DNA Mixture Interpretation:
Principles and Practice in Component Deconvolution and Statistical Analysis

Statistical Analysis of Mixtures

AAFS 2008 Workshop #16
Washington, DC
February 19, 2008

George R. Carmody
gcarmody@gmail.com

Carleton UNIVERSITY
Canada's Capital University

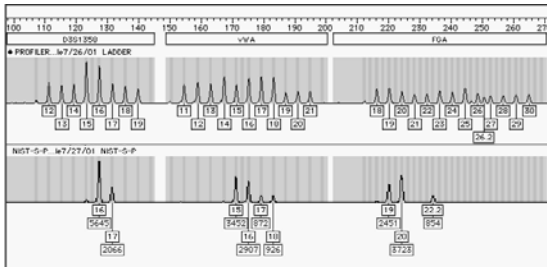
Statistics and DNA Mixture Interpretation

George Carmody

February 19, 2008
AAFS Workshop
Washington, DC

http://www.carleton.ca/~gcarmody

Common mixture scenarios



CROATIAN MEDICAL JOURNAL
CMJ
42(3):244-246,2001
FORENSIC SCIENCES

Interpretation of Complex Forensic DNA Mixtures

Carll Ladd, Henry C. Lee, Nicholas Yang, Frederick R. Bieber[1]

State of Connecticut Department of Public Safety, Division of Scientific Services, Forensic Science Laboratory, Conn; and [1]Departments of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Mass, USA

Ladd et al. 2001. Croatian Medical Journal 43(3): 244-246

Mixture Interpretation Possibilities

- 1. Qualitative statement ('..cannot exclude..')
- 2. Interpret as since source from peak height differences, differential extraction, etc. and calculate random match probability
- 3. Calculate probability of exclusion (CPE)
- 4. Calculate likelihood ratio

Mixture interpretation

- How many contributors?
- Exclusion probability (CPE)
- Likelihood ratio calculations comparing two alternative hypotheses

Estimate genotype frequency:

1. Frequency at each locus
   Hardy-Weinberg Equilibrium
2. Frequency across all loci
   Linkage Equilibrium
3. Corrections (NRC II)
   'θ' correction
   minimum allele frequency
   confidence interval

Hardy - Weinberg Equilibrium

A_1A_1 A_1A_2 A_2A_2

P_1^2 $2P_1P_2$ P_2^2

freq(A_1) = p_1

freq(A_2) = p_2

	A_1	A_2
A_1	P_1^2 A_1A_1	P_1P_2 A_1A_2
A_2	P_1P_2 A_1A_2	P_2^2 A_2A_2

$$(P_1 + P_2)^2 = P_1^2 + 2P_1P_2 + P_2^2$$

$$(P_1 + P_2 + P_3)^2 = P_1^2 + P_2^2 + P_3^2 + 2P_1P_2 + 2P_1P_3 + 2P_2P_3$$

What do the numbers mean?

Random match probability = .000001

Random match probability = 1/1,000,000

Exclusion probability = .999999

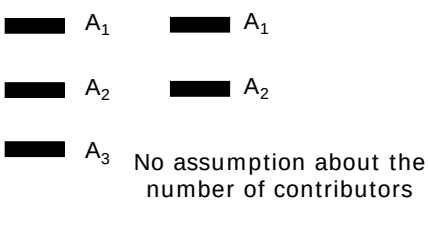
Exclusion probability = 99.9999%

Exclusion Probability =

the combined frequency of all genotypes that could be excluded from the mixture, assuming H-W equilibrium for the genotype frequencies.

All possible genotypes

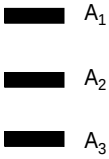
- 3 alleles: A_1, A_2, A_3
- A_1A_1, A_1A_2, A_1A_3, A_2A_2, A_2A_3, A_3A_3
- 4 alleles: A_1, A_2, A_3, A_4
- A_1A_1, A_1A_2, A_1A_3, A_1A_4, A_2A_2, A_2A_3, A_2A_4, A_3A_3, A_3A_4, A_4A_4
- k alleles: k(k+1)/2 genotypes; HWE



Three Allele Scenario

Evidence

Evidence



$p_{A1} + p_{A2} + p_{A3} = p \quad P_{inc} = p^2$

$P_E = 1 - P_{inc} = 1 - p^2$

Combined Exclusion Probability =

$1 - [P_{inc,\,1} \times P_{inc,\,2} \times P_{inc,\,3} \times P_{inc,\,4} \ldots]$

- multiply inclusion probabilities together across independent loci

Assumptions:

- Independence
- All contributors of same racial group
- All unrelated
- No allele dropout
- No intensity differences

Exclusion Probability

Not as powerful as LR

(ie: number not as large)

Possibilities:

Calculate using P_E for all loci

Calculate RMP for single contributor loci only

Use more definitive hypotheses

With STRs and intensity differences, an evidence profile may have...

some loci where the contributors can be determined as single sources

and others where this may not be possible

loci with peaks below the LOQ should not be used in the CPE calculation

J Forensic Sci, January 2007, Vol. 52, No. 1
doi:10.1111/j.1556-4029.2006.00318.x
Available online at: www.blackwell-synergy.com

TECHNICAL NOTE

Jason R. Gilder,[1] M.S.; Travis E. Doom,[2] Ph.D.; Keith Inman,[3] M. Crim.; and Dan E. Krane,[4] Ph.D.

Run-Specific Limits of Detection and Quantitation for STR-based DNA Testing

LOD = mean background + 3 s.d.

LOQ = mean background + 10 s.d.

peak height imbalance

stutter percentage

Bayes formula (odds form):

$$\left(\frac{P(G|E)}{P(\bar{G}|E)}\right)=\left(\frac{P(E|G)}{P(E|\bar{G})}\right)\times\left(\frac{P(G)}{P(\bar{G})}\right)$$

posterior odds = (likelihood ratio) x (prior odds)

G = guilt; E = DNA evidence $P(G)+P(\bar{G})=1$

mutually exclusive

Bayes formula (odds form):

$$\left(\frac{P(H_1 \mid E)}{P(H_2 \mid E)}\right) = \left(\frac{P(E \mid H_1)}{P(E \mid H_2)}\right) \times \left(\frac{P(H_1)}{P(H_2)}\right)$$

posterior odds = (likelihood ratio) x (prior odds)

E = DNA evidence

H_1 = hypothesis #1 H_2 = hypothesis #2

mutually exclusive

Likelihood ratio - interpretation

- LR can range from 0 ⟷ 1 ⟷ ∞
- LR < 1; the genetic evidence has more support from denominator hypothesis than from numerator hypothesis
- LR=1; the genetic evidence has equal support from both numerator/denominator hypotheses
- LR > 1; the genetic evidence has more support from numerator hypothesis than from denominator hypothesis.

Likelihood ratio - interpretation

LR	Verbal equivalent
1 - 10	'limited support'
10 - 100	'moderate support'
100 - 1,000	'strong support'
> 1,000	'very strong support'

- from:
- - Evett, I.W. & B.S. Weir. 1998. Interpreting DNA Evidence. (p. 226)

Likelihood ratio - interpretation

LR	Verbal equivalent
1 - 10	'limited support'
10 - 100	'moderate support'
100 - 1,000	'moderately strong support'
1,000 – 10,000	'strong support'
> 10,000	'very strong support'

- from:
- - Evett, I.W. et al. 2000. Science & Justice. 40:233-239.

Bayes formula (odds form):

$$\left(\frac{P(H_1 \mid E)}{P(H_2 \mid E)}\right) = \left(\frac{P(E \mid H_1)}{P(E \mid H_2)}\right) \times \left(\frac{P(H_1)}{P(H_2)}\right)$$

posterior odds = (likelihood ratio) x (prior odds)

E = DNA evidence

H_1 = hypothesis #1 H_2 = hypothesis #2

Mixtures - Likelihood Ratio Approach

- 1. H_p - prosecution hypothesis
- 2. H_d - defense hypothesis
- 3. Mutually exclusive hypotheses
- 4. LR = Prob(evidence|H_p) / Prob(evidence|H_d)

Likelihood Ratios for Two Person Mixtures

- Scenario 1. Victim is in the mixture
- (eg.: intimate sample)
- Scenario 2. Victim is not in the mixture

Available online at www.sciencedirect.com
SCIENCE DIRECT
Forensic Science International 160 (2006) 90–101
Forensic Science International
www.elsevier.com/locate/forsciint

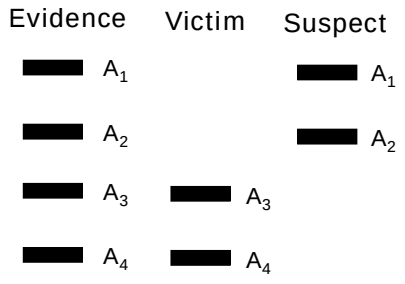
DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures

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
[d] Institute of Legal Medicine, Faculty of Medicine, University of Santiago de Compostela, 15705 Santiago de Compostela, Spain
[e] Institute of Medical Informatics and Statistics, Kiel, Germany
[f] Division of Blood Group Serology, Medical University of Vienna, Austria
[g] Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark
[h] Office of the Chief Medical Examiner, Department of Forensic Biology, 520 First Avenue, New York, NY 10016, USA
[i] Institute of Legal Medicine University Clinic of Cologne, Melatengürtel, 60-62 D 50823 Köln, Germany
[j] University of Washington, Department of Biostatistics, Box 357232, Seattle, WA 98195, USA

Received 4 April 2006; accepted 10 April 2006

'unrestricted combinatorial approach'



Four Allele Scenario

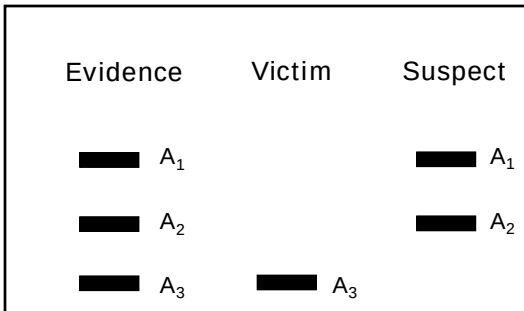
Four Alleles

Two Match the victim - A_3 A_4

Two Match the suspect - A_1 A_2

$LR = Prob(evidence|H_p)/Prob(evidence|H_d)$

$1/2p_1p_2$



Three Allele Scenario

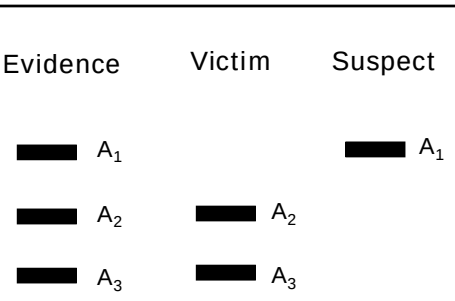
Three Alleles

Victim is homozygote - $A_3\,A_3$

Two Match the suspect - $A_1\,A_2$

$LR = P(E|H_p)/P(E|H_d)$

$1 / 2p_1p_2$



Three Allele Scenario

Three Alleles

Victim is heterozygote - $A_2 A_3$

Suspect is homozygote- $A_1 A_1$

Three possible genotypes can explain the evidence

Given that the victim is heterozygote - A_2A_3

The possible genotypes to explain the evidence:

A_1A_1, A_1A_2, A_1A_3

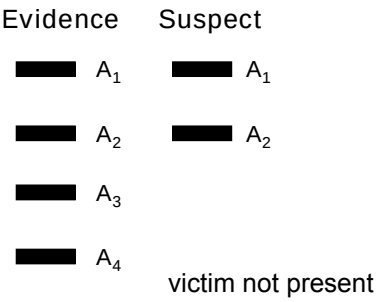
A_1A_1 A_1A_2 A_1A_3

$p_1^2 + 2p_1p_2 + 2p_1p_3$

$$LR = P(E|H_p)/P(E|H_d)$$

$$1/(p_1^2 + 2p_1p_2 + 2p_1p_3)$$

Evidence	Victim	Suspect	H_p	H_d
A_1A_1	A_1A_1	A_1A_1	1	p_1^2
A_1A_2	A_1A_2	A_1A_2	1	$p_1^2 + 2p_1p_2 + p_2^2$
A_1A_2	A_1A_1	A_2A_2	1	$2p_1p_2 + p_2^2$



Four Allele Scenario

Four Alleles

Two match the suspect - A_1A_2

Two match the unknown - A_3 & A_4

H_p

Suspect is A_1A_2

Unknown is A_3A_4

$2p_3p_4$

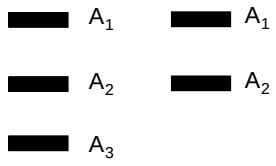
H_d

	Unk 1	Unk 2	
3	A_1A_2	A_3A_4	$2p_1p_2 \times 2p_3p_4$
	A_1A_3	A_2A_4	$2p_1p_3 \times 2p_2p_4$
	A_1A_4	A_2A_3	$2p_1p_4 \times 2p_2p_3$
3	A_3A_4	A_1A_2	$2p_3p_4 \times 2p_1p_2$
	A_2A_4	A_1A_3	$2p_2p_4 \times 2p_1p_3$
	A_2A_3	A_1A_4	$2p_2p_3 \times 2p_1p_4$
			$24p_1p_2p_3p_4$

$$LR = P(E|H_p)/P(E|H_d)$$

$$\frac{2p_3p_4}{24p_1\,p_2\,p_3\,p_4}$$

$$\frac{1}{12p_1p_2}$$



Three Allele Scenario

Three Alleles

Two match the suspect - A_1A_2

The unknown contributes - at least A_3

H_p

Suspect is A_1A_2

Unknown is A_3A_3 or A_1A_3 or A_2A_3

$p_3^2 + 2p_1p_3 + 2p_2p_3$

H_d

Unk 1	Unk 2	
A_1A_2	A_1A_3	$2p_1p_2 \times 2p_1p_3$
A_1A_2	A_2A_3	$2p_1p_2 \times 2p_2p_3$
A_1A_2	A_3A_3	$2p_1p_2 \times p_3^2$
A_1A_3	A_1A_2	$2p_1p_3 \times 2p_1p_2$
A_1A_3	A_2A_3	$2p_1p_3 \times 2p_2p_3$
A_1A_3	A_2A_2	$2p_1p_3 \times p_2^2$
A_2A_3	A_1A_2	$2p_2p_3 \times 2p_1p_2$
A_2A_3	A_1A_3	$2p_2p_3 \times 2p_1p_3$
A_2A_3	A_1A_1	$2p_2p_3 \times p_1^2$

A_1A_1	A_2A_3	$p_3^2 \times 2p_2p_3$
A_2A_2	A_1A_3	$p_2^2 \times 2p_1p_3$
A_3A_3	A_1A_2	$p_3^2 \times 2p_1p_2$

$12p_1p_2p_3(p_1+p_2+p_3)$

$$LR = P(E|H_p)/P(E|H_d)$$

$$\frac{p_3^2 + 2p_1 p_3 + 2p_2 p_3}{12p_1p_2p_3 (p_1 + p_2 + p_3)}$$

$$\frac{p_3 + 2p_1 + 2p_2}{12p_1p_2 (p_1 + p_2 + p_3)}$$

Number of Contributors

2 alleles - 2, 3, 4 contributors

3 alleles - 2, 3, 4 contributors

4 alleles - 2, 3, 4 contributors

Bayes formula (odds form):

$$\left(\frac{P(H_1|E)}{P(H_2|E)}\right) = \left(\frac{P(E|H_1)}{P(E|H_2)}\right) \times \left(\frac{P(H_1)}{P(H_2)}\right)$$

posterior odds = likelihood ratio x prior odds

E = DNA evidence

H_1 = hypothesis #1; H_2 = hypothesis #2