



Least-Square Deconvolution (LSD): A New Way of Resolving STR/DNA Mixture Samples

Presented by

Tsewei Wang*, Ning, Xue*, and Ray Wickenheiser⁺

*Department of Chemical Engineering, &
Laboratory for Information Technologies
The University of Tennessee

&

⁺Crime Lab, Acadiana, New Iberia, LA

At the

13th International Symposium on Human Identification

October 7-10, 2002

Phoenix, AZ



Acknowledgement

- This project is supported by the US Depart. Of Justice, Federal Bureau of Investigation under contract J-FBI-98-08.
- Doug Birdwell, Director of the LIT Lab, inception of the original idea
- Ning Xue, M.Sc. Thesis, Aug. 2002
- Mark Rader, Jay Flaherty, web-implementation of LSD
- P. Graham, Texas Department of Public Safety, Corpus Christi, TX for sending forensic mixture data.
- Ray Wickenheiser, Crime Lab, New Iberia, Acadiana, LA, for supplying data, and valuable consultations
- St. Louis Metro. Police Dep. Crime Lab
- Beta-testers of web-LSD around the countries, and abroad

*The views and conclusions contained in this presentation are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the US Government.



OUTLINE

- Formulation of the LSD algorithm
- Some examples using real mixture sample data
- Limitation of LSD and interpretation guidelines for DNA analysts
- Comparison of LSD to LMA (Linear Mixture Analysis)
 - Control run—our implementation of LMA
 - Two counter examples in which LSD yielded all correct resolution; and LMA yielded incorrect resolution at some loci in both samples



Underlying Assumptions

- The relative mass ratio is approximately preserved during PCR amplification across
 - loci
 - alleles within a locus
- The allele peak area is proportional to its relative DNA mass;
 - e.g. twice the DNA mass, twice the allele peak area

Possible Genotype Combinations for the 4-Allele Case

<u>Case</u>	<u>Person1</u>	<u>Person2</u>	<u>Matrix A Representation</u>	<u>A⁺, Pseudoinverse Of the Matrix</u>
1	A, B	C, D	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	$0.5 * \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$
2	A, C	B, D	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$	$0.5 * \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$
3	A, D	B, C	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$	$0.5 * \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$

Possible Genotype Combinations for the 3-allele Case

Case	Person1	Person2	Matrix A representation	A ⁺ Pseudoinverse Of the Matrix
1	A	B, C	$\begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	$0.5 * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$
2	B	A, C	$\begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix}$	$0.5 * \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$
3	C	A, B	$\begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 2 & 0 \end{bmatrix}$	$0.5 * \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

3-allele case (Continued)

Matrix A
representation

A⁺ Pseudoinverse
Of the Matrix

Case Person1 Person2

4 A, B B, C

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\frac{1}{3} * \begin{bmatrix} 2 & 1 & -1 \\ -1 & 1 & 2 \end{bmatrix}$$

5 A, B A, C

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\frac{1}{3} * \begin{bmatrix} 1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}$$

6 A, C B, C

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\frac{1}{3} * \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}$$

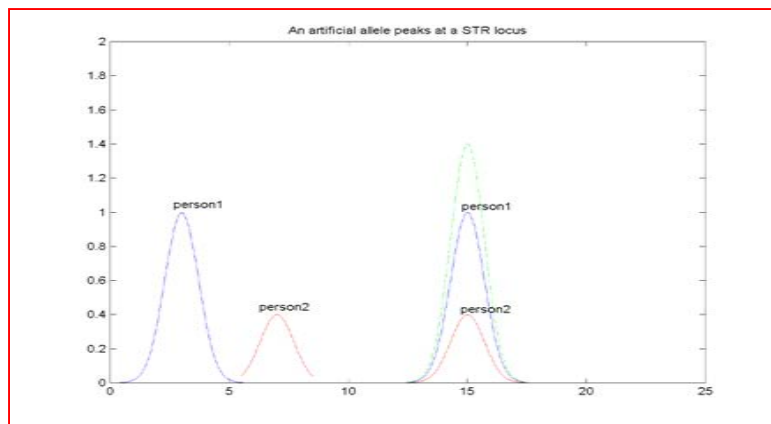
Possible Genotype Combinations for the 2-allele Case

Case	Person1	Person2	Matrix A Representation	Pseudoinverse Of the Matrix
1	A	B	$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$	$0.5 * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2	A, B	A	$\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$	$0.5 * \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix}$
3	A, B	B	$\begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$	$0.5 * \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}$
4	A, B	A, B	$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	$0.25 * \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

Central Theme of the Least Square Approach

The allele peak areas at a locus of a mixed STR/DNA sample are a **SUPERPOSITION** of the individual contributors' allele peak areas, weighted by its corresponding mass proportion

$$\begin{array}{l}
 \text{allele1} \longrightarrow \\
 \text{allele2} \longrightarrow \\
 \text{allele3} \longrightarrow
 \end{array}
 \begin{array}{c}
 A \\
 \left[\begin{array}{cc}
 1 & 0 \\
 0 & 1 \\
 1 & 1
 \end{array} \right]
 \end{array}
 \begin{array}{c}
 x \\
 \left[\begin{array}{c}
 \textit{mass1} \\
 \textit{mass2}
 \end{array} \right]
 \end{array}
 =
 \begin{array}{c}
 = \\
 \textit{mass1}
 \end{array}
 \begin{array}{c}
 \left[\begin{array}{c}
 1 \\
 0 \\
 1
 \end{array} \right]
 \end{array}
 +
 \begin{array}{c}
 + \\
 \textit{mass2}
 \end{array}
 \begin{array}{c}
 \left[\begin{array}{c}
 0 \\
 1 \\
 1
 \end{array} \right]
 \end{array}
 =
 \begin{array}{c}
 = \alpha \\
 \left[\begin{array}{c}
 \textit{allele1area} \\
 \textit{allele2area} \\
 \textit{allele3area}
 \end{array} \right]
 \end{array}
 \begin{array}{c}
 b
 \end{array}$$



The Least Square Fit of the Allele Peak Areas at a Locus

Set up A and b: $Ax=b$
where

$$A x = b$$

A = the matrix of genotype combinations, 2 columns

b = the vector of measured allele peak areas at a locus

x = the vector of unknown mass coefficients; 2 elements

A^+ = pseudoinverse of A

Answer



$$X = (A^+)b$$

To Compute the Least Square Solution for the Example

$$x = [A^+]b$$

$$\begin{bmatrix} mass1 \\ mass2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} peakarea1 \\ peakarea2 \\ peakarea3 \end{bmatrix}$$



Calculation of the Error of the Least-Square Fit

- First, calculate the fitted peak area vector:

$$b_f = A * x$$

- Calculate the fitting error:

$$error = b - b_f$$

- Calculate the magnitude of the error vector:

$$\|e\|^2 = \text{sum of the squares of the elements}$$

Mixture Sample 1 with 9 Loci Informaton, from TDPS, Corpus Christi

Locus	Alleles in the Mixture	Peak Areas	Genotypes	
			Victim	Suspect
D3S1358	15 16 17	1242 657 1546	15, 17	16, 17
VWA	16 17 18	616 2021 425	17, 17	16, 18
FGA	22 23 24 25	850 468 681 315	22, 24	23, 25
D8S1179	12 13 16	1842 490 461	12, 12	13, 16
D21S11	30 30.2 31.2 32.2	842 490 509 804	30, 32.2	30.2, 31.2
D18S51	12 13 15	440 1503 387	13, 13	12, 15
D5S818	7 12 13	486 512 1886	13, 13	7, 12
D13S317	8 9 11	734 1068 299	8, 9	9, 11
D7S820	10 11	614 1169	11, 11	10, 10

Source of data: P. Graham. "Allele Peak Data for DNA Mixture samples from real cases," Texas Department of Public Safety, Corpus Christi, 2001.

LSD Output Results for Mixture Sample 1

LSD Results

Locus	Person 1	Person 2	Fitting Error	Error Ratio	Mass Ratio	Comments
D3S1358						Three Alleles
1	16,17	15,17	9.6E-02	1.0	1.0:2.1	
2	16,16	15,17	0.11	1.1	1.0:4.2	←
3	17,17	15,16	0.40	4.1	1.0:1.2	
4	15,16	15,17	0.71	7.4	1.0:3.2	←
5	15,15	16,17	0.92	9.5	1.0:1.8	
6	15,16	16,17	3.51	36	1.0:1.6	
TH01						No Alleles
D21S11						Four Alleles
1	30.2,31.2	30,32.2	3.8E-03	1.0	1.0:1.6	
2	30.2,32.2	30,31.2	0.44	116	1.0:1.0	
3	31.2,32.2	30,30.2	0.44	117	1.0:1.0	
D18S51						Three Alleles
1	12,15	13,13	9.4E-03	1.0	1.0:1.8	
2	13,15	12,13	1.02	108	1.0:1.1	
3	15,15	12,13	3.77	402	1.0:5.0	
4	12,12	13,15	4.16	443	1.0:4.3	
5	12,15	12,13	4.68	499	1.0:-10.6	
6	12,15	13,15	5.39	575	1.0:-12.5	
D5S818						Three Alleles
1	7,12	13,13	1.4E-03	1.0	1.0:1.9	
2	7,13	12,13	1.11	778	1.0:1.0	
3	7,7	12,13	4.00	2793	1.0:4.9	
4	12,12	7,13	4.15	2899	1.0:4.6	
5	7,12	12,13	4.88	3412	1.0:-9.4	
6	7,12	7,13	5.16	3605	1.0:-10.0	
D13S317						Three Alleles
1	9,11	8,9	4.6E-03	1.0	1.0:2.4	
2	11,11	8,9	0.62	137	1.0:6.0	

3	8,11	9,9	1.06	232	1.0:1.0	
4	8,11	8,9	1.49	327	1.0:9.7	
5	8,8	9,11	3.31	724	1.0:1.9	
6	8,11	9,11	8.42	1844	1.0:2.4	

D7S820

Two Alleles
REVIEW the
LSD Results

1	10,10	11,11	0.0E+00		1.0:1.9	←
2	10,10	10,11	0.0E+00		1.0:-4.2	
3	11,11	10,11	0.0E+00		1.0:2.2	←
4	10,11	10,11	0.41		1.0:1.0	

D16S539

No Alleles

CSF1PO

No Alleles

VWA

Three Alleles

1	16,18	17,17	0.10	1.0	1.0:1.9	
2	17,18	16,17	1.77	18	1.0:1.3	
3	18,18	16,17	5.46	54	1.0:6.2	
4	16,18	16,17	6.18	61	1.0:-7.6	
5	16,16	17,18	7.05	70	1.0:4.0	
6	16,18	17,18	9.03	89	1.0:-10.6	

D8S1179

Three Alleles

1	13,16	12,12	2.0E-03	1.0	1.0:1.9	
2	12,16	12,13	1.25	629	1.0:1.0	
3	16,16	12,13	4.30	2173	1.0:5.1	
4	13,13	12,16	4.49	2268	1.0:4.7	
5	13,16	12,13	5.16	2606	1.0:-8.6	
6	13,16	12,16	5.49	2775	1.0:-9.1	

TPOX

No Alleles

FGA

Four Alleles

1	23,25	22,24	0.26	1.0	1.0:2.0	
2	24,25	22,23	1.41	5.4	1.0:1.3	
3	23,24	22,25	1.67	6.4	1.0:1.0	

**Best-Fit Mass Ratio of the Top-Ranked Genotype
Combination Case from All the 4-allele, and 3-allele Loci of
Sample 1**

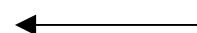
Loci	Mass Ratio of Top-Ranked Genotype Case
3-allele locus	2.1
4-allele	1.6
3-allele	1.8
3-allele	1.9
3-allele	2.4
3-allele	1.9
3-allele	1.9
4-allele	2.0
<i>Average</i>	<i>1.95</i>

Detailed LSD Results for the 3-alleleVWA Locus of Mixture Sample 1

Cases	Genotype Combinations		Matrix A	Error Ratio	Mass Ratio
	Person 1	Person 2			
1	16, 16	17, 18	$\begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	69.82	1: 3.97
2	16, 18	17, 17	$\begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix}$	1	1: 1.94 ←
3	18, 18	16, 17	$\begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 1 \end{bmatrix}$	54.11	1: 6.20
4	17, 18	16, 17	$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$	17.55	1: 1.25
5	16, 18	16, 17	$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$	61.20	1: -7.63
6	16, 18	17, 18	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$	89.42	1: -10.58

Ranked Genotype Combination Cases w.r.t. the Fitting Error of VWA Locus of Sample 1

Rank	Genotype Combinations		Matrix A	Ratio of Errors	Mass Ratio
	Person 1	Person 2			
1	16, 18	17, 17	$\begin{bmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 1 \end{bmatrix}$	1	1: 1.94
2	17, 18	16, 17	$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$	17.55	1: 1.25
3	18, 18	16, 17	$\begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 2 & 1 \end{bmatrix}$	54.11	1: 6.20
4	16, 18	16, 17	$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$	61.20	1: -7.63
5	16, 16	17, 18	$\begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	69.82	1: 3.97
6	16, 18	17, 18	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$	89.42	1: -10.58



Detailed Ranked LSD Result for D21S11 (4-alleles) of Sample 1

Rank	Genotype Combinations		Matrix A	Ratio of Errors	Mass Ratio Calculated
	Person 1	Person 2			
1	30.2, 31.2	30, 32.2	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$	1	1: 1.65
2	30.2, 32.2	30, 31.2	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$	116.06	1:1.04
3	31.2, 32.2	30, 30.2	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	116.86	1:1.01

Detailed LSD Results for D7S820 (2 alleles): two possible choices

Case s	Genotype Combinations		Matrix A	Error s	Mass Ratio
	Person 1	Person 2			
1	10, 10	11, 11	$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$	0	1: 1.90
2	10, 10	10, 11	$\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$	0	1: -4.21
3	11, 11	10, 11	$\begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$	0	1: 2.21
4	10, 11	10, 11	$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	0.41	1:1

Allele height: {614, 1169}

LSD-Guided Resolution Results for Mixture Sample 1: All Loci Are Correctly Resolved

Top-ranked genotype combination case of each non-2-allele locus turns out to give the correct genotypes for both contributors.

Locus	Alleles	Ratio of Peak Area	LSD guided Result			Comment
			Mass Ratio Calculated	Person 1 (suspect)	Person 2 (victim)	
D3S1385	15, 16, 17	1.89: 1: 2.35	1: 2.08	16, 17	15, 17	Correct
VWA	16, 17, 18	1.45: 4.76: 1	1: 1.94	16, 18	17, 17	Correct
FGA	22, 23, 24, 25	2.70: 1.49: 2.16: 1	1: 1.96	23, 25	22, 24	Correct
D8S1179	12, 13, 16	4.00: 1.06: 1	1: 1.94	13, 16	12, 12	Correct
D21S11	30, 30.2, 31.2, 32.2	1.72: 1: 1.04: 1.64	1: 1.65	30.2, 31.2	30, 32.2	Correct
D18S51	12, 13, 15	1.14: 3.88: 1	1: 1.82	12, 15	13, 13	Correct
D5S818	7, 12, 13	1: 1.06: 3.88	1: 1.89	7, 12	13, 13	Correct
D13S317	8, 9, 11	2.45: 3.57: 1	1: 2.40	9, 11	8, 9	Correct
D7S820	10, 11	1: 1.90	1: 1.90	10, 10	11, 11	← Two possible choices
			1: 2.21	11, 11	10, 11	

Mixture Sample 2, from Acadiana Crime Lab, Louisiana

Locus	Alleles in the Mixture	Allele Peak Heights Measured	Contributors' Genotypes	
			Suspect	Victim
D3S1358	14 16 17 19	616 688 3592 3479	14, 16	17, 19
VWA	14 16	3483 3534	14, 16	14, 16
FGA	22	5503	22, 22	22, 22
D8S1179	13 14	3111 3175	13, 14	13, 14
D21S11	28 29 30	3166 489 2964	29, 29	28, 30
D18S51	13 14 15 18	2383 380 2041 271	14, 18	13, 15
D5S818	11 12 13	538 2448 2240	11, 11	12, 13
D13S317	11 12 13 14	309 1892 366 1734	11, 13	12, 14
D7S820	8 10 12	1685 1409 144	8, 12	8, 10
D16S539	11 12 13 14	3047 460 562 2329	12, 13	11, 14
TH01	9 9.3	2276 2666	9, 9.3	9, 9.3
TPOX	8 11	2556 2447	8, 11	8, 11
CSF1PO	10 11 12	1693 339 2164	11, 12	10, 12

LSD-Guided Resolution Result for Sample 2

- Top-Ranked Genotype Case of All 4-allele and 3-allele Loci Gives True Genotype
- 2-allele Loci Can Be Resolved with Help of a Reference Genotype

Locus	Alleles	Measured Peak Area Ratio	LSD Suggested Genotype Resolution Result			
			Person 1 (Suspect)	Person 2 (Victim)	Mass Ratio Calculated	Remarks
D3S1385	14 16 17 19	1: 1.12: 5.83: 5.64	14, 16	17, 19	1: 5.42	Correct
VWA	14 16	1: 1.01	14, 16	14, 16	Indeterminate	2-allele locus
FGA	22	-	22, 22	22, 22	Indeterminate	One-allele case
D8S1179	13 14	1: 1.02	13, 14	13, 14	Indeterminate	2-allele locus
D21S11	28 29 30	6.47: 1: 6.06	29,29	28,30	1:12.5	Correct (smaller error)
			28,29	28,30	1:7.3	possible
D18S51	13 14 15 18	8.79: 1.40: 7.53: 1	14, 18	13, 15	1: 6.80	Correct
D5S818	11 12 13	1: 4.55: 4.16	11, 11	12, 13	1: 8.71	Correct
D13S317	11 12 13 14	1: 6.12: 1.18: 5.61	11, 13	12, 14	1: 5.37	Correct
D7S820	8 10 12	11.70: 9.78: 1	8, 12	8, 10	1: 7.73	Correct
D16S539	11 12 13 14	6.61: 1: 1.22: 5.06	12, 13	11, 14	1: 5.26	Correct
TH01	9 9.3	1: 1.17	9, 9.3	9, 9.3	Indeterminate	2-allele locus
TPOX	8 11	1.04: 1	8, 11	8, 11	Indeterminate	2-allele locus
CSF1PO	10 11 12	4.99: 1: 6.38	11, 12	10, 12	1: 4.54	Correct

Mixture Sample 3: 9 Loci from TDPS, Corpus Christi, with one 4-allele locus, five 3-allele loci, two 2-allele loci, and one 1-allele locus

Locus	Alleles in the Mixture	Allele Peak Areas Measured	Contributors' Genotypes	
			Female Victim	Male Suspect
D3S1358	15 16 17	2935 1791 999	15, 16	15, 17
VWA	16	5074	16, 16	16, 16
FGA	20 21 22	1973 2679 1919	20, 22	21, 21
D8S1179	10 12 13	3237 1297 2116	10, 13	10, 12
D21S11	29 30	2077 4266	29, 30	30, 30
D18S51	17 18 19	1637 1101 2468	17, 19	18, 19
D5S818	11 12 13	4825 2154 2600	11, 13	11, 12
D13S317	9 11 12 14	1774 2173 1566 2016	11, 14	9, 12
D7S820	10 12	3528 1559	10, 12	10, 10

LSD-Guided Resolution Result for Sample 3

- Top ranked genotype case for all non 2-allele loci yields the correct genotype
- Correct resolution at 2-allele loci with the help of reference genotype

Locus	Alleles in the Mixture	Peak Area Ratio Measured	LSD Suggested Genotype Resolution Result			
			Person 1 (Suspect)	Person 2 (Victim)	Mass Ratio Calculated	Remarks
D3S1385	15 16 17	2.94: 1.79: 1	15, 17	15, 16	1: 1.76	True one
VWA	16	--	16, 16	16, 16	Indeterminate	True one
FGA	20 21 22	1.02: 1.40: 1	21, 21	20, 22	1: 1.45	True one
D8S1179	10 12 13	2.50: 1: 1.63	10, 12	10, 13	1: 1.66	True one
D21S11	29 30	1; 2.05	30, 30	29, 30	1: 1.89	True one
			29, 29	30, 30	1: 2.05	
D18S51	17 18 19	1.49: 1: 2.24	18, 19	17, 19	1: 1.53	True one
D5S818	11 12 13	2.24: 1: 1.21	11, 12	11, 13	1: 1.20	True one
D13S317	9 11 12 14	1.13: 1.39: 1: 1.29	9, 12	11, 14	1: 1.25	True one
D7S820	10 12	2.263: 1	10, 10	10, 12	1: 1.58	True one
			12, 12	10, 10	1: 2.26	

Mixture Sample 4: Mother-Child Mixture (From Wickenheiser Four 3-allele, six 2-allele,, 3 1-allele loci); correct resolution results

Exhibits Locus	Mixture on towel	Note with baby (possible mother) Person 1	Bloodstain on towel (baby) Person 2	LSD Person 1 Best Fit	LSD Person 2 Best Fit	LSD Best Fit Mass Ratio	Comment
D3S1358	15 5944 17 3737	15	15 17	17,17 15,15* 17,17	15,15 15,17* 15,17	1.0:1.6 1.0:3.4* 1.0:-5.4	* Correct suspect with known contributor
VWA	15 3258 17 5723	17	15 17	15,15 15,15 17,17*	17,17 15,17 15,17*	1.0:1.8 1.0:-4.6 1.0:2.6*	* Correct suspect with known contributor
FGA	22 3317 23 4599	23	22 23	22,22 22,22 23,23*	23,23 22,23 22,23*	1.0:1.4 1.0:-7.2 1.0:5.2*	* Correct suspect with known contributor
Amelogenin	X 7090 Y 3934	X	X Y				
D8S1179	14 7419	14	14	14,14	14,14	1:1	LSD correct
D21S11	28 3708 31 2364 32.2 1011	28 32.2	28 31	28,32.2	28,31	1.0:2.2	LSD correct
D18S51	15 2627 16 2046 17 739	15 17	15 16	15,17	15,16	1.0:2.9	LSD correct
D5S818	11 4245 12 3960	11 12	11 12	12,12 11,11 12,12 11,12*	11,11 11,12 11,12 11,12*	1.0:1.1 1.0:27.8 1.0:-29.8 1.0:1.0*	* Correct suspect with known contributor
D13S317	11 3366 13 1775	11	11 13	13,13 11,11* 13,13	11,11 11,13* 11,13	1.0:1.9 1.0:2.2* 1.0:-4.2	* Correct suspect with known contributor
D7S820	10 3876	10	10	10,10	10,10	1:1	LSD correct
D16S539	9 6028	9	9	9,9	9,9	1:1	LSD correct
TH01	7 3879 8 875 9 2120	7 8	7 9	7,8	7,9	1.0:2.1	LSD correct
TPOX	9 1002 11 5198	9 11	11	9,9 9,9 9,11*	11,11 9,11 11,11*	1.0:5.2 1.0:-2.5 1.0:2.1*	* Correct suspect with known contributor
CSF1P0	10 845 11 2664 12 1916	10 11	11 12	10,11	11,12	1.0:2.3	LSD correct

Mixture Sample 5: from Acadiana Crime Lab, LA
13 loci, with NINE 2-allele loci, where DNA mass portions are ~ 1:1

Locus	Alleles in the Mixture	Allele Peak Heights Measured	Contributors' True Genotypes	
			Person 1 (Victim)	Person 2 (Suspect)
D3S1358	15 16 18	754 1253 551	16, 18	15, 16
VWA	14 17	489 1178	14, 17	17, 17
FGA	22 23	757 626	22, 23	22, 23
D8S1179	11 12 15	513 1013 612	11, 12	12, 15
D21S11	30 31.2	1108 332	30, 30	30, 31.2
D18S51	12 17 18	150 415 190	17, 18	12, 17
D5S818	10 11	1164 948	10, 11	10, 11
D13S317	9 12	339 957	9, 12	12, 12
D7S820	8 11	279 463	8, 11	11, 11
D16S539	11 12 13	506 765 304	11, 12	12, 13
TH01	7 9.3	1143 374	7, 7	7, 9.3
TPOX	9 10	809 645	9, 10	9, 10
CSF1PO	10 11	267 527	10, 11	11, 11

LSD-Guided Resolution Results of Sample 5:

Green cells: true genotypes

Orange cells: Permuted assignment

Red cell: wrong assignment

Mass ratio close to 1:1 → low confidence in LSD result

Locus	Alleles in the Mixture	Peak Area Ratio Calculated	LSD Suggested Genotype Resolution Result			
			Person 1 (Victim)	Person 2 (Suspect)	Mass Ratio	Remarks
D3S1385	15	1.37:	16, 18	15, 16	1: 1.4	True one
	16	2.27:	15, 16	16, 18	1.4: 1	
VWA	14	1:	17, 17	14, 17	1: 1.4	True one
	17	2.41	14, 17	17, 17	1.4: 1	
FGA	22	1.21:	23, 23	22, 22	1: 1.2	True one
	23	1	22, 22	23, 23	1.2: 1	
			22, 23	22, 23		
D8S1179	11	1:	11, 12	12, 15	1: 1.2	True one
	12	1.97:	12, 15	11, 12	1.2: 1	
	15	1.19	12, 12	11, 15	1: 1.1	
			11, 15	12, 12	1.1: 1	
D21S11	30	3.34:	30, 31.2	30, 30	1: 1.17	True one
	31.2	1	30, 30	30, 31.2	1.17: 1	
D18S51	12	1:	12, 18	17, 17	1: 1.2	True one
	17	2.77:	17, 17	12, 18	1.2: 1	
	18	1.27	12, 17	17, 18	1: 1.2	
			17, 18	12, 17	1.2: 1	
D5S818	10	1.23:	11, 11	10, 10	1: 1.2	True one
	11	1	10, 10	11, 11	1.2: 1	
			10, 11	10, 11	Indeterminate	
D13S317	9	1:	12, 12	9, 12	1: 1.1	True one
	12	2.82	9, 12	12, 12	1.1: 1	
D7S820	8	1:	8, 8	11, 11	1: 1.6	Wrong
	11	1.66	11, 11	8, 8	1.6: 1	
D16S539	11	1.66:	12, 13	11, 12	1: 1.70	True one
	12	2.52:	11, 12	12, 13	1.70: 1	
TH01	7	3.06:	7, 9.3	7, 7	1: 1.03	True one
	9.3	1	7, 7	7, 9.3	1.03: 1	
TPOX	9	1.25:	10, 10	9, 9	1: 1.2	True one
	10	1	9, 9	10, 10	1.2: 1	
			9, 10	9, 10		
CSF1PO	10	1:	10, 10	11, 11	1: 2.0	True one
	11	1.97	11, 11	10, 10	2.0: 1	
			11, 11	10, 11	1: 2.0	
			10, 11	11, 11	2.0: 1	

To Date Testing of LSD in Mixture Resolution

- Currently, LSD is being beta-tested by several DNA labs around the country, and abroad
- LSD can tolerate peak imbalance up to ~20+% level.
- 35+ forensic mixture samples have been tested by us → correct resolution at all 4-allele, and 3-allele loci, except for those few where allele peak data have imbedded errors
- At 2-allele loci, obtained correct resolution result almost all of the time, using the interpretation guidelines (sometimes need the help of reference genotype)
- Limitation of LSD applies when DNA mass portions are close to 1:1; and 1:2 (with 1:2 peak height ratio also)

Good and clean allele peak data → correct LSD resolution result, always

Comparison of LSD to Linear Mixture Analysis (LMA) of Perlin and Szabady⁺

Features	LMA	LSD
Necessary input information	1) Allele peak data and 2) one of the two contributors' genotypes	Only the allele peak data
Computation approach	Iterative searching through a large number of assumed mass fractions	Direct computation in one step for the best-fit mass ratio
Loci processing	Concatenate all loci into one entity	Locus by locus, independently
Optimal mass fraction	Formulated to be the same for all loci	Allow a different one for each locus; optimality with respect to each locus

⁺M. Perlin, and B., Szabady, 'Linear Mixture Analysis: 'A Mathematical Approach to Resolving Mixed DNA Samples,' *Journal of Forensic Science*, 46-6: 1372-1378, 2001.

Outline of Comparison of LSD and LMA

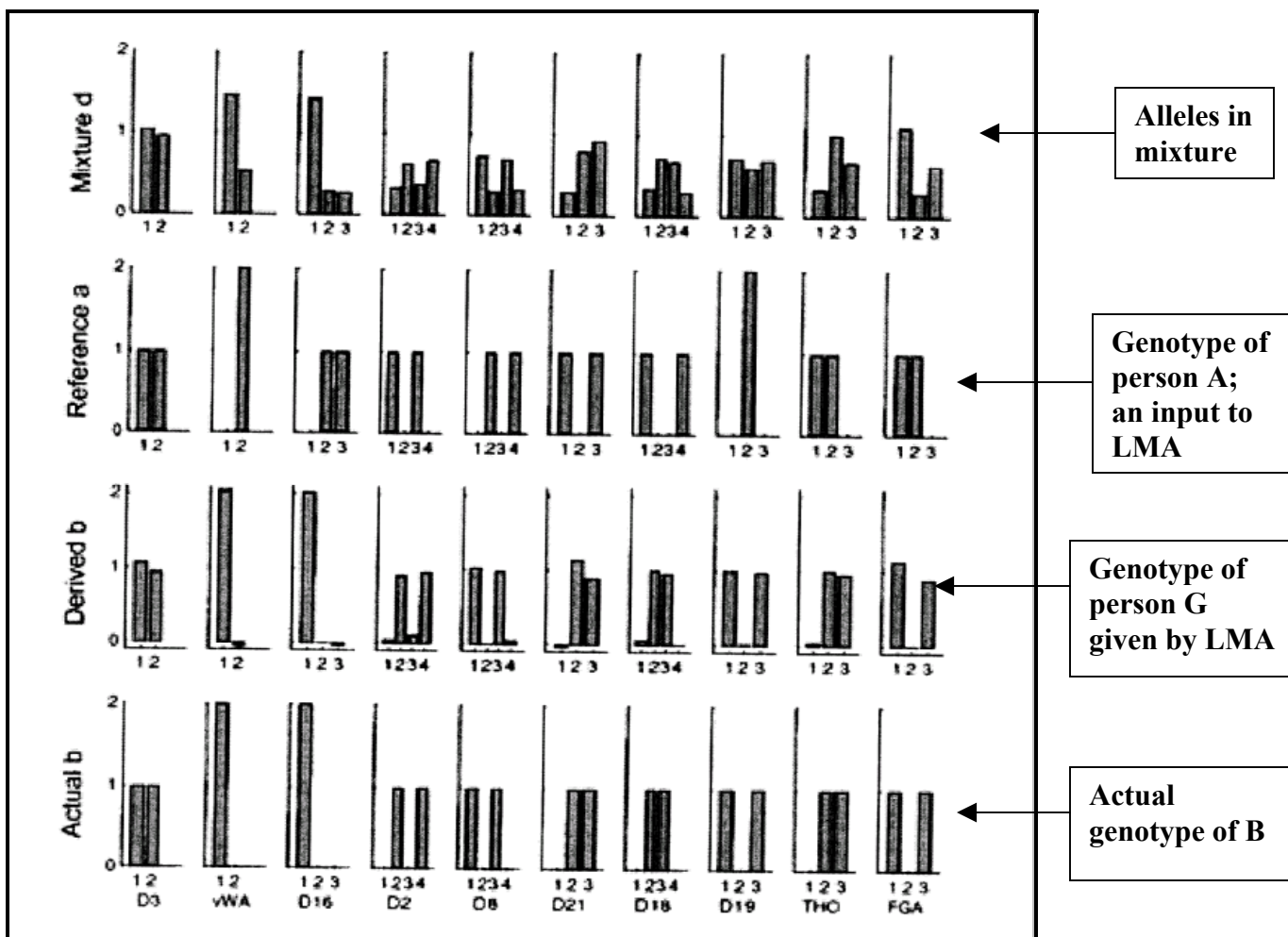
- Control run: Reproducing same results as that reported in the LMA paper using our implementation of LMA.
- Counter examples: Comparing LSD and LMA results of two sets of mixture samples
 - LSD yielded correct resolution results
 - LMA yielded incorrect resolution results at several loci in both samples

Mixture Sample Used in the Perlin and Szabady Paner

Locus	Alleles in the Mixture	Genotype Combination		Normalized Allele Peak Areas
		Person A	Person B	
D3S1358	14 15	14 15	14 15	1.04 0.96
VWA	17 18	18 18	17 17	1.48 0.52
D16S539	11 13 14	13 14	11 11	1.45 0.29 0.27
D2S1338	16 18 20 21	16 20	18 21	0.32 0.64 0.37 0.68
D8S1179	9 12 13 14	12 14	9 13	0.73 0.27 0.68 0.32
D21S11	27 29 30	27 30	29 30	0.28 0.79 0.93
D18S51	12 13 14 17	12 17	13 14	0.34 0.70 0.68 0.29
D19S433	12.2 14 15	14 14	12.2 15	0.70 0.61 0.69
TH01	6 7 9	6 7	7 9	0.32 1.01 0.67
FGA	19 24 25.2	19 24	19 25.2	1.06 0.28 0.66

Source of data: M. Perlin and B. Szabady, "Linear Mixture Analysis: A Mathematical Approach to Resolving Mixed DNA Samples," *Journal of Forensic Science*, 46-6(2001): 1372-1378.

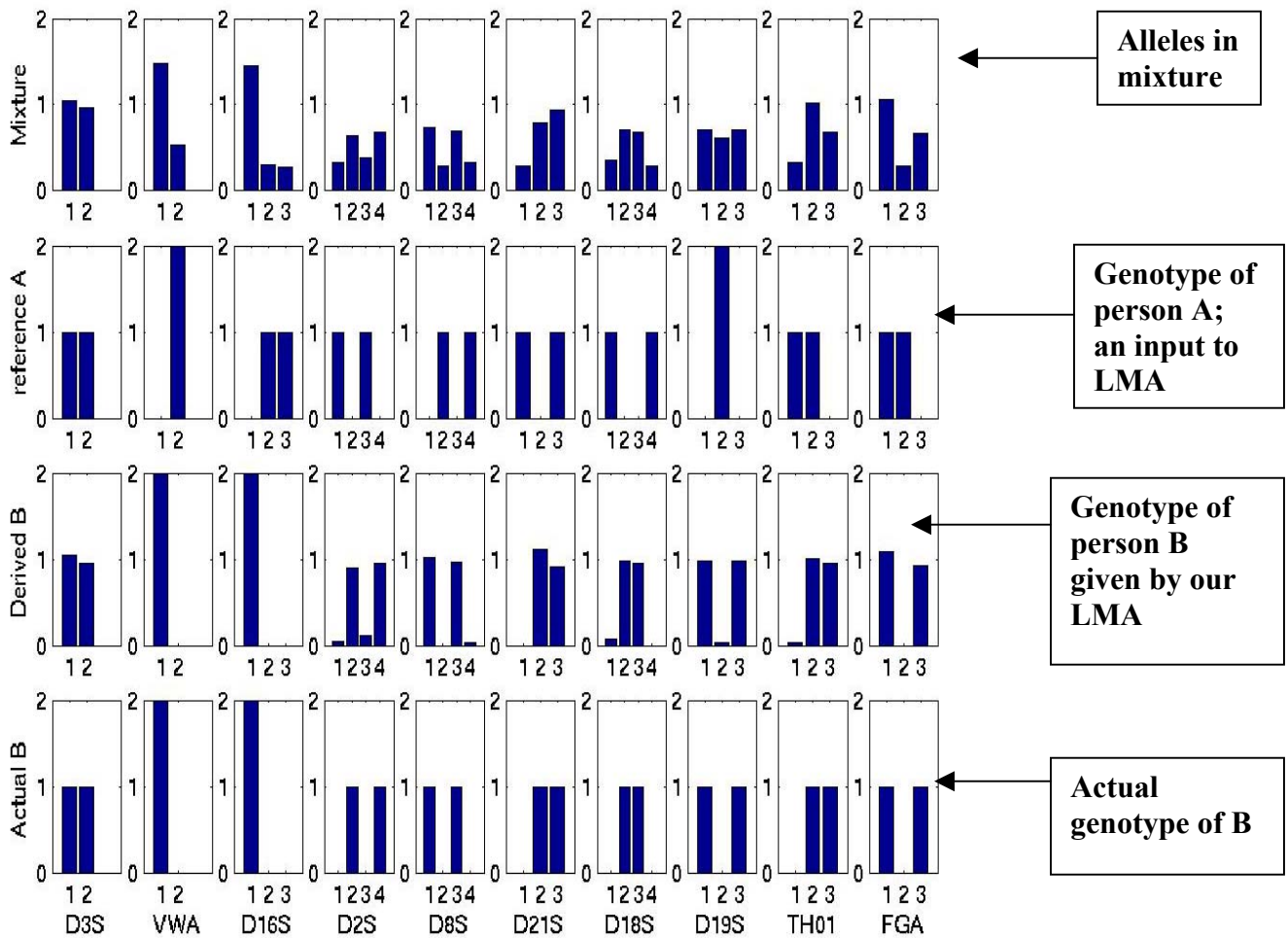
LMA Resolution Result of the Mixture Sample



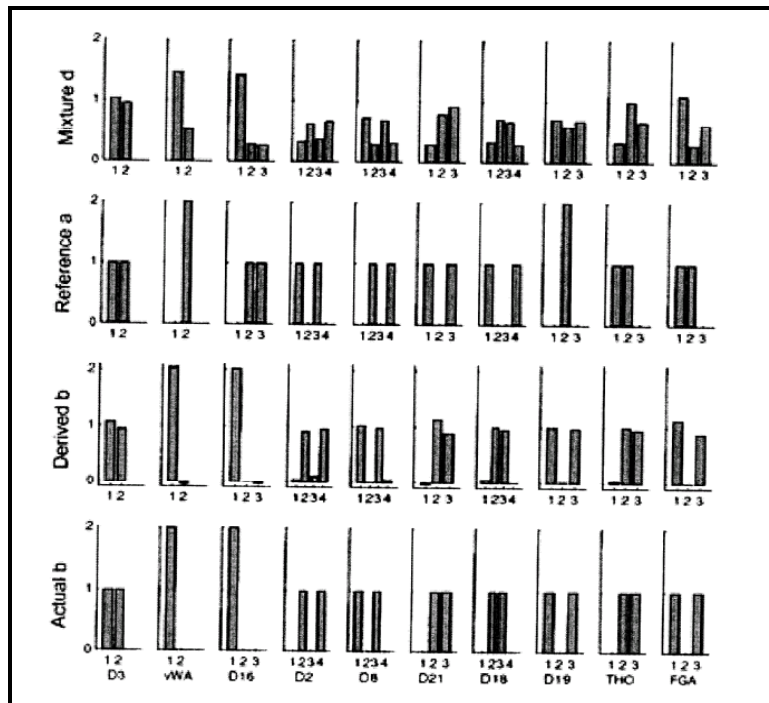
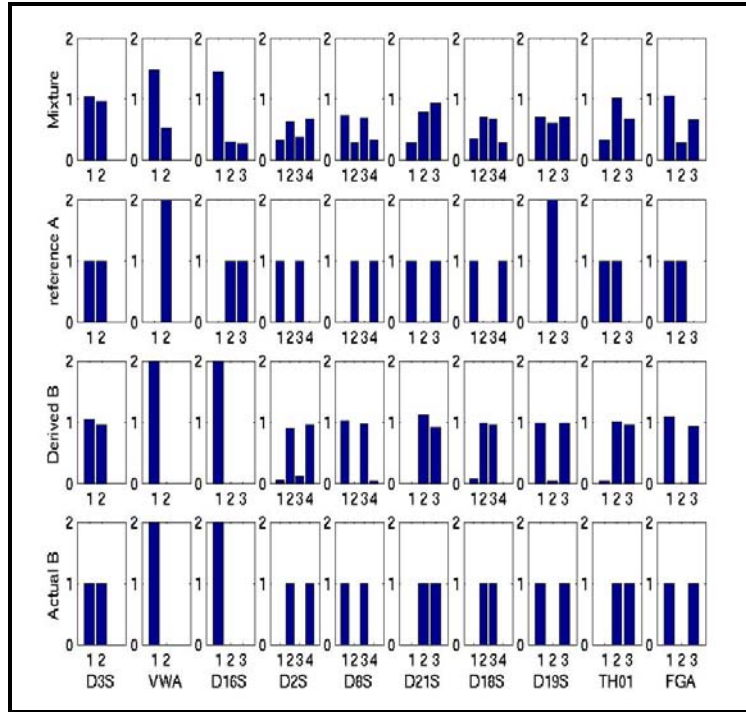
Note: Reprinted, with permission from the Journal of Forensic Sciences, Vol. 46, No. 6, copyright ASTM International.

Source: M. Perlin and B. Szabady, "Linear Mixture Analysis: A Mathematical Approach to Resolving Mixed DNA Samples," *Journal of Forensic Science*, 46-6(2001): 1372-1378.

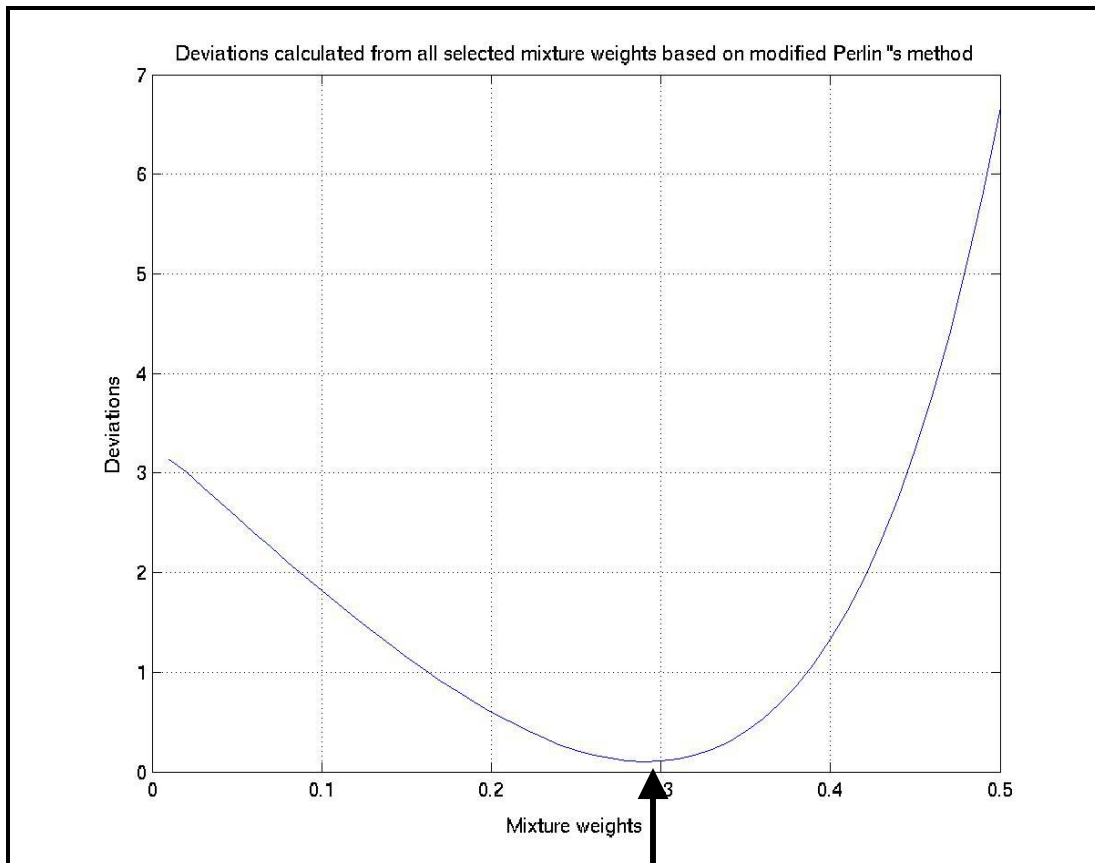
Identical Resolution Result Given by Our Implementation of LMA



At-a-Glance Comparison of Our Implementation of LMA



Our Curve of Fitting Deviation as a Function of Mass Weight in Mixture of the LMA Sample (identical plot)



Optimal mixture weight of 29%

Identical Results Achieved by Our Implementation of LMA

Locus	Alleles in the Mixture	LMA Result for Person B Presented in [18]	Genotype of Person B Yielded from Our Implementation of LMA	True Genotype of Person B				
		Vector g_B	Alleles	Vector g_B	Alleles			
D3S1358	14 15	$\begin{bmatrix} 1.0516 \\ 0.9484 \\ 2.0835 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$	14 15	$\begin{bmatrix} 1.05 \\ 0.95 \\ 2.08 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$	14 15	14 15
VWA	17 18	$\begin{bmatrix} -0.0835 \\ 2.0406 \\ -0.0041 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$	17 17	$\begin{bmatrix} -0.08 \\ 2.04 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$	17 17	17 17
D16S539	11 13 14	$\begin{bmatrix} -0.0365 \\ 0.0384 \\ 0.8951 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	11 11	$\begin{bmatrix} -0.03 \\ 0.04 \\ 0.89 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	11 11	11 11
D2S1338	16 18 20 21	$\begin{bmatrix} 0.1122 \\ 0.9543 \\ 1.0278 \\ -0.0239 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$	18 21	$\begin{bmatrix} 0.11 \\ 0.95 \\ 1.03 \\ -0.02 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$	18 21	18 21
D8S1179	9 12 13 14	$\begin{bmatrix} 0.9620 \\ 0.0341 \\ -0.0185 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$	9 13	$\begin{bmatrix} 0.96 \\ 0.04 \\ -0.02 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$	9 13	9 13
D21S11	27 29 30	$\begin{bmatrix} 1.1121 \\ 0.9064 \\ 0.0741 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$	29 30	$\begin{bmatrix} 1.11 \\ 0.91 \\ 0.08 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$	29 30	29 30
D18S51	12 13 14 17	$\begin{bmatrix} 0.9816 \\ 0.9538 \\ -0.0096 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$	13 14	$\begin{bmatrix} 0.98 \\ 0.95 \\ -0.01 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$	13 14	13 14
D19S433	12.2 14 15	$\begin{bmatrix} 0.9872 \\ 0.0361 \\ 0.9813 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$	12.2 15	$\begin{bmatrix} 0.98 \\ 0.04 \\ 0.98 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$	12.2 15	12.2 15
TH01	6 7 9	$\begin{bmatrix} 0.0366 \\ 1.0104 \\ 0.9530 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$	7 9	$\begin{bmatrix} 0.04 \\ 1.01 \\ 0.95 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$	7 9	7 9
FGA	19 24 25.2	$\begin{bmatrix} 1.0819 \\ -0.0124 \\ 0.9304 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$	19 25.2	$\begin{bmatrix} 1.08 \\ -0.01 \\ 0.93 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$	19 25.2	19 25.2

Counter Example 1: six loci data⁺

·Evetts, I., P. Gill, and J. Lambert, 'Taking Account of Peak Areas When Interpreting Mixed DNA Profiles,' *J. Forensic Science*, 43:6, pp.62-69, 1998.

Locus	Alleles in the Mixture	Genotype Combination		Allele Peak Areas
		Suspect	Victim	
TH01	8 9.3	8, 9.3	9.3, 9.3	17441 22368
D21S11	59 65 67 70	67, 70	59, 65	1226 1434 8816 8894
D18S51	13 16 17	13, 13	16, 17	38985 1914 1991
D8S1179	10 11 14	10, 14	10, 11	6416 383 5659
VWA	16 17 18 19	16, 18	17, 19	4669 931 4724 188
FGA	21 22 23	21, 22	21, 23	16099 10538 1014

LMA Resolution Results of Counter Example 1: Incorrect at four loci

Locus	Alleles in the Mixture	Normalized Allele Peak Areas	Genotype of Person A Assumed Known		LMA Suggested Result for Person B			
			A Alleles	Vector g_A	Vector g_B		LMA	Alleles
Th01	8 9.3	0.88 1.12	8 9.3	$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.14 \\ 1.86 \\ 0.86 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 2 \\ 1 \\ 1 \end{bmatrix}$	9.3 9.3	9.3 9.3
				$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.07 \\ 0.07 \\ 0.64 \\ 0.64 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$	59 65	59 65
D21S11	59 65 67 70	0.12 0.14 0.87 0.87	67 70	$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.71 \\ 0.64 \\ 0.64 \\ 1.21 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$	13 16 17	16 17
				$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.21 \\ 0.21 \\ 0.21 \\ 0.21 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$	10 11	10 11
D18S51	13 16 17	1.82 0.09 0.09	13	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0.43 \\ 0.36 \\ 0.21 \\ 1.29 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$	17 19	17 19
				$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 2.14 \\ -0.71 \\ 0.50 \end{bmatrix}$	$\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$	21 21 23	21 23
D8S1179	10 11 14	1.03 0.06 0.91	10 14	$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$				
				$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$				
VWA	16 17 18 19	0.89 0.18 0.90 0.03	16 18	$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$				
				$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$				
FGA	21 22 23	1.16 0.76 0.07	21 22	$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$				
				$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$				

} Incorrect resolution

LSD-Guided Resolution Result of Counter

Example 1: Correct at All Loci

Locus	Alleles in the Mixture	Peak Area Ratio Measured	LSD Suggested Genotype Resolution Result			
			Person 1(victim)	Person 2(suspect)	Mass Ratio	Remarks
TH01	8 9.3	1: 1.28	9.3, 9.3	8, 9.3	1: 7.08	With confidence
D21S11	59 65 67 70	1: 1.17: 7.19: 7.25	59, 65	67, 70	1: 6.66	With confidence
D18S51	13 16 17	20.37: 1: 1.04	16, 17	13, 13	1: 9.98	With confidence
D8S1179	10 11 14	16.75: 1: 1.476	10, 11	10, 14	1: 11.39	With confidence
VWA	16 17 18 19	24.84: 4.95: 25.13: 1	17, 19	16, 18	1: 8.39	With confidence
FGA	21 22 23	15.88: 10.39: 1	21, 23	21, 22	1: 4.76	With confidence

Counter Example 2: Nine loci data*

Data from the P. Neuman, The Promega Corporation, 2001

Locus	Alleles in the Mixture	Ratio of Allele Peak Data	True Genotypes	
			Person 1	Person 2
TH01	6, 9, 9.3	19.79:1:1.19	6, 6	9, 9.3
D18S51	12, 13, 15, 17	14.23:1.40:11.34:1	12, 15	13, 17
D5S818	10, 11, 12	3.08:1:2.97	10, 12	11, 11
D13S17	9, 10, 11	1:5.65:5.10	10, 11	9, 10
D7S820	9, 11, 12	54.10:1.94:1	9, 9	11, 12
CSF1PO	10, 11, 12	1:11.74:12.42	11, 12	10, 12
VWA	15, 17, 18	3.89:1:4.28	15, 18	17, 17
D8S1179	12, 13, 15	4.40:1:3.99	12, 15	13, 13
FGA	21, 24, 25	10.58:12.76:1	21, 24	24, 25

LMA-Guided Resolution Result of Counter Ex 2: Incorrect at 2 Loci (pink cells)

Locus	Alleles in the Mixture	Normalized Allele Peak Areas	Genotype of Person A Assumed Known		LMA Suggested Result for Person B		True Genotype of Person B
			Alleles	Vector g_A	Vector g_B	Alleles	
TH01	6 9 9.3	1.800: 0.091: 0.108	6	2 0 0	0.33 0.76 0.90	9 9.3	9 9.3
D18S51	12 13 15 17	1.017: 0.100: 0.811: 0.072	12 15	- 1 0 1 0	1.14 0.83 -0.58	12 13 17	13 17
D5S818	10 11 12	0.876: 0.284: 0.840	10 12	- 1 0 1	0.60 -0.03 2.37	11 11	11 11
D13S17	9 10 11	0.170: 0.962: 0.868	10 11	- 0 1 1	-0.33 1.42 0.68	9 10	9 10
D7S820	9 11 12	1.897: 0.068: 0.035	9	- 2 0 0	-0.10 1.14 0.57	9 11	11 12
CSF1PO	10 11 12	0.080: 0.933: 0.987	11 12	- 0 1 1	0.29 0.66 0.44	10 12	10 12
VWA	15 17 18	0.858: 0.220: 0.922	15 18	- 1 1 0	0.89 -0.18 1.83	17 17	17 17
D8S1179	12 13 15	0.937: 0.213: 0.850	12 15	1 - 1 0 0 1	0.35 0.48 1.78	13 13	13 13
FGA	21 24 25	0.869: 1.048: 0.082	21 24	1 - 1 1 0	-0.25 -0.09 1.40 0.68	24 25	24 25

LSD-Guided Resolution for Counter Ex. 2: Correct Resolution at all Loci

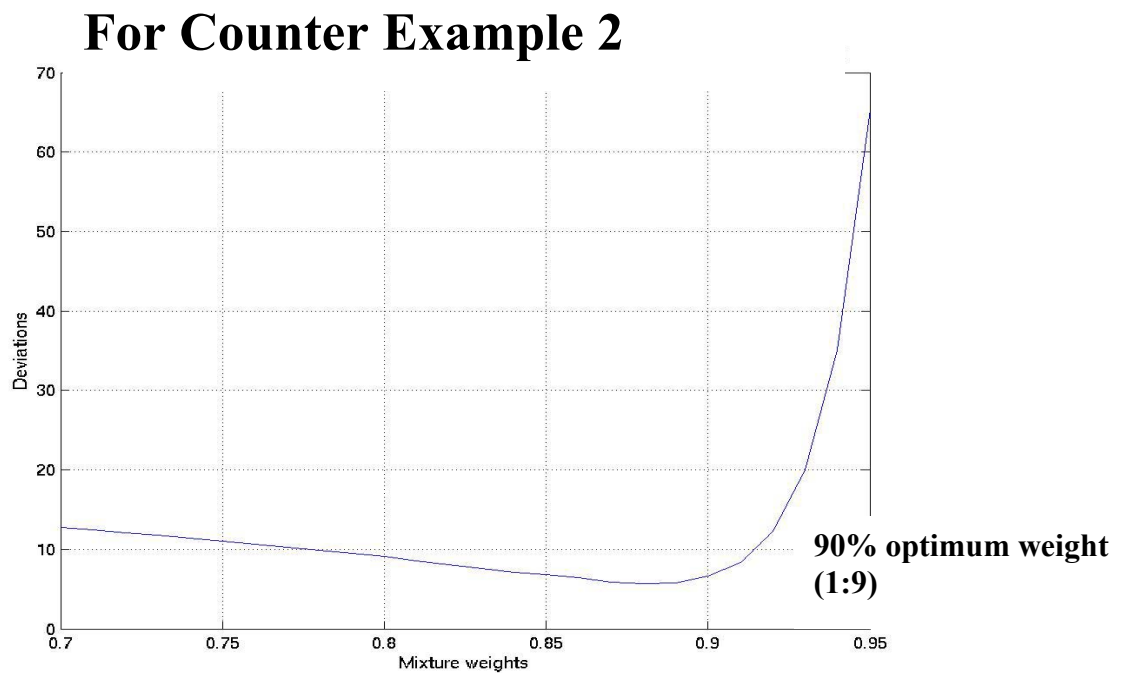
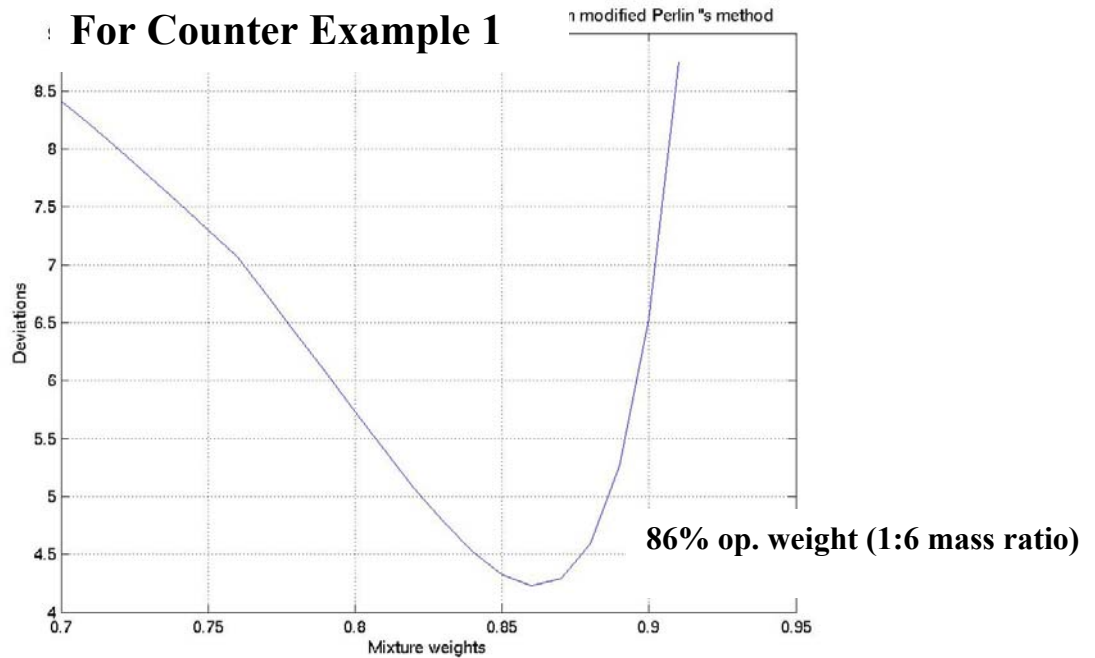
Locus	Alleles in the Mixture	LSD Suggested Genotype Resolution Result		
		Person 1	Person 2	Mass Ratio Calculated
TH01	6, 9, 9.3	6, 6	9, 9.3	9.0:1
D18S51	12, 13, 15, 17	12, 15	13, 17	10.7:1
D5S818	10, 11, 12	10, 12	11, 11	6.0:1
D13S17	9, 10, 11	10, 11	9, 10	5.8:1
D7S820	9, 11, 12	9, 9	11, 12	18.4:1
CSF1PO	10, 11, 12	11, 12	10, 12	13.0:1
VWA	15, 17, 18	15, 18	17, 17	8.1:1
D8S1179	12, 13, 15	12, 15	13, 13	8.4:1
FGA	21, 24, 25	21, 24	24, 25	7.9:1

Summary of LSD—LMA Comparison in Resolving Mixture Samples

LSD is more effective in resolving mixture samples because

- LSD operates on each locus independently of others, allowing a different optimum mass ratios to fit the given allele peak data
- LSD allows compilation of partial composite profile for the contributors, by picking and choosing those loci where resolution is confident;
- Search in DNA database can be done with a target profile consisted of the partial resolved loci, and mixed alleles at the remaining loci
- Probabilities can be calculated for matching to the partial profile

Our Fitting Deviation vs. Mixture Weight for the 2 Counter Example LMA Runs



Summary of LSD in Mixture Resolution

- For each locus, LSD uses allele peak height/area data to fit each of the possible genotype combination cases to see which fits the ‘best’ in the least square sense.
- LSD yields good results when the data is relatively consistent; but will tolerate peak imbalance up to about 20+% level.
- Provide analytical result to extend DNA analyst’s manual capability
- LSD followed by searching resolved profile in CODIS
- More beta-testing in progress.