using RMNE. *Bioinformatics 25*, 225-229.

van Nieuwerburgh, F., et al. (2009). RMNE probability of forensic DNA profiles with allelic drop-out. *Forensic Science International: Genetics Supplement Series, 2*, 462-463.

Weir, B.S., et al. (1997). Interpreting DNA mixtures. *Journal of Forensic Sciences 42*, 213-222.

Other Topics

Separating Cells to Avoid Mixtures

Li, C.-X., et al. (2011). New cell separation technique for the isolation and analysis of cells from biological mixtures in forensic caseworks. *Croatian Medical Journal, 52(3),* 293-298.

Rothe, J., et al. (2011). Individual specific extraction of DNA from male mixtures--First evaluation studies. *Forensic Science International: Genetics, 5*(2), 117-121.

Schneider, H., et al. (2011). Hot flakes in cold cases. *International Journal of Legal Medicine, 125,* 543-548.

Software

Bill, M., et al. (2005). PENDULUM-a guideline-based approach to the interpretation of STR mixtures. *Forensic Science International, 148*, 181-189.

Mortera, J., et al. (2003). Probabilistic expert system for DNA mixture profiling. *Theoretical and Population Biology, 63*, 191-205.

Oldroyd, N., & Shade, L.L. (2008) Expert assistant software enables forensic DNA analysts to confidently process more samples. *Forensic Magazine Dec 2008/Jan 2009*, 25-28; available at http://www.forensicmag.com/articles.asp?pid=240.

Perlin, M.W. (2006). Scientific validation of mixture interpretation methods. *Proceedings of Promega's Seventeenth International Symposium on Human Identification.* Available at http://www.promega.com/geneticidproc/ussymp17proc/oralpresentations/Perlin.pdf.

Probabilistic Genotyping Approach

Cowell, R.G., et al. (2010). Probabilistic expert systems for handling artifacts in complex DNA mixtures. *Forensic Science International: Genetics, 5(3),* 202-209.

Curran, J.M. (2008). A MCMC method for resolving two person mixtures. *Science & Justice, 48,* 168-177.

Gill, P., & Buckleton, J. (2010). Commentary on: Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forensic Sci 2009;54(4):810-21. *Journal of Forensic Sciences, 55(1),* 265-268.

Perlin, M.W., & Sinelnikov, A. (2009). An information gap in DNA evidence interpretation. *PloS ONE*, *4(12),* e8327.

Perlin, M.W., et al. (2009). Match likelihood ratio for uncertain genotypes. *Law, Probability and Risk, 8*, 289-302.

Perlin, M.W., et al. (2011). Validating TrueAllele DNA mixture interpretation. *Journal of Forensic Sciences,* (in press). doi: 10.1111/j.1556-4029.2011.01859.x.

General Information

Clayton, T., & Buckleton, J. (2005). Mixtures. Chapter 7 in *Forensic DNA Evidence Interpretation* (Eds.: Buckleton, J., Triggs, C.M., Walsh, S.J.), CRC Press, pp. 217-274.

Dror, I.E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice,* (in press), doi:10.1016/j.scijus.2011.08.004

Nurit, B., et al. (2011). Evaluating the prevalence of DNA mixtures found in fingernail samples from victims and suspects in homicide cases. *Forensic Science International: Genetics,* 5, 532-537.

Tomsey, C.S., et al. (2001). Case work guidelines and interpretation of short tandem repeat complex mixture analysis. *Croatian Medical Journal, 42*, 276-280.

Torres, Y., et al. (2003). DNA mixtures in forensic casework: a 4-year retrospective study. *Forensic Science International*, 134, 180-186.

Wetton, J.H., et al. (2011). Analysis and interpretation of mixed profiles generated by 34 cycle SGM Plus amplification. *Forensic Science International: Genetics, 5(5)*, 376-380.