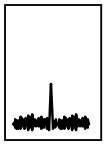
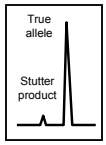
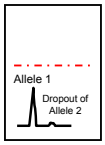
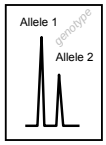
To Review…

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

---

Promega

ISHI 2010 Mixture Interpretation Workshop: Principles, Protocols, and Practice
October 11, 2010 – San Antonio, TX

NIJ National Institute of Justice

Number of Contributors

John M. Butler

BOSTON UNIVERSITY

NIST

Outline for Number of Contributors

- **GUIDELINES**
  - **SWGDAM Guideline 3.4**
  - **Related SWGDAM Guidelines 3.5, 3.5.2.2, 3.6.5, 4.2, 5.1, …**

Many statistical sections!

- **PRINCIPLES**
  - **Approaches to estimating**
  - **Role of assumptions ("conditional" vs. "unconditional")**
  - **Importance to statistical calculations**
- **PROTOCOLS**
  - **Simulation studies from the literature**
  - **Rare situations: tri-alleles and amelogenin deletions**
- **PRACTICE**
  - **How is this topic applied?**

GUIDELINES

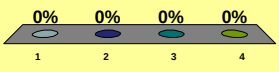
What constitutes a mixture?

SWGDAM Interpretation Guideline 3.4:

A sample is generally considered to have originated from more than one individual **if three or more alleles are present at one or more loci** (excepting tri-allelic loci) **and/or the peak height ratios between a single pair of allelic peaks for one or more loci are below the empirically determined heterozygous peak height ratio expectation**.

Do you currently attempt to determine the number of contributors in a DNA mixture?

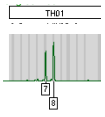
1. Yes
2. No
3. It depends on the case and how complicated the mixture is.
4. We use CPE/CPI statistics and therefore don't need to estimate the number of contributors.



PRINCIPLES

How do you distinguish between a single source sample and a mixture?

- **Don't focus on a single locus** – must **evaluate the entire profile** (or at least multiple loci that are available if a partial profile)!



Look to the most polymorphic loci to see if there are >2 alleles present…

PRACTICE — UPDATED SLIDE

Mixture Case Summaries

Collection organized by Ann Gross (July 2007 – Feb 2008)

	minimum # of contributors						
Crime Class	**1**	**2**	**3**	**4**	**>4**	**N**	
Sexual Assault	884	787	145	11	0	1827	40.2%
Major Crime	1261	519	182	32	0	1994	43.9%
High Volume	344	220	140	11	5	720	15.9%
Total	2489	1526	467	54	5	4541	
	54.8%	33.6%	10.3%	1.2%	0.1%		

This initial data compilation performed by Michelle Burns (NIST 2008 summer intern)

We are in the process of further data analysis and plan to publish this information

GUIDELINES

Minimum number of contributors

SWGDAM Interpretation Guideline 3.4:

Generally, the minimum number of contributors to a mixed sample **can be determined based on the locus that exhibits the greatest number of allelic peaks**. As an example, if at most five alleles are detected per locus, then the DNA typing results are consistent with having arisen from at least three individuals.

PRINCIPLES

STR Loci Ranking by Variability in 1426 U.S. samples

23 loci currently present in commercial STR kits

STR Locus	Alleles Observed	Genotypes Observed	H(obs)	PIC	P_i(total) n = 1426
SE33	**58**	**341**	**0.9383**	**0.9424**	**0.0063**
Penta_E*	20	113	0.8779	0.8992	0.0175
D2S1338	13	73	0.8752	0.8818	0.0221
D1S1656	17	99	0.8871	0.8806	0.0229
D18S51	**23**	**102**	**0.8696**	**0.8694**	**0.0263**
D12S391	24	120	0.8654	0.8646	0.0279
FGA	29	111	0.8702	0.8599	0.0299
Penta_D*	16	70	0.8733	0.8486	0.0360
D21S11	32	98	0.8331	0.8300	0.0399
D19S433	16	83	0.8100	0.7987	0.0534
D8S1179	11	48	0.7966	0.7965	0.0553
vWA	11	42	0.8000	0.7863	0.0624
D16S539	9	30	0.7812	0.7650	0.0723
D13S317	9	30	0.7749	0.7637	0.0724
D7S820	12	35	0.7826	0.7627	0.0745
TH01	9	27	0.7518	0.7578	0.0752
D2S441	14	46	0.7777	0.7490	0.0807
D10S1248	12	41	0.7812	0.7458	0.0828
D3S1358	11	31	0.7489	0.7309	0.0904
D22S1045	11	45	0.7567	0.7305	0.0935
D5S818	9	34	0.7225	0.7033	0.1057
CSF1PO	10	33	0.7567	0.7024	0.1071
TPOX	**10**	**30**	**0.6830**	**0.6549**	**0.1351**

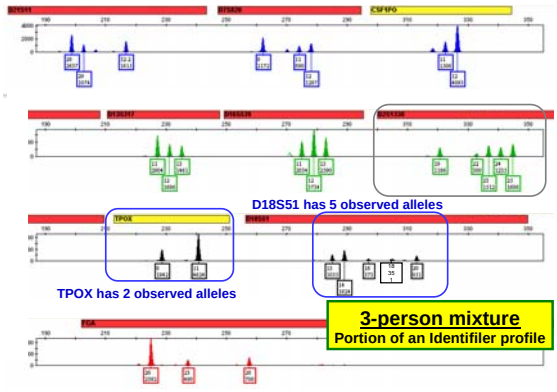
Butler, J.M., et al. (2010) Characterization of Additional STR Loci for Possible U.S. Core Expansion: D12S391, D1S1656, D2S441, D10S1248, D22S1045, and SE33. Poster #40 at ISHI 2010 (San Antonio, TX)

The more polymorphic (variable) the locus, the greater the chance of having non-overlapping alleles between contributors in a mixture leading to a greater ability to accurately determine the number of contributors

D18S51 (with 23 observed alleles in a population set; 87% heterozygosity) is more likely to exhibit 4 alleles with a two person mixture than **TPOX** (with only 10 observed alleles in the same population set; 68% heterozygosity)

PRINCIPLES

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



PROTOCOLS

ELSEVIER

Forensic Science International
91 (1998) 55–70

Forensic Science International

Analysis and interpretation of mixed forensic stains using DNA STR profiling

T.M. Clayton[a,*], J.P. Whitaker[a], R. Sparkes[b], P. Gill[b]

[a]Forensic Science Service, Wetherby Laboratory, Sandbeck Way, Audby Lane, Wetherby, West Yorkshire LS22 4DN, UK
[b]Forensic Science Service, Priory House, Gooch Street North, Birmingham B56QQ, UK

Received 13 May 1997; received in revised form 9 October 1997; accepted 27 October 1997

The FSS approach is cited by ISFG (2006) Recommendation #4

PROTOCOLS

Steps in the interpretation of mixtures

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

- 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

Figure 7.4, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

GUIDELINES

Impact of Assumptions on Statistical Calculations

SWGDAM Guidelines Section 4. Statistical Analysis of DNA Typing Results (introduction):

- While **the RMP** is commonly thought of in terms of single-source profiles, the application of this formula to evidentiary profiles inherently includes **an assumption of the number of contributors to the DNA sample**. As such, this document also applies the term RMP to mixture calculations where the number of contributors is assumed (this has sometimes been referred to as a "modified RMP"). By using the RMP nomenclature, these calculations are distinguished from **the CPI** nomenclature which is commonly thought of in terms of **a mixture calculation that makes no assumption as to the number of contributors**.

GUIDELINES

Terminology

SWGDAM Guidelines glossary:

- **Conditional**: an interpretation category that **incorporates assumption(s) as to the number of contributors**.
- **Restricted**: referring to a statistical approach conditioned on the number of contributors and with consideration of quantitative peak height information and inference of contributor mixture ratios; used to limit the genotypic combinations of possible contributors.
- **Unrestricted**: referring to a statistical approach performed without consideration of quantitative peak height information and inference of contributor mixture ratios; for CPE/CPI this may or may not be conditioned on the number of contributors.

GUIDELINES

Different Statistical Approaches Can Be Used

Table 1 – Suitable Statistical Analyses for DNA Typing Results
The statistical methods listed in the table cannot be combined into one calculation. For example, combining RMP at one locus with a CPI calculation at a second locus is not appropriate. However, an RMP may be calculated for the major component of a mixture and a CPE/CPI for the entire mixture (as referred to in section 4.6.2).

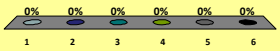
Category of DNA Typing Result	RMP	CPE/CPI	LR (1)
Single Source	✔		✔
Single Major Contributor to a Mixture	✔		✔
Multiple Major Contributors to a Mixture	✔ (2)	✔ (2)	✔
Single Minor Contributor to a Mixture	✔	✔ (3)	✔
Multiple Minor Contributors to a Mixture	✔ (2)	✔ (3)	✔
Indistinguishable Mixture	✔ (1)	✔	✔

(1) Restricted or unrestricted
(2) Restricted
(3) All potential alleles identified during interpretation are included in the statistical calculation

http://www.fbi.gov/hq/lab/html/codis_swgdam.pdf

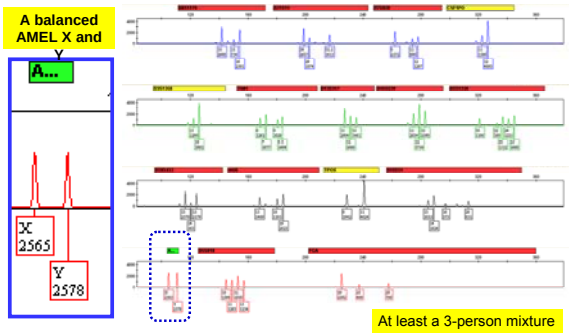
Primary means by which you determine the number of contributors?

1. Amelogenin X/Y ratio
2. Number of alleles present at a single locus
3. Assess the number of alleles present at multiple loci
4. Peak height ratio imbalance
5. Both #3 and #4
6. Our lab does not attempt to determine the number of contributors



PRACTICE

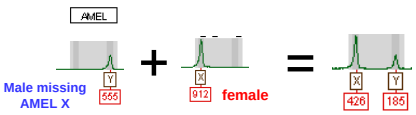
Amelogenin X/Y ratio can be a poor determinant of contributor number



PRACTICE

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)



PRINCIPLES

Possible genotype combinations in 2-person mixtures

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

♀ ♂

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

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

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

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

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

PRINCIPLES

All possible 2-contributor combinations

14 total combinations

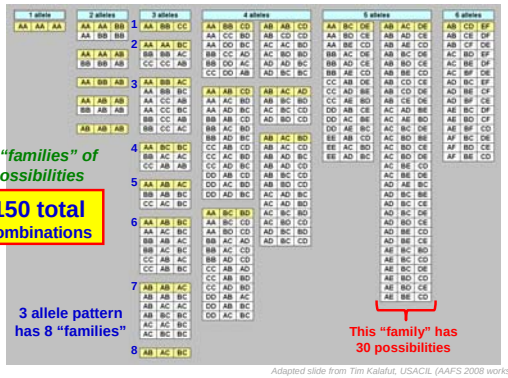
4 alleles			3 alleles			2 alleles			1 allele		
I	AB	CD 1	II	AA	BC 4	IV	AA	BB 10	VII	AA	AA 14
	AC	BD 2		BB	AC 5						
	BC	AD 3		CC	AB 6	V	AA	AB 11			
			III	AB	AC 7		BB	AB 12			
				AC	BC 8	VI	AB	AB 13			
				BC	AB 9						

7 "families" (groups or classes) of possibilities

Adapted slide from Tim Kalafut, USACIL (AAFS 2008 workshop)

PRINCIPLES

All possible 3-contributor combinations



23 "families" of possibilities

150 total combinations

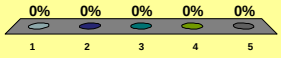
3 allele pattern has 8 "families"

This "family" has 30 possibilities

Adapted slide from Tim Kalafut, USACIL (AAFS 2008 workshop)

If 5 alleles are observed at a single locus in a DNA mixture profile, what conclusions would you draw?

1. I am still freaked out by the last slide showing how many allele combinations are possible with 3 contributors
2. At least 3 people are present in the mixture
3. 2-person mixture with a potential tri-allele at the one locus
4. Both #2 and #3 are possibilities and would be stated as such in the written report
5. I would not attempt to draw any conclusion



PROTOCOLS

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

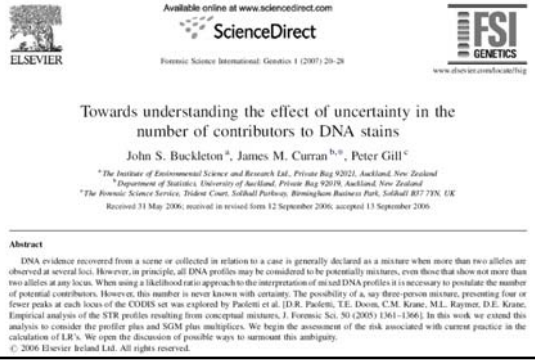
Results from Simulation Studies:

- 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

PROTOCOLS

Simulation Study Regarding Detecting the Number of Contributors to a Mixture



Available online at www.sciencedirect.com

ScienceDirect

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

ELSEVIER

FSI GENETICS

www.elsevier.com/locate/fsig

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

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

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

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

Abstract

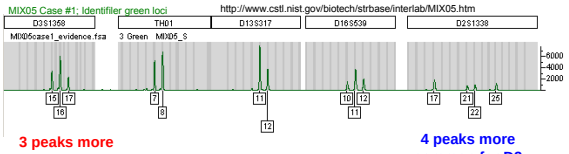
DNA evidence recovered from a scene or collected in relation to a case is generally declared as a mixture when more than two alleles are observed at several loci. However, in principle, all DNA profiles may be considered to be potentially mixtures, even those that show not more than two alleles at any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

PROTOCOLS

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

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

PRACTICE NEW SLIDE

Two Contributors and Observed Alleles

Maximum: 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

PRACTICE NEW SLIDE

Simulations with 2-person Mixtures

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

Loci	No. of alleles			
	1	2	3	4
D3	0.011	0.240	0.559	0.190
vWA	0.008	0.194	0.548	0.250
D16	0.016	0.287	0.533	0.164
D2	0.003	0.094	0.462	0.441
D8	0.011	0.194	0.521	0.274
D21	0.007	0.147	0.505	0.341
D18	0.003	0.095	0.472	0.430
D19	0.020	0.261	0.516	0.203
THO	0.016	0.271	0.547	0.166
FGA	0.003	0.116	0.500	0.381

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

PRACTICE NEW SLIDE

Three Contributors and Observed Alleles

Maximum: 6 alleles
All heterozygotes and non-overlapping alleles

PRACTICE NEW SLIDE

Simulations with 3-person Mixtures

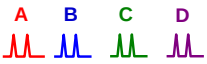
Table 2
The probability of observing a given number of alleles in a three-person mixtures for simulated profiles at the SGM^{+TM} loci

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

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

PRACTICE NEW SLIDE

Four Contributors and Observed Alleles



Maximum: 8 alleles
All heterozygotes and non-overlapping alleles

PRACTICE NEW SLIDE

Simulations with 4-person Mixtures

Table 3
The probability of observing a given number of alleles in a four person mixtures for simulated profiles at the SGM^{+TM} loci

Loci	No. of alleles showing							
	1	2	3	4	5	6	7	8
D3	0.000	0.011	0.178	0.497	0.291	0.023	0.001	0.000
vWA	0.000	0.008	0.107	0.406	0.377	0.097	0.005	0.000
D16	0.000	0.027	0.240	0.458	0.238	0.036	0.001	0.000
D2	0.000	0.001	0.020	0.148	0.363	0.345	0.112	0.012
D8	0.000	0.009	0.103	0.340	0.377	0.151	0.019	0.001
D21	0.000	0.005	0.058	0.262	0.392	0.231	0.049	0.003
D18	0.000	0.000	0.023	0.166	0.382	0.321	0.101	0.008
D19	0.000	0.025	0.199	0.399	0.282	0.086	0.010	0.000
THO	0.000	0.020	0.222	0.501	0.241	0.016	0.000	0.000
FGA	0.000	0.001	0.034	0.215	0.398	0.281	0.068	0.004

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

GUIDELINES

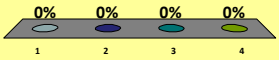
Methods Needed for Determining the Minimum Number of Contributors to a Mixture

SWGDAM Guidelines 3.4.1 and 3.4.2:

- 3.4.1. For DNA mixtures, **the laboratory should establish guidelines for determination of the minimum number of contributors to a sample**. Alleles need not meet the stochastic threshold to be used in this assessment.
- 3.4.2. The laboratory should define the number of alleles per locus and the relative intra-locus peak height requirements for assessing whether a DNA typing result is consistent with originating from one or more sources. **The minimum number of loci should be defined for determination of whether a sample is a mixture**.

What do you think are the minimum number of alleles and loci needed to determine if **a sample is a mixture**?

1. >2 alleles at ≥ 2 loci
2. >2 alleles at ≥ 1 locus
3. A single PHR imbalance of <60%
4. I haven't thought about this issue before!



Summary

- A common method to estimate the number of contributors is to examine the loci in a DNA profile and look for the maximum number of alleles observed
- The most variable loci (e.g., D18S51) will typically have the greatest number of alleles in mixtures
- The amelogenin X/Y ratio is only helpful in some limited situations for aiding assessment of the number of contributors and mixture ratios
- Simulations have shown that only a small percentage of the time would a 3-person mixture be incorrectly designated a 2-person mixture based solely on a maximum count of four alleles (two non-overlapping heterozygous allele pairs) across all 13 CODIS STRs