Basic STR Interpretation Workshop

John M. Butler, Ph.D.

Simone N. Gittelson, Ph.D.

National Institute of Standards and Technology

Gaithersburg, Maryland, USA

31 August 2015

Basic STR Interpretation Workshop

John M. Butler, Ph.D.
Simone N. Gittelson, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

Intended Audience and Objectives

- **Intended audience:** students and beginning forensic DNA scientists (less than 5 years of experience)
- **Objectives:** To provide easy-to-follow, basic-to-intermediate level information and to introduce key concepts and fundamental literature in STR data and statistical interpretation.
- Participants should expect to come away with **an understanding of key concepts related to interpreting single-source samples and simple two-person DNA mixtures** and foundational literature to support work with STR data and statistical interpretation.

Instructor: **John M. Butler**

NIST Fellow and Special Assistant to the Director for Forensic Science at the U.S. National Institute of Standards and Technology (NIST) where he has worked for the past two decades to advance use and understanding of STR typing methods. His Ph.D. research, conducted at the FBI Laboratory, involved developing capillary electrophoresis for forensic DNA analysis. The most recent of his five textbooks forms the basis for this workshop.

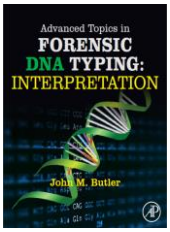
Phone: +1-301-975-4049
Email: john.butler@nist.gov

Instructor: **Simon N. Gittelson**

Forensic statistician in the NIST Statistical Engineering Division. She conducted her Ph.D. research at the University of Lausanne (Switzerland) in applying probability and decision theory to inference and decision problems in forensic science. She then specialized in the interpretation of DNA evidence during her postdoc at NIST and the University of Washington.

Phone: +1-301-975-4892
Email: simone.gittelson@nist.gov

Resources

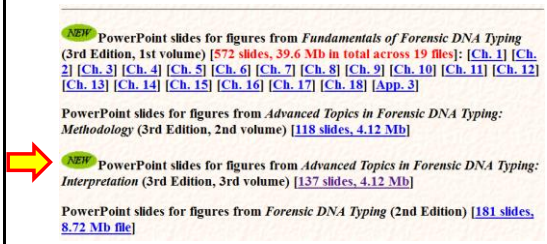


Available since October 2014

- Butler, J.M. (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*. Elsevier Academic Press: San Diego.
  - All figures available on STRBase: http://www.cstl.nist.gov/strbase/training.htm
- Boston University DNA Mixture Training: http://www.bu.edu/dnamixtures/
- STRBase DNA Mixture Information: http://www.cstl.nist.gov/strbase/mixture.htm

Slides Available for Use from *Forensic DNA Typing* Books

http://www.cstl.nist.gov/strbase/training.htm



NEW PowerPoint slides for figures from *Fundamentals of Forensic DNA Typing* (3rd Edition, 1st volume) [572 slides, 39.6 Mb in total across 19 files]: [Ch. 1] [Ch. 2] [Ch. 3] [Ch. 4] [Ch. 5] [Ch. 6] [Ch. 7] [Ch. 8] [Ch. 9] [Ch. 10] [Ch. 11] [Ch. 12] [Ch. 13] [Ch. 14] [Ch. 15] [Ch. 16] [Ch. 17] [Ch. 18] [App. 3]

PowerPoint slides for figures from *Advanced Topics in Forensic DNA Typing: Methodology* (3rd Edition, 2nd volume) [118 slides, 4.12 Mb]

NEW PowerPoint slides for figures from *Advanced Topics in Forensic DNA Typing: Interpretation* (3rd Edition, 3rd volume) [137 slides, 4.12 Mb]

PowerPoint slides for figures from *Forensic DNA Typing* (2nd Edition) [181 slides, 8.72 Mb file]

Workshop Schedule

Time	Module (Instructor)	Topics
0900-0930	Welcome & Introductions	Review expectations and questions from participants
0930 – 1100	Data Interpretation 1 (John)	STR kits, loci, alleles, genotypes, profiles Data interpretation thresholds and models Simple PCR and CE troubleshooting
1100 – 1130	Break	
1130 – 1300	Statistical Interpretation 1 (Simone)	Introduction to probability and statistics STR population data collection, calculations, and use Approaches to calculating match probabilities
1300 – 1430	Lunch	
1430 – 1600	Data Interpretation 2 (John)	Mixture interpretation: Clayton rules, # contributors Stochastic effects and low-template DNA challenges Worked examples
1600 – 1630	Break	
1630 – 1800	Statistical Interpretation 2 (Simone)	Approaches to calculating mixture statistics Likelihood ratios and formulating propositions Worked examples

What We Will Not Cover…

- Handling low-template DNA information
- Probabilistic genotyping for DNA mixtures
- Complicated mixtures with >2 contributors
- Y-STRs or other lineage markers
- Kinship analysis and dealing with relatives

Other ISFG 2015 workshops, being held tomorrow, cover more advanced aspects of DNA interpretation:

a) **The interpretation of complex DNA profiles using open-source software LRmix Studio and EuroForMix (EFM)** - Peter Gill, Hinda Haned, Corina Benschop, Oskar Hansson, Oyvind Bleka

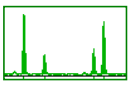
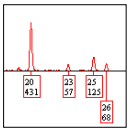
b) **Interpretation of complex DNA profiles using a continuous model – an introduction to STRmix™** - John Buckleton, Jo-Anne Bright, Catherine McGovern, Duncan Taylor

c) **Kinship analysis** - Thore Egeland, Klaas Slooten

DNA Interpretation Training Workshops

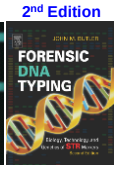
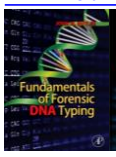
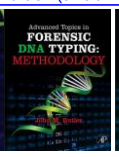
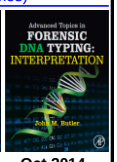
Handouts and reference list available at
http://www.cstl.nist.gov/strbase/training/ISFG2013workshops.htm

Math Analogy to DNA Evidence

$2 + 2 = 4$	$2x^2 + x = 10$	$\int_{x=0}^{\infty} f(x)dx$
Basic Arithmetic	**Algebra**	**Calculus**
		
Single-Source DNA Profile (DNA databasing)	**Sexual Assault Evidence** (2-person mixture with high-levels of DNA)	**Touch Evidence** (>2-person, low-level, complex mixtures perhaps involving relatives)
Morning discussion	Afternoon discussion	

http://www.cstl.nist.gov/strbase/pub_pres/Butler-DNA-interpretation-AAFS2015.pdf

DNA Mixture Information Coverage
in *Forensic DNA Typing* Textbooks

1st Edition	2nd Edition	3rd Edition (3 volumes)		
				
Jan 2001	Feb 2005	Sept 2009	Aug 2011	Oct 2014
335 pages	688 pages	520 pages	704 pages	608 pages
13 pages	25 pages	10 pages	1 page p. 235	**126 pages** Chapters 6, 7, 12, 13 Appendix 4 (low-level, 2-person example)

Introduce Real-Time Response Clickers

- Questions will be provided throughout our presentations
- **Click on response that best represents your opinion (do not take too long to respond)**
- We will review results obtained from the group
- Please leave your clicker behind at the end of class

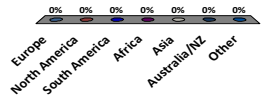
Do Not Take a Clicker Home with You!

TEST QUESTION

Where are you from?

A. Europe
B. North America
C. South America
D. Africa
E. Asia
F. Australia/NZ
G. Other

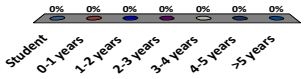
Response Counter



Your experience with forensic DNA?

A. Student
B. 0-1 years
C. 1-2 years
D. 2-3 years
E. 3-4 years
F. 4-5 years
G. >5 years

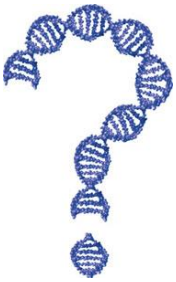
Response Counter



Background of Participants…

Without Your Clicker…

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



Ask Questions!

- If you feel uncomfortable asking questions in front of everyone, please write your question down and give it to us at a break
- We will read the question and attempt to answer it in front of the group
- Or raise your hand and ask a question at any time!

Greg Matheson on Forensic Science Philosophy

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

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

D.N.A. Approach to Understanding

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

Adapted from David A. Bednar, *Increase in Learning* (Deseret Book, 2011)

Basic STR Interpretation Workshop
John M. Butler & Simone N. Gittelson
Krakow, Poland
31 August 2015

Data Interpretation 1:

STR kits, loci, alleles, genotypes, profiles
Data interpretation thresholds and models
Simple PCR and CE troubleshooting

John M. Butler, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

Workshop Schedule

Time	Module (Instructor)	Topics
0900-0930	Welcome & Introductions	Review expectations and questions from participants
0930 – 1100	Data Interpretation 1 (John)	STR kits, loci, alleles, genotypes, profiles Data interpretation thresholds and models Simple PCR and CE troubleshooting
1100 – 1130	Break	
1130 – 1300	Statistical Interpretation 1 (Simone)	Introduction to probability and statistics STR population data collection, calculations, and use Approaches to calculating match probabilities
1300 – 1430	Lunch	
1430 – 1600	Data Interpretation 2 (John)	Mixture interpretation: Clayton rules, # contributors Stochastic effects and low-template DNA challenges Worked examples
1600 – 1630	Break	
1630 – 1800	Statistical Interpretation 2 (Simone)	Approaches to calculating mixture statistics Likelihood ratios and formulating propositions Worked examples

Acknowledgment and Disclaimers

I will quote from my recent book entitled "Advanced Topics in Forensic DNA Typing: Interpretation" (Elsevier, 2015). I do not receive any royalties for this book. Completing this book was part of my job last year at NIST.

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

I have been fortunate to have had discussions with numerous scientists on interpretation issues including Mike Coble, Bruce Heidebrecht, Robin Cotton, Charlotte Word, Catherine Grgicak, Peter Gill, Ian Evett …

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

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

Steps in Forensic DNA Analysis

Gathering the Data

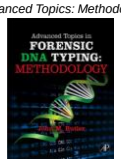
Understanding Results Obtained & Sharing Them

Collection/Storage/ Characterization → Extraction/ Quantitation → Amplification/ Marker Sets → Separation/ Detection → Data → Stats → Report

Interpretation

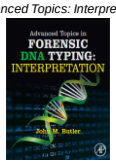
>1300 pages of information with >5000 references cited in these two books

Advanced Topics: Methodology



August 2011

Advanced Topics: Interpretation



October 2014

Understanding Results Obtained & Sharing Them

Data → Stats → Report

Interpretation

Presentation Outline

1. Data interpretation overview
2. Data collection with ABI Genetic Analyzers
3. STR alleles and PCR amplification artifacts
4. STR genotypes and heterozygote balance
5. STR profiles and tri-allelic patterns
6. …
7. …
8. Troubleshooting data collection

These points correspond to chapter numbers in
Advanced Topics in Forensic DNA Typing: Interpretation (2015)

How Book Chapters Map to Data Interpretation Process

Chapter	Input Information	Decision to be made	How decision is made
2	Data file	**Peak** or **Noise**	Analytical threshold
3	Peak	**Allele** or **Artifact**	Stutter threshold; precision sizing bin
4	Allele	**Heterozygote** or **Homozygote** or **Allele(s) missing**	Peak heights and peak height ratios; stochastic threshold
5	Genotype/full profile	**Single-source** or **Mixture**	Numbers of peaks per locus
6	Mixture	Deconvolution or not	Major/minor mixture ratio
7	Low level DNA	Interpret or not	Complexity threshold
In Data Interp2 presentation			
8	Poor quality data	Replace CE components (buffer, polymer, array) or call service engineer	Review size standard data quality with understanding of CE principles

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Table 1.1**, p. 6

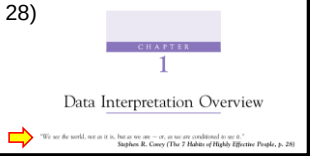
Data Interpretation Overview

Advanced Topics in Forensic DNA Typing: Interpretation, Chapter 1

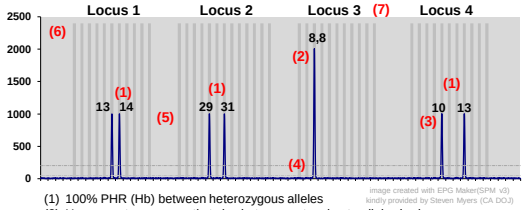
Our Backgrounds Influence Our Interpretation

We see the world, not as it is, but as we are – or, as we are conditioned to see it.

- Stephen R. Covey (*The 7 Habits of Highly Effective People*, p. 28)



Using **Ideal Data** to Discuss Principles



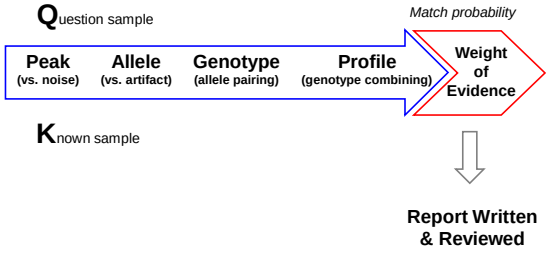
(1) 100% PHR (Hb) between heterozygous alleles
(2) Homozygotes are exactly twice heterozygotes due to allele sharing
(3) No peak height differences exist due to size spread in alleles (any combination of resolvable alleles produces 100% PHR)
(4) No stutter artifacts enabling mixture detection at low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

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 1.5**, p. 11

Challenges in Real-World Data

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

Steps in DNA Interpretation



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 1.3**, p. 7

Overview of the SWGDAM 2010 Interp Guidelines

See http://www.swgdam.org/

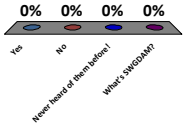
1. Preliminary evaluation of data – **is something a peak and is the analysis method working properly?**
2. Allele designation – **calling peaks as alleles**
3. Interpretation of DNA typing results – **using the allele information to make a determination about the sample**
   1. Non-allelic peaks
   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

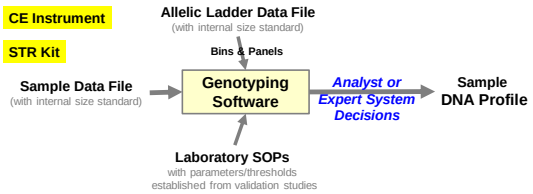
Have you read the 2010 SWGDAM STR Interpretation Guidelines?

1. Yes
2. No
3. Never heard of them before!
4. What's SWGDAM?

Response Counter



Overview of Data Interpretation Process



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 1.2**, p. 5

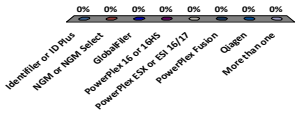
Questions for Workshop Participants

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

Autosomal STR Kit(s) in Your Laboratory?

1. Identifiler or ID Plus
2. NGM or NGM Select
3. GlobalFiler
4. PowerPlex 16 or 16HS
5. PowerPlex ESX or ESI 16/17
6. PowerPlex Fusion
7. Qiagen
8. More than one

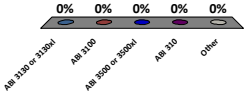
Response Counter



CE Instrumentation in Your Laboratory?

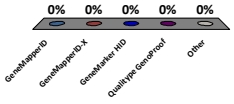
A. ABI 3130 or 3130xl
B. ABI 3100
C. ABI 3500 or 3500xl
D. ABI 310
E. Other

Response Counter



Analysis Software in Your Laboratory?

A. GeneMapper*ID*
B. GeneMapper*ID-X*
C. GeneMarker HID
D. Qualitype GenoProof
E. Other



Response Counter

Types of STR Repeat Units

Requires size based DNA separation to resolve different alleles from one another

High stutter

- **Di**nucleotide (CA)(CA)(CA)(CA)
- **Tri**nucleotide (GCC)(GCC)(GCC)
- **Tetra**nucleotide (AATG)(AATG)(AATG)
- **Penta**nucleotide (AGAAA)(AGAAA)
- **Hexa**nucleotide (AGTACA)(AGTACA)

Low stutter

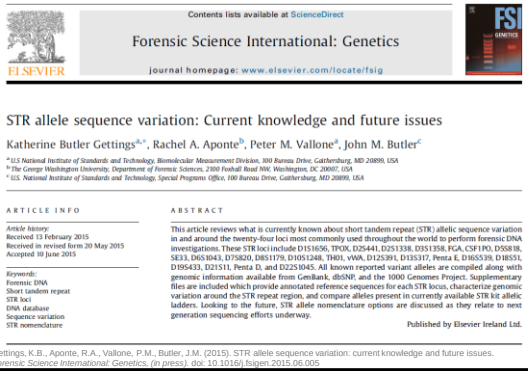
Short tandem repeat (STR) = microsatellite = simple sequence repeat (SSR)

Categories for STR Markers

Category	Example Repeat Structure	13 CODIS Loci
Simple repeats – contain units of identical length and sequence	(GATA)(GATA)(GATA)	TPOX, CSF1PO, D5S818, D13S317, D16S539
Simple repeats with non-consensus alleles (e.g., TH01 9.3)	(GATA)(GAT-)(GATA)	TH01, D18S51, D7S820
Compound repeats – comprise two or more adjacent simple repeats	(GATA)(GATA)(GACA)	VWA, FGA, D3S1358, D8S1179
Complex repeats – contain several repeat blocks of variable unit length	(GATA)(GACA)(CA)(CATA)	D21S11

These categories were first described by Urquhart *et al.* (1994) *Int. J. Legal Med.* 107:13-20

Review of STR Allele Sequence Variation



Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

STR allele sequence variation: Current knowledge and future issues

Katherine Butler Gettings[a,*], Rachel A. Aponte[b], Peter M. Vallone[a], John M. Butler[c]

[a] US National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD 20899, USA
[b] The George Washington University, Department of Forensic Sciences, 2300 Foxhall Road NW, Washington, DC 20007, USA
[c] U.S. National Institute of Standards and Technology, Special Programs Office, 100 Bureau Drive, Gaithersburg, MD 20899, USA

ARTICLE INFO

Article history:
Received 13 February 2015
Received in revised form 20 May 2015
Accepted 10 June 2015

Keywords:
Forensic DNA
Short tandem repeat
STR loci
DNA database
Sequence variation
STR nomenclature

ABSTRACT

This article reviews what is currently known about short tandem repeat (STR) allelic sequence variation in and around the twenty-four loci most commonly used throughout the world to perform forensic DNA investigations. These STR loci include D1S1656, TPOX, D2S441, D2S1338, D3S1358, FGA, CSF1PO, D5S818, SE33, D6S1043, D7S820, D8S1179, D10S1248, TH01, vWA, D12S391, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D, and D22S1045. All known reported variant alleles are compiled along with genomic information available from GenBank, dbSNP, and the 1000 Genomes Project. Supplementary files are included which provide annotated reference sequences for each STR locus, characterize genomic variation around the STR repeat region, and compare alleles present in currently available STR kit allelic ladders. Looking to the future, STR allele nomenclature options are discussed as they relate to next generation sequencing efforts underway.

Published by Elsevier Ireland Ltd.

Gettings, K.B., Aponte, R.A., Vallone, P.M., Butler, J.M. (2015). STR allele sequence variation: current knowledge and future issues. *Forensic Science International: Genetics, (in press)*. doi: 10.1016/j.fsigen.2015.06.005

STR Loci Currently in Use

*Currently there are **29 autosomal STR markers** present in commercial kits*

13 CODIS loci

+5 additional loci
In PowerPlex CS7
F13B
FES/FPS
F13A01
LPL
Penta C

U.S.	Europe
TPOX	
CSF1PO	
D5S818	
D7S820	
D13S317	
FGA	FGA
vWA	vWA
D3S1358	D3S1358
D8S1179	D8S1179
D18S51	D18S51
D21S11	D21S11
TH01	TH01
D16S539	D16S539
D2S1338	D2S1338
D19S433	D19S433
Penta D	
Penta E	
	D12S391
	D1S1656
	D2S441
	D10S1248
	D22S1045
	SE33

ESS = European Standard Set

7 ESS loci

3 miniSTR loci developed at NIST

5 loci adopted in 2009 to expand to 12 ESS loci

Locus used in China ← D6S1043

SE33 → Core locus for Germany

European Expansion Efforts

More loci added as databases grew…

European Standard Set (ESS) of Loci
1998: 4 STRs
1999: 7 STRs
2009: 12 STRs
(but 15 with most kits)

1995 UK began with 6 STRs (SGM)
UK National Database launched

1998 Martin et al. (2001) Multiple countries have established national databases

1999 Leriche et al. (1998) Initial Interpol ESS selection (4 STRs)

2001 ESS increased to 7 STRs; UK went to 10 STRs (SGM Plus)

2004 Degraded DNA interlab study (EDNAP)

2005 **Prüm treaty** (data sharing)

2006 **EDNAP/ENFSI recommend new loci**

2007 Gill et al. (2006a, 2006b) Letters to editor (1) & (2) announcing proposed new loci - miniSTRs advocated

2009 **ESS expanded to 12 loci**

2010 Dixon et al. (2006) EDNAP interlab published

2011 Implementation required

Prototype kits developed & tested

NGM and NGM SElect; Qiagen kits

PowerPlex ESI/ESX 16/17 kits released

With expanded loci selections, focus is on casework capabilities (miniSTRs, increased sensitivity)

U.S. Core Loci Expansion Efforts

More loci added as databases grew…

Implementation to be required 2 years after announcement

U.S. Core Loci Goals Announced
1990: 4 VNTRs (RFLP)
1997: 13 STRs (PCR)
2011: 20+ STRs

1990 U.S. began with 4 RFLP VNTRs
1994 DNA Identification Act (federal law)
1996 Early STR kits
1997 Baechtel et al. (1991) Combined DNA Index System (CODIS) proposed
1998 Launch of U.S. National DNA Database (NDIS) *Initial QAS released*
2000 Budowle et al. (1998) Initial CODIS core loci (13 STRs)
2002 NDIS exceeds 1 million profiles
2003 16plex STR kits
2009 Europe expands from 7 to 12 loci
2011 NDIS exceeds 10 million profiles
2012 President's DNA Initiative Debbie Smith Act increases funding for DNA databases
2015 PowerPlex Fusion and GlobalFiler **24plex kits available**

Hares (2012a, 2012b) Letters to editor (1) & (2) announcing proposed new loci

Butler et al. (2012) NIST population data published covering all 29 kit STR loci

CODIS Core Loci WG recommend new loci

QAS: Quality Assurance Standards

U.S. is Moving to 20 Core Loci

Locus
CSF1PO
D3S1358
D5S818
D7S820
D8S1179
D13S317
D16S539
D18S51
D21S11
FGA
TH01
TPOX
vWA
D1S1656
D2S441
D2S1338
D10S1248
D12S391
D19S433
D22S1045

Red is for original CODIS Core 13 Loci.
Blue is for new additional CODIS Core Loci.

Forensic Science International: Genetics 17 (2015) 33–34
Contents lists available at ScienceDirect
Forensic Science International: Genetics
journal homepage: www.elsevier.com/locate/fsig

Letter to the Editor **Required in U.S. starting January 1, 2017**

Selection and implementation of expanded CODIS core loci in the United States

"The CODIS Core Loci Working Group selected a consortium of 11 CODIS laboratories…these laboratories performed validation experiments…

With the assistance of the National Institute of Standards and Technology (NIST), the data generated through these validation studies were compiled, reviewed and analyzed."

Hares, D.R. (2015) Selection and implementation of expanded CODIS core loci in the United States. *Forensic Sci. Int. Genet.* 17:33-34

Position of Forensic STR Markers on Human Chromosomes

Core STR Loci for the United States

13 Core U.S. STR Loci

TPOX, D3S1358, D2S441, D5S818, D1S1656, D2S1338, FGA, CSF1PO, D7S820, D8S1179, TH01, VWA, D10S1248, D12S391

1997 (13 loci) → 2017 (20 loci)

15 STR loci overlap between U.S. and Europe

AMEL *Sex-typing*

D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, AMEL

Value of STR Kits

Advantages

- Quality control of materials is in the hands of the manufacturer (saves time for the end-user)
- Improves consistency in results across laboratories – same allelic ladders used
- Common loci and PCR conditions used – aids DNA databasing efforts
- Simpler for the user to obtain results

Disadvantages

- Contents may not be completely known to the user (e.g., primer sequences)
- Higher cost to obtain results

Different DNA Tests from Various STR Kits

Kit Name	# STR Loci Tested	Manufacturer	Why Used?
Identifiler, **Identifiler Plus***	**15 autosomal STRs** (aSTRs) & amelogenin	ThermoFisher (Applied Biosystems)	Covers the 13 core CODIS loci plus 2 extra
PowerPlex 16 **PowerPlex 16 HS***	**15 aSTRs** & amelogenin	Promega Corporation	Covers the 13 core CODIS loci plus 2 extra
Profiler Plus & COfiler (2 different kits)	**13 aSTRs** [9 + 6 with 2 overlapping] & amelogenin	ThermoFisher (Applied Biosystems)	Original kits used to provide 13 CODIS STRs
Yfiler	**17 Y-chromosome STRs**	ThermoFisher (Applied Biosystems)	Male-specific DNA test
MiniFiler	**8 aSTRs** & amelogenin	ThermoFisher (Applied Biosystems)	Smaller regions examined to aid degraded DNA recovery
GlobalFiler*	**21 aSTRs**, DYS391, Y indel, & amelogenin	ThermoFisher (Applied Biosystems)	Covers expanded US core loci
PowerPlex Fusion*	**22 aSTRs**, DYS391, & amelogenin	Promega Corporation	Covers expanded US core loci; 5-dye
Investigator 24plex*	**21 aSTRs**, DYS391, quality sensor, & amelogenin	Qiagen	Covers expanded US core loci; quality sensors

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

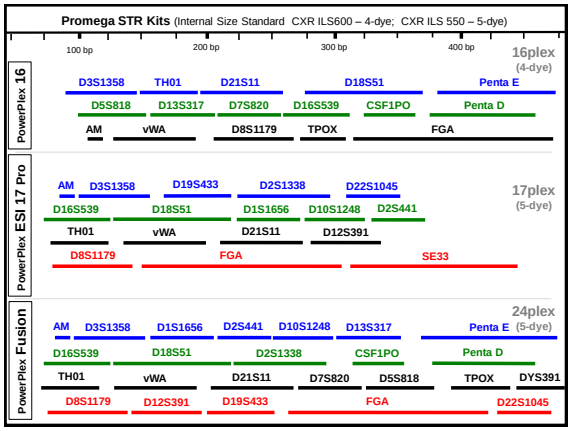
ThermoFisher/ABI STR Kits (Internal Size Standard LIZ GS500 – 5-dye; LIZ GS600 – 6-dye)

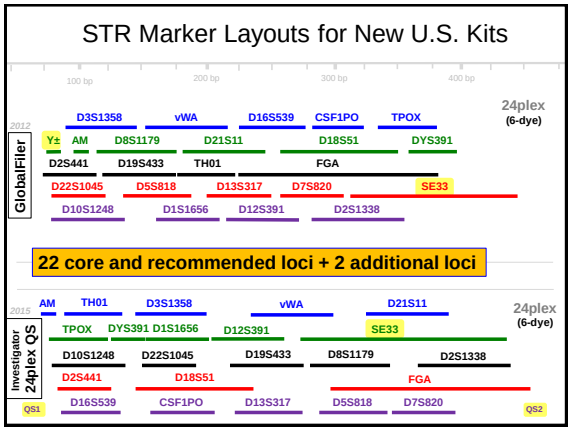
100 bp 200 bp 300 bp 400 bp

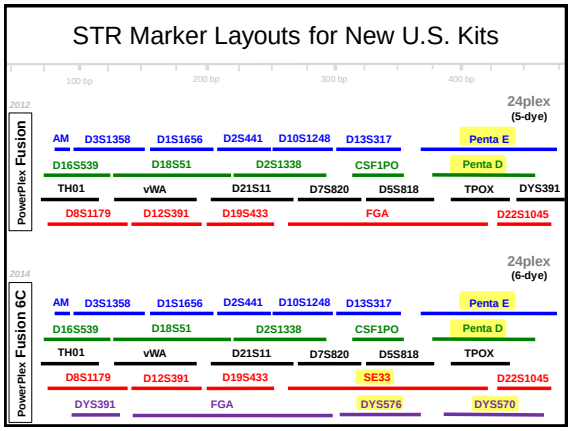
Identifiler (16plex, 5-dye)
D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, AM, D5S818, FGA

NGM SElect (17plex, 5-dye)
D10S1248, vWA, D16S539, D2S1338, AM, D8S1179, D21S11, D18S51, D22S1045, D19S433, TH01, FGA, D2S441, D3S1358, D1S1656, D12S391, SE33

GlobalFiler (24plex, 6-dye)
D3S1358, vWA, D16S539, CSF1PO, TPOX, Y±, AM, D8S1179, D21S11, D18S51, DYS391, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S391, D2S1338







STR Kits and Dye Sets Used

Sets virtual filter

Example STR Kits	Dye Labels	Dye Set
Profiler Plus, SGM Plus, COfiler, Profiler	5-FAM, JOE, NED, ROX	F
Identifiler, MiniFiler, NGM, NGM SElect	6-FAM, VIC, NED, PET, LIZ	G5
GlobalFiler	6-FAM, VIC, NED, TAZ, SID, LIZ	J6 (3500)
PowerPlex 16, 16HS	FL, JOE, TMR, CXR	F
PowerPlex ESI 16/17, ESX 16/17, 18D, 21, Fusion	FL, JOE, TMR-ET, CXR-ET, CC5	G5
Qiagen Investigator Kits	B, G, Y, R, O	G5
Research assays	6-FAM, TET, HEX, ROX	C

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, Table 5.1, p. 111

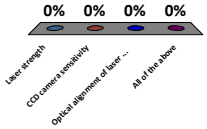
ABI Genetic Analyzer Data Collection

Advanced Topics in Forensic DNA Typing: Interpretation, Chapter 2

Advanced Topics in Forensic DNA Typing: Methodology, Chapter 6

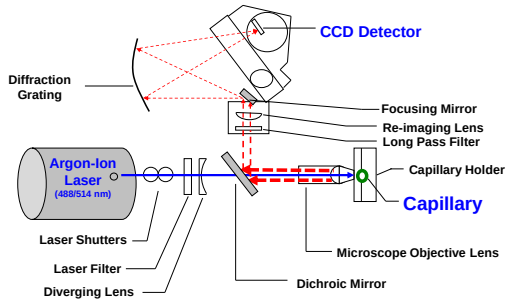
What is the primary reason for CE instrument sensitivity variation?

A. Laser strength
B. CCD camera sensitivity
C. Optical alignment of laser and detector
D. All of the above



Response Counter

Optics for ABI 310 Genetic Analyzer



J.M. Butler (2012) *Advanced Topics in Forensic DNA Typing: Methodology*, Figure 6.6

Key Points

- **On-scale data** of STR allele peaks are important to interpretation (both lower and upper limits exist for reliable data)
- Data signals from ABI Genetic Analyzers are processed by **proprietary algorithms** that include variable binning (adjustment for less sensitive fluorescent dyes), baselining, smoothing, and multi-componenting for separating color channels
- **Instrument sensitivities vary** due to different lasers, detectors, and optical alignment (remember that signal strength is in "relative fluorescence units", RFUs)

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

Off-scale data with flat-topped peaks

(a)

Too much DNA amplified

Or injected onto CE

Allele dropout due to stochastic effects

(b)

Too little DNA amplified

"Just right"

(c)

Within optimal range

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

Applied Biosystems (ABI)

A BRIEF HISTORY OF NAME CHANGES FOR APPLIED BIOSYSTEMS

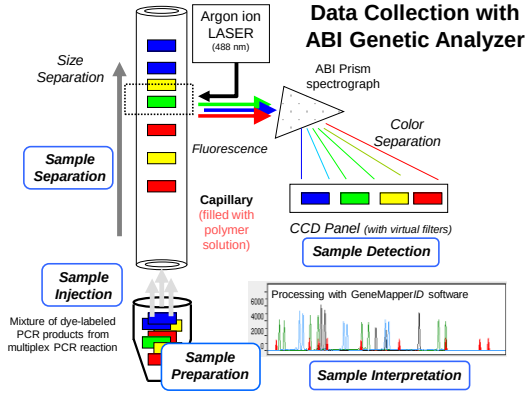
Applied Biosystems, Inc. (ABI) has dominated the genetic instrument business for the past three decades. What began as a small biotechnology company in 1981 named GeneCo, for Genetic Systems Company, has grown through mergers and acquisitions over the years. Along with this growth has come a number of name changes, as noted in the table below. In February 2014, Life Technologies became a premier brand of Thermo Fisher Scientific.

Year	Company Name
1981	Genetic Systems Company (GeneCo)
1982	Applied Biosystems, Inc. (ABI)
1993	Applied Biosystems Division of Perkin-Elmer
1996	PE Applied Biosystems
1998	PE Biosystems
1999	PE Biosystems Group of PE Corporation
2000	Applied Biosystems Group of Applera Corporation
2002	Applied Biosystems (AB)
2008	Life Technologies (after Applied Biosystems merged with Invitrogen)
2014	Thermo Fisher Scientific (Life Technologies is a premier brand)

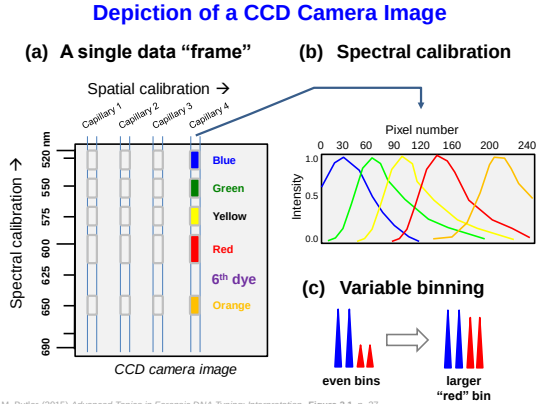
Sources: http://www.lifetechnologies.com/us/en/home/about-us/news-gallery/company-fact-sheet/company-history.html; http://en.wikipedia.org/wiki/Applied_Biosystems; author's personal experience over the past 20 years.

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, Box 2.1, p. 26

Data Collection with ABI Genetic Analyzer

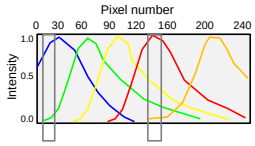


Depiction of a CCD Camera Image



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, Figure 2.1, p. 27

Even Bins



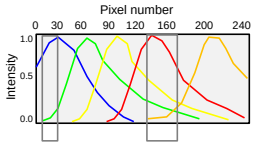
"blue" channel **"red" channel**

The red dye naturally has a lower signal output

Relative Signal Produced

even bins

Variable Bins



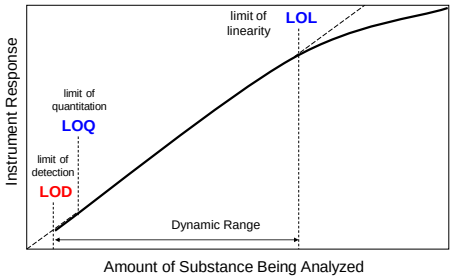
"blue" channel **wider "red" channel**

More light signal is collected in the red region using a wider virtual bin to increase apparent peak heights

Relative Signal Produced

with larger "red" bin

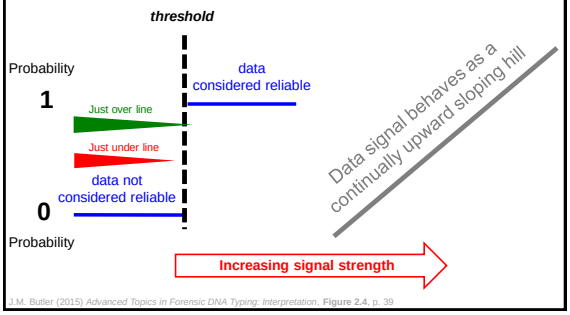
Useful Range of an Analytical Method



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 2.3**, p. 31

Dilemma of a Threshold in a Continuous World

(a) Binary Threshold Applied **(b) Continuous Data**



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 2.4**, p. 39

Impact of Setting Thresholds Too High or Too Low

If	Then
Threshold is set too high…	Analysis may miss low-level legitimate peaks (**false negative conclusions produced**)
Threshold is set too low…	Analysis will take longer as artifacts and baseline noise must be removed from consideration as true peaks during data review (**false positive conclusions produced**)

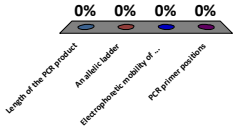
J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Table 2.3**, p. 44

STR Alleles and PCR Amplification Artifacts

Advanced Topics in Forensic DNA Typing: Interpretation, Chapter 3

What answer best describes what determines the "size" of an STR allele?

A. Length of the PCR product
B. An allelic ladder
C. Electrophoretic mobility of labeled DNA molecules relative to an internal size standard
D. PCR primer positions

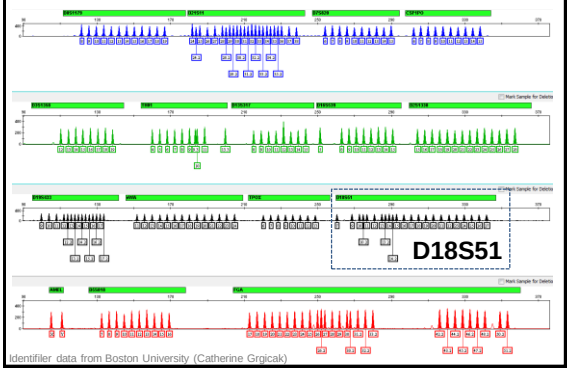


Key Points

- STR allele designations are made by comparing the relative size of sample peaks to allelic ladder allele sizes
- A common, calibrated STR allele nomenclature is essential in order to compare data among laboratories
- STR allele sizes are based on a measure of the relative electrophoretic mobility of amplified PCR products (defined by primer positions) compared to an internal size standard using a specific sizing algorithm
- STR alleles can vary in their overall length (number of repeat units), with their internal sequence of repeats, and in the flanking region

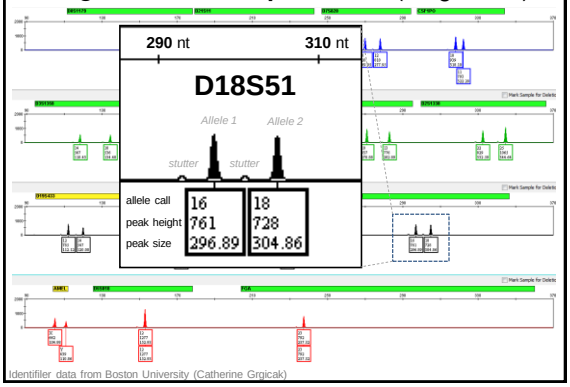
Identifiler STR Kit Allelic Ladder

(internal size standard not shown)



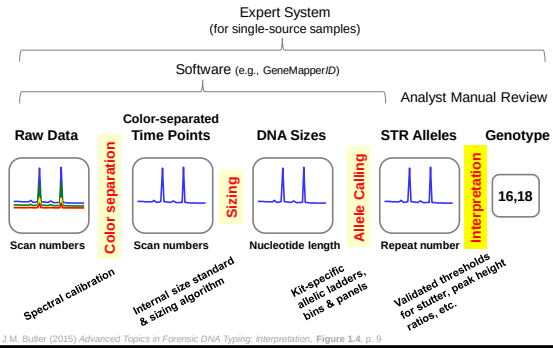
Identifiler data from Boston University (Catherine Grgicak)

Single-Source Sample Profile (1 ng of "C")



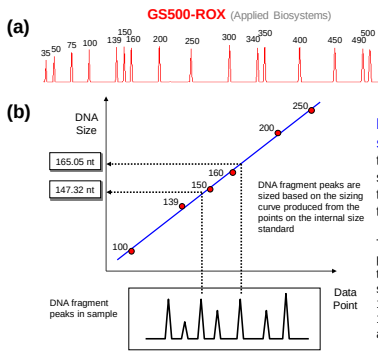
Identifiler data from Boston University (Catherine Grgicak)

Transformation of Information at a Single STR Locus during Data Processing



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 1.4**, p. 9

DNA Size Standard and Sizing Algorithm

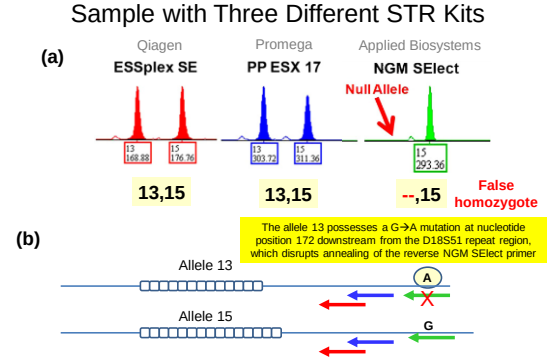


Local Southern sizing algorithm uses two peaks (from the size standard) above and two peaks below below the unknown peak

The 147.32 nt example peak would be defined by the relative positions of the size standard 100 nt and 139 nt peaks below and the 150 nt and 160 nt peaks above

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Box 3.1**, p. 54

D18S51 Typing Results on the Same DNA Sample with Three Different STR Kits



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 4.8**, p. 100

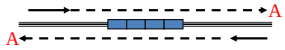
Null Alleles

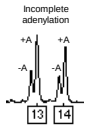
- Allele is present in the DNA sample but fails to be amplified due to a **nucleotide change in a primer binding site**
- Allele dropout is a problem because a heterozygous sample appears falsely as a homozygote
- Two PCR primer sets can yield different results on samples originating from the same source
- This phenomenon impacts DNA databases
- Large concordance studies are typically performed prior to use of new STR kits

Non-Template Addition

- Taq polymerase will often add an extra nucleotide to the end of a PCR product; most often an "A" (termed "adenylation")
- **Dependent on 5'-end of the reverse primer; a "G" can be put at the end of a primer to promote non-template addition**
- Can be enhanced with extension soak at the end of the PCR cycle (e.g., 15-45 min @ 60 or 72 ºC) – to give polymerase more time
- Excess amounts of DNA template in the PCR reaction can result in incomplete adenylation (not enough polymerase to go around)

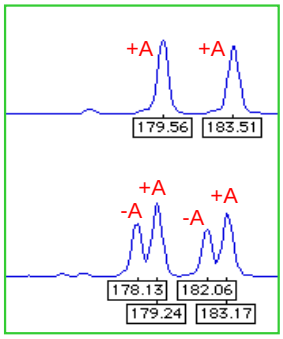
Best if there is NOT a mixture of "+/- A" peaks
(desirable to have full adenylation to avoid split peaks)





D8S1179

Impact of the 5' Nucleotide on Non-Template Addition



5'-ACAAG…

Last Base for Primer Opposite Dye Label

(PCR conditions are the same for these two samples)

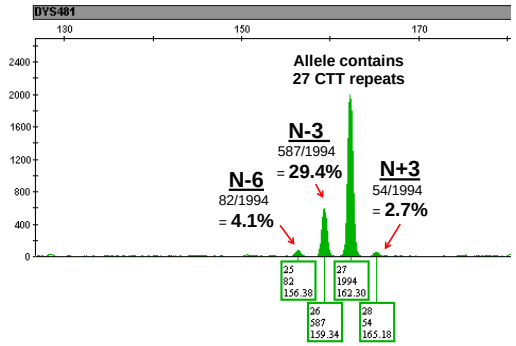
5'-CCAAG…

Promega includes an ATT sequence on the 5'-end of many of their unlabeled PP16 primers to promote adenylation
see Krenke *et al.* (2002) *J. Forensic Sci.* 47(4): 773-785
http://www.cstl.nist.gov/biotech/strbase/PP16primers.htm

Stutter Products

- Peaks that show up primarily one repeat less than the true allele as a result of strand slippage during DNA synthesis
- Stutter is less pronounced with larger repeat unit sizes (dinucleotides > tri- > tetra- > penta-)
- Longer repeat regions generate more stutter
- Each successive stutter product is less intense (allele > repeat-1 > repeat-2)
- Stutter peaks make mixture analysis more difficult

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



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 3.10**, p. 71

Stutter Data from a Set of 345 D18S51 Alleles

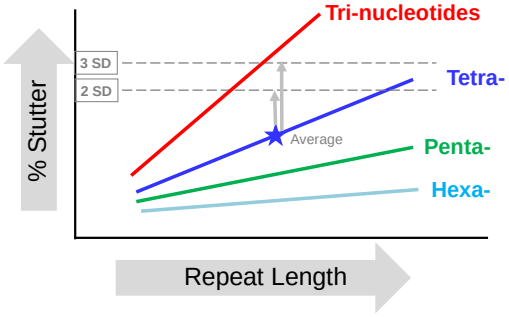
Measured at NIST Using the PowerPlex 16 Kit

Allele	Allele Size (nucleotides)	# Measured	Median (%)	Standard Deviation
12	296.9	43	4.8	0.4
13	300.7	27	5.7	0.5
14	304.6	35	6.2	0.5
15	308.5	55	6.9	0.6
16	312.4	46	7.7	0.5
17	316.2	47	8.3	0.4
18	320.2	38	9.0	0.9
19	324.0	30	9.6	0.9
20	328.0	24	10.6	0.8
		345	Average 7.7 ± 1.9	

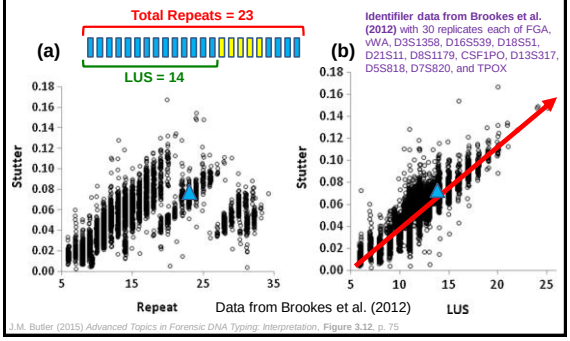
Locus Stutter Filter: **Average + 3 standard deviations** = 7.7 + (3×1.9) = 7.7 + 5.7 = **13.4%**

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Table 3.4**, p. 73

Simplified Illustration of Stutter Trends



Stutter Ratios Model Better with Longest Uninterrupted Stretch (LUS) Compared to Total Repeat Length



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 3.12**, p. 75

STR Genotypes

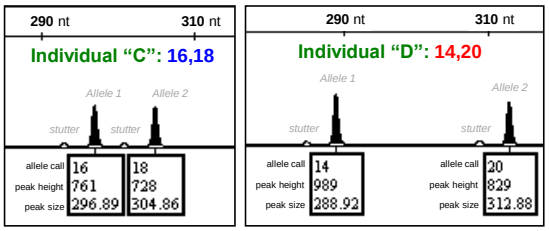
Heterozygote Balance, Stochastic Effects, etc.

Advanced Topics in Forensic DNA Typing: Interpretation, Chapter 4

Key Points

- In heterozygous loci, the two alleles should be equal in amount; however, stochastic effects during PCR amplification (especially when the amount of DNA being amplified is limited) create an imbalance in the two detected alleles
- Heterozygote balance (Hb) or peak height ratios (PHRs) measure this level of imbalance
- Under conditions of extreme imbalance, one allele may "drop-out" and not be detected
- Stochastic thresholds are sometimes used to help assess the probability of allele drop-out in a DNA profile

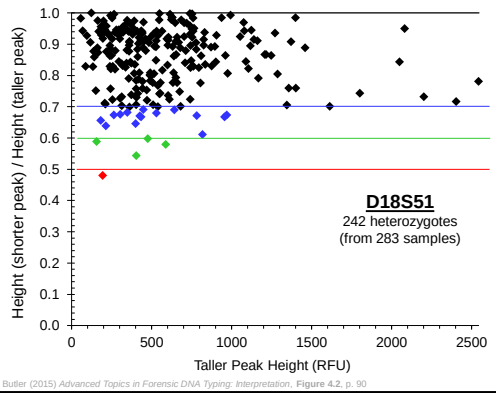
D18S51 Results from Two Samples



Peak Height Ratios (PHRs) or Heterozygote balance (*Hb*)

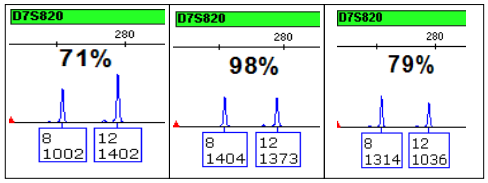
728/761 = 0.957= **95.7%**

829/989 = 0.838 = **83.8%**



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 4.2**, p. 90

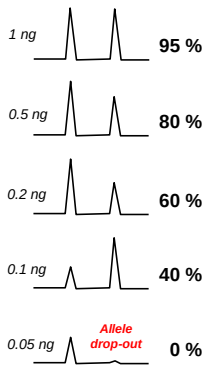
Natural Variation in Peak Height Ratio During Replicate PCR Amplifications



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

Slide from Charlotte Word (ISHI 2010 mixture workshop)

Hypothetical Heterozygote Alleles



Heterozygote balance typically decreases with DNA template level

In the extreme, one of the alleles fails to be amplified (this is known as allele drop-out)

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 4.3**, p. 92

STR Profiles

Multiplex PCR, Tri-Alleles, Amelogenin, and Partial Profiles

Advanced Topics in Forensic DNA Typing: Interpretation, Chapter 5

Key Points

- Tri-allelic patterns occasionally occur at STR loci (~1 in every 1000 profiles) and are due to copy number variation (CNVs) in the genome
- The amelogenin gene is found on both the X and Y chromosomes and portions of it can be targeted to produce assays that enable gender identification as part of STR analysis using commercial kits
- Due to potential deletions of the amelogenin Y region, additional male confirmation markers are used in newer 24plex STR kits
- Partial profiles can result from low amounts of DNA template or DNA samples that are damaged or broken into small pieces or contain PCR inhibitors

Multiplex PCR

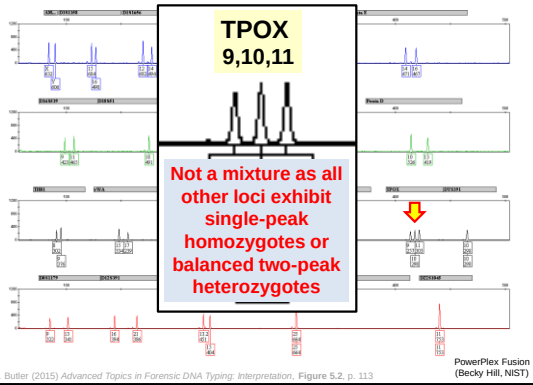
(Parallel Sample Processing)

- **Compatible primers are the key to successful multiplex PCR**
- **STR kits are commercially available**
- **15 or more STR loci can be simultaneously amplified**

Challenges to Multiplexing
- primer design to find compatible primers (no program exists)
- reaction optimization is highly empirical often taking months

Advantages of Multiplex PCR
- Increases information obtained per unit time (increases power of discrimination)
- Reduces labor to obtain results
- Reduces template required (smaller sample consumed)

Single-Source DNA Sample Exhibiting a TPOX Tri-Allelic Pattern



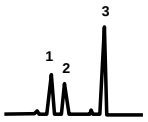
PowerPlex Fusion (Becky Hill, NIST)

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 5.2**, p. 113

Types of Tri-Allelic Patterns

More common

(a) Type 1



(1+2≈3)

(b) Type 2

(1≈2≈3)

This classification scheme was developed by Tim Clayton and colleagues at the UK Forensic Science Service (Clayton et al. 2004, *J. Forensic Sci.* 49: 1207-1214)

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 5.3**, p. 114

Tri-Allelic Patterns Occur about 1 in 1000 Profiles

but the frequency varies across STR loci

D.N.A. BOX 5.2

ESTIMATION OF AVERAGE FREQUENCY FOR TRI-ALLELIC PATTERNS IN STR PROFILES

Several years ago, DNA analyst Malena Jimenez from the Missouri State Highway Patrol supplied the NIST STRBase website with information on tri-alleles observed during analysis of 69,600 convicted offender samples. Below is a summary of the number of reported tri-alleles for each locus examined with their PowerPlex 16 single-source profiles.

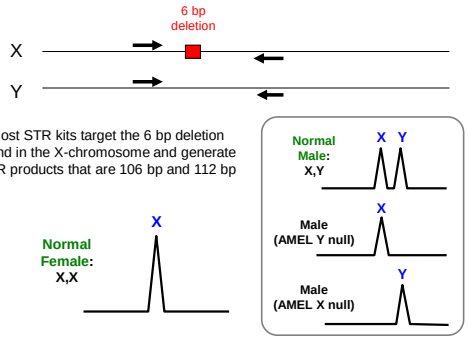
Noting a total of 68 tri-alleles observed across these 69,600 samples, a tri-allelic pattern is expected to occur about once every one thousand samples on average. However, as can be seen with this data, the distribution of tri-allelic patterns is not even across loci.

Source: Steven Myers, California Department of Justice presentation provided to the author based on data collected from the NIST STRBase website at http://www.cstl.nist.gov/strbase/tri_tab.htm.

STR Locus	# Tri-Alleles	STR Locus	# Tri-Alleles	STR Locus	# Tri-Alleles
CSF1PO	1	D3S1358	2	D16S539	3
FGA	11	D5S818	1	D18S51	3
TH01	0	D7S820	0	D21S11	9
TPOX	9	D8S1179	2	Penta D	3
vWA	10	D13S317	4	Penta E	10

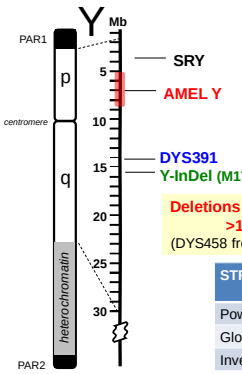
J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Box 5.2**, p. 114

Amelogenin Sex-Typing Assay



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 5.4**, p. 119

Relative Positions Along the Y-Chromosome of Amelogenin (AMEL Y) and Male Confirmation Markers Used in Newer STR Kits

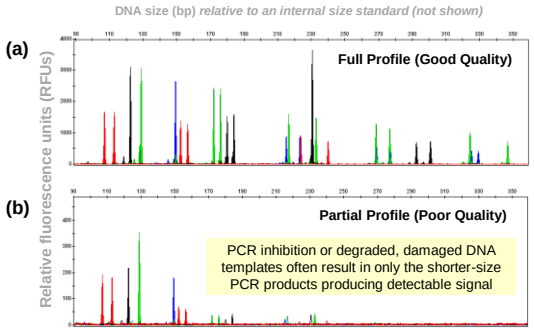


Deletions of the Y-chromosome can encompass >1 Mb around the AMEL Y region
(DYS458 from Y-STR kits is often lost in these situations)

STR Kit	Male Confirmation Marker(s)
PowerPlex Fusion	DYS391
GlobalFiler	DYS391, Y-InDel
Investigator 24plex	DYS391, Y-InDel

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 5.5**, p. 122

Partial Profiles Can Occur from Poor Quality DNA or Low Amounts of DNA Template



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 5.6**, p. 122

Troubleshooting Data Collection

Advanced Topics in Forensic DNA Typing: Interpretation, Chapter 8

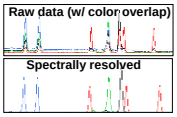
Key Points

- The better you understand your instrument(s) and how DNA typing data are generated during the PCR process, the better you will be able to troubleshoot problems that arise
- Three key analytical requirements for capillary electrophoresis instruments are (1) spectral (color) resolution, (2) size (spatial) resolution, and (3) run-to-run precision
- Salt levels need to be low in samples in order to effectively inject them into a CE instrument

Analytical Requirements for STR Typing

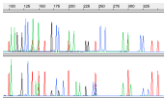
Butler *et al.* (2004) Electrophoresis 25: 1397-1412

- Fluorescent dyes must be **spectrally resolved** in order to distinguish different dye labels on PCR products



- PCR products must be **spatially resolved** – desirable to have single base resolution out to >350 bp in order to distinguish variant alleles

- High **run-to-run precision** – an internal sizing standard is used to calibrate each run in order to compare data over time

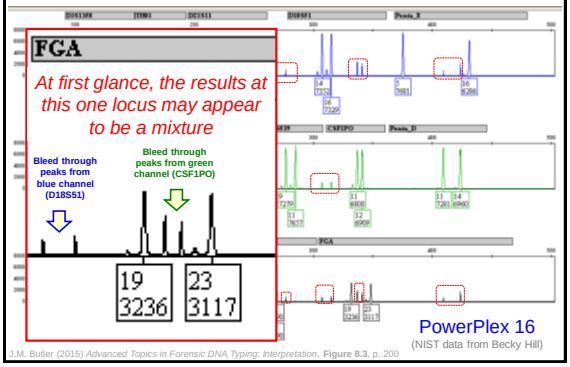


Potential Issues and Solutions
with Multicolor Capillary Electrophoresis

Issue	Cause/Result with Failure	Potential Solutions
Spectral resolution (color separation)	High RFU peaks result in bleed through or pull-up that create artificial peaks in adjacent dye channel(s)	Inject less DNA into the CE capillary to avoid overloading the detector
Analytical size resolution	Inner capillary wall coating failures result in an inability to resolve closely spaced STR alleles and in some cases incorrect allele calls can be made	Reinject sample (if a bubble causes poor polymer filling for a single run) or replace the pump (if polymer is not being routinely delivered to fully fill the capillaries)
Run-to-run precision	Room temperature changes result in sample alleles running differently compared to allelic ladder alleles and false "off-ladder" alleles are generated	Make adjustments to improve room temperature consistency or reinject samples with an allelic ladder run in an adjacent capillary or a subsequent run

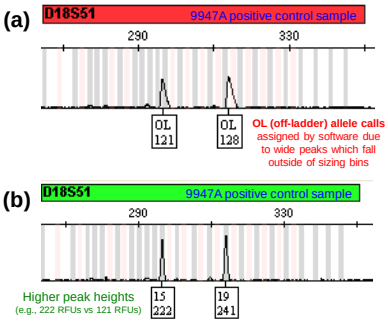
J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Table 8.1**, p. 191

Single-Source DNA Profile Exhibiting Pull-Up
Due to Off-Scale Data at Several Loci



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 8.3**, p. 200

Impact of Formamide Quality
on Peak Shape and Height



Old, poor-quality formamide used for denaturing this sample

Fresh, high-quality formamide used for denaturing this sample

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 8.2**, p. 189

Sample Conductivity Impacts Amount Injected

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

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

E is the electric field applied

t is the injection time

r is the radius of the capillary

μ_{ep} is the mobility of the sample molecules

μ_{eof} is the electroosmotic mobility

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

λ_{buffer} is the buffer conductivity

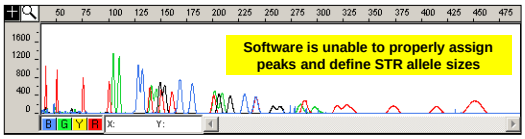
λ_{sample} is the sample conductivity

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

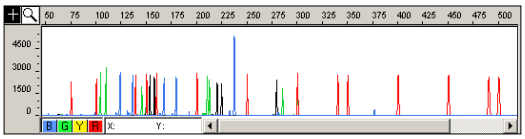
Butler *et al.* (2004) Electrophoresis 25: 1397-1412

Impact of Polymer Filling the Capillary

(a) Incomplete fill → a "meltdown" → poor-quality data and resolution loss



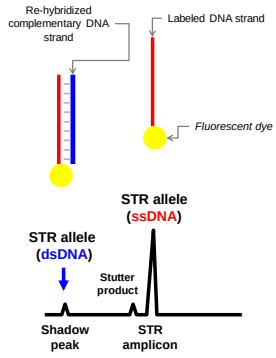
(b) Appropriate fill → high-quality data and sharp peaks → reliable STR typing



ABI 310 data from Margaret Kline (NIST)

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 8.6**, p. 205

Impact of Sample Renaturation



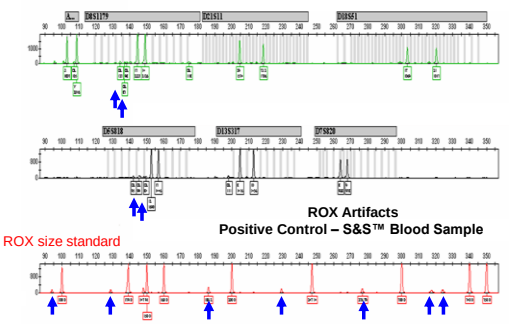
Double-stranded DNA (dsDNA) molecules, which are more rigid than their corresponding single-stranded DNA (ssDNA) counterparts, migrate more quickly through the network of polymer strands inside of a capillary.

When CE conditions permit re-hybridization of the complementary strand, then a shadow peak occurs in front of its corresponding labeled STR allele (or internal size standard DNA fragment).

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 8.5**, p. 203

Extra Peaks Due to Sample Renaturation

(issue mostly like due to the CE instrument temperature control)

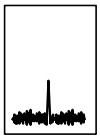
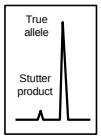
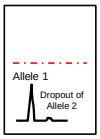
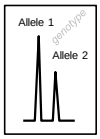


Profiler Plus data from Peggy Philion (RCMP) for a 2008 ISHI Troubleshooting Workshop

Data Interpretation Overview

The Steps of Data Interpretation

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

Analytical Threshold	Expected Stutter %	Stochastic Threshold	Peak Height Ratio (PHR)
			

Next step: Examine feasible genotypes to deduce possible contributor profiles

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

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

Population allele frequencies

DNA Profile (with specific alleles) → **Rarity estimate of DNA profile**

Genetic formulas and assumptions

There are different ways to express the profile rarity

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 9.1**, p. 214

Acknowledgments

$ NIST Special Programs Office
Simone Gittelson
Becky (Hill) Steffen

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

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

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

Basic STR Interpretation Workshop
John M. Butler & Simone N. Gittelson
Krakow, Poland
31 August 2015

Statistical Interpretation 1:

Introduction to probability and statistics
STR population data collection, calculations, and use
Approaches to calculating match probabilities

Simone N. Gittelson, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

Workshop Schedule

Time	Module (Instructor)	Topics
0900-0930	Welcome & Introductions	Review expectations and questions from participants
0930 – 1100	Data Interpretation 1 (John)	STR kits, loci, alleles, genotypes, profiles Data interpretation thresholds and models Simple PCR and CE troubleshooting
1100 – 1130	Break	
1130 – 1300	Statistical Interpretation 1 (Simone)	Introduction to probability and statistics STR population data collection, calculations, and use Approaches to calculating match probabilities
1300 – 1430	Lunch	
1430 – 1600	Data Interpretation 2 (John)	Mixture interpretation: Clayton rules, # contributors Stochastic effects and low-template DNA challenges Worked examples
1600 – 1630	Break	
1630 – 1800	Statistical Interpretation 2 (Simone)	Approaches to calculating mixture statistics Likelihood ratios and formulating propositions Worked examples

Acknowledgement and Disclaimers

I thank John Butler for the discussions and advice on preparing this presentation. I also acknowledge John Buckleton and Bruce Weir for all their helpful explanations on forensic genetics topics.

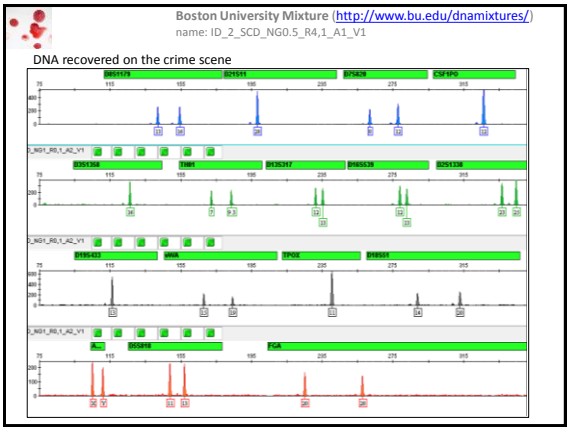
Points of view in this presentation are mine and do not necessarily represent the official position or policies of the National Institute of Standards and Technology.

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

Presentation Outline

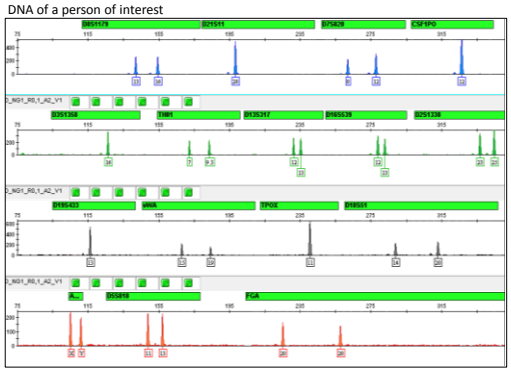
1. Why do we need to do a statistical (probabilistic) interpretation?
2. How do we do a statistical interpretation?
   a. Hardy-Weinberg Equilibrium (HWE)
   b. Recombination and Linkage
   c. Subpopulations
   d. Linkage Equilibrium (LE)
   e. NRC II Report Recommendations
   f. Population Allele Frequencies
   g. Logical Approach for Evidence Interpretation

Why do we need to do a statistical (probabilistic) interpretation?



Boston University Mixture (http://www.bu.edu/dnamixtures/)
name: ID_2_SCD_NG0.5_R4,1_A1_V1

DNA of a person of interest



DNA recovered on the crime scene

DNA of a person of interest

Does the DNA recovered on the crime scene come from the person of interest?

DNA recovered on the crime scene

DNA of a person of interest

If the DNA recovered on the crime scene comes from the person of interest, we would expect to see peaks for the same genotypes.

DNA recovered on the crime scene

DNA of a person of interest

Does the DNA recovered on the crime scene come from the person of interest?

DNA recovered on the crime scene

DNA of a person of interest

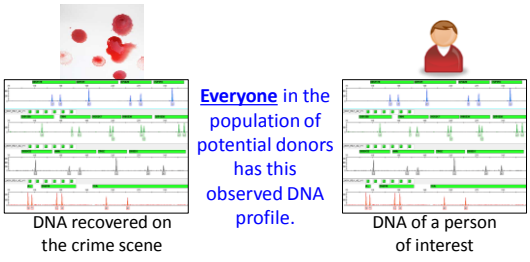
If the DNA recovered on the crime scene does not come from the person of interest, we need to know how rare it is to observe the peaks in the EPG of the DNA recovered on the crime scene.

DNA recovered on the crime scene

The observed DNA profile is **very rare** in the population of potential donors.

DNA of a person of interest

In this case, the observations support the proposition that the DNA recovered on the crime scene came from the person of interest.



In this case, the observations provide no information on whom the DNA recovered on the crime scene comes from.

A **statistical interpretation** tells us what our observations mean in a particular case, with regard to a particular question of interest to the court.

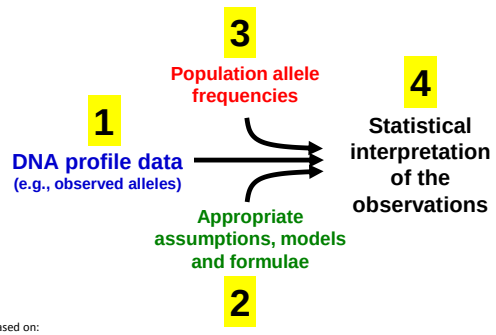
statistical interpretation

synonym: probabilistic interpretation

definition: A quantitative expression of the value of the evidence.

How do we do a statisitcal interpretation?

Elements required for a statistical interpretation



Based on:
J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*: **Figure 9.1**, page 214.

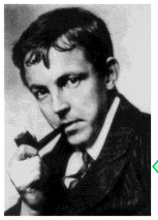
2

Hardy-Weinberg Equilibrium (HWE)

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 10: pages 240-243 and 257-259.

2

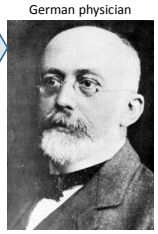
Godfrey Harold Hardy
British mathematician



January 13, 1908: Weinberg's lecture to the Society for the Natural History of the Fatherland in Württemberg *(Verein für vaterländische Naturkunde in Württemberg)*, entitled *Über den Nachweis der Vererbung beim Menschen* (On the Proof of Heredity in Humans). Printed in the *Jahreshefte*, Vol. 64: 368-382 (1908).

April 5, 1908: Date of Hardy's signature in his July 10, 1908 publication in Science 28 (706): 49-50, entitled *Mendelian Proportions in a Mixed Population.*

Wilhelm Weinberg
German physician



C. Stern. (1943). The Hardy-Weinberg Law. *Science, 97 (2510)*: 137-138.

2

Hardy Weinberg Equilibrium (HWE)

Assumptions:

1. size of population is infinite
2. no migration
3. random mating
4. no mutations
5. no natural selection

Why is HWE important?
Allele and genotype frequencies in this population remain constant from one generation to the next.

2

Hardy Weinberg Equilibrium (HWE)

Assumptions:

1. size of population is infinite

Reality:
world ≈ 7.3 *billion*
Poland ≈ 38.5 *million*
Krakow ≈ 760,000

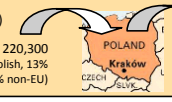
2

Hardy Weinberg Equilibrium (HWE)

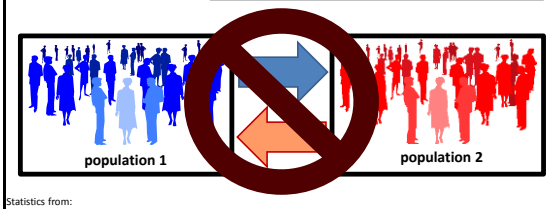
Assumptions:

2. no migration

Reality:
Poland (2013)
220,300
(60% Polish, 13% EU, 27% non-EU)



2.1 *million*



Statistics from:
http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Immigration_by_citizenship,_2013_YB15.png
http://www.economist.com/blogs/easternapproaches/2013/11/poland-and-eu

2

Hardy Weinberg Equilibrium (HWE)

Assumptions:

3. random mating

Reality:
Poland (2003)
``The most common model of marriage is between people from the same age group (49.1%) and also similar economical status and especially similar education level (53.4%).''

father

mother

Quote and statistics from:
Urban-Klaehn J. Polish Marriages and Families, Some Statistics, II. 23 February 2003 (article #87), available at:
http://culture.polishsite.us/articles/art87fr.htm

2

Hardy Weinberg Equilibrium (HWE)

Assumptions:

4. no mutations

Reality:
mutation rate for locus D21S11: 0.19%

D21S11: father {28,28}

child {29,...}

Mutation rate from:
Butler J.M. STRBase website at: http://www.cstl.nist.gov/strbase/mutation.htm

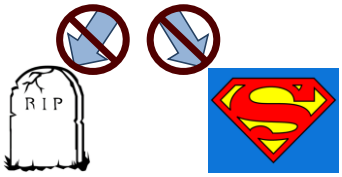
2 Hardy Weinberg Equilibrium (HWE)

Assumptions:

5. no natural selection

Reality:
Some genes are more likely to lead to diseases than others.
However, STR loci used in forensic science come from regions that are not used for coding genes (i.e., they are introns).

D21S11: allele 28



2 Hardy Weinberg Equilibrium (HWE)

Assumptions:

1. size of population is infinite
2. no migration
3. random mating
4. no mutations
5. no natural selection

Why is HWE important?
Allele and genotype frequencies in this population remain constant from one generation to the next.

If a population is in Hardy-Weinberg Equilibrium, the Hardy-Weinberg Law predicts the genotype frequencies.

2 Laws of Mendelian Genetics

Punnett Square:

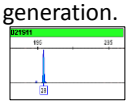
		mother	
		a	b
father	A	Aa	Ab
	B	Ba	Bb

Law of Segregation
The genotype at a locus consists of one maternal allele and one paternal allele. Each child receives a randomly selected allele from each parent.

Law of Independent Assortment
The allele transmitted from parent to child at one locus is independent of the allele transmitted from parent to child at a different locus.

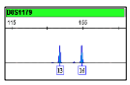
2 Hardy Weinberg Equilibrium (HWE)

If a population is in Hardy-Weinberg Equilibrium, we can predict the genotype frequencies after one generation.



$Pr(28,28)$
probability that a person has genotype {28,28}

homozygote {28,28}



$Pr(13,16)$
probability that a person has genotype {13,16}

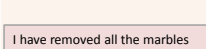
heterozygote {13,16}

frequency

the counted number of occurrences in a known set of events

The events have already been realized and I count the results.

I have removed all the marbles from the urn and counted the number of red ones and blue ones:

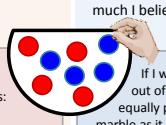


frequency of a red marble = 4
frequency of a blue marble = 4

probability

a degree of belief in the occurrence of an unknown event

The event has not been realized or is unknown to me and I describe how much I believe in it occurring.



If I were to randomly pick a marble out of this urn, I believe that it is equally probable for me to pick a red marble as it is for me to pick a blue marble.

probability of picking a red marble = 0.5
probability of picking a blue marble = 0.5

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 11: page 301

Laws of Probability

Law #1: A probability can take any value between 0 and 1, including 0 and 1.

certain

1 certainty that statement is true

0.75
0.66
0.5
0.33
0.25

0 certainty that statement is false

impossible

EXAMPLE: rolling a 6-sided die

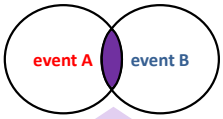
$Pr(1,2,3,4,5\ or\ 6) = 1$
$Pr(7) = 0$

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 9: pages 222-224.

Laws of Probability *and*

Law #3 (**independent** events): The probability of event A ***and*** event B occurring is equal to the probability of event A ***times*** the probability of event B.

$$\Pr(A \text{ and } B) = \Pr(A) \times \Pr(B)$$



we want the probability of the overlapping region

EXAMPLE: rolling two 6-sided dice
A: rolling a 5 with die 1
B: rolling a 5 with die 2

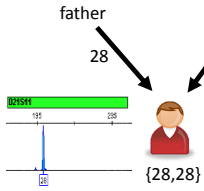
$Pr(A \text{ and } B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 9: pages 222-224.

2 Hardy-Weinberg Law

Homozygote

$Pr(28,28)$
$= Pr(paternal\ allele = 28) \overset{and}{\times} Pr(maternal\ allele = 28)$
$= p_{28} \times p_{28}$
$= p_{28}^2$



3 Population allele frequencies

D21S11:

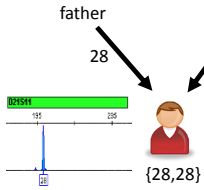
Allele	2N = 722 Caucasian	2N = 684 Black	2N = 472 Hispanic	2N = 194 Asian
24.2	-	-	0.002	-
25.2	0.001	-	-	-
26	-	0.001	-	-
26.2	-	-	0.002	-
27	0.022	0.075	0.028	-
28	0.159	0.246	0.100	0.057
28.2	-	-	-	0.005
29	0.202	0.205	0.208	0.201
⋮	⋮	⋮	⋮	⋮
39	-	0.001	-	-

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Appendix 1: STR Allele Frequencies from U.S. Population Data, page 510.

2 Hardy-Weinberg Law

Homozygote

$Pr(28,28)$
$= Pr(paternal\ allele = 28) \overset{and}{\times} Pr(maternal\ allele = 28)$
$= p_{28} \times p_{28}$
$= p_{28}^2$



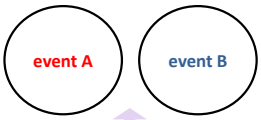
$p_{28} = 0.159$

$Pr(28,28) = (0.159)^2$
$= 0.025$

Laws of Probability *or*

Law #2 (**mutually exclusive** events): The probability of event A ***or*** event B occurring is equal to the probability of event A ***plus*** the probability of event B.

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$$



mutually exclusive = no overlap

EXAMPLE: rolling a 6-sided die
A: rolling an odd number
B: rolling a 2

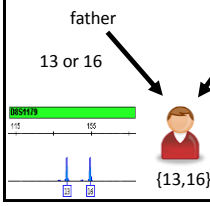
$Pr(A \text{ or } B) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 9: pages 222-224.

2 Hardy-Weinberg Law

Heterozygote

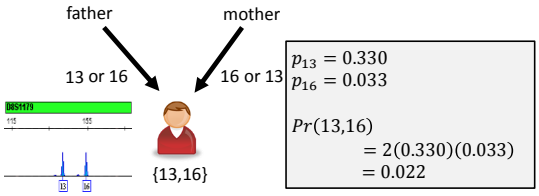
$Pr(13,16)$
$= Pr(paternal\ allele = 13) \overset{and}{\times} Pr(maternal\ allele = 16)$
$\overset{or}{+} Pr(paternal\ allele = 16) \overset{and}{\times} Pr(maternal\ allele = 13)$
$= p_{13} \times p_{16} + p_{16} \times p_{13}$
$= p_{13}p_{16} + p_{16}p_{13}$
$= 2p_{13}p_{16}$



2

Hardy-Weinberg Law

Heterozygote

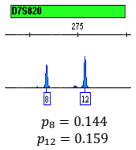


$p_{13} = 0.330$
$p_{16} = 0.033$

$Pr(13,16)$
$= 2(0.330)(0.033)$
$= 0.022$

2

According to the Hardy-Weinberg law, what is the probability that a person has **genotype {8,12}**?

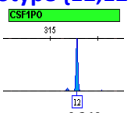


$p_8 = 0.144$
$p_{12} = 0.159$

A. $0.144 \times 0.159 = 0.023$
B. $2 \times 0.144 \times 0.159 = 0.046$
C. $2 \times 0.159 \times 0.159 = 0.051$
D. $0.144 + 0.159 = 0.303$
E. $2 \times 8 \times 12 = 192$
F. ???

Response Counter

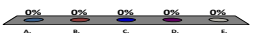
According to the Hardy-Weinberg law, what is the probability that a person has **genotype {12,12}**?



$p_{12} = 0.360$

A. $(0.36)^2 = 0.130$
B. 0.360
C. $0.36 + 0.36 = 0.720$
D. $(12)^2 = 144$
E. ???

Response Counter



2

Recombination and Linkage

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 10: pages 259-260.

2

Recombination

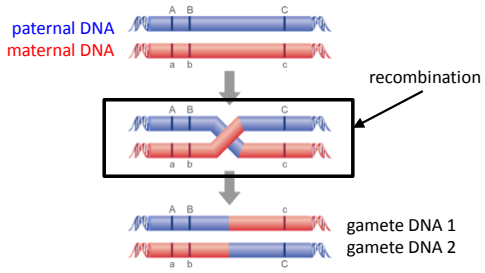
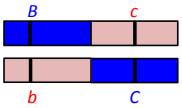


Image from:
Wellcome Trust Website. The Human Genome. http://genome.wellcome.ac.uk/doc_WTD020778.html

2

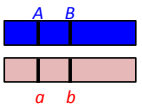
Linkage

independence between loci

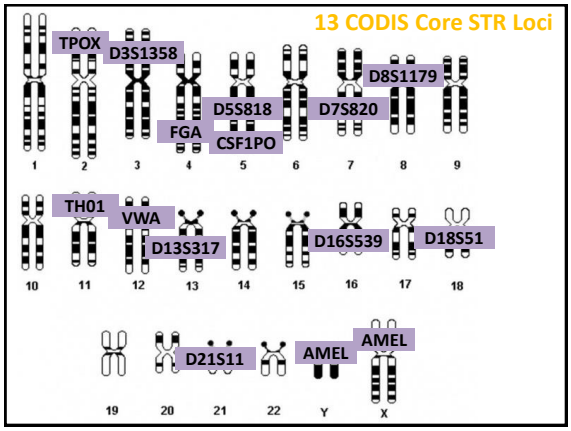


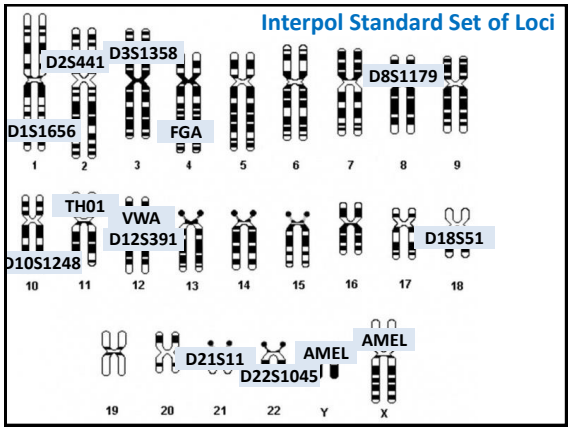
If the child inherits *B*, there is a probability of 0.5 that the child inherits *C*, and a probability of 0.5 that the child inherits *c*.

dependence between loci = **linkage**



If the child inherits *B*, there is a probability >0.5 that the child inherits *A*, and a probability of <0.5 that the child inherits *a*.





2

Is there linkage?

Loci on different chromosomes, or on different arms of the same chromosome:

No, there is no linkage.

Loci on the same arm of the same chromosome:

Linkage is possible. This has no impact on unrelated individuals, but should be taken into account for related individuals by incorporating the probability of recombination into the statistical interpretation.

2

Subpopulations

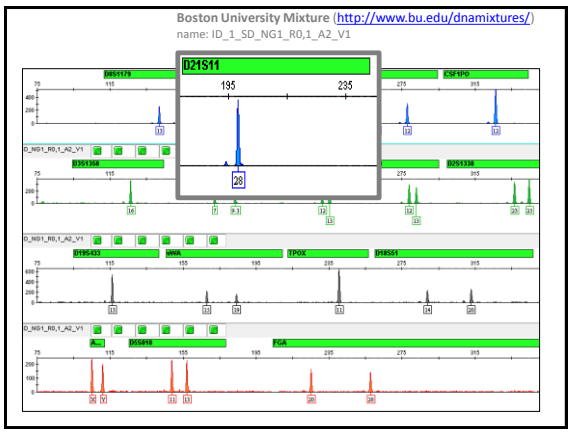
J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 10: pages 260-262, and Chapter 11, D.N.A. Box 11.2, page 291.

2

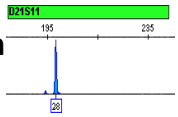
Hardy-Weinberg Equilibrium (HWE)

Assumptions:

- size of population is infinite ~
- no migration ✗
- random mating ✗
- no mutations ✗
- no natural selection ~



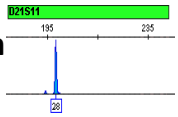
2 Subpopulation



General Population

$p_{28} = 0.5$

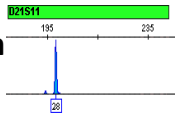
2 Subpopulation



50% Subpopulation 1	50% Subpopulation 2
mates only with members of subpopulation 1	mates only with members of subpopulation 2
$p_{28} = 0.4$	$p_{28} = 0.6$

no random mating

2 Subpopulation

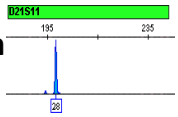


Taking into account subpopulations:

50% Subpopulation 1	50% Subpopulation 2
$\Pr(28,28) = 0.4^2 = 0.16$	$\Pr(28,28) = 0.6^2 = 0.36$

$$Pr(28,28) = \frac{1}{2} \times 0.16 + \frac{1}{2} \times 0.36 = 0.26$$

2 Subpopulation



Not taking into account subpopulations:

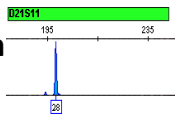
General Population

needs correction!

$p_{28} = 0.5$

$$Pr(28,28) = 0.5^2 = 0.25 < \mathbf{0.26}$$

2 Subpopulation



General Population

We can use the **coancestry coefficient** F_{ST}, also called θ, to take into account the effect of subpopulations when we use the proportion $p_{28} = 0.5$ of the general population.

2 Subpopulations

allele 28 is

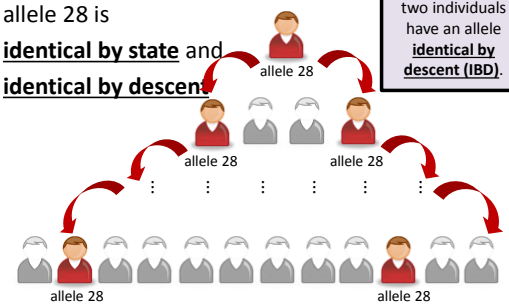
identical by state

allele 28 allele 28

2 Subpopulations

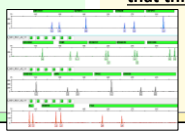
The **coancestry coefficient** F_{ST}, also called θ, is the probability that two individuals have an allele **identical by descent (IBD)**.

allele 28 is **identical by state** and **identical by descent**



---

profile probability	match probability
probability of observing this profile in a population	probability of observing this profile in a population **knowing that this profile has already been observed in one individual in this population**



What is the probability of observing this profile in this population?

this individual has this profile

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*. Chapter 11: pages 301-302.

---

profile probability	match probability
probability of observing this profile in a population	probability of observing this profile in a population knowing that this profile has already been observed in one individual in this population

If $\theta = 0$: profile probability = match probability — **no** relatives, **no** coancestors

If $\theta > 0$: profile probability < match probability — relatives, coancestors

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*. Chapter 11: pages 301-302.

---

2 Subpopulations

The **coancestry coefficient** F_{ST}, also called θ, is the probability that two individuals have an allele **identical by descent (IBD)**.

What is the probability of seeing **allele 28** in this population given that we have already observed one copy of ***allele 28***?

---

2 Subpopulations

The **coancestry coefficient** F_{ST}, also called θ, is the probability that two individuals have an allele **identical by descent (IBD)**.

We have seen: ***allele 28***

The probability of observing an **allele 28** is:

either θ + or $(1-\theta)p_{28}$

allele 28 is IBD with ***28***

allele 28 is not IBD with any of the alleles already seen, it is observed by chance

---

2 Subpopulations

Rule of Thumb

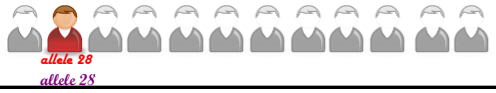
If the allele in question has not been seen previously, then it is seen by chance.

If the allele in question has already been seen, then it could be observed again by chance **or because it is IBD with an allele that has already been seen**.

2

Subpopulations

What is the probability of seeing **allele 28** in this population given that we have already observed *allele 28* and *allele 28*?



2

Subpopulations

We have seen: *allele 28* and *allele 28*

The probability of observing an **allele 28** is:

$$\underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{(1-\theta)p_{28}}_{\text{allele 28 is not IBD with any of the alleles already seen, it is observed by chance}}$$

2

Subpopulations

We have seen: *allele 28* and *allele 28*

The probability of observing an **allele 28** is:

$$\underbrace{2\theta + (1-\theta)p_{28}}_{1+\theta}$$

2

Subpopulations

We have seen: *allele 28* and *allele 28*

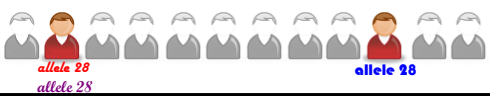
The probability of observing an **allele 28** is:

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

2

Subpopulations

What is the probability of seeing **allele 28** in this population given that we have already observed *allele 28*, *allele 28* and **allele 28**?



2

Subpopulations

We have seen: *allele 28*, *allele 28* and **allele 28**

The probability of observing an **allele 28** is:

$$\underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{(1-\theta)p_{28}}_{\text{allele 28 is not IBD with any of the alleles already seen, it is observed by chance}}$$

2 Subpopulations

We have seen: ***allele 28***, *allele 28* and **allele 28**

The probability of observing an **allele 28** is:

$$\frac{3\theta + (1-\theta)p_{28}}{1+2\theta}$$

2 Subpopulations

We have seen: ***allele 28***, *allele 28* and **allele 28**

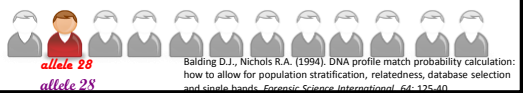
The probability of observing an **allele 28** is:

$$\frac{3\theta + (1-\theta)p_{28}}{1+2\theta}$$

2 Subpopulations

What is the probability of seeing genotype {**28**,*28*} in this population given that we have already observed a genotype {***28***,*28*}?

$$\frac{2\theta + (1-\theta)p_{28}}{1+\theta} \times \frac{3\theta + (1-\theta)p_{28}}{1+2\theta}$$

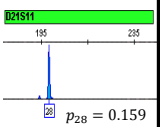


Balding D.J., Nichols R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International*, *64*: 125-40

2 Subpopulations

What is the probability of seeing genotype {**28**,*28*} in this population given that we have already observed a genotype {***28***,*28*}?

$$\frac{2\theta + (1-\theta)p_{28}}{1+\theta} \times \frac{3\theta + (1-\theta)p_{28}}{1+2\theta}$$



$p_{28} = 0.159$

if $\theta = 0.02$:

$$\frac{2(0.02) + (1-0.02)(0.159)}{1+0.02} \times \frac{3(0.02) + (1-0.02)(0.159)}{1+2(0.02)}$$

$$= 0.040$$

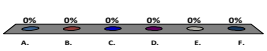
What is the genotype probability

$$\frac{2\theta+(1-\theta)p_{28}}{1+\theta} \times \frac{3\theta+(1-\theta)p_{28}}{1+2\theta}$$

equal to if $\theta = 0$?

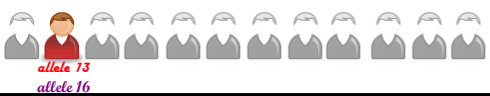
A. 0
B. θ
C. p_{28}^2
D. $2p_{28}$
E. 1
F. ???

Response Counter



2 Subpopulations

What is the probability of seeing **allele 13** in this population given that we have already observed ***allele 13*** and *allele 16*?



2 Subpopulation

We have seen	Divide by
1 allele	1
2 alleles	$1+\theta$
3 alleles	$1+2\theta$

We have seen: ***allele 13*** and ***allele 16***

The probability of observing an **allele 13** is:

$$\frac{\underbrace{1\times\theta}_{\text{allele 13 is IBD with 13}} + \underbrace{0\times\theta}_{\text{allele 13 is IBD with 16}} + \underbrace{(1-\theta)p_{13}}_{\text{allele 13 is not IBD with any of the alleles already seen, it is seen by chance}}}{1+\theta}$$

2 Subpopulations

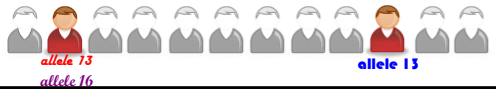
We have seen: ***allele 13*** and ***allele 16***

The probability of observing an **allele 13** is:

$$\frac{\theta+(1-\theta)p_{13}}{1+\theta}$$

2 Subpopulations

What is the probability of seeing **allele 16** in this population given that we have already observed ***allele 13***, ***allele 16*** and **allele 13**?



2 Subpopulation

We have seen	Divide by
1 allele	1
2 alleles	$1+\theta$
3 alleles	$1+2\theta$

We have seen: ***allele 13***, ***allele 16*** and **allele 13**

The probability of observing an **allele 16** is:

$$\frac{\underbrace{0\times\theta}_{\text{allele 16 is IBD with 13}} + \underbrace{0\times\theta}_{\text{allele 16 is IBD with 16}} + \underbrace{1\times\theta}_{\text{allele 16 is IBD with 13}} + \underbrace{(1-\theta)p_{16}}_{\text{allele 16 is not IBD with any of the alleles already seen}}}{1+2\theta}$$

2 Subpopulations

We have seen: ***allele 13***, ***allele 16*** and **allele 13**

The probability of observing an **allele 16** is:

$$\frac{\theta+(1-\theta)p_{16}}{1+2\theta}$$

2 Subpopulations

What is the probability of seeing genotype {**13**,**16**} in this population given that we have already observed a genotype {***13***,***16***}?

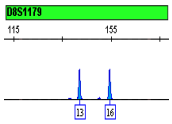
$$2\times\frac{\theta+(1-\theta)p_{13}}{1+\theta}\times\frac{\theta+(1-\theta)p_{16}}{1+2\theta}$$

Balding D.J., Nichols R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International*. 64: 125-40.

2 Subpopulations

What is the probability of seeing genotype {13,16} in this population given that we have already observed a genotype {13,16}?

$$2 \times \frac{\theta + (1-\theta)p_{13}}{1+\theta} \times \frac{\theta + (1-\theta)p_{16}}{1+2\theta}$$



$p_{13} = 0.330$
$p_{16} = 0.033$

if $\theta = 0.02$:

$$2 \times \frac{0.02 + (1-0.02)(0.33)}{1+0.02} \times \frac{0.02 + (1-0.02)(0.033)}{1+2(0.02)}$$

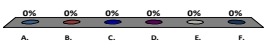
What is the genotype probability

$$2 \times \frac{\theta+(1-\theta)p_{13}}{1+\theta} \times \frac{\theta+(1-\theta)p_{16}}{1+2\theta}$$

equal to if $\boldsymbol{\theta = 0}$?

A. 0
B. 2θ
C. p_{13}^2
D. $2p_{13}p_{16}$
E. 1
F. ???

Response Counter



2 Linkage Equilibrium (LE)

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 10: pages 240-241 and 257-259.

2

Linkage Equilibrium

In a population, the alleles at one locus are **independent** of the alleles at a different locus.

Linkage Disequilibrium

In a population, the alleles at one locus are **not independent** of the alleles at a different locus.

2 Linkage ≠ Linkage Disequilibrium

CAUSES: Linkage, Population Subdivision

EFFECT: Linkage Disequilibrium

2 Linkage Equilibrium (LE)

Assumptions:

1. size of population is infinite
2. no migration
3. random mating
4. no mutations
5. no selection
6. **number of generations is infinite**

Why is LE important? Mendel's Law of Independent Assortment holds.

If a population is in Linkage Equilibrium, the product rule predicts the genotype frequencies.

2

Linkage Equilibrium

In a population, the alleles at one locus are **independent** of the alleles at a different locus.

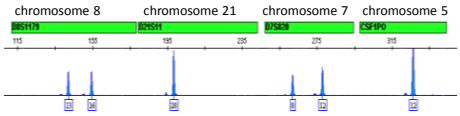
Product rule:

probability of genotype at multiple loci
= product of genotype probabilities at each locus

2

Assumption: Linkage Equilibrium

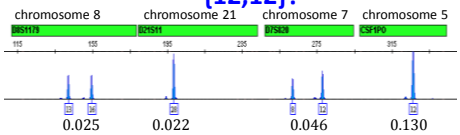
Product rule:



$$Pr(\{13,16\},\{28,28\},\{8,12\}\ and\ \{12,12\})$$

$$= Pr(13,16) \times Pr(28,28) \times Pr(8,12) \times Pr(12,12)$$

What is the probability that a person has genotype {13,16}, {28,28}, {8,12} and {12,12}?

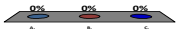


A. $0.025 \times 0.022 \times 0.046 \times 0.130 = 3.3 \times 10^{-6}$

B. $0.130 - 0.025 - 0.022 - 0.046 = 0.037$

C. $0.025 + 0.022 + 0.046 + 0.130 = 0.223$

Response Counter



2

NRC II Report Recommendations

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Appendix 2: NRC I & NRC II Recommendations, pages 525-526.

National Research Council Committee on DNA Forensic Science. The Evaluation of Forensic DNA Evidence. National Academy Press, Washington D.C., 1996.

2

Fixation indices (F-statistics)

F-statistics	alternative notation	Meaning
F_{IS}	f	**Individual to Subpopulation:** the correlation of alleles within an individual within a subpopulation
F_{IT}	F	**Individual to Total population:** the correlation of alleles within an individual (``inbreeding'')
F_{ST}	θ	**Subpopulation to Total population:** the correlation of alleles of different individuals in the same subpopulation (``coancestry'')

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 10: pages 260-262.

2

NRC II Report Recommendations

		Assumptions
	Hardy-Weinberg Law:	Assumes Hardy-Weinberg Equilibrium and Linkage Equilibrium in the population
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding''):	Corrects for Hardy-Weinberg Disequilibrium in the population caused by population subdivision. Assumes Linkage Equilibrium in the population.
Recommendation 4.2	includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (``coancestry''):	Corrects for Hardy-Weinberg Disequilibrium and Linkage Disequilibrium in the population caused by population subdivision. Assumes Hardy-Weinberg Equilibrium and Linkage Equilibrium in the sub-populations.

J. Buckleton, C.M. Triggs, S.J. Walsh. (2005). *Forensic DNA Evidence Interpretation.* CRC Press, London: pages 84-98.

2 NRC II Report Recommendations

		Homozygotes	Heterozygotes
	Hardy-Weinberg Law:	p_{28}^2	$2p_{13}p_{16}$
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding´´):	$Fp_{28} + (1-F)p_{28}^2$	$2p_{13}p_{16}$
Recommendation 4.2	includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (``coancestry´´):	$\frac{[2\theta + (1-\theta)p_{28}][3\theta + (1-\theta)p_{28}]}{(1+\theta)(1+2\theta)}$	$\frac{2[\theta + (1-\theta)p_{13}][\theta + (1-\theta)p_{16}]}{(1+\theta)(1+2\theta)}$

2 NRC II Report Recommendations

		Homozygotes	Heterozygotes
	Hardy-Weinberg Law:	0.025	0.022
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding´´):	$F = 0.02$: 0.028	0.022
Recommendation 4.2	includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (``coancestry´´):	$\theta = 0.02$: 0.040	$\theta = 0.02$: 0.034

2 NRC II Report Recommendations

		match probability for 15 loci
	Hardy-Weinberg Law:	8.9×10^{-23}
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding´´):	$F = 0.02$: 1.2×10^{-22}
Recommendation 4.2	includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (``coancestry´´):	$\theta = 0.02$: 3.7×10^{-20}

2 NRC II Report Recommendations

		Consequences
	Hardy-Weinberg Law:	The profile seems **more rare** than it actually is.
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding´´):	 The profile seems **a little more rare** than it actually is.
Recommendation 4.2	includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (``coancestry´´):	The profile seems **more common** than it actually is. 

J. Buckleton, C.M. Triggs, S.J. Walsh. (2005). *Forensic DNA Evidence Interpretation.* CRC Press, London: pages 84-98.

3

Population Allele Frequencies

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 10: pages 245-257.

3 Population allele frequencies

D8S1179:

Allele	Total 2N=2072		2N = 722	2N = 684	2N = 472	2N = 194
	Total #	Total %	Caucasian	Black	Hispanic	Asian
8	22	1.06	0.014	0.007	0.0148	-
9	10	0.48	0.006	0.004	0.006	-
10	163	7.87	0.102	0.031	0.093	0.124
11	139	6.71	0.076	0.053	0.053	0.119
12	294	14.20	0.168	0.130	0.129	0.119
13	556	26.80	0.330	0.219	0.273	0.201
14	484	23.40	0.166	0.294	0.263	0.201
15	291	14.00	0.104	0.190	0.129	0.129
16	101	4.87	0.033	0.064	0.032	0.093
17	8	0.39	0.001	0.004	0.004	0.010
18	4	0.19	-	0.003	0.002	0.005

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Appendix 1: STR Allele Frequencies from U.S. Population Data, page 505.

3 Steps in generating and validating a population database

Decide on number of samples and ethical/racial grouping

↓

Gather samples

↓

Analyze samples at desired genetic loci

↓

Determine allele frequencies for each locus

Excerpt from:
J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*: **Figure 10.4,** page 247.

Locus D18S51

Allele 1 \ Allele 2	10	11	12	13	14	14.2	15	16	16.2	17	18	19	20	21	22	count	frequency
10			3	2	1											6	0.008
11			1	1	1		1	1		1				1		7	0.010
12			7	6	9		15	13		11	5	2		2	1	82	0.114
13				5	11		17	10		18	6	3	4	1		89	0.123
14					10		12	12		17	4	5	3	1	1	97	0.134
14.2							1									1	0.001
15							9	14		20	19	3	2		1	123	0.170
16								12		9	13	7	1	1	1	106	0.147
16.2										1						1	0.001
17										7	4	4	1	1		100	0.139
18											1	2				56	0.078
19												1	1			29	0.040
20															1	13	0.018
21																7	0.010
22																5	0.007

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*: D.N.A. Box 10.2, page 248. total: 722 1

3 Steps in generating and validating a population database

Decide on number of samples and ethical/racial grouping

↓

Gather samples

↓

Analyze samples at desired genetic loci

↓

Determine allele frequencies for each locus

↓

Perform statistical tests on population data

Excerpt from:
J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*: **Figure 10.4,** page 247.

3 Validating a Population Database

genotype frequencies observed in the population ⇔ genotype frequencies expected in the population

allele frequencies observed in the population

Population Data Collection

Statistical Tests (e.g., for HWE and LE)

3 Validating a Population Database

Key question: What are you going to do with the database?

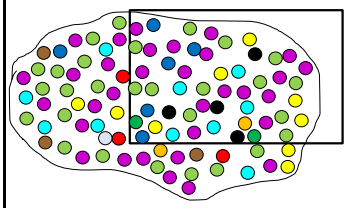
Validating a database requires validating the population genetic model that will be used.

Statistical tests examine whether the data in the database performs as expected according to the population genetic model.

Buckleton, J., Triggs, C., Walsh, S.J. (2005). *Forensic DNA evidence interpretation*. Boca Raton, FL: CRC Press, page 152.

3 Sampling Variation

Slide courtesy to Dr. John Buckleton



If we profiled everyone, the true frequencies are:
$f_A = 0.015$
$f_B = 0.550$
$f_C = 0.075$
$f_D = 0.025$
⋮

close, but not the same

Population database frequencies:
$f_A = 0$
$f_B = 0.500$
$f_C = 0.044$
$f_D = 0.022$
⋮

3

Options

Factor of 10 and minimum allele frequency of 5/2N

NRC II – National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. National Academy Press, Washington D.C., 1996.

- Multiply match probability by 10
- Use 5/2N as the minimum allele frequency

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*: Chapters 10 and 11, pages 251, 255 and 283-284.

3

Population allele frequencies

D8S1179:

Allele	Total 2N=2072		2N = 722	2N = 684	2N = 472	2N = 194
	Total #	Total %	Caucasian	Black	Hispanic	Asian
8	22	1.06	0.014	0.007	0.0148	$\frac{5}{194} = 0.026$
9	10	0.48	$\frac{5}{722} = 0.007$	0.004	0.006	$\frac{5}{194} = 0.026$
10	163	7.87	0.102	0.031	0.093	0.124
11	139	6.71	0.076	0.053	0.053	0.119
12	294	14.20	0.168	0.130	0.129	0.119
13	556	26.80	0.330	0.219	0.273	0.201
14	484	23.40	0.166	0.294	0.263	0.201
15	291	14.00	0.104	0.190	0.129	0.129
16	101	4.87	0.033	0.064	0.032	0.093
17	8	0.39	$\frac{5}{722} = 0.007$	0.004	0.004	$\frac{5}{194} = 0.026$
18	4	0.19	$\frac{5}{722} = 0.007$	0.003	0.002	$\frac{5}{194} = 0.026$

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Appendix 1: STR Allele Frequencies from U.S. Population Data, page 505.

3

Options

Factor of 10 and minimum allele frequency of 5/2N

NRC II – National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. National Academy Press, Washington D.C., 1996.

Normal approximation

Chakraborty R., Srinivasan M.R., Daiger S.F. (1993). Evaluation of standard errors and confidence intervals of estimated multilocus genotype probabilities and their implications in DNA. *Am. J. Hum. Genet., 52*: 60-70.
Good I.J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika, 40*: 237-264.

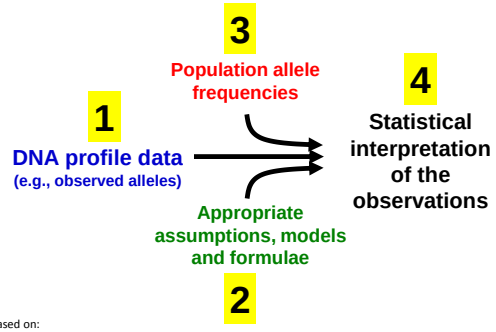
Size bias correction

Balding D.J. (1995). Estimating products in forensic identification using DNA profiles. *J. Am. Stat. Assoc., 90*: 839-844.

Highest posterior density

Curran J.M., Buckleton J.S., Triggs C.M., Weir B.S. (2002). Assessing uncertainty in DNA evidence caused by sampling effects. *Sci. Justice, 42*: 29-37.
Curran J.M., Buckleton J.S. (2011). An investigation into the performance of methods for adjusting for sampling uncertainty in DNA likelihood ratio calculations. *Forensic Sci. Int.: Genet., 5*: 512-516.

Elements required for a statistical interpretation

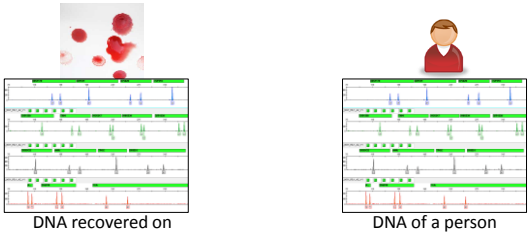


Based on:
J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*: **Figure 9.1**, page 214.

4

Logical Appraoch for Evidence Interpretation

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapters 9 and 11: pages 224-228, 295-297, and 302-303.



DNA recovered on the crime scene

DNA of a person of interest

Does the DNA recovered on the crime scene come from the person of interest?

4

Framing the question

Different questions have different answers.

Question 1 *profile probability*
What is the probability of observing this profile in the population?

Question 2 *match probability*
What is the probability of observing this profile in the population if we have already observed one person with this profile in this population?

Question 3 *combined probability of inclusion*
What is the probability that a person selected randomly in the population would be included (or not excluded) as a possible donor of the DNA?

Question 4 *likelihood ratio*
By how much do the DNA typing results support the ***person of interest*** being the donor?

4

Framing the question

A court of law is interested in what these DNA typing results mean in **this particular case**, with regard to **this particular person of interest** and the **case circumstances**.

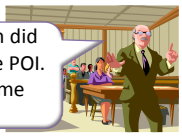
Question 4 is the question that addresses this issue.

4

There are two sides to every story…

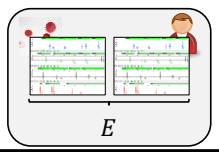
H_p: The crime stain came from the person of interest (POI).

H_d: The crime stain did not come from the POI. It came from some other person.



prosecution's proposition

defense's proposition



4

Likelihood Ratio (LR)

given or **if**

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$

The probability of observing the DNA typing results given that the prosecution's proposition is true

divided by

the probability of observing the DNA typing results given that the defense's proposition is true.

4

Likelihood Ratio (LR)

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$

The probability of observing the DNA typing results given that the prosecution's proposition is true

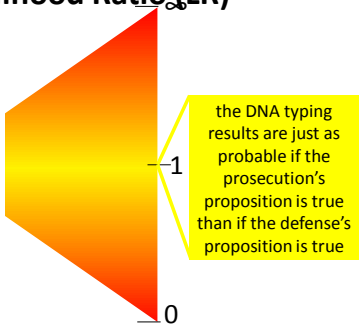
divided by

the probability of observing the DNA typing results given that the defense's proposition is true.

4

Likelihood Ratio (LR)

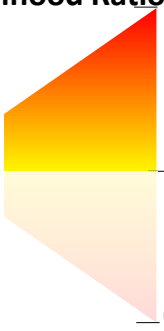
$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$



4

Likelihood Ratio (LR)

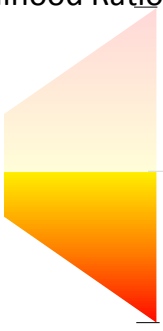
$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$



4

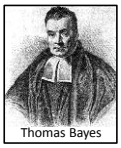
Likelihood Ratio (LR)

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$



4

Logical Framework for Updating Uncertainty


Thomas Bayes

Odds form of **Bayes' theorem**:

$$\underbrace{\frac{\Pr(H_p|E)}{\Pr(H_d|E)}}_{\text{posterior odds}} = \underbrace{\frac{\Pr(E|H_p)}{\Pr(E|H_d)}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{\Pr(H_p)}{\Pr(H_d)}}_{\text{prior odds}}$$

4

Logical Framework for Updating Uncertainty

$$\underbrace{\frac{\Pr(H_p|E)}{\Pr(H_d|E)}}_{\text{posterior odds}} = \underbrace{\frac{\Pr(E|H_p)}{\Pr(E|H_d)}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{\Pr(H_p)}{\Pr(H_d)}}_{\text{prior odds}}$$


factfinder

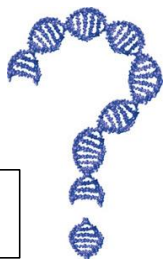

expert witness


factfinder

Acknowledgements

John Butler

Slides and discussions on forensic genetics
John Buckleton and Bruce Weir



Contact Info.:
simone.gittelson@nist.gov
+1-301-975-4892

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

Basic STR Interpretation Workshop
John M. Butler & Simone N. Gittelson
Krakow, Poland
31 August 2015

Data Interpretation 2:

Mixture interpretation: Clayton rules, # contributors
Stochastic effects and low-template DNA challenges
Worked examples

John M. Butler, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

ISFG

FORENSIC SCIENCES

Workshop Schedule

Time	Module (Instructor)	Topics
0900-0930	Welcome & Introductions	Review expectations and questions from participants
0930 – 1100	Data Interpretation 1 (John)	STR kits, loci, alleles, genotypes, profiles Data interpretation thresholds and models Simple PCR and CE troubleshooting
1100 – 1130	Break	
1130 – 1300	Statistical Interpretation 1 (Simone)	Introduction to probability and statistics STR population data collection, calculations, and use Approaches to calculating match probabilities
1300 – 1430	Lunch	
1430 – 1600	Data Interpretation 2 (John)	Mixture interpretation: Clayton rules, # contributors Stochastic effects and low-template DNA challenges Worked examples
1600 – 1630	Break	
1630 – 1800	Statistical Interpretation 2 (Simone)	Approaches to calculating mixture statistics Likelihood ratios and formulating propositions Worked examples

Information that goes into a DNA rarity estimate (i.e., where errors can occur)

3 Estimates are derived from testing a small subsection of a population

1 **Evidentiary DNA Profile** (with specific alleles/genotypes)

Population allele frequencies

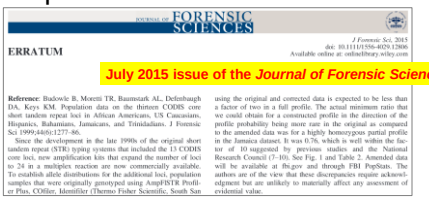
Genetic formulas and assumptions made

Rarity estimate of DNA profile (e.g., RMP or LR)

The risk of error goes up with complexity of the DNA profile (e.g., >2 person mixture or low-quality, low-template DNA sample)

2 "All models are wrong – but some are useful" (George Box, 1979)

Recent FBI Erratum on Allele Frequencies **Errors Made in 1999**


JOURNAL OF FORENSIC SCIENCES

ERRATUM

J Forensic Sci, 2015
doi: 10.1111/1556-4029.12806
Available online at: onlinelibrary.wiley.com

July 2015 issue of the *Journal of Forensic Sciences*

Reference: Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J Forensic Sci 1999;44(6):1277–86.

Since the development in the late 1990s of the original short tandem repeat (STR) typing systems that included the 13 CODIS core loci, new amplification kits that expand the number of loci to 24 in a multiplex reaction are now commercially available. To establish allele distributions for the additional loci, population samples that were originally genotyped using AmpFlSTR Profiler Plus, COfiler, Identifiler (Thermo Fisher Scientific, South San

using the original and corrected data is expected to be less than a factor of two in a full profile. The actual minimum ratio that we could obtain for a constructed profile in the direction of the profile probability being more rare in the original as compared to the amended data was for a highly homozygous partial profile in the Jamaica dataset. It was 0.76, which is well within the factor of 10 suggested by previous studies and the National Research Council (7–10). See Fig. 1 and Table 2. Amended data will be available at fbi.gov and through FBI PopStats. The authors are of the view that these discrepancies require acknowledgment but are unlikely to materially affect any assessment of evidential value.

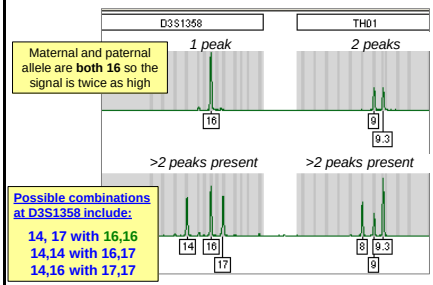
- **Genotyping errors were made in 27 samples, affecting the reported frequencies of 51 alleles** **In Table 1, 255 allele frequencies are impacted**
- For alleles requiring a frequency correction, the magnitude of the change in frequencies ranged from 0.000012 to 0.018 (average 0.0020 ± 0.0025)
- "The authors are of the view that these discrepancies require acknowledgment but are **unlikely to materially affect any assessment of evidential value**"

How Book Chapters Map to Data Interpretation Process

Chapter	Input Information	Decision to be made	How decision is made
2	Data file	**Peak** or **Noise**	Analytical threshold
3	Peak	**Allele** or **Artifact**	Stutter threshold; precision sizing bin
4	Allele	**Heterozygote** or **Homozygote** or **Allele(s) missing**	Peak heights and peak height ratios; stochastic threshold
5	Genotype/full profile	**Single-source** or **Mixture**	Numbers of peaks per locus
6	Mixture	Deconvolution or not	Major/minor mixture ratio
7	Low level DNA	Interpret or not	Complexity threshold
In Data Interp2 presentation 8	Poor quality data	Replace CE components (buffer, polymer, array) or call service engineer	Review size standard data quality with understanding of CE principles

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Table 1.1**, p. 6

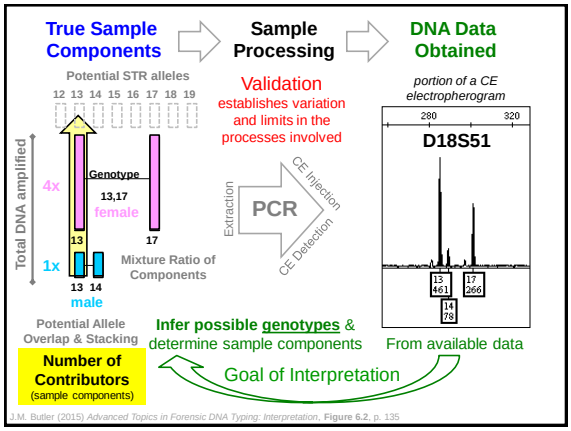
Single-Source Sample vs Mixture Results



Single-Source

Mixture

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



The Forensic Institute

Forensic e-Symposium™
The Forensic Institute

April 14, 2005

"If you show 10 colleagues a mixture, you will probably end up with 10 different answers."
- Dr. Peter Gill

"Don't do mixture interpretation unless you have to"
- Dr. Peter Gill (1998)

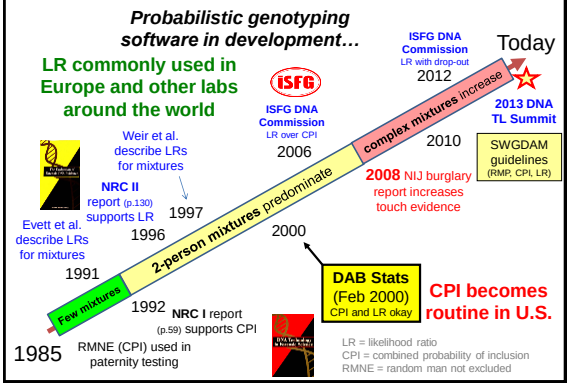
A Brief History of DNA Mixtures (1)

- **1991**– Ian Evett article (with single-locus RFLP probes)
- **1995** – Mixtures presented in OJ Simpson trial
- **1996** – 9plex STR kits (Profiler Plus, PowerPlex 1.1)
- **1997** – Weir et al using Likelihood Ratios (LRs) for mixture statistics
- **1998** – **Clayton et al (FSS) DNA mixture deconvolution**
- **2000** – initial SWGDAM Interpretation Guidelines published
- **2000** – Combined Probability of Inclusion (CPI) statistic is allowed by DNA Advisory Board and pushed by the FBI
- **2000** – 16plex STR kits (PP16 and Identifiler)
- **2005** – NIST Interlaboratory Mixture Study **(MIX05) finds extensive variation in laboratory approaches**

A Brief History of DNA Mixtures (2)

- **2006** – ISFG Mixture Recommendations published emphasizing that LRs are a better method over CPI
- **2007** – informal SWGDAM study finds most labs doing 2-person mixtures (committee begins writing guidelines)
- **2008** – NIJ study shows value of DNA in burglary cases and more touch DNA samples with complex mixtures begin being processed
- **2010** – SWGDAM Interpretation Guidelines emphasize need for statistics and stochastic thresholds with CPI; probabilistic genotyping approach is mentioned
- **2012** – ISFG publishes LR with probability of dropout to cope with potential of allele dropout
- **2013** – Another NIST Interlaboratory Study **(MIX13) finds extensive variation in laboratory approaches**
- ***Present*** – a number of software programs exist to help with calculations but no universal approach exists

Historical Perspective on DNA Mixture Approaches



Statistical Approaches with Mixtures

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

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

2. **Combined Probability of Exclusion/Inclusion** – CPE/CPI (RMNE) – Calculation of the probability that a random (unrelated) person would be excluded/included as a contributor to the observed DNA mixture
   RMNE = Random Man Not Excluded (same as CPI)
   CPE = Combined Probability of Exclusion (CPE = 1 – CPI)
   CPI = Combined Probability of Inclusion (CPI = 1 – CPE)

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

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

The FBI DNA Advisory Board (DAB) Recommendations on Statistics

February 23, 2000

Forensic Sci. Comm. 2(3); available on-line at https://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2000/index.htm/dnastat.htm

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

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

NIST Interlaboratory Mixture Studies

http://www.cstl.nist.gov/strbase/interlab.htm

- **Provide a big-picture view of the community**
  - not graded proficiency tests
  - offers laboratories an opportunity to directly compare themselves to others in an anonymous fashion
- Some lessons learned:
  - instrument sensitivities can vary significantly
  - amount of input DNA plays important role in ability to detect minor component(s)
  - **protocols and approaches are often different between forensic labs**
- Studies Conducted

Study	Year	# Labs	# Samples	Mixture Types
MSS 1	1997	22	11 stains	ss, 2p, 3p
MSS 2	1999	45	11 stains	ss, 2p, 3p
MSS 3	2000-01	74	7 extracts	ss, 2p, 3p
MIX05	2005	69	4 cases (.fsa)	only 2p
MIX13	2013	108	5 cases (.fsa)	2p, 3p, 4p

MSS: mixed stain study

ss: single-source; 2p: 2-person, etc.

Elements of DNA Mixture Interpretation

Principles (theory)

↓

Protocols (validation)

↓

Practice (training & experience)

Consistency across analysts

ISFG Recommendations
SWGDAM Guidelines

↓

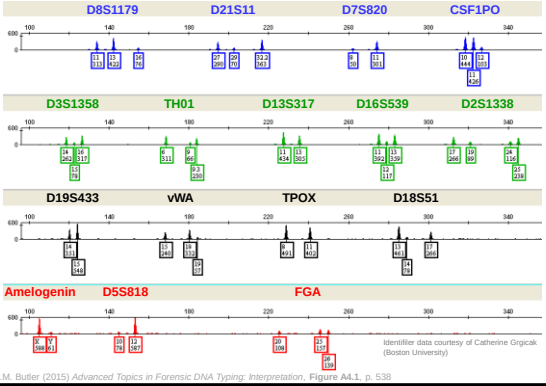
Your Laboratory SOPs

↓

Training within Your Laboratory

Periodic training will aid accuracy and efficiency within your laboratory

Worked Example Mixture in Interpretation (2015) Book



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, Figure A4.1, p. 538

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

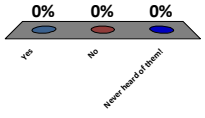
Received 4 April 2006; accepted 10 April 2006
Available online 5 June 2006

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

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

Have you read the 2006 ISFG DNA Commission Recommendations on Mixture Interpretation?

A. Yes
B. No
C. Never heard of them!



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

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

"Clayton Rules": Steps in the Interpretation of DNA Mixtures (Clayton *et al.* 1998)

Step #1 Identify the Presence of a Mixture

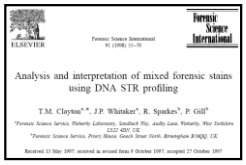
Step #2 Designate Allele Peaks

Step #3 Identify the Number of Potential Contributors

Step #4 Estimate the Relative Ratio of the Individuals Contributing to the Mixture

Step #5 Consider All Possible Genotype Combinations

Step #6 Compare Reference Samples


Analysis and interpretation of mixed forensic stains using DNA STR profiling

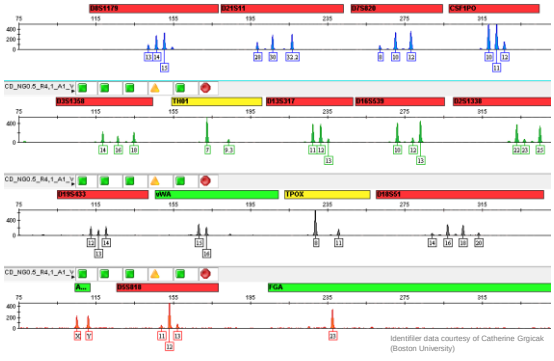
T.M. Clayton, J.P. Whitaker, R. Sparkes, P. Gill

Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70

Step #1: Is a Mixture Present in an Evidentiary Sample?

- Examine the number of peaks present in a locus
  - More than 2 peaks at a locus (except for tri-allelic patterns at perhaps one of the loci examined)
- Examine relative peak heights
  - Heterozygote peak imbalance <60%
  - Peak at stutter position >15%
- Consider all loci tested

DNA Mixture Example for this Workshop



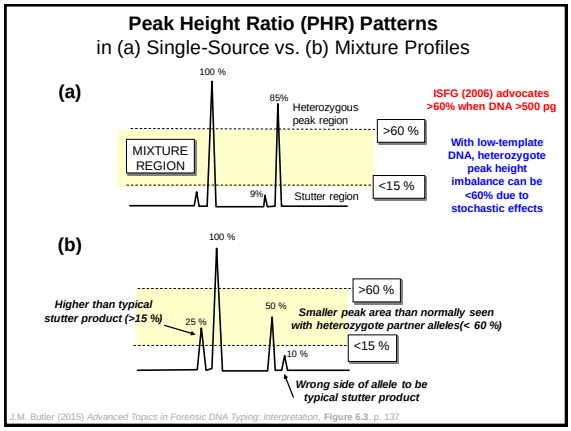
Identifiler data courtesy of Catherine Grgicak (Boston University)

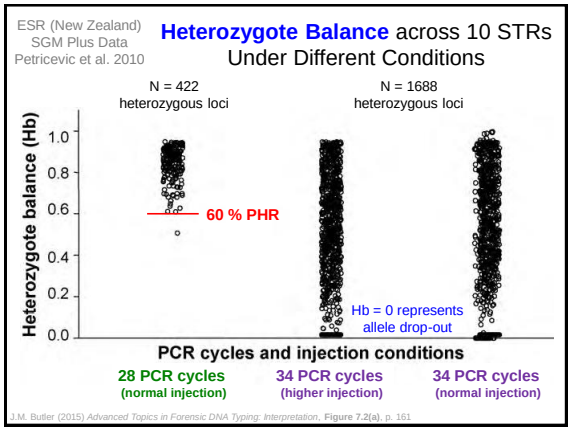
Is a DNA Profile Consistent with Being a Mixture?

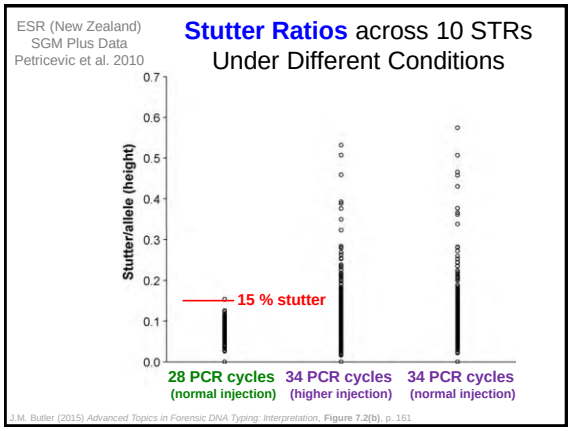
From J.M. Butler (2005) *Forensic DNA Typing, 2nd Edition*, pp. 156-157

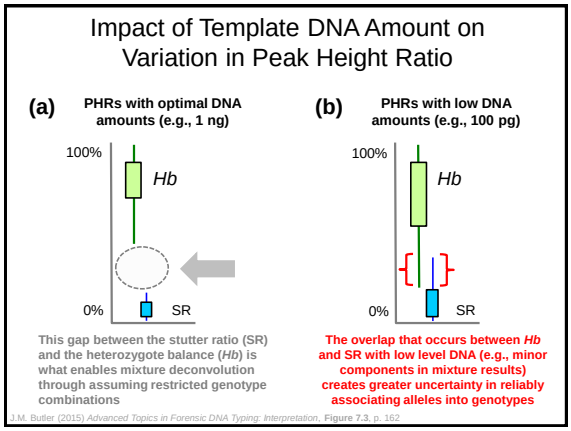
If the answer to any one of the following three questions is yes, then the DNA profile may very well have resulted from a mixed sample:

- Do any of the loci show more than two peaks in the expected allele size range?
- Is there a severe peak height imbalance between heterozygous alleles at a locus?
- Does the stutter product appear abnormally high (e.g., >15-20%)?



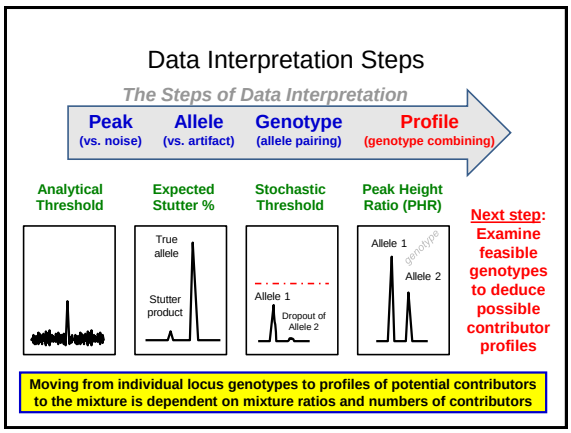






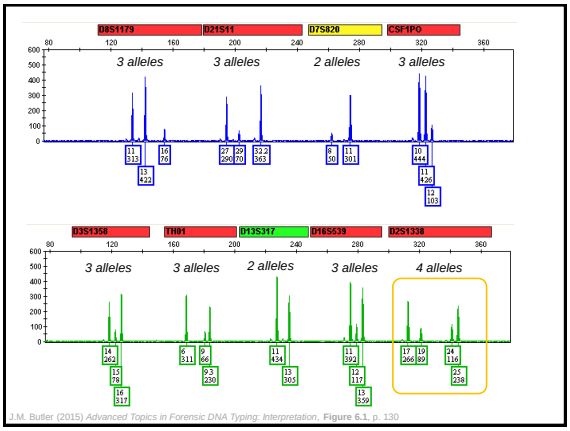
Step #2: Designate Allele Peaks

- Use regular data interpretation rules to decipher between true alleles and artifacts
- Use stutter filters to eliminate stutter products from consideration (although stutter may hide some of minor component alleles at some loci)
- Consider heterozygote peak heights that are highly imbalanced (<60%) as possibly coming from two different contributors

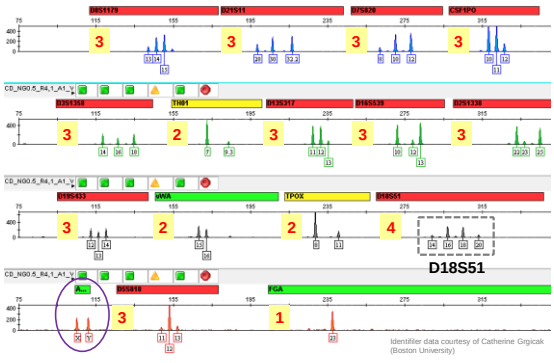


Step #3: Identifying the Potential Number of Contributors

- **Important for statistical calculations**
- Typically if 2, 3, or 4 alleles then 2 contributors
- If 5 or 6 alleles per locus then 3 contributors
- If >6 alleles in a single locus, then >4 contributors
- Also pay attention to relative peak heights and potential genotype combinations

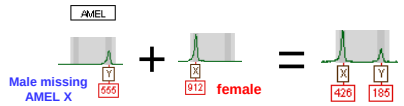


DNA Mixture Example for this Workshop



Potential Problems with Amelogenin

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



Comparison of Expected and Simulated Mixture Results

Expected Results when estimating # of contributors:

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

Results from Simulation Studies:

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

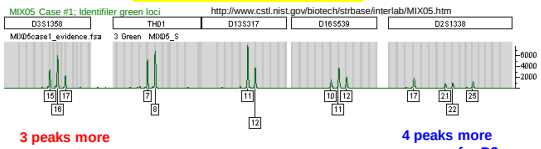
Buckleton et al. (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *FSI Genetics* 1:20-28

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

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

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

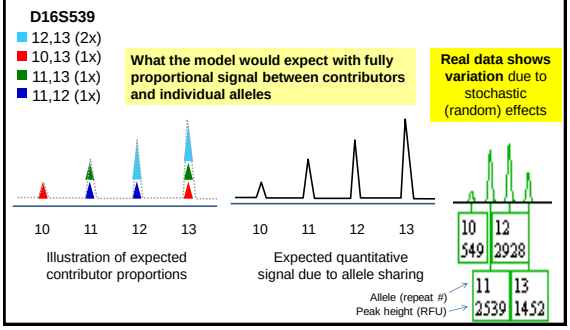
Results from a 2-Person Mixture



3 peaks more common for D3

4 peaks more common for D2

Impact of Allele Sharing on Ability to Clearly Determine the Number of Contributors



Impact of Additional STR Loci on Mixture Assumptions

Probability of incorrectly assigning the specific number of contributors based on observed alleles
(not considering peak height imbalances)

True # of contributors	Using NIST Caucasians (Hill et al. 2013)	1	2	3	4	5
6	CODIS13	1.75E-40	6.34E-09	0.161242	0.945657	0.999873
	CODIS22	0 (< E-99)	9.59E-21	5.32E-05	0.188138	0.859901
5	CODIS13	9.78E-33	2.10E-06	0.41432	0.989651	
	CODIS22	6.36E-61	7.01E-15	0.004837	0.610149	
4	CODIS13	7.02E-25	0.000515	0.785495		
	CODIS22	3.50E-46	3.49E-09	0.16523		
3	CODIS13	8.42E-17	0.059486			
	CODIS22	5.77E-31	0.000433			
2	CODIS13	1.70E-08				
	CODIS22	2.05E-15				

0.05%

With 13 CODIS loci, **5.9% of 3-person contributors could falsely be considered a 2-person mixture** based on observed alleles (using NIST Caucasian allele frequencies)

With expanded CODIS loci, this drops to **0.04%**

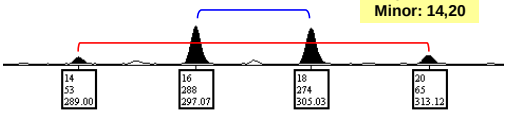
Coble, Bright, Buckleton, Curran (2015) Uncertainty in the number of contributors in the proposed new CODIS set. *FSI Genetics, in press*

Step #4: Estimation of Relative Ratios for Major and Minor Components to a Mixture

- Mixture studies with known samples have shown that the mixture ratio between loci is fairly well preserved during PCR amplification
- Thus it is generally thought that the peak heights (areas) of alleles present in an electropherogram can be related back to the initial component concentrations
- For 2-person mixtures, **start with loci possessing 4 alleles…**

Mixture Ratio and Mixture Proportion Determined Using D18S51 Peak Heights

If we assume:
Major: 16,18
Minor: 14,20



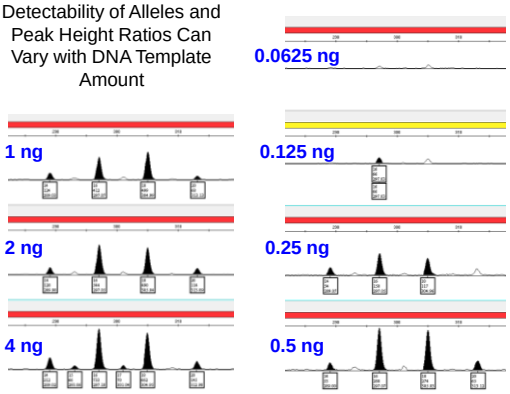
Mixture Ratio

$$M_R = \frac{\varphi_{16}+\varphi_{18}}{\varphi_{14}+\varphi_{20}} = \frac{288+274}{53+65} = 4.76$$

Mixture Proportion

$$M_x = \frac{\varphi_{16}+\varphi_{18}}{\varphi_{14}+\varphi_{16}+\varphi_{18}+\varphi_{20}} = \frac{288+274}{53+288+274+65} = 0.826 = 83\%$$

Detectability of Alleles and Peak Height Ratios Can Vary with DNA Template Amount



Step #5: Consider All Possible Genotype Combinations

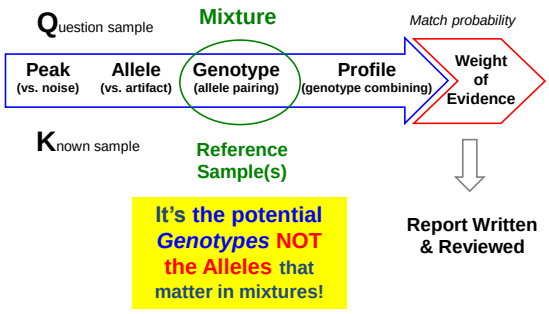
Table 3
Pairwise combinations of two, three and four alleles

Four alleles (a,b,c,d)		Three alleles (a,b,c)		Two alleles (a,b)	
a,b	c,d	a,a	b,c	a,a	a,b
a,c	b,d	b,b	a,c	a,b	a,b
a,d	b,c	c,c	a,b	a,a	b,b
c,d	a,b	a,b	a,c	a,b	b,b
b,d	a,c	b,c	a,c	a,b	a,a
b,c	a,d	a,b	b,c	b,b	a,a
		b,c	a,a	b,b	a,b
		a,c	b,b		
		a,b	c,c		
		a,c	a,b		
		a,c	b,c		
		b,c	a,b		

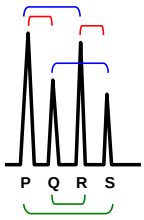
Key: bold entries represent reciprocal combinations.

Clayton *et al. Forensic Sci. Int.* 1998; 91:55-70

Steps in DNA Interpretation



Considering Genotype Combinations

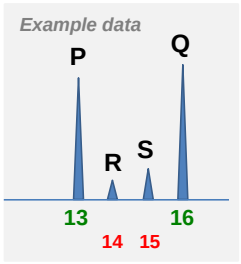


PR
QS
PQ
RS
QR
PS

Depends on PHR

Peak Height Ratios (PHR)
Minimum Peak Height (mPH)
Proportion (p) or mixture proportion (M_x)

Restricted vs Unrestricted Genotype Combinations



Unrestricted
genotype combinations

PQ & RS 13,16 & 14,15	**RS & PQ** 14,15 & 13,16
PR & SQ 13,14 & 15,16	**SQ & PR** 15,16 & 13,14
PS & RQ 13,15 & 14,16	**RQ & PS** 14,16 & 13,15

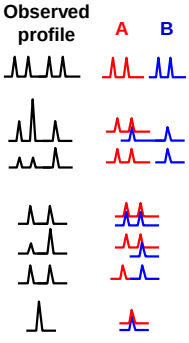
Restricted
genotype combinations

PQ & RS 13,16 & 14,15	**RS & PQ** 14,15 & 13,16

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 6.4**, p. 146

Possible genotype combinations in 2-person mixtures

Observed profile



4 alleles
All heterozygotes and non-overlapping alleles

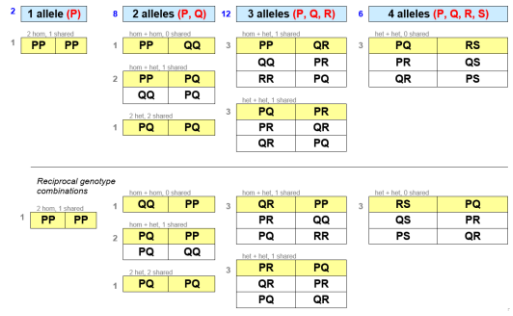
3 alleles
Heterozygote + heterozygote, one overlapping allele
Heterozygote + homozygote, no overlapping alleles

2 alleles
Heterozygote + heterozygote, two overlapping alleles
Heterozygote + homozygote, one overlapping allele
Homozygote + homozygote, no overlapping alleles

1 allele
Homozygote + homozygote, overlapping allele

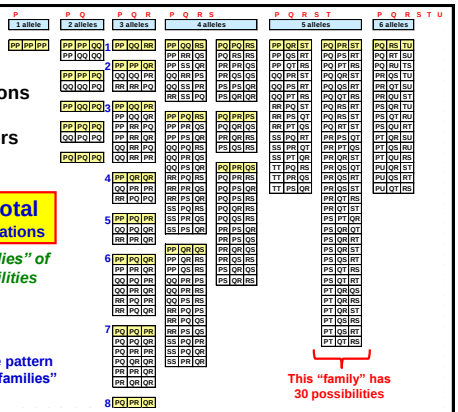
J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 6.4**, p. 139

Possible Genotype Combinations with Two-Person Mixtures



J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Table 6.1**, p. 139

Potential Genotype Combinations with Three Contributors

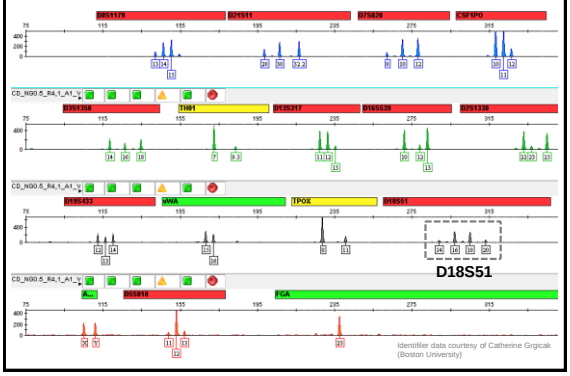


J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 7.4**, p. 168

Step #6: Compare Reference Samples

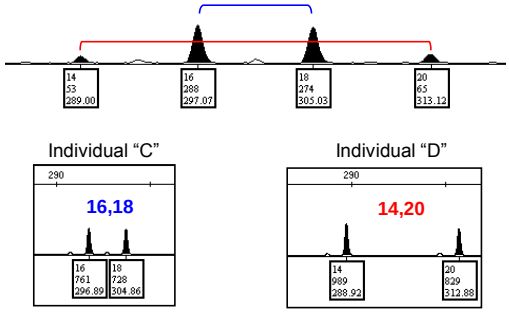
- If there is a suspect, a laboratory must ultimately decide to include or exclude him…
- **If no suspect is available for comparison, does your laboratory still work the case?** (Isn't this a primary purpose of the national DNA database?)
- Victim samples can be helpful to eliminate their allele contributions to intimate evidentiary samples and thus help deduce the perpetrator

DNA Mixture Example for this Workshop

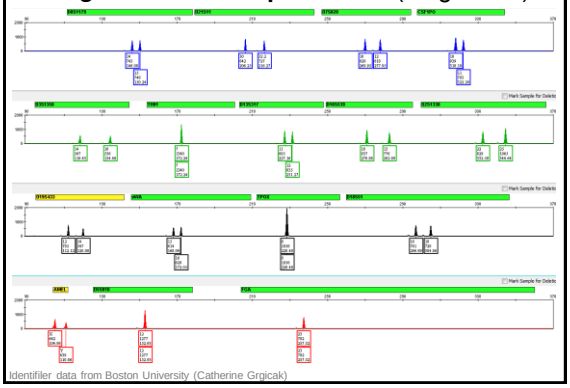


D18S51 Results

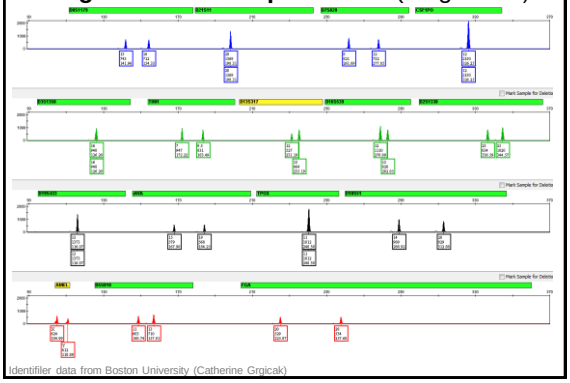
Major: 16,18
Minor: 14,20



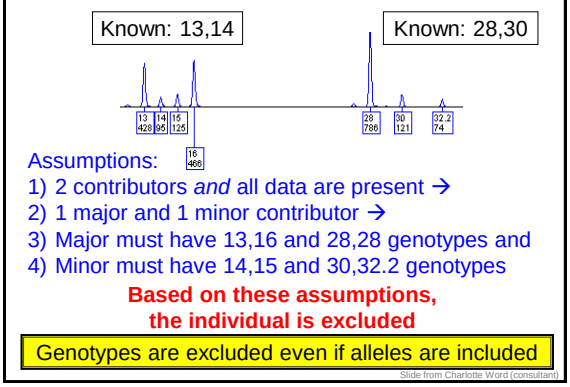
Single-Source Sample Profile (1 ng of "C")



Single-Source Sample Profile (1 ng of "D")



Is the Known Individual Included or Excluded?



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 the Individuals Contributing to the Mixture

Step #5 Consider All Possible Genotype Combinations

Step #6 Compare Reference Samples

Step #7 **Determine statistical weight-of-evidence**

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 6.5**, p. 143

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"

Mixture Classification Flowchart

Is the sample a mixture?

>2 alleles at a locus, except tri-alleles? — NO → Single Source DNA Sample → Determine STR profile and compute RMP

YES → Mixed DNA Sample

Differentiate a Major/Minor Component? — YES → TYPE B → Determine component profile(s) and compute RMP for major

NO → Stochastic Effects ? Possible Low Level DNA) ? — YES → TYPE C → A biostatistical analysis should not be performed

NO → TYPE A → A biostatistical analysis must be performed

Probability of Exclusion [CPE] "RMNE" ← NO ← Assume number of contributors ? — YES → Likelihood Ratio [LR]

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Box 6.3**, p. 141

Information from Chapter 7 of my New Book

Advanced Topics in Forensic DNA Typing: Interpretation

CHAPTER 7

Low-Level DNA and Complex Mixtures

"The limits of each DNA typing procedure should be understood, especially when the DNA sample is small, is a mixture of DNA from multiple sources, or is contaminated with interfering substances."
NRC I, 1992, p. 8

"For the complex DNA profile, there is no predominant or overarching standard interpretation method."
Peter Gill (Gill et al. 2012, report to the UK Forensic Science Regulator, p. 18)

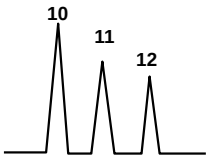
"The limits of each DNA typing procedure should be understood, especially when the DNA sample is small, is a mixture of DNA from multiple sources…" (NRC I, 1992, p. 8)

Butler, J.M. (2015) *Advanced Topics in Forensic DNA Typing: Interpretation* (Elsevier Academic Press: San Diego), pp. 159-182

Allele stacking with overlapping genotypes

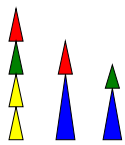
(a)

observed data



(b)

overlapping genotypes result in shared, stacked alleles



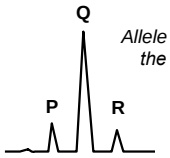
contributor genotypes

10,11
10,12
10,10
11,12

1:1:1:2

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 7.1**, p. 160

Potential Impact of Allele Sharing



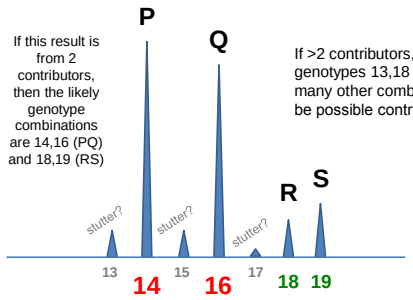
Allele Q may not represent the "major" contributor

Contributor A **PQ**
Contributor B **QQ**
Contributor C **QR**

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 6.6**, p. 154

Peaks in the Stutter Position of Major Alleles May Need to Be Considered as Pairing with Alleles from Minor Contributors

If this result is from 2 contributors, then the likely genotype combinations are 14,16 (PQ) and 18,19 (RS)



If >2 contributors, then genotypes 13,18 or 15,19 (and many other combinations) may be possible contributors…

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Figure 6.7**, p. 154

http://www.cstl.nist.gov/strbase/pub_pres/Butler-DNA-interpretation-AAFS2015.pdf

5 Reasons that DNA Results Are Becoming More Challenging to Interpret

1. **More sensitive DNA test results**
2. **More touch evidence samples** that are poor-quality, low-template, complex mixtures
3. **More options exist** for statistical approaches involving probabilistic genotyping software
4. **Many laboratories are not prepared** to cope with complex mixtures
5. **More loci being added** because of the large number of samples in DNA databases

More Sensitive Assays and Instruments

- **Superb sensitivity is available** with DNA amplification using the polymerase chain reaction and laser-induced fluorescence detection with capillary electrophoresis
- Since 2007 (beginning with the release of the MiniFiler STR kit), **improved buffers and enzymes** have been used to boost DNA sensitivities in all STR kits
  - In 2010 the ABI 3500 Genetic Analyzer was released with 4X signal over the previous ABI 3100 and ABI 310 instruments
  - Energy-transfer dyes are used with some of the STR kits
  - Some labs increase the sensitivity dial with additional PCR cycles
- **So what is wrong with have improved sensitivity?**

Improved Sensitivity is a Two-Edged Sword

"As sensitivity of DNA typing improves, laboratories' abilities to examine smaller samples increases. This improved sensitivity is a two-edged sword. **With greater capabilities comes greater responsibilities to report meaningful results.** Given the possibility of DNA contamination and secondary or even tertiary transfer in some instances, **does the presence of a single cell (or even a few cells) in an evidentiary sample truly have meaning?**…"

Butler, J.M. (2015) *Advanced Topics in Forensic DNA Typing: Interpretation* (Elsevier Academic Press: San Diego), p. 458

Ian Evett and Colleagues' **Case Assessment and Interpretation:**

Hierarchies of Propositions

TABLE 16.2 Hierarchical Levels of Propositions Originally Developed by the UK Forensi 1998a, 1998b, Evett et al. 2000a, 2000b, Gill 2001)

Hierarchy Levels		Propositions	Decision Maker
Level III	Offense	Supplies the probability that a suspect has committed a criminal offense	Responsibility of the jury or judge
Level II	Activity	Informs regarding the kinds of activities which may have produced the forensic evidence	Jury or possibly scientist if given adequate case circumstances
Level I	Source	Addresses the source of the sample	Scientist
Sub-level I	Sub-source	With low amounts of DNA, the scientist may not be able to infer how the DNA arrived at the site where the DNA sample was collected	Scientist

Butler, J.M. (2015) *Advanced Topics in Forensic DNA Typing: Interpretation* (Elsevier Academic Press: San Diego), p. 458

More Touch Evidence Samples

https://www.ncjrs.gov/pdffiles1/nij/grants/222318.pdf

NIJ April 2008 Research Report

http://www.nij.gov/journals/261/pages/dna-solves-property-crimes.aspx

NIJ Journal October 2008 (vol. 261, pp. 2-12)

- **More poor-quality samples are being submitted**
  - Samples with <100 pg of DNA submitted in Belgium:
    19% (2004) → 45% (2008)
    (Michel 2009 FSIGSS 2:542-543)
- AAFS 2014 presentations showed poor success rates
  - NYC (**A110**): **only 10% of >9,500 touch evidence swabs from 2007 to 2011 produced usable DNA results**
  - Allegheny County (**A114**): examined touch DNA items processed from 2008 to 2013 across different evidence types (e.g., 6 of 56 car door handles yielded "resolvable profiles")

New Options Exist for Statistical Analysis

- Increase in approaches to try and cope with potential allele dropout → number of **probabilistic genotyping** methods have grown since Balding & Buckleton 2009 article
- Many possible choices for **probabilistic genotyping software** with commercial interests at stake

Balding, D.J. & Buckleton, J. (2009) Interpreting low template DNA profiles. *Forensic Sci. Int. Genet.* 4(1):1-10.

Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci. Int.* 112(1):17-40.

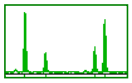
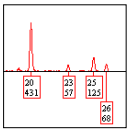
TABLE 13.1 **Probabilistic Genotyping Software Programs (as of March 2014)**

Program Name	Type	Creator(s)	Availability
LRmix	Discrete (semi-continuous)	Hinda Haned & Peter Gill	Open-source https://sites.google.com/site/forensicdnastatistics/PCR-simulation/lrmix
Lab Retriever	Discrete (semi-continuous)	Developed by David Balding and maintained by Norah Rudin and colleagues	Open-source http://www.scieg.org/lab_retriever.html
likeLTD	Discrete (semi-continuous)	David Balding	Open-source https://sites.google.com/site/baldingstatisticalgenetics/software/likeltd-r-forensic-dna-r-code
FST	Discrete (semi-continuous)	Adele Mitchell	Proprietary to the NYC OCME Forensic Biology Laboratory
Armed Xpert	Discrete (semi-continuous)	Developed by USACIL and maintained and improved by NicheVision	Commercial product http://www.armedxpert.com/
TrueAllele	Fully-continuous	Mark Perlin	Commercial product http://www.cybgen.com/
STRmix	Fully-continuous	Duncan Taylor, Jo-Anne Bright, John Buckleton	Commercial product http://strmix.esr.cri.nz/
DNA View Mixture Solution	Fully-continuous	Charles Brenner	Commercial product http://dna-view.com/

Discrete (semi-continuous) methods use only the allele information in conjunction with probabilities of drop-out and drop-in.
Fully-continuous methods use peak height data and other parameters in addition to the allele information.

Butler, J.M. (2015) *Advanced Topics in Forensic DNA Typing: Interpretation* (Elsevier Academic Press: San Diego), p. 341

Math Analogy to DNA Evidence

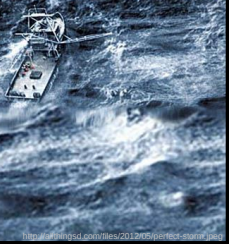
2 + 2 = 4	$2x^2 + x = 10$	$\int_{x=0}^{\infty} f(x)dx$
Basic Arithmetic	**Algebra**	**Calculus**
		
Single-Source DNA Profile (DNA databasing)	**Sexual Assault Evidence** (2-person mixture with high-levels of DNA)	**Touch Evidence** (>2-person, low-level, complex mixtures perhaps involving relatives)

Many laboratories are not prepared to cope with complex mixtures

- Have **appropriate validation studies** been performed to inform proper interpretation protocols? (curriculum & classroom instruction)
- Are **appropriately challenging proficiency tests** being given? (graded homework assignments)
- **Would we want to go into a calculus exam only having studied algebra and having completed homework assignments involving basic arithmetic?**

Are We Facing a "Perfect Storm" for DNA Testing and Interpretation?

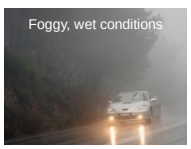
- Increase in assay and instrument sensitivity
- Increase in challenging casework samples (touch evidence)
- Increase in possible statistical tools for use with complex mixtures
- Increase in number of loci examined with new STR kits



http://allthingsd.com/files/2012/05/perfect-storm.jpeg

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

Poor Quality Conditions



Large Numbers of Contributors

Decisions during Data Interpretation

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

J.M. Butler (2015) *Advanced Topics in Forensic DNA Typing: Interpretation*, **Table 1.1**, p. 6

Results Depend on Assumptions

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

Ian Evett on Interpretation

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

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

Acknowledgments

$ NIST Special Programs Office
Simone Gittelson

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

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

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

Basic STR Interpretation Workshop
John M. Butler & Simone N. Gittelson
Krakow, Poland
31 August 2015

Statistical Interpretation 2:

Approaches to calculating mixture statistics
Likelihood ratios and formulating propositions
Worked examples

Simone N. Gittelson, Ph.D.
U.S. National Institute of Standards and Technology
31 August 2015

Acknowledgement and Disclaimers

I thank John Butler for the discussions and advice on preparing this presentation. I also acknowledge John Buckleton for all his helpful explanations on DNA mixture interpretation.

Points of view in this presentation are mine and do not necessarily represent the official position or policies of the National Institute of Standards and Technology.

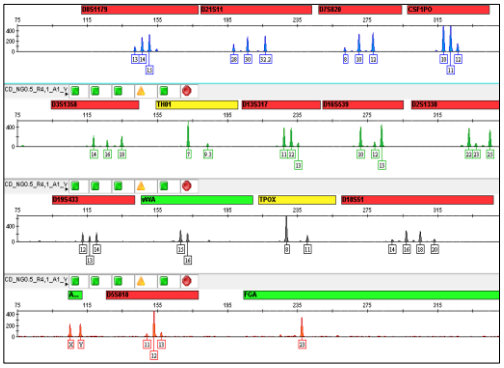
Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that any of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

Presentation Outline

1. Combined Probability of Inclusion (CPI)
2. modified Random Match Probability (mRMP)
3. Likelihood Ratio (LR)
4. Formulating Propositions for Likelihood Ratios

Boston University Mixture (http://www.bu.edu/dnamixtures/)
name: ID_2_SCD_NG0.5_R4,1_A1_V1

EPG of the crime stain:



Combined Probability of Inclusion (CPI)

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 12 and Appendix 4: pages 312-316, 320-322, 335-338, and 549-552.

Combined Probability of Inclusion (CPI)

the probability that a random person would be included as a contributor to the mixture

the probability that a random person would not be exluded as a contributor to the mixture

‘‘random man not excluded’’

Combined Probability of Inclusion (CPI)

answers the question:

What proportion of the population is expected to **be included** as a possible contributor to this mixture?

or

What proportion of the population is expected to **not be excluded** as a possible contributor to this mixture?

Information it takes into account

- presence of alleles

Assumptions

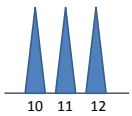
The population is in Hardy-Weinberg and Linkage equilibrium.

All genotypes considered are **equally likely**.

We do not take into account the number of contributors in the calculation.

Combined Probability of Inclusion (CPI)

CSF1PO
$p_{10} = 0.220$
$p_{11} = 0.309$
$p_{12} = 0.360$



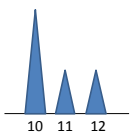
$$PI = p_{10}^2 + 2p_{10}p_{11} + 2p_{10}p_{12} + p_{11}^2 + 2p_{11}p_{12} + p_{12}^2$$
$$= (p_{10} + p_{11} + p_{12})^2$$
$$= (0.220 + 0.309 + 0.360)^2$$
$$= 0.79$$

the probability that a random person would be included as a contributor to this mixture

random person:
10,10
10,11
10,12
11,11
11,12
12,12

Combined Probability of Inclusion (CPI)

CSF1PO
$p_{10} = 0.220$
$p_{11} = 0.309$
$p_{12} = 0.360$



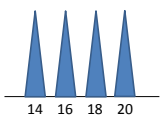
$$PI = p_{10}^2 + 2p_{10}p_{11} + 2p_{10}p_{12} + p_{11}^2 + 2p_{11}p_{12} + p_{12}^2$$
$$= (p_{10} + p_{11} + p_{12})^2$$
$$= (0.220 + 0.309 + 0.360)^2$$
$$= 0.79$$

the probability that a random person would be included as a contributor to this mixture

random person:
10,10
10,11
10,12
11,11
11,12
12,12

Combined Probability of Inclusion (CPI)

D18S51
$p_{14} = 0.134$
$p_{16} = 0.147$
$p_{18} = 0.078$
$p_{20} = 0.018$



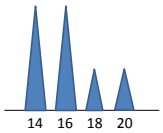
$$PI = p_{14}^2 + 2p_{14}p_{16} + 2p_{14}p_{18} + 2p_{14}p_{20} + p_{16}^2 + 2p_{16}p_{18} + 2p_{16}p_{20} + p_{18}^2 + 2p_{18}p_{20} + p_{20}^2$$
$$= (p_{14} + p_{16} + p_{18} + p_{20})^2$$
$$= (0.134 + 0.147 + 0.078 + 0.018)^2$$
$$= 0.14$$

the probability that a random person would be included as a contributor to this mixture

random person:
14,14
14,16
14,18
14,20
16,16
16,18
16,20
18,18
18,20
20,20

Combined Probability of Inclusion (CPI)

D18S51
$p_{14} = 0.134$
$p_{16} = 0.147$
$p_{18} = 0.078$
$p_{20} = 0.018$



random person:
14,14
14,16
14,18
14,20
16,16
16,18
16,20
18,18
18,20
20,20

$$PI = p_{14}^2 + 2p_{14}p_{16} + 2p_{14}p_{18} + 2p_{14}p_{20} + p_{16}^2 + 2p_{16}p_{18} + 2p_{16}p_{20} + p_{18}^2 + 2p_{18}p_{20} + p_{20}^2$$

$$= (p_{14} + p_{16} + p_{18} + p_{20})^2$$

$$= (0.134 + 0.147 + 0.078 + 0.018)^2$$

$$= 0.14$$

the probability that a random person would be included as a contributor to this mixture

Combined Probability of Inclusion (CPI)

$$CPI = PI_{CSF1PO} \times PI_{D18S51}$$

$$= 0.79 \times 0.14$$

$$= 0.11$$

HOWEVER, CPI can only be applied if…

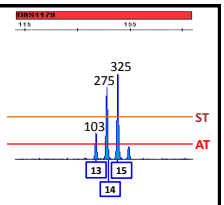
…there is **no possibility of allele drop-out**.

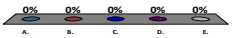
If there is a possibility of allele drop-out, then everyone would be included as a possible contributor. In this case, the probability that a random person would be included is equal to 1.

Can CPI be applied at this locus?

A. Yes
B. Maybe
C. No, because the peaks are too high
D. No, because allele drop-out is possible
E. No, because it's a mixture

Response Counter





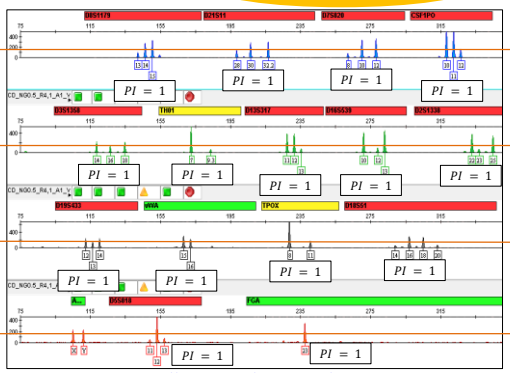
Can CPI be applied?

stochastic threshold = 150 rfu

Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

Probabilities of Inclusion

stochastic threshold = 150 rfu



Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

Combined Probability of Inclusion (CPI)

$$CPI = PI_{D8S1179} \times PI_{D21S11} \times PI_{D7S820} \times PI_{CSF1PO} \times PI_{D3S1358} \times PI_{TH01} \times PI_{D13S317} \times PI_{D16S539} \times PI_{D2S1338} \times PI_{D19S433} \times PI_{vWA} \times PI_{TPOX} \times PI_{D18S51} \times PI_{D5S818} \times PI_{FGA}$$

$$= 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1$$

$$= 1$$

Everyone is included.
No one is excluded.

modified Random Match Probability (mRMP)

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 12 and Appendix 4: pages 314-315, 325 and 553-558.

modified Random Match Probability (mRMP)

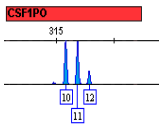
answers the question:

What proportion of the population is expected to have a particular **genotype**, or a particular set of **genotypes**?

These genotypes are inferred from the mixture based on the observed peaks, peak heights, and inferred number of contributors.

Information it takes into account

- presence of genotypes



presence of alleles and peak heights

list of genotype combinations that are possible

- peaks below stochastic threshold (where allele drop-out is possible)

Assumptions

The population is in Hardy-Weinberg and Linkage equilibrium (because we are doing the calculation with $\theta = 0$).

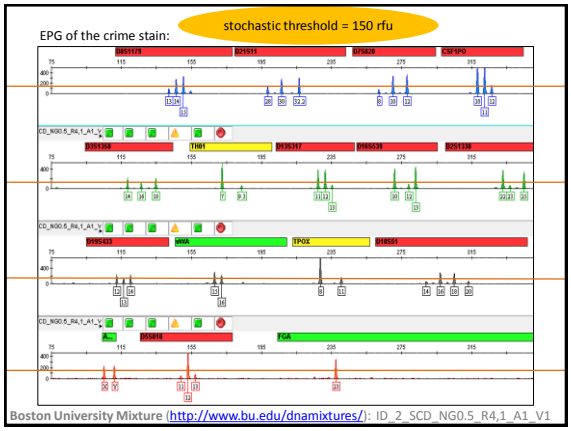
The number of contributors is ___.

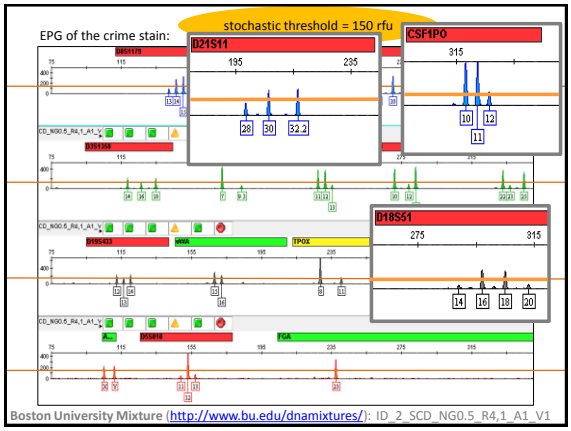
Only the genotypes satisfying the mixture deconvolution rules are possible.

modified Random Match Probability (mRMP)

Two Steps:

1. Make a list of the possible genotypes of each contributor to the mixture.
2. For each contributor, assign the probability of randomly selecting someone in the population of potential donors who has one of the listed genotypes.

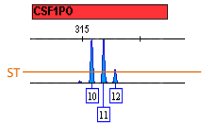




modified Random Match Probability (mRMP)

CSF1PO

$p_{10} = 0.220$
$p_{11} = 0.309$
$p_{12} = 0.360$



*peak at 12 is **above** the stochastic threshold*

1. major contributor: 10,11

2. $\text{mRMP} = 2p_{10}p_{11}$
$= 2(0.22)(0.309)$
$= 0.136$

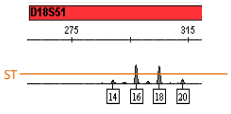
1. minor contributor: 12,12
10,12
11,12

2. $\text{mRMP} = p_{12}^2 + 2p_{10}p_{12} + 2p_{11}p_{12}$
$= p_{12}(p_{12} + 2p_{10} + 2p_{11})$
$= 0.36(0.36 + 2(0.22) + 2(0.309))$
$= 0.510$

modified Random Match Probability (mRMP)

D18S51

$p_{14} = 0.134$
$p_{16} = 0.147$
$p_{18} = 0.078$
$p_{20} = 0.018$



1. major contributor: 16,18

2. $\text{mRMP} = 2p_{16}p_{18}$
$= 2(0.147)(0.078)$
$= 0.023$

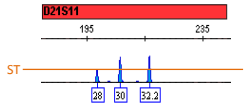
1. minor contributor: 14,20

2. $\text{mRMP} = 2p_{14}p_{20}$
$= 2(0.134)(0.018)$
$= 0.005$

modified Random Match Probability (mRMP)

D21S11

$p_{28} = 0.159$
$p_{30} = 0.283$
$p_{32.2} = 0.090$



*peak at 28 is **below** the stochastic threshold*

1. major contributor: 30,32.2

2. $\text{mRMP} = 2p_{30}p_{32.2}$
$= 2(0.283)(0.09)$
$= 0.051$

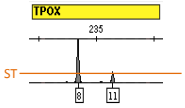
1. minor contributor: 28,F

2. $\text{mRMP} = 2p_{28}(1 - p_{28}) + p_{28}^2$
$= 2p_{28} - 2p_{28}^2 + p_{28}^2$
$= 2p_{28} - p_{28}^2$
$= 2(0.159) - 0.159^2$
$= 0.293$

modified Random Match Probability (mRMP)

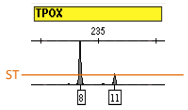
TPOX

$p_8 = 0.525$
$p_{11} = 0.252$



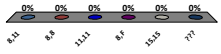
*peak at 11 is **above** the stochastic threshold*

Genotype of major contributor?

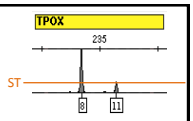


A. 8,11
B. 8,8
C. 11,11
D. 8,F
E. 15,15
F. ???

Response Counter



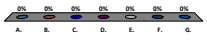
mRMP for major contributor?



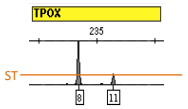
$p_8 = 0.525$
$p_{11} = 0.252$

A. $(0.252)^2 = 0.064$
B. $2(0.525)(0.252) = 0.265$
C. $(0.525)^2 = 0.276$
D. 0.525
E. 1
F. ???
G. I forgot to bring my math skills to this workshop

Response Counter



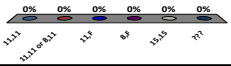
Genotype of minor contributor?



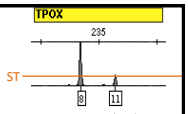
*The peak at 11 is **above** the stochastic threshold.*

A. 11,11
B. 11,11 or 8,11
C. 11,F
D. 8,F
E. 15,15
F. ???

Response Counter



mRMP for minor contributor?



$p_8 = 0.525$
$p_{11} = 0.252$

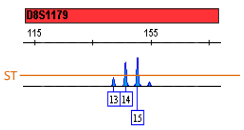
A. $(0.252)^2 = 0.064$
B. $(0.252)^2 + 2(0.525)(0.252) = 0.328$
C. $2(0.252) - (0.252)^2 = 0.440$
D. $2(0.525) - (0.525)^2 = 0.774$
E. 1
F. ???
G. All answers look the same to me
(I drank too much vodka last night)

Response Counter

modified Random Match Probability (mRMP)

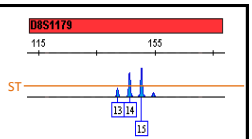
D8S1179

$p_{13} = 0.330$
$p_{14} = 0.166$
$p_{15} = 0.104$



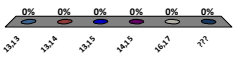
*peak at 13 is **below** the stochastic threshold*

Genotype of major contributor?

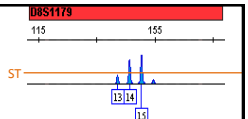


A. 13,13
B. 13,14
C. 13,15
D. 14,15
E. 16,17
F. ???

Response Counter



mRMP for major contributor?

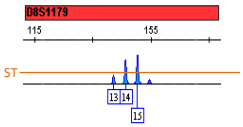


$p_{13} = 0.330$
$p_{14} = 0.166$
$p_{15} = 0.104$

A. $(0.166)^2 = 0.028$
B. $2(0.166)(0.104) = 0.035$
C. $2(0.330)(0.104) = 0.069$
D. $2 - 0.166 - 0.104 = 1.73$
E. $2 + 0.166 + 0.104 = 2.27$
F. ???
G. All answers look the same to me
(I drank vodka for breakfast)

Response Counter

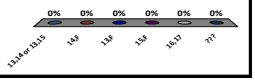
Genotype of minor contributor



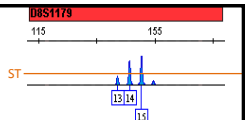
*The peak at 13 is **below** the stochastic threshold.*

A. 13,14 or 13,15
B. 14,F
C. 13,F
D. 15,F
E. 16,17
F. ???

Response Counter



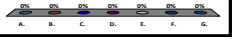
mRMP for minor contributor?



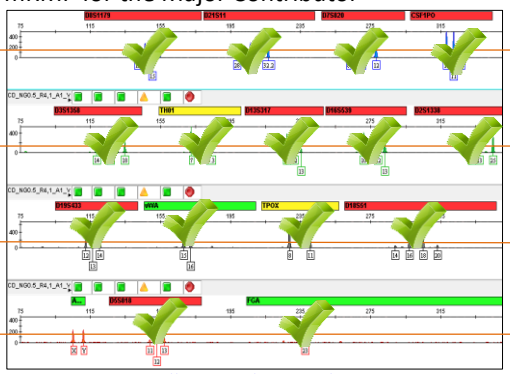
$p_{13} = 0.330$
$p_{14} = 0.166$
$p_{15} = 0.104$

A. $2(0.330)(0.166 + 0.104) = 0.178$
B. $2(0.166) - (0.166)^2 = 0.304$
C. 0.330
D. $2(0.330) - (0.330)^2 = 0.551$
E. 1
F. ???
G. I used up all my math skills on the previous questions

Response Counter



mRMP for the Major Contributor



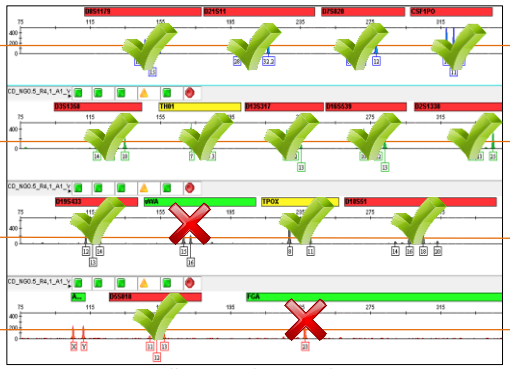
Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

mRMP for the Major Contributor

Locus	mRMP (minor)
D8S1179	0.035
D21S11	0.046
D7S820	0.081
CSF1PO	0.136
D3S1358	0.032
TH01	0.038
D13S317	0.175
D16S539	0.019
D2S1338	0.007
D19S433	0.051
vWA	0.042
TPOX	0.276
D18S51	0.023
D5S818	0.151
FGA	0.023

All Loci: 2.6×10^{-20}

mRMP for the Minor Contributor



Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

mRMP for the Minor Contributor

Locus	mRMP (minor)
D8S1179	0.551
D21S11	0.293
D7S820	0.267
CSF1PO	0.510
D3S1358	0.419
TH01	0.571
D13S317	0.219
D16S539	0.529
D2S1338	0.199
D19S433	0.445
vWA	1
TPOX	0.328
D18S51	0.005
D5S818	0.266
FGA	1

All Loci: 2.3×10^{-8}

Likelihood Ratio (LR)

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 12 and Appendix 4: pages 315-317, 322-325 and 558-565.

Likelihood Ratio (LR)

person of interest

answers the question:

What do the DNA typing results mean with regard to the **person of interest** being a contributor?

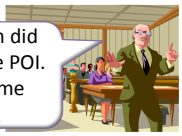
What is the value of the DNA typing results with regard to the **person of interest** being a contributor?

By how much do the DNA typing results support the **person of interest** being a contributor?

There are two sides to every story…

H_p: The crime stain came from the person of interest (POI).

H_d: The crime stain did not come from the POI. It came from some other person.



prosecution's proposition

defense's proposition

The likelihood ratio gives the value of the findings with regard to the prosecution's and defense's standpoints in the case.

Information it takes into account

- presence of genotypes
- peaks below stochastic threshold (where allele drop-out is possible)

and

- the standpoints of the prosecution and the defense (i.e., the two competing propositions)

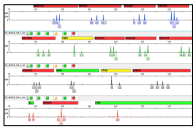
Assumptions

DEPEND ON THE MODEL USED

Here, we will use the classical binary model and make the same assumptions as we did for mRMP:

- The population is in Hardy-Weinberg and Linkage equilibrium.
- The number of contributors is ___.
- Only the genotypes satisfying the mixture deconvolution rules are possible.

Standpoints of the prosecution and the defense

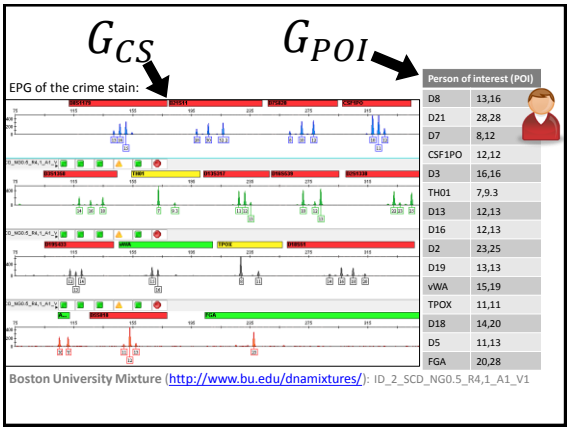


EPG of the crime stain

person of interest (POI)

H_p: The DNA came from the POI and an unknown contributor.

H_d: The DNA came from two unknown contributors.



Likelihood Ratio (LR)

The probability of observing the DNA typing results of the crime stain given the POI's genotype and that the DNA came from the POI and one unknown contributor

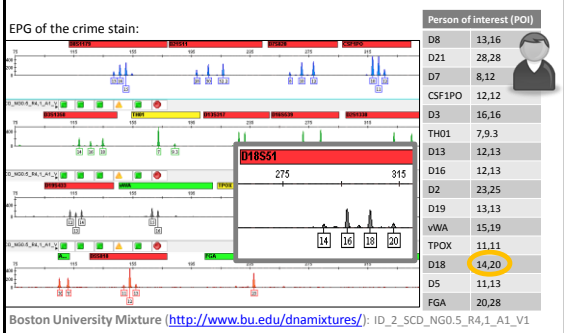
$$LR = \frac{\Pr(G_{CS}|G_{POI}, H_p)}{\Pr(G_{CS}|G_{POI}, H_d)}$$

numerator

divided by

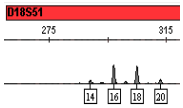
denominator

the probability of observing the DNA typing results of the criem stain given the POI's genotype and that the DNA came from two unknown contributors.



Likelihood Ratio (LR)

D18S51
$p_{14} = 0.134$
$p_{16} = 0.147$
$p_{18} = 0.078$
$p_{20} = 0.018$



$G_{POI} = \{14,20\}$

Numerator:

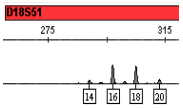
What is the probability of obtaining these DNA typing results for the crime stain if the POI **is** a contributor and the POI has genotype {14,20}?

Major	Minor
16,18	14,20

$$Pr(16,18) \times Pr(14,20)$$
$$= 2p_{16}p_{18} \times 1$$
$$= 2p_{16}p_{18}$$

Likelihood Ratio (LR)

D18S51
$p_{14} = 0.134$
$p_{16} = 0.147$
$p_{18} = 0.078$
$p_{20} = 0.018$



$G_{POI} = \{14,20\}$

Denominator:

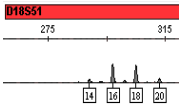
What is the probability of obtaining these DNA typing results for the crime stain if the POI **is not** a contributor?

Major	Minor
16,18	14,20

$$Pr(16,18) \times Pr(14,20)$$
$$= 2p_{16}p_{18} \times 2p_{14}p_{20}$$
$$= 4p_{14}p_{16}p_{18}p_{20}$$

Likelihood Ratio (LR)

D18S51
$p_{14} = 0.134$
$p_{16} = 0.147$
$p_{18} = 0.078$
$p_{20} = 0.018$



$G_{POI} = \{14,20\}$

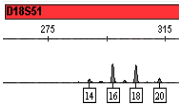
$$LR = \frac{2p_{16}p_{18}}{4p_{14}p_{16}p_{18}p_{20}}$$

$$= \frac{1}{2p_{14}p_{20}}$$

$$= 207.30$$

Likelihood Ratio (LR)

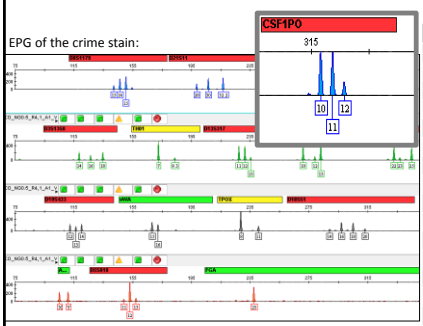
D18S51
$p_{14} = 0.134$
$p_{16} = 0.147$
$p_{18} = 0.078$
$p_{20} = 0.018$



$G_{POI} = \{14,20\}$

The DNA typing results are 207 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.

EPG of the crime stain:

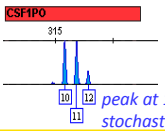


Person of interest (POI)	
D8	13,16
D21	28,28
D7	8,12
CSF1PO	12,12
D3	16,16
TH01	7,9.3
D13	12,13
D16	12,13
D2	23,25
D19	13,13
vWA	15,19
TPOX	11,11
D18	14,20
D5	11,13
FGA	20,28

Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

Likelihood Ratio (LR)

CSF1PO
$p_{10} = 0.220$
$p_{11} = 0.309$
$p_{12} = 0.360$



$G_{POI} = \{12,12\}$

*peak at 12 is **above** the stochastic threshold*

Numerator:

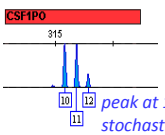
What is the probability of obtaining these DNA typing results for the crime stain if the POI **is** a contributor and the POI has genotype {12,12}?

Major	Minor
10,11	12,12
10,11	10,12
10,11	11,12

$$Pr(10,11) \times Pr(12,12)$$
$$= 2p_{10}p_{11} \times 1$$
$$= 2p_{10}p_{11}$$

Likelihood Ratio (LR)

CSF1PO
$p_{10} = 0.220$
$p_{11} = 0.309$
$p_{12} = 0.360$



$G_{POI} = \{12,12\}$

*peak at 12 is **above** the stochastic threshold*

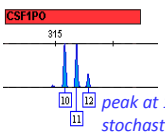
Denominator:

What is the probability of obtaining these DNA typing results for the crime stain if the POI **is not** a contributor?

Major	Minor
10,11	12,12
10,11	10,12
10,11	11,12

Likelihood Ratio (LR)

CSF1PO
$p_{10} = 0.220$
$p_{11} = 0.309$
$p_{12} = 0.360$



$G_{POI} = \{12,12\}$

*peak at 12 is **above** the stochastic threshold*

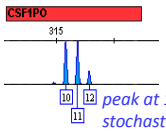
Denominator:

$$Pr(10,11) \times Pr(12,12) + Pr(10,11) \times Pr(10,12) + Pr(10,11) \times Pr(11,12)$$
$$= 2p_{10}p_{11} \times p_{12}^2 + 2p_{10}p_{11} \times 2p_{10}p_{12}$$
$$+2p_{10}p_{11} \times 2p_{11}p_{12}$$
$$= 2p_{10}p_{11}p_{12}(p_{12} + 2p_{10} + 2p_{11})$$

Major	Minor
10,11	12,12
10,11	10,12
10,11	11,12

Likelihood Ratio (LR)

CSF1PO
$p_{10} = 0.220$
$p_{11} = 0.309$
$p_{12} = 0.360$



$G_{POI} = \{12,12\}$

*peak at 12 is **above** the stochastic threshold*

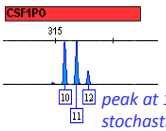
$$LR = \frac{2p_{10}p_{11}}{2p_{10}p_{11}p_{12}(p_{12} + 2p_{10} + 2p_{11})}$$

$$= \frac{1}{p_{12}(p_{12} + 2p_{10} + 2p_{11})}$$

$$= 1.96$$

Likelihood Ratio (LR)

CSF1PO
$p_{10} = 0.220$
$p_{11} = 0.309$
$p_{12} = 0.360$

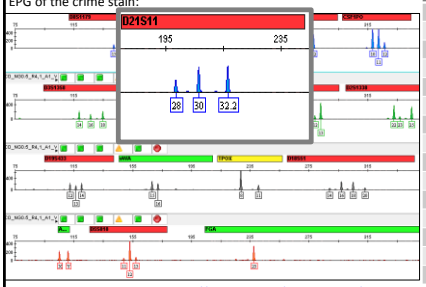


$G_{POI} = \{12,12\}$

*peak at 12 is **above** the stochastic threshold*

The DNA typing results are about 2 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.

EPG of the crime stain:

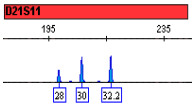


Person of interest (POI)	
D8	13,16
D21	28,28
D7	8,12
CSF1PO	12,12
D3	16,16
TH01	7,9.3
D13	12,13
D16	12,13
D2	23,25
D19	13,13
vWA	15,19
TPOX	11,11
D18	14,20
D5	11,13
FGA	20,28

Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

Likelihood Ratio (LR)

D21S11
$p_{28} = 0.159$
$p_{30} = 0.283$
$p_{32.2} = 0.090$



$G_{POI} = \{28,28\}$

*peak at 28 is **below** the stochastic threshold*

Numerator:

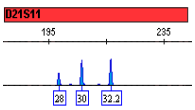
What is the probability of obtaining these DNA typing results for the crime stain if the POI **is** a contributor and the POI has genotype {28,28}?

Major	Minor
30,32.2	28,F

$$Pr(30,32.2) \times Pr(28,F)$$
$$= 2p_{30}p_{32.2} \times 1$$
$$= 2p_{30}p_{32.2}$$

Likelihood Ratio (LR)

D21S11
$p_{28} = 0.159$
$p_{30} = 0.283$
$p_{32.2} = 0.090$



$G_{POI} = \{28,28\}$

*peak at 28 is **below** the stochastic threshold*

Denominator:

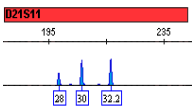
What is the probability of obtaining these DNA typing results for the crime stain if the POI **is not** a contributor?

Major	Minor
30,32.2	28,F

$$Pr(30,32.2) \times Pr(28,F)$$
$$= 2p_{30}p_{32.2} \times [2p_{28}(1 - p_{28}) + p_{28}^2]$$
$$= 2p_{30}p_{32.2}(2p_{28} - p_{28}^2)$$

Likelihood Ratio (LR)

D21S11
$p_{28} = 0.159$
$p_{30} = 0.283$
$p_{32.2} = 0.090$



$G_{POI} = \{28,28\}$

*peak at 28 is **below** the stochastic threshold*

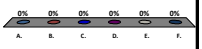
$$LR = \frac{2p_{30}p_{32.2}}{2p_{30}p_{32.2}(2p_{28} - p_{28}^2)}$$

$$= \frac{1}{(2p_{28} - p_{28}^2)}$$

$$= 3.42$$

What does a $LR \approx 3$ mean?

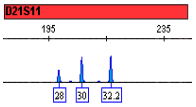
A. The person of interest committed the crime.
B. A total of 3 peaks were observed at this locus.
C. It is about 3 times more probable that the DNA came from the person of interest and an unknown contributor than that the DNA came from two unknown contributors.
D. There are 3 contributors to this DNA mixture.
E. The DNA typing results are about 3 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.
F. ???



Response Counter

Likelihood Ratio (LR)

D21S11
$p_{28} = 0.159$
$p_{30} = 0.283$
$p_{32.2} = 0.090$



*peak at 28 is **below** the stochastic threshold*

$G_{POI} = \{28,28\}$

Transposed Conditional

$$\underbrace{\frac{\Pr(H_p|E)}{\Pr(H_d|E)}}_{100} = \underbrace{\frac{\Pr(E|H_p)}{\Pr(E|H_d)}}_{100} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

posterior odds | likelihood ratio | prior odds

These DNA typing results indicate that the probability of the prosecution's proposition being true is 100 times greater than the probability of the defense's proposition being true.

The probability of these DNA typing results is 100 times greater if the prosecution's proposition is true than if the defense's proposition is true.

Transposed Conditional

H: the animal is an elephant
E: the animal has four legs

given: 4 legs

$\Pr(H|E)$
$\frac{1}{5,000}$

$\neq$

1 leg
2 legs
3 legs
4 legs

given:

$\Pr(E|H)$
~ 1

Transposed Conditional

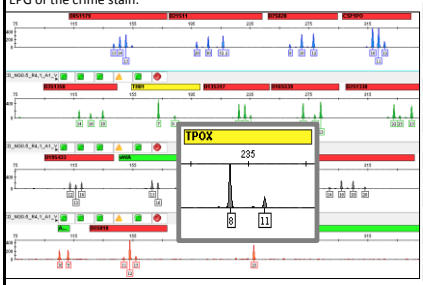
$$\underbrace{\frac{\Pr(H_p|E)}{\Pr(H_d|E)}}_{\frac{1}{1,000}} = \underbrace{\frac{\Pr(E|H_p)}{\Pr(E|H_d)}}_{100} \times \underbrace{\frac{\Pr(H_p)}{\Pr(H_d)}}_{\frac{1}{100,000}}$$

posterior odds | likelihood ratio | prior odds

These DNA typing results indicate that the probability of the **defense's proposition** being true is **1,000 times greater** than the probability of the **prosecution's proposition** being true.

The probability of these DNA typing results is 100 times greater if the prosecution's proposition is true than if the defense's proposition is true.

EPG of the crime stain:

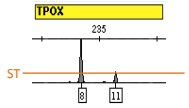


Person of interest (POI)	
D8	13,16
D21	28,28
D7	8,12
CSF1PO	12,12
D3	16,16
TH01	7,9.3
D13	12,13
D16	12,13
D2	23,25
D19	13,13
vWA	15,19
TPOX	11,11
D18	14,20
D5	11,13
FGA	20,28

Boston University Mixture (http://www.bu.edu/dnamixtures/): ID_2_SCD_NG0.5_R4,1_A1_V1

Likelihood Ratio (LR)

TPOX
$p_8 = 0.525$
$p_{11} = 0.252$



$G_{POI} = \{11,11\}$

*The peak at 11 is **above** the stochastic threshold.*

Numerator:

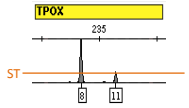
What is the probability of obtaining these DNA typing results for the crime stain if the POI **is** a contributor and the POI has genotype {11,11}?

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

$$Pr(8,8) \times Pr(11,11)$$
$$= \cdots$$

Likelihood Ratio (LR)

TPOX
$p_8 = 0.525$
$p_{11} = 0.252$



$G_{POI} = \{11,11\}$

*The peak at 11 is **above** the stochastic threshold.*

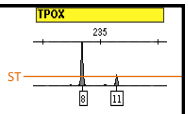
Denominator:

What is the probability of obtaining these DNA typing results for the crime stain if the POI **is not** a contributor?

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

$$Pr(8,8) \times Pr(11,11) + Pr(8,8) \times Pr(8,11)$$
$$= \cdots$$

What is the likelihood ratio?



*The peak at 11 is **above** the stochastic threshold.*

A. $\frac{p_8^2}{p_8^2(p_{11}^2+2p_8p_{11})} = \frac{1}{p_{11}(p_{11}+2p_8)}$

B. $\frac{1}{p_{11}+2p_8}$

C. 1

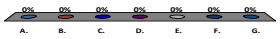
D. $\frac{1}{2p_8p_{11}}$

E. $\frac{1}{p_{11}^2}$

F. infinity

G. ???

Response Counter



Likelihood Ratio (LR)

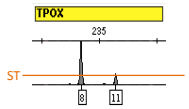
TPOX
$p_8 = 0.525$
$p_{11} = 0.252$

$G_{POI} = \{11,11\}$

$LR = 3.05$

Likelihood Ratio (LR)

TPOX
$p_8 = 0.525$
$p_{11} = 0.252$



$G_{POI} = \{11,11\}$

The DNA typing results are about 3 times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.

Likelihood Ratio (LR) for all loci

H_p: The DNA came from the POI and an unknown contributor.

H_d: The DNA came from two unknown contributors.

If H_p is true, is the POI the major contributor or the minor contributor?

If H_p is true, the POI could be either the major contributor or the minor contributor. Let us consider these possibilities to be equally probable. So if H_p is true, there is a probability of $\frac{1}{2}$ that the POI is the major contributor and a probability of $\frac{1}{2}$ that the POI is the minor contributor.

We can only observe these DNA typing results if the POI is the minor contributor.

	Major	Minor	
D18S51:	16,18	14,20	$G_{POI} = \{14,20\}$
CSF1PO:	10,11	12,12	$G_{POI} = \{12,12\}$
	10,11	10,12	
	10,11	11,12	
D21S11:	30,32.2	28,F	$G_{POI} = \{28,28\}$
TPOX:	8,8	11,11	$G_{POI} = \{11,11\}$
	8,8	8,11	

Likelihood Ratio (LR) for all loci

H_p: The DNA came from the POI and an unknown contributor.

H_d: The DNA came from two unknown contributors.

Numerator:

Because these DNA typing results are only possible when the POI is the minor contributor, and the POI is the minor contributor with a probability of $\frac{1}{2}$, we multiply the numerator of the likelihood ratio for the entire profile by $\frac{1}{2}$.

Locus	Likelihood Ratio
D8S1179	3.66
D21S11	3.42
D7S820	3.74
CSF1PO	1.96
D3S1358	2.39
TH01	1.75
D13S317	4.58
D16S539	1.89
D2S1338	5.03
D19S433	1.29
vWA	1
TPOX	3.05
D18S51	207.30
D5S818	3.77
FGA	1

Likelihood Ratio (LR)

All Loci: $\boldsymbol{LR = 2.5 \times 10^7}$

True or false?

A likelihood ratio of $\mathbf{2.5 \times 10^7}$ means that it is $\mathbf{2.5 \times 10^7}$ times more probable that the DNA came from the person of interest and an unknown contributor than that the DNA came from two unknown contributors.

A. True
B. False

Response Counter

Likelihood Ratio (LR)

$\boldsymbol{LR = 2.5 \times 10^7 = 25\ million}$

The DNA typing results are about 25 million times more probable if the DNA came from the person of interest and an unknown contributor than if the DNA came from two unknown contributors.

Formulating Propositions for Likelihood Ratios

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 12: pages 323-324.

S. Gittelson, T. Kalafut, S. Myers, D. Taylor, T. Hicks, F. Taroni, I.W. Evett, J.-A. Bright, J. Buckleton. (2015). A practical guide for the formulation of propositions in the Bayesian approach to DNA evidence interpretation in an adversarial environment. *Journal of Forensic Sciences,* doi: 10.1111/1556-4029.12907

Formulating Propositions for Likelihood Ratios

If the propositions change, the likelihood ratio changes.

The propositions depend on the case circumstances and the standpoints of the prosecution and the defense.

Consider the following 4 cases for a 2-person mixture.

2-person mixture

Case 1: Alleged Rape Case

The crime sample is a vaginal swab taken from the complainant V.

Standpoints of the prosecution and the defense:

prosecution: "POI raped V."

defense: "POI did not rape V. Someone else raped V."

case circumstances: V had no consensual partner at the time of this event.

What is H_p? What is H_d?

2-person mixture

Case 1: Alleged Rape Case

H_p: The DNA came from the complainant V and the POI.

H_d: The DNA came from the complainant V and an unknown contributor.

2-person mixture

Case 2: Stabbing Case

A person V is found stabbed to death. The crime sample is taken from the POI's shirt sleeve shortly after the discovery of V.

Standpoints of the prosecution and the defense:

prosecution: "POI stabbed V."

defense: "POI did not stab V. POI has never seen V before. Someone else stabbed V."

What is H_p? What is H_d?

2-person mixture

Case 2: Stabbing Case

H_p: The DNA came from the victim V and the POI.

H_d: The DNA came from the POI and an unknown contributor.

2-person mixture

Case 3: Assault Case

A person V is found unconscious in an alleyway. There are indications that V was hit on the head with a hard object. The crime sample is taken from a metal bar found on the ground nearby.

Standpoints of the prosecution and the defense:

prosecution: "POI hit V with the metal bar."

defense: "POI did not hit V. Someone else hit V."

case circumstances: The metal bar is associated with neither V nor POI.

What is H_p? What is H_d?

2-person mixture

Case 3: Assault Case

H_p: The DNA came from the victim V and the POI.

H_d: The DNA came from two unknown contributors.

2-person mixture

Case 4: Shooting

The crime sample is taken from the trigger of a handgun found on the crime scene.

Standpoints of the prosecution and the defense:

prosecution: "POI shot this gun."

defense: "Someone else shot this gun. POI never touched this gun."

What is H_p? What is H_d?

2-person mixture

Case 4: Assault Case

H_p: The DNA came from POI and an unknown contributor.

H_d: The DNA came from two unknown contributors.

2-person mixture

If **V** has the same profile as the **major contributor** to the mixture, and the **POI**'s profile is a possible profile of the **minor contributor**, then we obtain the following values for the different pairs of propositions:

	H_p	H_d	LR
Case 1	V and POI	V and unknown	2.2×10^7
Case 2	V and POI	POI and unknown	1.9×10^{19}
Case 3	V and POI	2 unknowns	9.8×10^{26}
Case 4	POI and unknown	2 unknowns	2.5×10^7

Note that for a same H_p (i.e., Cases 1,2 and 3) the likelihood ratio is larger when H_d postulates 2 unknown contributors (i.e., Case 3) than when H_d postulates 1 known and 1 unknown contributor (i.e., Cases 1 and 2).

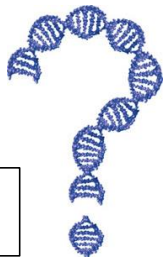
Summary

		Takes into account:		Models:	
		presence/ absence of alleles	possible genotypes based on peak heights	allele drop-out and allele drop-in	peak heights
Binary	CPI	X			
	mRMP	X	X		
	LR (binary)	X	X		
Probabilistic genotyping	LR (semi-continuous)	X		X	
	LR (fully continuous)	X	X	X	X

Acknowledgements

John Butler

Discussions on DNA mixture interpretation
John Buckleton



Contact Info.:
simone.gittelson@nist.gov
+1-301-975-4892

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

Additional Training Resources

Boston University DNA Mixture Training: http://www.bu.edu/dnamixtures/

NIST DNA Analyst Training on Mixture Interpretation: http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation.cfm

NIST 2013 webcast: http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation-webcast.cfm

NIST DNA Analyst Webinar Series: Probabilistic Genotyping and Software Programs (Part 1): http://www.nist.gov/forensics/nist-dna-analyst-webinar-series-pt1.cfm

NIST DNA Analyst Webinar Series: Probabilistic Genotyping and Software Programs (Part 2): http://www.nist.gov/forensics/nist-dna-analyst-webinar-series-part-2.cfm

NIST DNA Analyst Webinar Series: Validation Concepts and Resources – Part 1: http://www.nist.gov/forensics/nist-dna-analyst-webinar-series-validation-concepts-and-resources-part-one-webinar-archive.cfm

NIST STRBase Mixture Information: http://www.cstl.nist.gov/strbase/mixture.htm

Simone Gittelson workshop on Mixture Interpretation & Statistics at the Bode East 14th Annual DNA Technical Conference (Orlando, FL), May 29, 2015

Guidance for DNA Interpretation

Butler, J.M. (2013). Forensic DNA advisory groups: DAB, SWGDAM, ENFSI, and BSAG. *Encyclopedia of Forensic Sciences, 2nd Edition.* Elsevier Academic Press: New York.

DNA Commission of the ISFG: http://www.isfg.org/Publications/DNA+Commission

European Network of Forensic Science Institutes (ENFSI) DNA Working Group: http://www.enfsi.eu/about-enfsi/structure/working-groups/dna?uid=98

ENFSI Guide for Evaluative Reporting in Forensic Science. Available at http://www.enfsi.eu/sites/default/files/afbeeldingen/enfsi_booklet_m1.pdf

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.

Gill, P., Guiness, J., Iveson, S. (2012). The interpretation of DNA evidence (including low-template DNA). Available at http://www.homeoffice.gov.uk/publications/agencies-public-bodies/fsr/interpretation-of-dna-evidence

Gill, P., et al. (2012). DNA commission of the International Society of Forensic Genetics: recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forensic Science International: Genetics, 6,* 679-688.

Hobson, D., et al. (1999). STR analysis by capillary electrophoresis: development of interpretation guidelines for the Profiler Plus and COfiler systems for use in forensic science. *Proceedings of the 10th International Symposium on Human Identification.* Available at http://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/10th-ishi-oral-presentations/.

Puch-Solis, R., Roberts, P., Pope, S., Aitken, C. (2012). Assessing the probative value of DNA evidence: *Guidance for judges, lawyers, forensic scientists and expert witnesses*. Available at http://www.maths.ed.ac.uk/~cgga/Guide-2-WEB.pdf.

QAS (2011). Quality Assurance Standards for Forensic DNA Testing Laboratories effective 9-1-2011. See http://www.fbi.gov/about-us/lab/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011.

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).

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

SWGDAM (2010). SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Available at http://www.swgdam.org/#!publications/c1mix.

SWGDAM (2012). Validation Guidelines for DNA Analysis Methods. Available at http://www.swgdam.org/#!publications/c1mix.

SWGDAM (2013). SWGDAM Training Guidelines. Available at http://www.swgdam.org/#!publications/c1mix.

SWGDAM (2014). SWGDAM Interpretation Guidelines for Y-Chromosome STR Testing. Available at http://www.swgdam.org/#!publications/c1mix.

SWGDAM (2014). SWGDAM Guidelines for STR Enhanced Detection Methods. Available at http://www.swgdam.org/#!publications/c1mix.

SWGDAM (2015). SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. Available at http://www.swgdam.org/#!publications/c1mix.

Software for DNA Analysis and Interpretation

Armed Xpert (NicheVision): http://www.armedxpert.com/

BatchExtract: ftp://ftp.ncbi.nih.gov/pub/forensics/BATCHEXTRACT

DNAMIX (Bruce Weir): http://www.biostat.washington.edu/~bsweir/DNAMIX3/webpage/

DNA Mixture Separator (Torben Tvedebrink): http://people.math.aau.dk/~tvede/mixsep/

EPG Maker program (Steven Myers): http://www.cstl.nist.gov/strbase/tools/EPG-Maker(SPMv.3,Dec2-2011).xlt (13 Mb Excel file)

Forensic DNA Statistics (Peter Gill): https://sites.google.com/site/forensicdnastatistics/

Forensim (Hinda Haned): http://forensim.r-forge.r-project.org/

GeneMapper*ID-X* (from Applied Biosystems): http://www.lifetechnologies.com/us/en/home/technical-resources/software-downloads/genemapper-id-x-software.html

GeneMarker HID (from Soft Genetics): http://www.softgenetics.com/GeneMarkerHID.html

Genetic Analysis Data File Format, Sept 2009. Available at http://www.appliedbiosystems.com/absite/us/en/home/support/software-community/tools-for-accessing-files.html

GenoProof Mixture (Qualitype): http://www.qualitype.de/en/qualitype/genoproof-mixture

ISFG Software Resources Page: http://www.isfg.org/software

Lab Retriever (Scientific Collaboration, Innovation & Education Group): http://www.scieg.org/lab_retriever.html

likeLTD (David Balding): https://sites.google.com/site/baldingstatisticalgenetics/software/likeltd-r-forensic-dna-r-code

LRmix (Hinda Haned): https://sites.google.com/site/forensicdnastatistics/PCR-simulation/lrmix

LRmix Studio (Hinda Haned): http://lrmixstudio.org/

OSIRIS (Open Source Independent Review and Interpretation System): http://www.ncbi.nlm.nih.gov/projects/SNP/osiris/

STRmix (Ducan Taylor, Jo-Anne Bright, John Buckleton): http://strmix.com/

TrueAllele Casework (Cybergenetics): http://www.cybgen.com/systems/casework.shtml

STR Kits, Loci, and Population Data

Butler, J.M. & Hill, C.R. (2013). Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. Chapter 9 in Shewale, J. (ed.), *Forensic DNA Analysis: Current Practices and Emerging Technologies*. Taylor & Francis/CRC Press: Boca Raton. pp. 181-198.

FBI (2012). Planned process and timeline for implementation of additional CODIS core loci. Available at http://www.fbi.gov/about-us/lab/codis/planned-process-and-timeline-for-implementation-of-additional-codis-core-loci

Gettings, K.B., Aponte, R.A., Vallone, P.M., Butler, J.M. (2015). STR allele sequence variation: current knowledge and future issues. *Forensic Science International: Genetics, (in press).* doi: 10.1016/j.fsigen.2015.06.005

GlobalFiler information: http://www.lifetechnologies.com/us/en/home/industrial/human-identification/globalfiler-str-kit/resources.html

NIST U.S. Population Data: http://www.cstl.nist.gov/strbase/NISTpop.htm

PowerPlex Fusion System. http://www.promega.com/products/pm/genetic-identity/powerplex-fusion

Qiagen Investigator 24plex: https://www.qiagen.com/us/landing-pages/applied-testing/investigator-24plex/

STRBase: http://www.cstl.nist.gov/strbase

Setting Thresholds

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

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.

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

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

Statistical Approaches

Bille, T., Bright, J.-A., Buckleton, J. (2013). Application of random match probability calculations to mixed STR profiles. *Journal of Forensic Sciences, 58(2)*, 474-485.

Buckleton, J., Triggs, C.M., & Walsh, S.J. (2005). *Forensic DNA Evidence Interpretation*. London: CRC Press.

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.

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*.*

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. (1995). Avoiding the transposed conditional. *Science & Justice, 35*, 127-131.

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

Evett, I.W., Gill, P., Jackson, G., Whitaker, J., Champod C. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences, 47(3)*, 520-530.

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.

Pascali, V.L., & Merigioli, S. (2012). Joint Bayesian analysis of forensic mixtures. *Forensic Science International: Genetics, 6,* 735-748.

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*.*

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

Statistical Interpretation for Linked Loci

Bright, J.-A., Curran J.M., Buckleton, J. (2013). Relatedness calculations for linked loci incorporating subpopulation effects. *Forensic Science International: Genetics, 7*, 380-383.

Bright, J.-A., Hopwood, A., Curran J.M., Buckleton, J. (2014). A Guide to Forensic DNA Interpretation and Linkage. Available from Promega's *Profiles in DNA* website: http://www.promega.com/resources/profiles-in-dna/2014/a-guide-to-forensic-dna-interpretation-and-linkage/ (accessed August 4, 2015).

Buckleton, J. & Triggs, C. (2006). The effect of linkage on the calculation of DNA match probabilities for siblings and half siblings. *Forensic Science International, 160*, 193-199.

Gill, P., Phillips, C., McGovern, C., Bright, J.-A., Buckleton, J. (2012). An evaluation of potential allelic association between the STRs vWA and D12S391: Implications in criminal casework and applications to short pedigrees. *Forensic Science International: Genetics, 6*, 477-486.

Kling, D., Egeland, T., Tillmar, A.O. (2012). FamLink – A user friendly software for linkage calculations in family genetics. *Forensic Science International: Genetics, 6*, 616-620.

O'Connor, K.L. & Tillmar A.O. (2012). Effect of linkage between vWA and D12S391 in kinship analysis. *Forensic Science International: Genetics, 6*, 840-844.

Subpopulations and *F*-Statistics

Anderson, A.D., & Weir, B.S. (2006). An assessment of the behavior of the population structure parameter, θ, at the CODIS loci. *Progress in Forensic Genetics, 11, ICS, 1288,* 495-497.

Balding, D.J. & Nichols, R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International, 64*, 125-140.

Balding, D.J., Greenhalgh, M., Nichols, R.A. (1996). Population genetics of STR loci in Caucasians. *Int. J. Legal Med., 108*, 300-305.

Buckleton J., Curran J.M., Walsh, S.J. (2006). How reliable is the sub-population model in DNA testimony? *Forensic Science International, 157*, 144-148.

Curran, J.M., Buckleton, J., Triggs, C.M. (2003). What is the magnitude of the sub-population effect? *Forensic Science International, 135*, 1-8.

Curran, J.M. & Buckleton, J. (2007). The appropriate use of subpopulation corrections for differences in endogamous communities. *Forensic Science International, 168*, 106-111.

Harbison, S.A. & Buckleton, J. (1998). Applications and extensions of subpopulation theory: caseworkers guide. *Science & Justice, 38*, 249-254.

Holsinger, K.E., & Weir, B.S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics, 10*, 639-650.

Tvedebrink, T. (2010). Overdispersion in allelic counts and θ-correction in forensic genetics. *Theoretical Population Biology, 78*, 200-210.

Weir, B.S. & Cockerham, C.C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution, 38(6)*, 1358-1370.

Weir, B.S. (1994). The effects of inbreeding on forensic calculations. *Annual Review of Genetics, 28*, 597-621.

Stutter Products & Peak Height Ratios

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.

Bright, J.-A., Curran, J.M., Buckleton, J.S. (2013). Investigation into the performance of different models for predicting stutter. *Forensic Science International: Genetics, 7(4),* 422-427.

Bright, J.-A., et al. (2013). Developing allelic and stutter peak height models for a continuous method of DNA interpretation. *Forensic Science International: Genetics, 7(2),* 296-304.

Brookes, C., Bright, J.-A., Harbison, S., Buckleton, J. (2012). Characterising stutter in forensic STR multiplexes. *Forensic Science International: Genetics, 6(1),* 58-63.

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.

Kelly, H., et al. (2012). Modelling heterozygote balance in forensic DNA profiles. *Forensic Science International: Genetics, 6,* 729-734.

Kirkham, A., et al. (2013). High-throughput analysis using AmpFlSTR Identifiler with the Applied Biosystems 3500xl Genetic Analyzer. *Forensic Science International: Genetics, 7,* 92-97.

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.

Manabe, S., et al. (2013). Mixture interpretation: experimental and simulated reevaluation of qualitative analysis. *Legal Medicine, 15,* 66-71.

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.

Petricevic, S., et al. (2010). Validation and development of interpretation guidelines for low copy number (LCN) DNA profiling in New Zealand using the AmpFlSTR SGM Plus multiplex. *Forensic Science International: Genetics, 4,* 305-310.

Wallin, J.M., et al. (1998). TWGDAM validation of the AmpFISTR 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.

Estimating the Number of Contributors

Biedermann, A., et al. (2012). Inference about the number of contributors to a DNA mixture: comparative analyses of a Bayesian network approach and the maximum allele count method. *Forensic Science International: Genetics, 6*, 689-696.

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.

Bright, J.-A., Curran, J.M., Buckleton, J. (2014). The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation. *Forensic Science International: Genetics, 12,* 208-214.

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.

Coble, M.D., et al. (2015). Uncertainty in the number of contributors in the proposed new CODIS set. *Forensic Science International: Genetics, (in press).* doi: http://dx.doi.org/10.1016/j.fsigen.2015.07.005

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

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. (2012). Inferring the number of contributors to mixed DNA profiles. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 9(1),* 113-122.

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.

Mixture Ratios & Deconvolution

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.

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.

Tvedebrink, T., et al. (2012). Identifying contributors of DNA mixtures by means of quantitative information of STR typing. *Journal of Computational Biology, 19(7),* 887-902.

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

Stochastic Effects & Allele Dropout

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

Benschop, C.C.G., et al. (2011). Low template STR typing: effect of replicate number and consensus method on genotyping reliability and DNA database search results. *Forensic Science International: Genetics, 5,* 316-328.

Bright, J.-A., et al. (2012). A comparison of stochastic variation in mixed and unmixed casework and synthetic samples. *Forensic Science International: Genetics, 6(2),* 180-184.

Bright, J.-A., et al. (2012). Composite profiles in DNA analysis. *Forensic Science International: Genetics, 6(3),* 317-321.

Buckleton, J., Kelly, H., Bright, J.-A., Taylor, D., Tvedebrink, T., Curran, J.M. (2014), Utilising allelic dropout probabilities estimated by logistic regression in casework. *Forensic Science International: Genetics, 9*, 9-11.

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.

Gill, P., et al. (2008). Interpretation of complex DNA profiles using empirical models and a method to measure their robustness. *Forensic Science International: Genetics, 2,* 91-103.

Gill, P., et al. (2008). Interpretation of complex DNA profiles using Tippett plots. *Forensic Science International: Genetics Supplement Series, 1,* 646-648.

Gill, P., & Haned, H. (2013). A new methodological framework to interpret complex DNA profiles using likelihood ratios. *Forensic Science International: Genetics, 7,* 251-263.

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., et al. (2012). The interpretation of low level DNA mixtures. *Forensic Science International: Genetics, 6(2),* 191-197.

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.

Puch-Solis, R., et al. (2013). Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. *Forensic Science International: Genetics, 7,* 555-563.

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. (2012). Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. *Forensic Science International: Genetics, 6(1),* 97-101.

Tvedebrink, T., et al. (2012). Allelic drop-out probabilities estimated by logistic regression – further considerations and practical implementation. *Forensic Science International: Genetics, 6(2),* 263-267.

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. (2012). Extending PCR conditions to reduce drop-out frequencies in low template STR typing including unequal mixtures. *Forensic Science International: Genetics, 6(1),* 102-107.

Low Template DNA Mixtures

Bekaert, B., et al. (2012). Automating a combined composite-consensus method to generate DNA profiles from low and high template mixture samples. *Forensic Science International: Genetics, 6(5),* 588-593.

Benschop, C.C.G., et al. (2012). Assessment of mock cases involving complex low template DNA mixtures: a descriptive study. *Forensic Science International: Genetics, 6,* 697-707.

Benschop, C.C.G., et al. (2013). Consensus and pool profiles to assist in the analysis and interpretation of complex low template DNA mixtures. *International Journal of Legal Medicine, 127,* 11-23.

Budimlija, Z.M., & Caragine, T.A. (2012). Interpretation guidelines for multilocus STR forensic profiles from low template DNA samples. *DNA Electrophoresis Protocols for Forensic Genetics (Methods in Molecular Biology, volume 830),* pp. 199-211.

Caragine, T., et al. (2009). Validation of testing and interpretation protocols for low template DNA samples using AmpFlSTR Identifiler. *Croatian Medical Journal, 50(3),* 250-267.

Haned, H., et al. (2012). Exploratory data analysis for the interpretation of low template DNA mixtures. *Forensic Science International: Genetics, 6,* 762-774.

Mitchell, A.A., et al. (2011). Likelihood ratio statistics for DNA mixtures allowing for drop-out and drop-in. *Forensic Science International: Genetics Supplement Series, 3,* e240-e241.

Mitchell, A.A., et al. (2012). Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. *Forensic Science International: Genetics, 6,* 749-761.

Pfeifer, C., et al. (2012). Comparison of different interpretation strategies for low template DNA mixtures. *Forensic Science International: Genetics, 6,* 716-722.

Westen, A.A., et al. (2012). Assessment of the stochastic threshold, back- and forward stutter filters and low template techniques for NGM. *Forensic Science International: Genetics, 6,* 708-715.

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.

Probabilistic Genotyping

Bille, T.W., et al. (2014). Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. *Electrophoresis, 35*, 3125-3133.

Bright, J.-A., et al. (2014). Searching mixed DNA profiles directly against profile databases. *Forensic Science International: Genetics, 9,* 102-110.

Bright, J.-A., Evett, I.W., Taylor, D., Curran, J.M., Buckleton, J.S. (2015). A series of recommended tests when validating probabilistic DNA profile interpretation software. *Forensic Science International: Genetics, 14*, 125-131.

Bright, J.-A., et al. (2015). The variability in likelihood ratios due to different mechanisms. *Forensic Science International: Genetics, 14*,187-190.

Cowell, R.G., et al. (2008). Probabilistic modeling for DNA mixture analysis. *Forensic Science International: Genetics Supplement Series, 1,* 640-642.

Cowell, R.G., et al. (2011). 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. (2011). Validating TrueAllele DNA mixture interpretation. *Journal of Forensic Sciences, 56(6)*, 1430-1447.

Perlin, M.W., et al. (2015). TrueAllele genotype identification on DNA mixtures containing up to five unknown contributors. *Journal of Forensic Sciences, 60,* 857-868.

Taylor, D., Bright, J.-A., Buckleton, J. (2013). The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics, 7(5),* 516-528.

Taylor, D. (2014). Using continuous DNA interpretation methods to revisit likelihood ratio behavior. *Forensic Science International: Genetics*, *11,* 144-153.

Taylor, D., Bright, J.-A., Buckleton, J.S. (2014). Considering relatives when assessing the evidential strength of mixed DNA profiles. *Forensic Science International: Genetics, 13*, 259-263.

Taylor, D., Bright, J.-A., Buckleton, J. (2014). Interpreting forensic DNA profiling evidence without specifying the number of contributors. *Forensic Science International: Genetics, 13*, 269-280.

Wilson, M. (2015). Letter to the Editor in response to Todd W. Bille, Steven M. Weitz, Michael D. Coble, John Buckleton, Jo-Anne Bright: 'Comparison of the performance of different models for the interpretation of low level mixed DNA profiles', Electrophoresis 2014, 35, 3125-3133. *Electrophoresis 36,* 489-494.

Basic STR Interpretation Workshop
John M. Butler & Simone N. Gittelson
Krakow, Poland
31 August 2015

U.S. Caucasian Population Data

number of individuals: N = 361
number of alleles: 2N =722

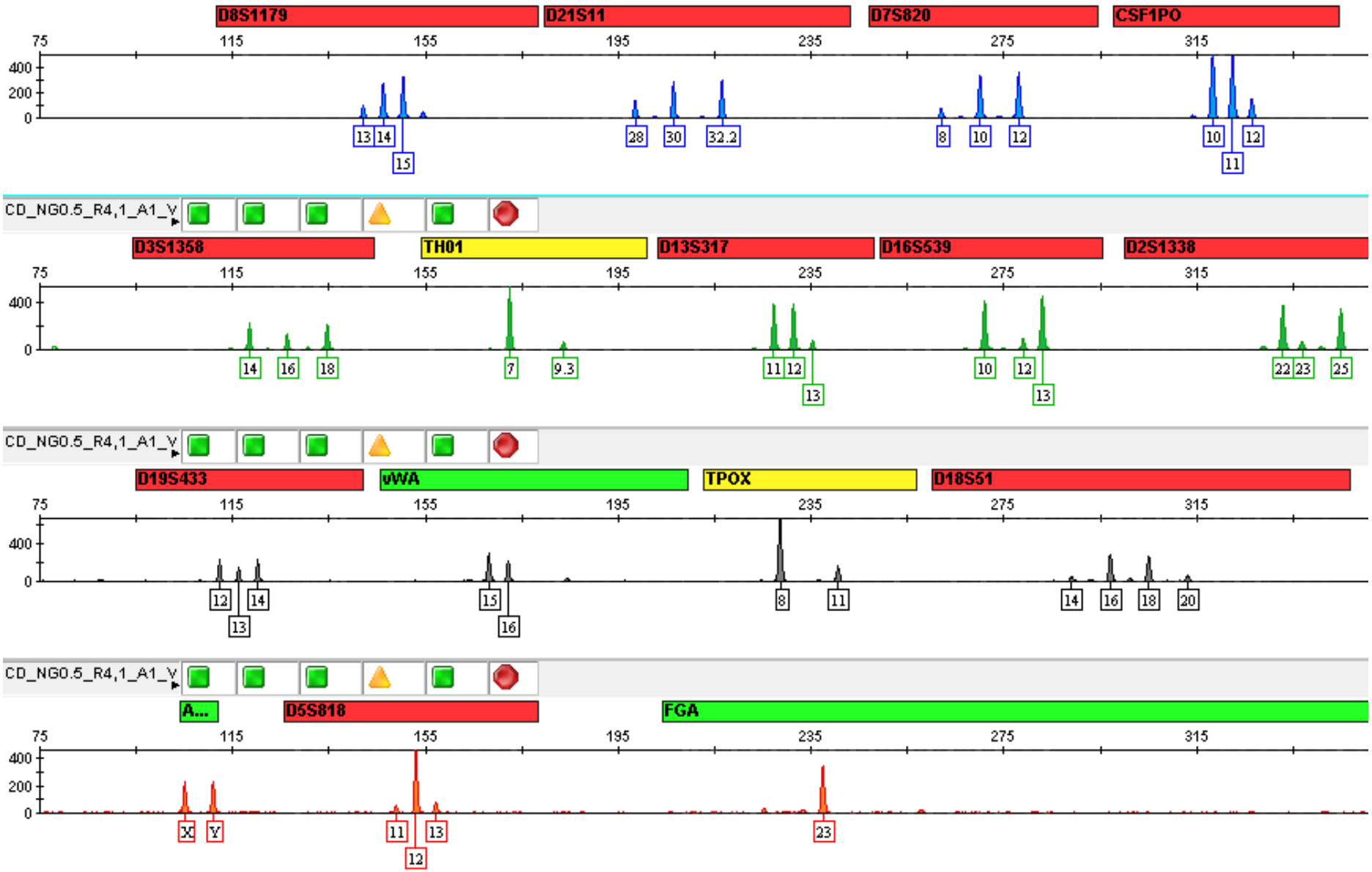
	Locus				
Allele	**CSF1PO**	**TPOX**	**D7S820**	**D8S1179**	**D18S51**
5		0.001			
6		0.001	-		
7	-	-	0.028		
8	0.006	0.525	0.144	0.014	
8.1			0.001		
9	0.014	0.127	0.168	0.006	-
10	0.22	0.05	0.256	0.102	0.008
10.3			-		
11	0.309	0.252	0.205	0.076	0.01
12	0.36	0.042	0.159	0.168	0.114
13	0.082	0.001	0.035	0.33	0.123
13.2					-
14	0.01		0.004	0.166	0.134
14.2					0.001
15	-			0.104	0.17
15.2					-
16				0.033	0.147
16.2					0.001
17				0.001	0.139
18				-	0.078
19					0.04
20					0.018
21					0.01
21.2					-
22					0.007
23					-
24					-
28					-

	Locus
Allele	**D21S11**
24.2	-
25.2	0.001
26	-
26.2	-
27	0.022
28	0.159
28.2	-
29	0.202
29.2	0.003
29.3	-
30	0.283
30.2	0.029
30.3	-
31	0.072
31.2	0.098
32	0.006
32.2	0.09
33	0.001
33.1	-
33.2	0.026
34	-
34.2	0.004
35	0.001
36	0.001
37	-
38	-
39	-

Reference:
Butler J.M. Advanced Topics in Forensic DNA Typing: Interpretation. Academic Press, Elsevier Inc., San Diego, 2015.

Basic STR Interpretation Workshop

John M. Butler & Simone N. Gittelson
Krakow, Poland



Basic STR Interpretation Workshop

John M. Butler & Simone N. Gittelson

Krakow, Poland