



Least Square Deconvolution (LSD) of STR/DNA Mixtures

by

Tsewei Wang*, Ning Xue*, M. Rader, and J. D. Birdwell
Laboratory of Information Technology

*Department of Chemical Engineering & Department of Electrical and
Computer Engineering

The University of Tennessee, Knoxville, TN
twang@utk.edu

At the

**Association of Forensic DNA Analysts
and Administrators Winter Meeting**

January 17-18

Austin, Texas

Research supported by FBI Contract 98-083

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

Problem Statement

Which alleles go with which contributor?

Locus	Alleles detected	Unknown contributor profile	
		Person 1	Person 2
D3S1385	15, 16, 17	16, 17	15, 16
VWA	14, 19, 20	14, 20	14, 19
FGA	21, 24, 25	21, 21	24, 25
D8S1179	11, 12, 13, 14	11, 12	13, 14
D21S11	29, 30, 31.2	31.2, 31.2	29, 30
D18S51	15, 16, 17	15, 16	15, 17
D5S818	10, 11, 12	11, 12	10, 11
D13S317	9, 11, 12, 13	9, 12	11, 13
D7S820	8, 10, 11	8, 8	10, 11

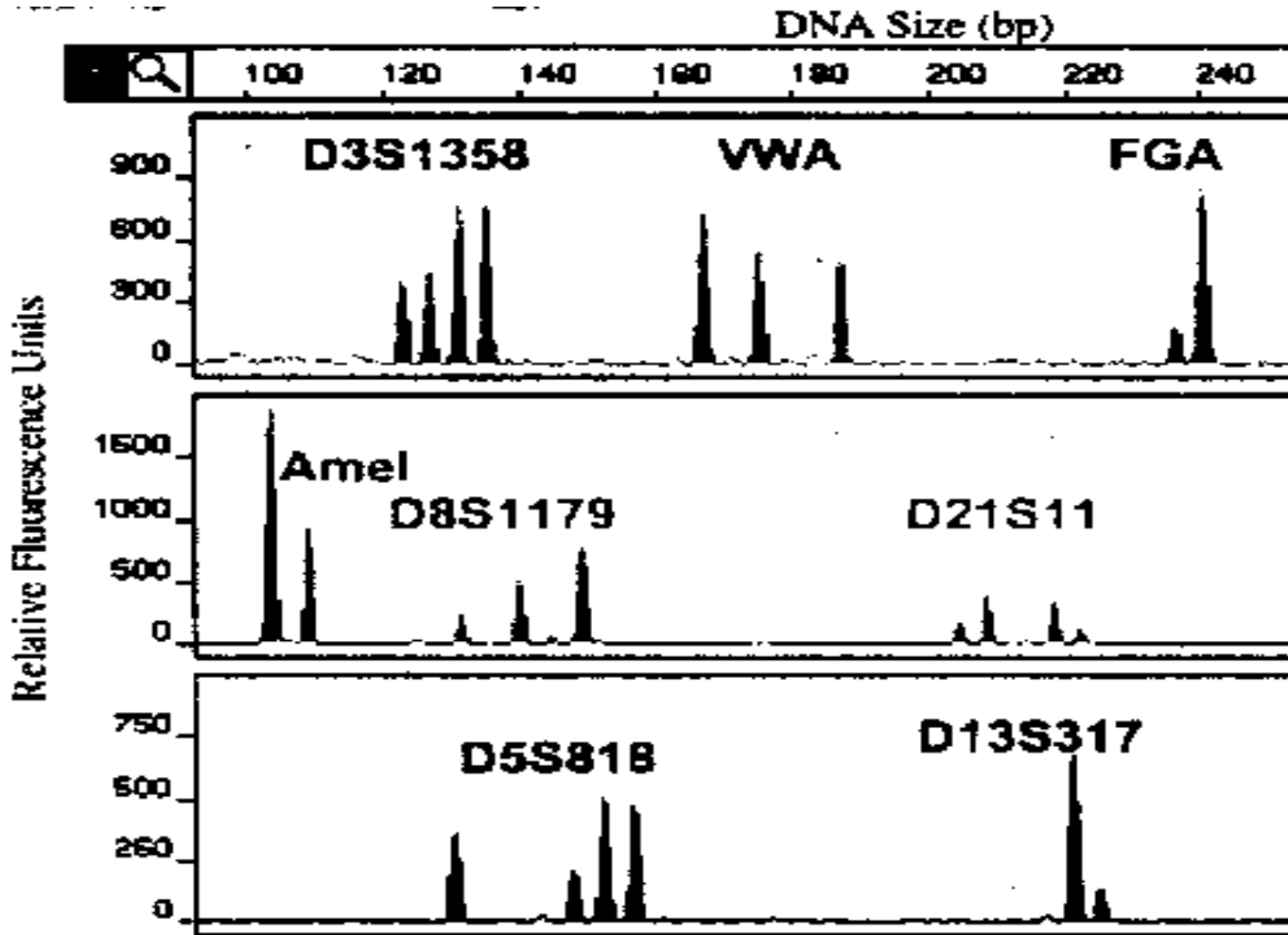


Motivation of this Work

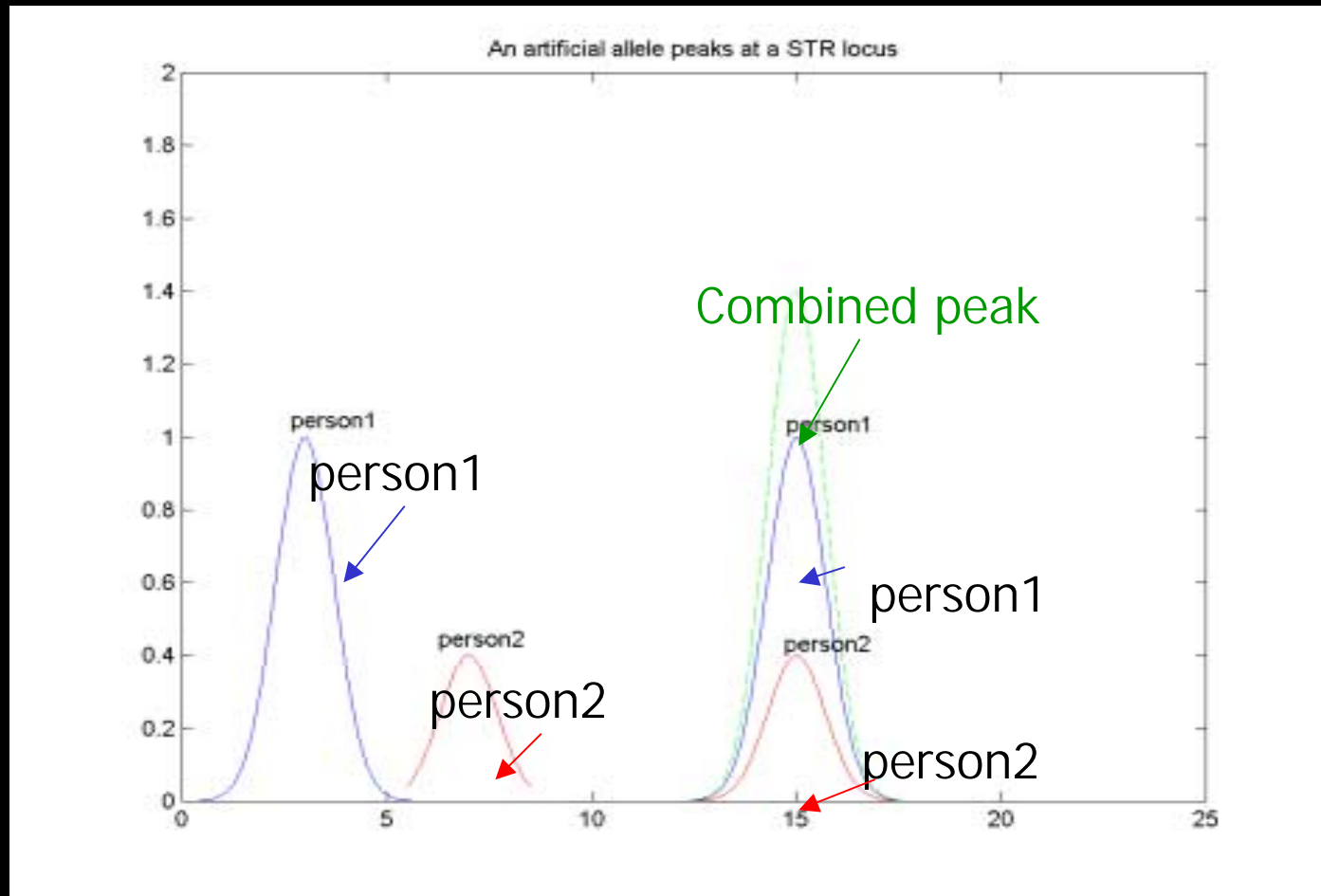
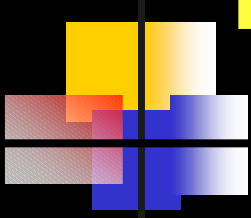
- To provide a systematic mathematical approach to resolve a mixture STR/DNA sample into its individual contribution components
- To cut down time consuming qualitative trial-and-error fitting to the peak data
- Application of LSD results depend on the quality of the original allele peak data.

Sample Allele Peaks at Seven Loci from a Mixed-Gender STR/DNA Profile

Source: M. John, Forensic DNA Typing ISBN 0-12-147951-x



A Sample 3-Allele Peaks at a STR Locus



OUTLINE

- Formulation of the LSD algorithm
- Interpreting the DNA mixture with LSD (a partial example)
- LSD analysis result for real DNA mixture data provided by Mr. Wickenheiser of the Acadiana Criminalistics Lab
- LSD analysis result for real DNA mixture data from Corpus Christi, Texas, provided by Ms. Graham
- Limitation of LSD and interpretation guidelines for DNA analysts
- Summary
- Two numerical Examples: a 3-allele and a 4-allele case
- LSD software under development
- LSD software demonstration



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

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

Set up A and b:

$$A x = b$$

■ where

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

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

Possible Genotype Combinations for the 4-Allele Case

Matrix A
Representation

Pseudoinverse
Of the Matrix

<u>Case</u>	<u>Person1</u>	<u>Person2</u>	Matrix A Representation	Pseudoinverse Of the Matrix
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)

Case_	Person1	Person2	Matrix A representation	A ⁺ Pseudoinverse Of the Matrix
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	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$	$\frac{1}{3} * \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}$

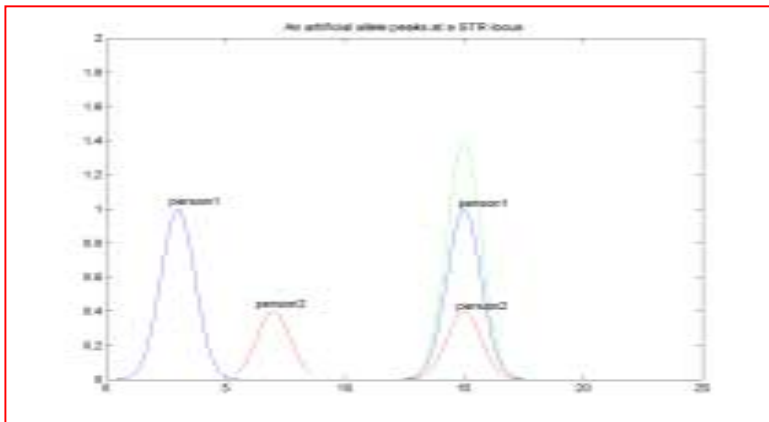
Possible Genotype Combinations for the 2-allele Case

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

$$\begin{array}{l}
 \text{allele1} \longrightarrow \\
 \text{allele2} \longrightarrow \\
 \text{allele3} \longrightarrow
 \end{array}
 \begin{bmatrix}
 1 & 0 \\
 0 & 1 \\
 1 & 1
 \end{bmatrix}
 \begin{bmatrix}
 \text{mass1} \\
 \text{mass2}
 \end{bmatrix}
 = \text{mass1} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \text{mass2} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \alpha \begin{bmatrix} \text{allele1area} \\ \text{allele2area} \\ \text{allele3area} \end{bmatrix}$$



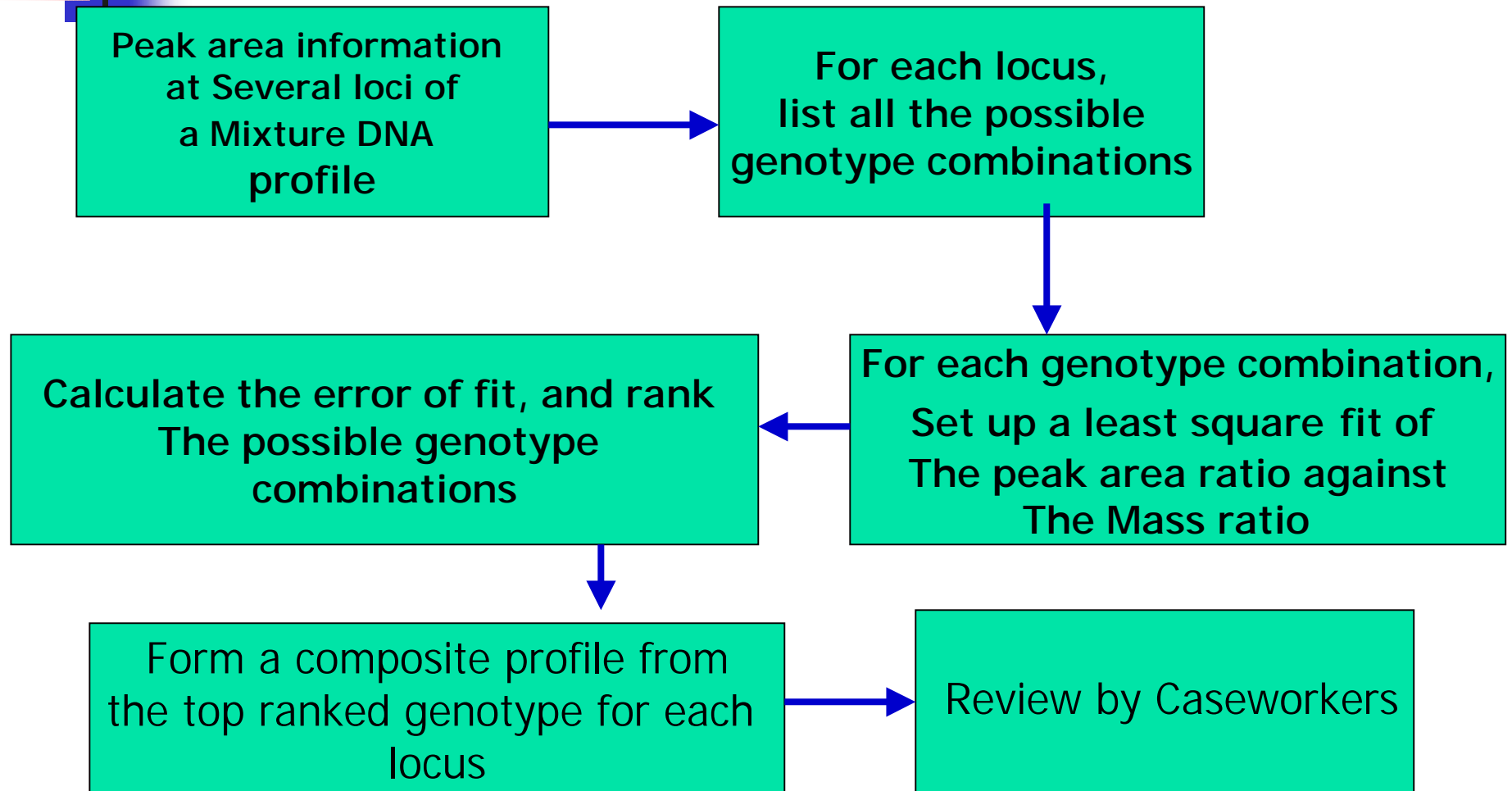
To Compute the Least Square Solution for the Example

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

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

- The x vector contains the mass ratio coefficients

Least Square Deconvolution Approaches



Interpreting the DNA Mixture with LSD

Example of a Mixture STR/DNA at 3 Loci

Source: Taking account of peak areas when interpreting mixed DNA profiles.
Journal of Forensic Science 1998; 43(1): 62-69; Evett IW, Gill P, Lambert JA

Locus	Alleles detected	Peak Areas
THO1	8	17441
	9.3	22368
D21	59	1226 1
	65	1434 1.17
	67	8816 7.19
	70	8894 7.25
D18	13	38985
	16	1914
	17	1991

Least Square Fitting of the Allele Peak Areas for D21S11 with Alleles {59, 65, 67, 70} : Setting up the Allele Matrix

Person1	Person2	Combinations	Matrix A	b vector: peak areas
59, 65	67, 70	[1 1 0 0]	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	1.0 1.2 7.2 7.3
67, 70	59, 65	[0 0 1 1]		
59, 67	65, 70	[1 0 1 0]	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$	
65, 70	59, 67	[0 1 0 1]		
59, 70	65, 67	[1 0 0 1]	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$	
65, 67	59, 70	[0 1 1 0]		

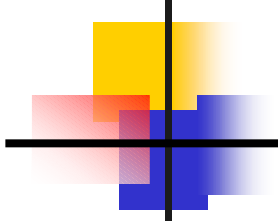
Least Square Analysis Results of Locus D21S11 with Alleles {59, 65, 67, 70}

Known Mass Ratio: 1:10

Ratio of peak areas	Rank	Person1	Person2	Combination	Error	Least-square coefficients	Ratio of peak areas predicted by XIs	Ratio of mass
Correct genotype	1	59, 65	67, 70	[1 1 0 0] [0 0 1 1]	0.016	1.0: 7.2	1.0: 1.0: 7.2: 7.2	1: 7.2
		67, 70	59, 65					
1: 1.2: 7.2: 7.3	2	59, 67	65, 70	[1 0 1 0] [0 1 0 1]	37.6	4.1: 4.2	4.1: 4.2: 4.1: 4.2	1: 1.0
		65, 70	59, 67					
	3	59, 70	65, 67	[1 0 0 1] [0 1 1 0]	37.6	4.1: 4.2	4.1: 4.2: 4.2: 4.1	1: 1.1
		65, 67	59, 70					

Composite Least Square Analysis Results for all Three Loci, with Known Mass Ratio of 1:10

Locus	Person1	Person2	Predicted Mass ratio in the mixture	Calculated peak area ratios (normalized)
TH01	9.3, 9.3	8, 9.3	1:7.0	1: 1.3
D21S11	59, 65	67, 70	1:6.6	1: 1: 6.6: 6.6
D18S51	16, 17	13, 13	1:9.9	19.9: 1: 1



LSD Analysis Result for Real DNA Mixture
Data provided by The Acadiana
Criminalistics Laboratory, Mr. Ray
Wickenheiser, Director

Real DNA Mixture Peak Area Data LA2648/01(#6)

Approximate mass ratio is 1:7

Locus	Allele	Peak Height	Locus	Allele	Peak Height
D3S1358	14, 16, 17, 19	616, 688, 3592, 3479	D13S317	11, 12, 13, 14	309, 1892, 366, 1734
VWA	14, 16	3483, 3534	D7S820	8, 10, 12	1685, 1409, 144
FGA	22	5503	D16S539	11, 12, 13, 14	3047, 460, 562, 2329
D8S1179	13, 14	3111, 3175	THO1	9, 9.3	2276, 2666
D21S11	28, 29, 30	3166, 489, 2964	TPOX	8, 11	2556, 2447
D5S818	11, 12, 13	538, 2448, 2240	CSF1PO	10, 11, 12	1693, 339, 2164
D18S51	13, 14, 15, 18	2383, 380, 2041, 271			

Analysis Result of Mixture Data LA2648/01(#6) Using LSD (all correct)

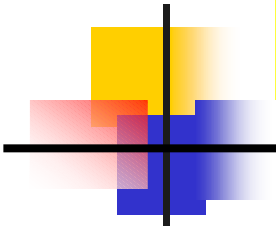
Locus	Alleles Detected	Analysis Result		
		Person1	Person2	Remarks
D3S1385	14, 16, 17, 19	14, 16	17, 19	With confidence (All correct)
FGA	22	22, 22	22, 22	
D18S51	13, 14, 15, 18	14, 18	13, 15	
D13S317	11, 12, 13, 14	11, 13	12, 14	
D7S820	8, 10, 12	8, 12	8, 10	
D16S539	11, 12, 13, 14	12, 13	11, 14	
CSF1PO	10, 11, 12	11, 12	10, 12	

(Continued table in next page)

LSD results of LA2648/01 (#6) data

cont'd

Locus	Alleles detected	Analysis Result		
		Person1	Person2	Remarks
TPOX	8, 11	8, 11	8, 11	High confidence.
		11, 11	8, 8	Low confidence
VWA	14, 16	14, 16	14, 16	High confidence.
		14, 14	16, 16	Low confidence
D8S1179	13, 14	13, 14	13, 14	High confidence.
		13, 13	14, 14	20% confidence
D5S818	11, 12, 13	11, 11	12, 13	Equal confidence
		11, 12	12, 13	Equal confidence
THO1	9, 9.3	9, 9	9.3, 9.3	Equal confidence
		9, 9.3	9, 9.3	Equal confidence
D21S11	28, 29, 30	29, 29	28, 30	Equal confidence
		28, 29	28, 30	Equal confidence



Comment of the Quality of LA2648/01 (#6) Data and Validity of LSD Results

- This set of data is the best one among all the data obtained from Acadiana Crime Lab. All 13 loci can be ranked correctly and the approximate mass ratio is 1:7.04.
- Most of the peak areas measured are less than 4000, less than the saturation threshold.

Real DNA Mixture Peak Area Data LA2649/01 (#5)

*All peak data are very small (low 100's) 6 loci have only 2 alleles; 2 outliers;
approximate mass ratio is 1:2




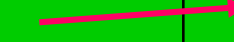


Locus	Allele	Peak Height	Locus	Allele	Peak Height
D3S135 8	15, 17	1021, 604	D13S31 7	8, 12	105, 3507
VWA	15, 17	545, 340	D7S820	8, 10	162, 212
FGA	19.2, 22, 24	324, 121, 390	D16S53 9	11, 12	535, 480
D8S117 9	13, 14	835, 145	Tho1	6, 7, 9.3	402, 156, 237
D21S11	29, 30, 35	151, 379, 376	TPOX	6, 8, 11	784, 131, 152
D18S51	10.2, 12, 15	186, 150, 83	CSF1PO	10, 11, 12	4754, 493, 4834
D5S818	10, 11, 12, 13	86, 140, 307, 335			

Analysis Result of Mixture Data LA2649/01 (#5) Using LSD (victim's known)

*All peak data are very small (low 100's) 6 loci have only 2 alleles; 2 outlier loci relative peak data
Fitted mass ratio is close to 1:2

Locus	Allele Detected	LSD Analysis Result		
		Person 1	Person 2	Remarks
FGA	19.2, 22, 24	22, 24	19.2, 24	High confidence (All correct)
D8S1179	13, 14	13, 14	13, 13	
D21S11	29, 30, 35	29, 29	30, 35	
D5S818	10, 11, 12, 13	10, 11	12, 13	
TH01	6, 7, 9.3	6, 7	6, 9.3	
TPOX	6, 8, 11	8, 11	6, 6	
CSF1PO	10, 11, 12	11, 11	10, 12	

Correct choice,  (LA2649/01 #5Data)

Locus	Allele detected	LSD Analysis Result		
		Person 1	Person 2	Remarks
D3S138 5	15, 17	17, 17	15, 15	Equal confidence
		 15, 15	15, 17	Equal confidence
VWA	15, 17	17, 17	15, 15	Equal confidence
		 15, 15	15, 17	Equal confidence
D18S51	10.2, 12, 15	 10.2, 15	10.2, 12	Equal confidence
		15, 15	10.2, 12	Equal confidence
D7S820	8, 10	8, 8	10, 10	Equal confidence
		 8, 10	8, 10	Equal confidence
D16S53 9	11, 12	12, 12	11, 11	Equal confidence
		 11, 12	11, 12	Equal confidence
D13S31 7	8, 12	12, 12	12, 12	Very high confidence(if allele 8 is noise)
		 8, 12	12, 12	little confidence
		8, 8	12, 12	little confidence



Comment of the Quality of Data LA2649/01 (#5) and Validity of LSD Results

- Correct genotype combinations were obtained from the LSD results at all 13 loci.
- Some peak areas reported are too small (less than 100; are these too small to be indicative of true proportionality?)
- Six loci have two alleles each; two are clear outlier (D13S317) and CSF1PO which lead to erroneous LSD ranking
- The mass ratio is close to 1:2, and LSD limitation rules applies

Real DNA Mixture Peak Area Data LA 3345/01(#2)

Overloading of DNA; saturated peaks (>4000);

Locus	Allele	Peak Height	Locus	Allele	Peak Height
D3S1358	14, 15, 17	5786, 1426, 5578	D13S317	8, 11, 12	5227, 1136, 5208
VWA	15, 16, 19	5899, 5890, 1169	D7S820	10, 12	5913, 385
FGA	22, 24, 25	4636, 4765, 866	D16S539	8, 10, 11	7082, 6995, 1300
D5S818	11, 12	5354, 1153	Tho1	6, 7, 9.3	6388, 6370, 1287
D21S11	28, 30.2, 32.2	6479, 1239, 5717	TPOX	8, 9, 12	1393, 7149, 1219
D18S51	16, 17, 20	1323, 3076, 2748	CSF1PO	8, 10, 11	729, 4251, 4599
D8S1179	11, 13, 14, 15	6069, 1959, 4769, 1425			

LSD Analysis Result of Mixture Data LA 3345/01(#2) (7 correct)

Overloading of DNA; saturated peaks (>4000)

—————> correct profiles;

Locus	Allele detected	LSD Analysis Result		
		Person 1	Person 2	Remarks
VWA	15, 16, 19	19, 19	15, 16	With confidence (but only 3 are correct)
FGA	22, 24, 25	25, 25	22, 24	
D8S1179	11, 13, 14, 15	13, 15	11, 14	
D5S818	11, 12	12, 12	11, 11	
D13S317	8, 11, 12	11, 11	8, 12	
D16S539	8, 10, 11	11, 11	8, 10	
THO1	6, 7, 9.3	9.3, 9.3	6, 7	
TPOX	8, 9, 12	8, 12	9, 9	

LSD Results of LA 3345/01(#2) Data

Locus	Allele detected	LSD Analysis Result		
		Person 1	Person 2	Remarks
D3S138 5	14, 15, 17 (15,17) & (14, 17)	15, 15	14, 17	High confidence
		14, 15	14, 17	Low confidence
D18S51	16, 17, 20	16, 16	17, 20	High confidence
		16, 17	17, 20	Low confidence
D21S11	28, 30.2, 32.2	30.2, 30.2	28, 32.2	Equal confidence
		28, 30.2	28, 32.2	Equal confidence
D7S820	10, 12	12, 12	10, 10	Equal confidence
		10, 12	10, 10	Equal confidence
CSF1PO	8, 10, 11	8, 11	10, 11	Equal confidence
		8, 8	10, 11	Equal confidence



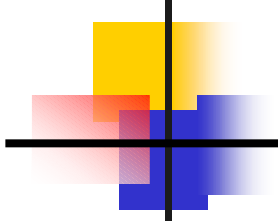
Comment of the Quality of LA 3345/01 (#2) Data and Validity of the LSD Results

- This set of data is the least consistent one among all 5 data sets. Only 7 loci were able to be correctly ranked using LSD.
- Many peak areas measured at this loci are larger than 4000 units. Because of the peak area saturation, the peak areas are not consistently proportional to the mass of the corresponding alleles.
- LSD results are questionable.



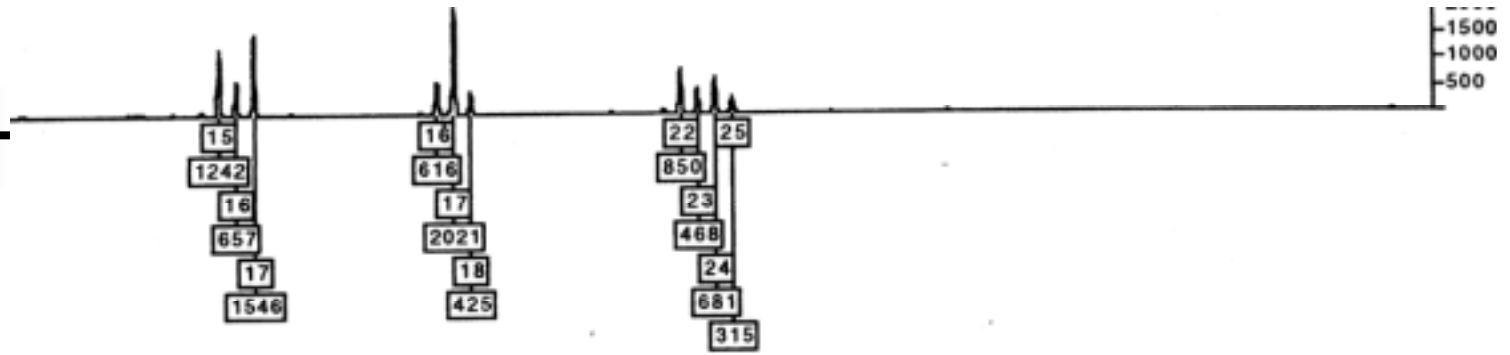
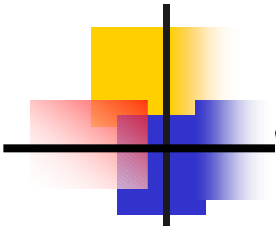
Characteristics of 'Good' Allele Peak Data for LSD Application

- Mass ratio different from 1:1 for all loci
- Mass ratio different from 1:2 for 2-allele loci
- No overloading of DNA; allele peaks not saturated
- High degree of consistent mass ratio across loci

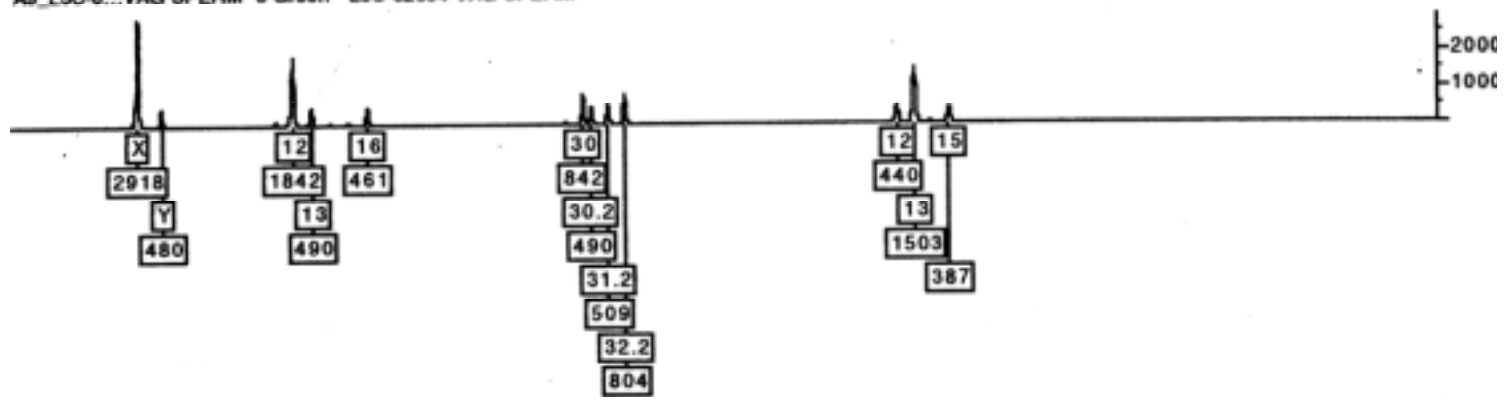


**LSD Analysis Result for Real DNA
Mixture Data from Corpus Christi,
Texas, provided by Ms. Graham**

Real DNA Mixture Peak Height Data (Texas 1:62334)



A9_L3C-6...VAG SPERM 6 Green L3C-62334 VAG SPERM



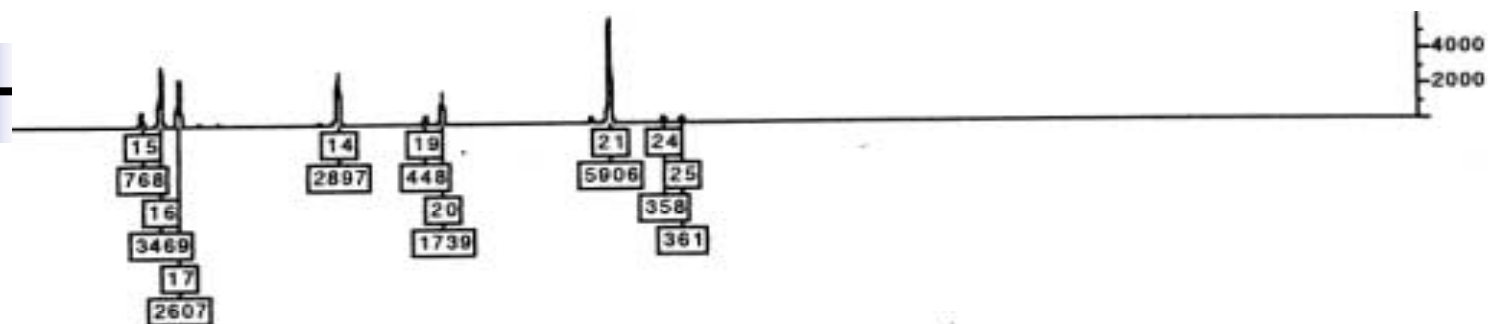
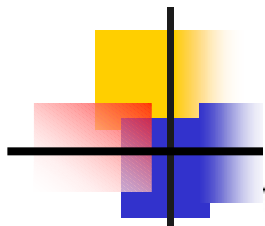
A9_L3C-6...VAG SPERM 6 Yellow L3C-62334 VAG SPERM



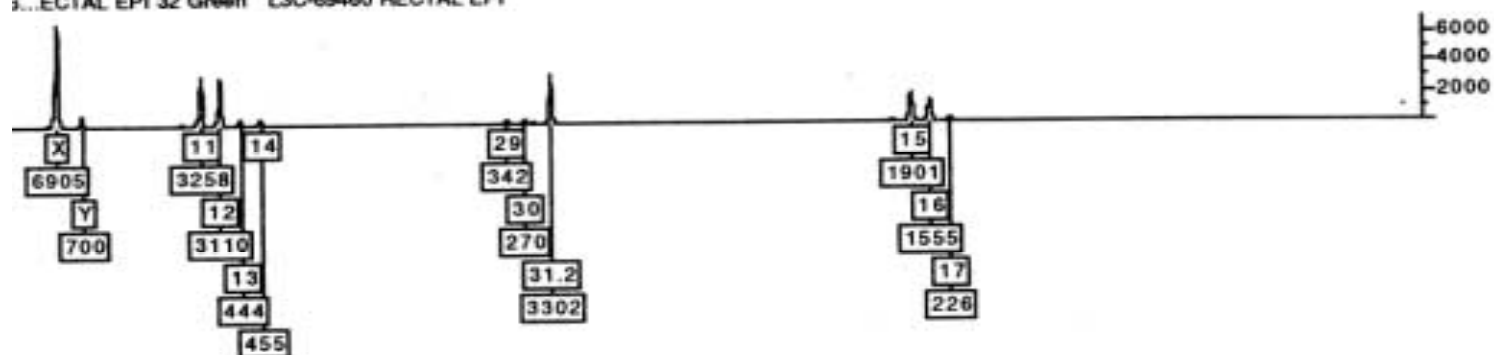
LSD Analysis Result of Real DNA Mixture Using Peak Height Data (Texas 1:62334); (all correct)

Locus	Alleles detected	Analysis Result			True Genotype	
		Mass ratio Calculated	Person1	Person2	Victim	Suspect
D3S1385	15, 16, 17	1: 2.08	16, 17	15, 17	15, 17	16, 17
		1: 4.24	16, 16	15, 17 (maybe)		
VWA	16, 17, 18	1: 1.94	16, 18	17, 17	17, 17	16, 18
FGA	22, 23, 24, 25	1: 1.96	23, 25	22, 24	22, 24	23, 25
D8S1179	12, 13, 16	1: 1.94	13, 16	12, 12	12, 12	13, 16
D21S11	30, 30.2, 31.2, 32.2	1: 1.65	30.2, 31.2	30, 32.2	30, 32.2	30.2, 31.2
D18S51	12, 13, 15	1: 1.82	12, 15	13, 13	13, 13	12, 15
D5S818	7, 12, 13	1: 1.89	7, 12	13, 13	13, 13	7, 12
D13S317	8, 9, 11	1: 2.40	9, 11	8, 9	8, 9	9, 11
D7S820	10, 11	1: 1.90	10, 10	11, 11 (maybe)	11, 11	10, 10
		1: 2.21	11, 11	10, 11 (maybe)		

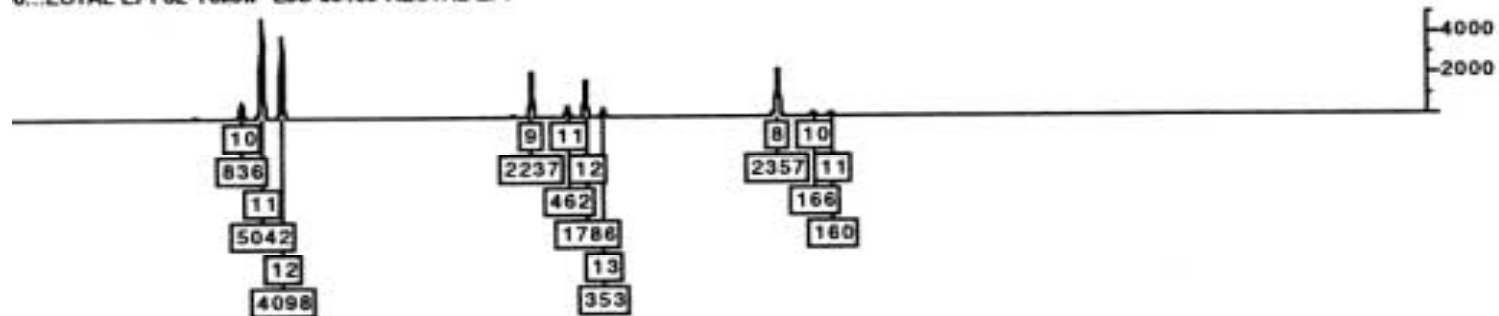
Real DNA Mixture Peak Height Data (Texas 2:69460)



3...ECTAL EPI 32 Green L3C-69460 RECTAL EPI



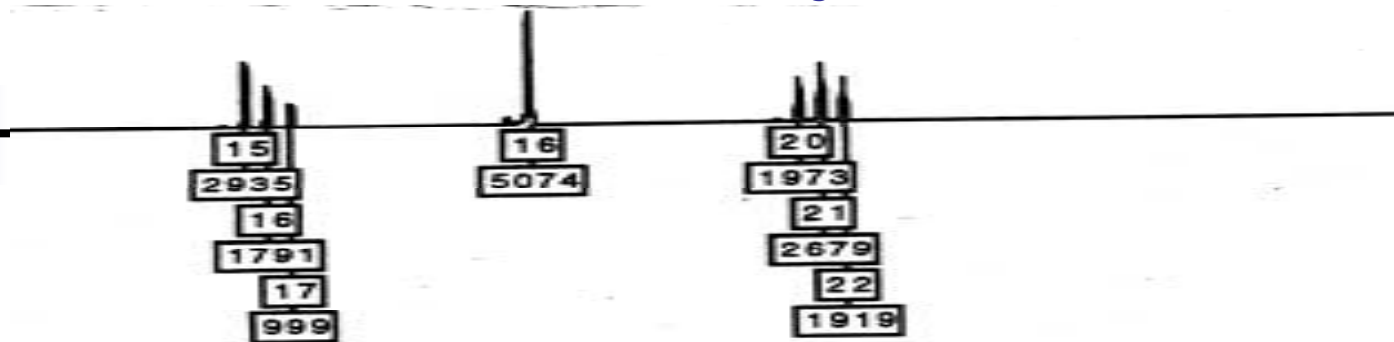
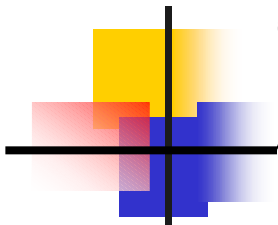
6...ECTAL EPI 32 Yellow L3C-69460 RECTAL EPI



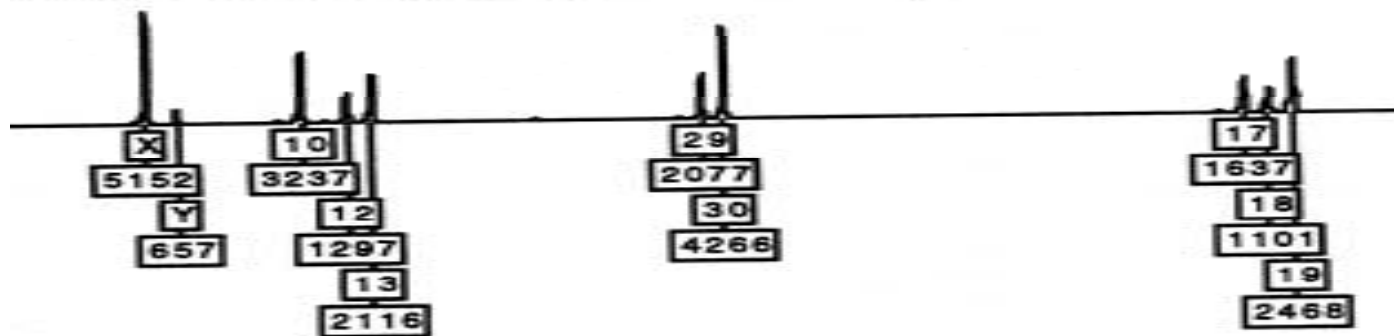
LSD Analysis Result of Real DNA Mixture Peak Height Data (Texas2:69460); (all correct)

Locus	Alleles detected	Analysis Result			True Genotype	
		Mass ratio Calculated	Person1	Person2	Victim	Suspect
D3S1385	15, 16, 17	1: 3.30	15, 16	16, 17	16, 17	15, 16
VWA	14,19, 20	1: 2.89	14, 19	14, 20	14, 20	14, 19
FGA	21, 24, 25	1: 8.21	24, 25	21, 21	21, 21	24, 25
D8S1179	11, 12, 13, 14	1: 7.08	13, 14	11, 12	11, 12	13, 14
D21S11	29, 30, 31.2	1: 5.40	29, 30	31.2, 31.2	31.2, 31.2	29, 30
D18S51	15, 16, 17	1: 6.00	15, 17	15, 16	15, 16	15, 17
D5S818	10, 11, 12	1: 4.74	10, 11	11, 12	11, 12	10, 11
D13S317	9, 11, 12, 13	1: 4.94	11, 13	9, 12	9, 12	11, 13
D7S820	8, 10, 11	1: 7.23	10, 11	8, 8	8, 8	10, 11

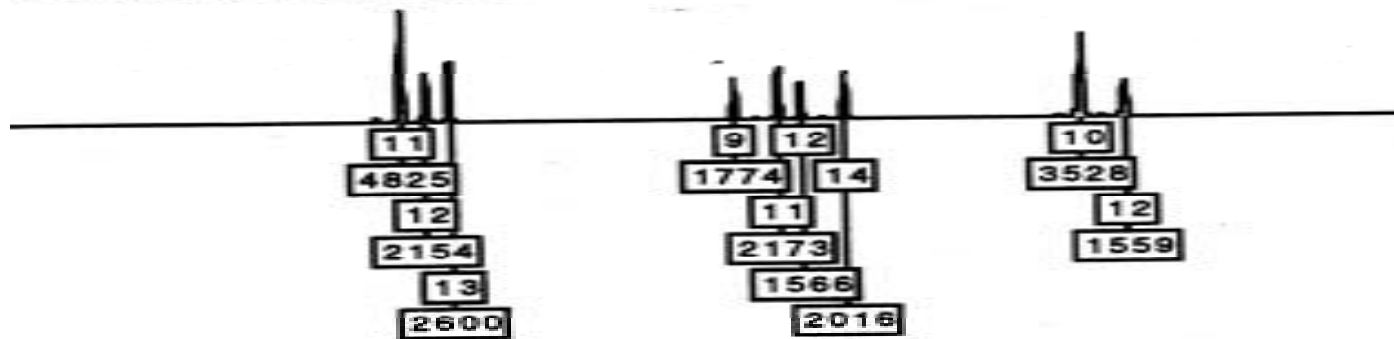
Real DNA Mixture Peak Height Data (Texas 3:69017)



.3C-6901...S (B) EPI 28 Green L3C-69017 PANTIES (B) EPI



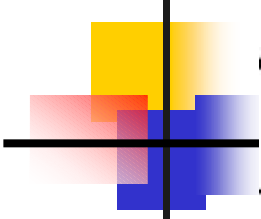
L3C-6901...S (B) EPI 28 Yellow L3C-69017 PANTIES (B) EPI



LSD Analysis Result of Real DNA Mixture Peak Height Data (all correct) (Texas 3:69017)

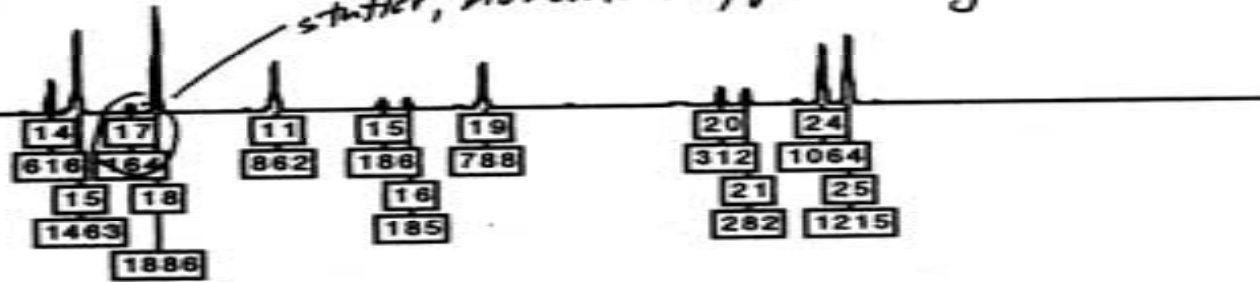
Locus	Alleles detected	Analysis Result			True Genotype	
		Mass ratio Calculated	Person1	Person2	Victim	Suspect
D3S1385	15, 16, 17	1: 1.76	15, 17	15, 16	15, 16	15, 17
VWA	16, 16	-----	16, 16	16, 16	16, 16	16, 16
FGA	20, 21, 22	1: 1.45	21, 21	20, 22	20, 22	21, 21
D8S1179	10, 12, 13	1: 1.66	10, 12	10, 13	10, 13	10, 12
D21S11	29, 30	1: 1.89	30, 30	29, 30	29, 30	30, 30
		1: 2.05	29, 29	30, 30 (maybe)		
D18S51	17, 18, 19	1: 1.53	18, 19	17, 19	17, 19	18, 19
D5S818	11, 12, 13	1: 1.20	11, 12	11, 13	11, 13	11, 12
D13S317	9, 11, 12, 14	1: 1.25	9, 12	11, 14	11, 14	9, 12
D7S820	10, 12	1: 1.58	10, 10	10, 12	10, 12	10, 10
		1: 2.26	12, 12	10, 10 (maybe)		

Real DNA Mixture Peak Height Data (Texas 4:68268)

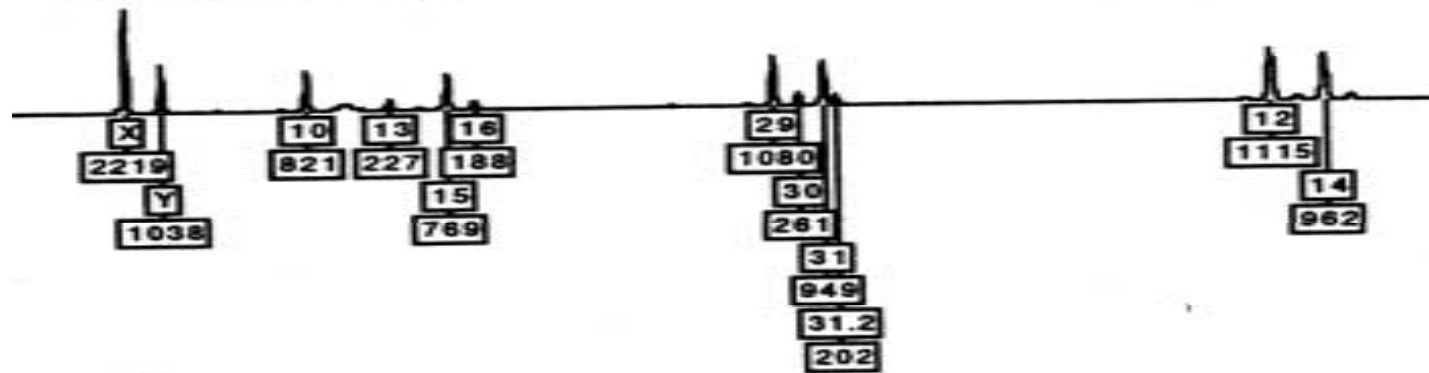


blue L3C-68268 PANTIES (E)

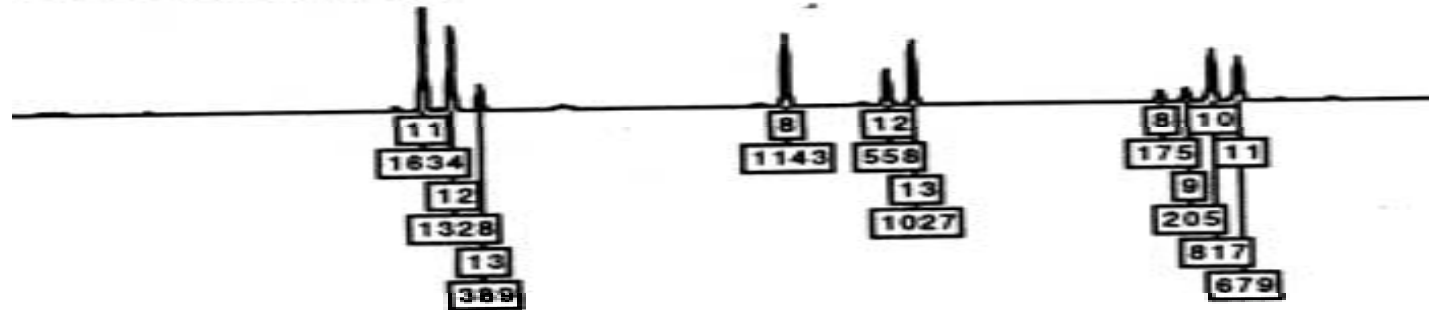
stutter, not called w/ profiles only



green L3C-68268 PANTIES (E)

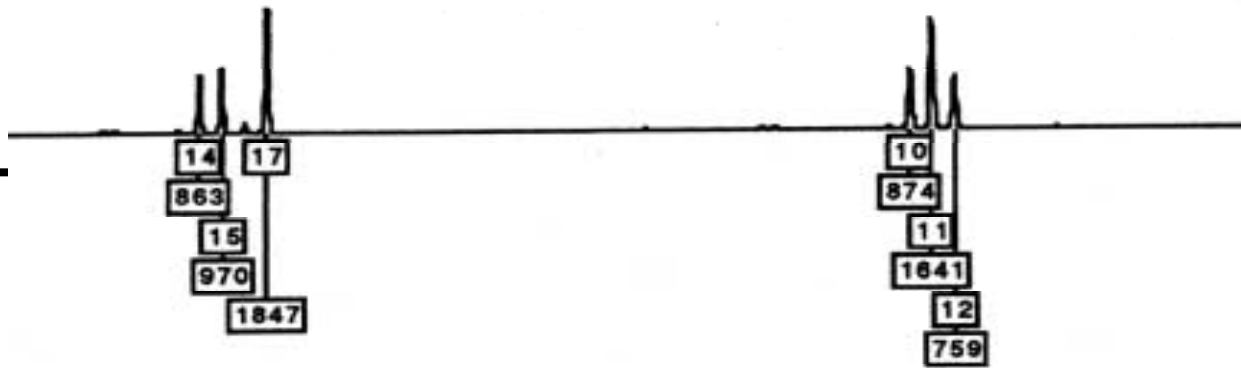
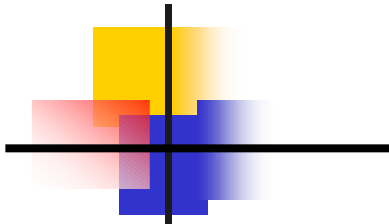


yellow L3C-68268 PANTIES (E)

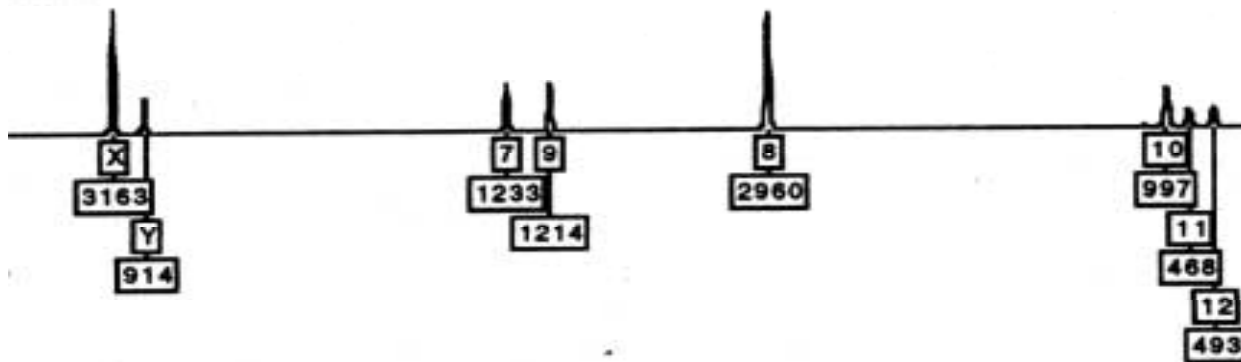


LSD Analysis Result of Real DNA Mixture Peak Area Data (all correct) (Texas 4 :68268)

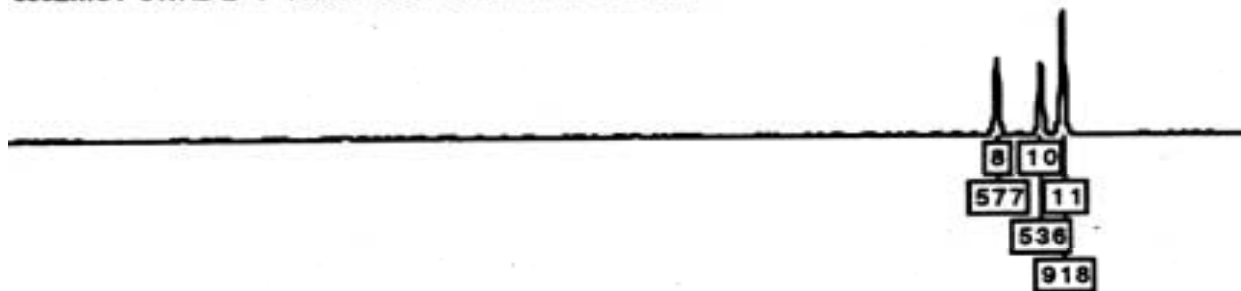
Locus	Alleles detected	Analysis Result			True Genotype	
		Mass ratio Calculated	Person1	Person2	Victim	Suspect
D3S1385	14, 15, 18	1: 2.54	14, 18	15, 18	14, 18	15, 18
VWA	11, 15, 16, 19	1: 4.45	15, 16	11, 19	15, 16	11, 19
FGA	20, 21, 24, 25	1: 3.84	20, 21	24, 25	20, 21	24, 25
D8S1179	10, 13, 15, 16	1: 3.83	13, 16	10, 15	13, 16	10, 15
D21S11	29, 30, 31, 31.2	1: 4.38	30, 31.2	29, 31	30, 31.2	29, 31
D18S51	12, 14	1: 1	12, 14	12, 14	12, 14	12, 14
D5S818	11, 12, 13	1: 3.60	11, 13	11, 12	11, 13	11, 12
D13S317	8, 12, 13	1: 3.89	12, 12	8, 13	12, 12	8, 13
D7S820	8, 9, 10, 11	1: 3.94	8, 9	10, 11	8, 9	10, 11



6852...ST SWAB-2 7 Green L3C-68523 BREAST SWAB



6852...ST SWAB-2 7 Yellow L3C-68523 BREAST SWAB



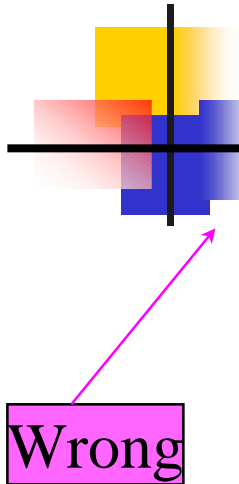
Analysis Result of the Real DNA Mixture Peak Area Data (10 out of 12 correct)(Texas 5 :68523)

Green cells indicate the correct genotype combinations

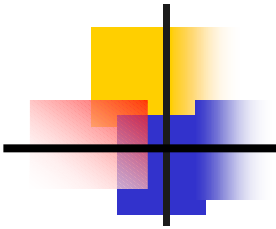
Locus	Alleles Detected	Peak Area Ratio Measured	Analysis Result			Comment
			Mass Ratio Calculated	Person1	Person2	
D3S1385	14, 15, 17	1.07: 1: 1.84	1: 1.12	17, 17 14, 15	14, 15 17, 17	All are likely
			1: 1.07	15, 17 14, 17	14, 17 15, 17	
			1: 1.08	17, 17 15, 19	15, 19 17, 17	
			1: 1.01	17, 19 15, 17	15, 17 17, 19	
VWA	15, 17, 19	1.01: 1.86: 1	1: 1.08	17, 17 15, 19	15, 19 17, 17	All are likely
			1: 1.01	17, 19 15, 17	15, 17 17, 19	
			1: 1.32	23, 24 20, 20	20, 20 23, 24	
			1: 1.36	20, 24 20, 23	20, 23 20, 24	
FGA	20, 23, 24	3.24: 1.46: 1	1: 1.32	23, 24 20, 20	20, 20 23, 24	All are likely
			1: 1.36	20, 24 20, 23	20, 23 20, 24	
			1: 1.13	14, 14 13, 13	13, 13 14, 14	
			1: 1	13, 14 13, 14	13, 14 13, 14	
D8S1179	13, 14	1.13: 1	1: 1.13	14, 14 13, 13	13, 13 14, 14	Less likely
			1: 1	13, 14 13, 14	13, 14 13, 14	More likely
D21S11	29, 30, 31.2, 32.2	1.39: 1.33: 1.09: 1	1: 1.30	31.2, 32.2 29, 30	29, 30 31.2, 32.2	Either is likely
			1: 1.30	29, 30 31.2, 32.2	31.2, 32.2 29, 30	
D18S51			Peak area data is not available at the locus			

Wrong

Green cells indicate the correct genotype combinations



D5S818	10, 12, 13	1: 2.29: 1.08	1: 1.10	10, 13	12, 12	All are likely
			1: 1.07	10, 12	12, 13	
D13S317	8, 10, 11, 12	1.16: 1.26: 1: 1.10	1: 1.16	11, 12	8, 10	Either is likely
				8, 10	11, 12	
D7S820	8, 10, 11	1.08: 1: 1.71	1: 1.26	11, 11	8, 10	All are likely
			1: 1.09	8, 10	11, 11	
D16S539	10, 11, 12	1.15: 2.16: 1	1: 1.15	11, 12	10, 11	All are likely
			1: 1.00	10, 11	11, 12	
THO1	7, 9	1.02: 1	1: 1.01	9, 9	7, 7	All are likely
			1: 1	7, 7	9, 9	
TPOX	8, 8			7, 9	7, 9	The only choice
CSFIPO	10, 11, 12	2.13: 1: 1.05	1: 1.04	8, 8	8, 8	All are likely
			1: 1.05	11, 12	10, 10	
				10, 10	11, 12	
				10, 11	10, 12	
				10, 12	10, 11	



Why Are the Results at Two Loci Not Correct in Texas 5 :68523?

- For the loci of D13S317 and D21S11, four alleles were present and the mass ratio is close to 1:1. In this case, the LSD is very sensitive to any noise present in the allele signals, such that correct profile combinations does not correspond to the one with the smallest fitting error; so no clear cut assignment can be made if the genotype of the victim is unknown.



Summary of the Analysis Result of Real DNA Mixture Data from Texas

- Five sets of DNA mixture data from Texas were analyzed and successfully resolved in 4 data sets.
- The top ranked genotype combinations coincide with that of true combinations in all but that of 2 loci in dataset 5.
- Dataset 5 has a mass ratio close to 1:1, which always signals potential nonresolvability.

Limitations of LSD

- **For all loci regardless of the number of alleles**

If the mass ratio is around 1:1 and the genotype of the victim is unknown, then both of these two genotype combinations are possible:

Genotype Combination 1			Genotype Combination 2		
Person 1	Person 2	Mass Ratio	Person 1	Person 2	Mass Ratio
A, B	A, C	1: 1.03	A, C	A, B	1.03: 1



Limitations of LSD (continued)

- **For those loci with four alleles in the mixture**
 - and if the mass ratio is around 1:1, then
 - no clear cut assignment can be made unless the genotype of the victim is known. The assignment is very sensitive to slight error in the allele peak measurements

Limitations of LSD (continued)

For those loci with three alleles

- If the mass ratio is around 1:1
- and if the ratio of the peak area is around 2:1:1

There are two possible genotype combinations, both of which fit equally well to the peak area ratio. The genotype assignment cannot be unambiguously resolved.

Genotype Combination 1			Genotype Combination 2		
Person 1	Person 2	Mass Ratio	Person 1	Person 2	Mass Ratio
A, A	B, C	1: 1	A, B	A, C	1: 1

Limitations of LSD (continued)

For loci with two alleles

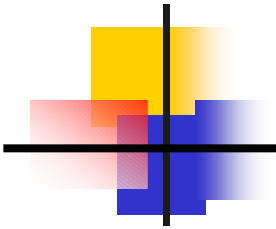
And if the two peak areas are almost the same (around 1:1)

if mass ratio is close to 1:1

if mass ratio is not close to 1:1

Genotype combo 1		Genotype combo 2	
Person 1	Person 2	Person 1	Person 2
A, A	B, B	A, B	A, B

Genotype	
Person 1	Person 2
A, B	A, B



Guidelines for Case Workers in Using the Least Square Analysis Results for Analyzing Mixture STR/DNA Data

1. Use LSD to get the ranking results for all genotype combinations for all loci.
2. First, identify those loci with 3 or 4 alleles which have a clear-cut minimum fitting-error and with consistent mass ratios.
3. For the less clear-cut loci with three or four alleles, pick those combinations with still relatively low error but have more consistent mass ratio as those in 1 above.
4. For those loci with only two alleles present, when the approximate mass ratio is close to 1:1 or 1:2, LSD limitations apply. If the mass ratio is not close to 1:1 or 1:2, the genotype combination with zero error and consistent mass ratio as those in 1 above, should be picked.



Summary of LSD

- Using the allele peak area or height information, the least square formulation and solution to the STR/DNA mixture interpretation problem **works if the peak areas (or height) are relatively consistent across all loci.**
- The technique is direct, elegant, and simple.
- Up to date, in our lab, about 18 sets of peak area (height) data of real DNA mixture samples have been analyzed. Results are very encouraging.
- Software implementation is being finalized.



Appendix

Two Numerical Examples Illustrating
LSD, the Least Square Deconvolution
Technique: a 3-allele and a 4-allele case

A Numerical Example of Using LSD To Resolve the Individual Genotype Profiles: a 3-Allele Example

- Alleles in the mixture at a locus: {6, 9, 9.3}
 - Peak Areas: [11086 1197 1348]
1. Normalize the peak areas by dividing each area by the smallest area: $\rightarrow [9.26, 1, 1.13]$
 2. Take the first genotype combination for the 3-allele locus: $[\{A\} \{B, C\}] \rightarrow [\{6\} \{9, 9.3\}]$
 3. Set up the A matrix for this combination: $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$
 4. Set up the normalized measured peak area vector b: $b = \begin{bmatrix} 9.26 \\ 1 \\ 1.13 \end{bmatrix}$

3-Allele example continued

6. Look up the pseudoinverse (A^+) for this A , in the pseudoinverse table for the 3-allele case:

$$A^+ = \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

5. Solve for the vector of mass coefficients by the least square technique: $x = A^+b$

$$x_{ls} = A^+b = \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9.26 \\ 1 \\ 1.13 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1*9.26+0*1+0*1.13 \\ 0*9.26+1*1+1*1.13 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 9.26 \\ 2.13 \end{bmatrix} = \begin{bmatrix} 4.63 \\ 1.065 \end{bmatrix}$$



3-Allele example continued

8. Calculate the fitted peak area vector: b_f

$$b_f = Ax = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4.63 \\ 1.065 \end{bmatrix} = \begin{bmatrix} 2 * 4.63 + 0 * 1.065 \\ 0 * 4.63 + 1 * 1.065 \\ 0 * 4.63 + 1 * 1.065 \end{bmatrix} = \begin{bmatrix} 9.26 \\ 1.065 \\ 1.065 \end{bmatrix}$$

9. Calculate the error between the measured peak-area vector, b , and the predicted area vector, b_p

$$e = b - b_f = \begin{bmatrix} 9.26 \\ 1 \\ 1.13 \end{bmatrix} - \begin{bmatrix} 9.26 \\ 1.065 \\ 1.065 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.065 \\ 0.065 \end{bmatrix}$$

10. Calculate the square of the magnitude of the error vector, $magsq_e$: (the sum of the squares of the terms of the e vector)

$$magsq_e = \|e\|^2 = (0^2 + (-0.065)^2 + 0.065^2) = 0.00845$$

A second example: the 4-allele case

Data: alleles: {59, 65, 67, 70}

peak areas: [1226 1434 8816 8894]

1. Select the genotype combo case 1: {A,B} { C, D}

2. Set up the A matrix: $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$

3. Set up the peak area vector and normalize the areas with respect to the smallest area:

$$\text{area} = \begin{bmatrix} 1226 \\ 1434 \\ 8816 \\ 8894 \end{bmatrix} \xrightarrow{\text{normalize}} \text{b} = \begin{bmatrix} 1 \\ 1.17 \\ 7.19 \\ 7.25 \end{bmatrix}$$



4-allele example

4. Calculate the least square mass coefficient vector: x

$$x = A^+b = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1.17 \\ 7.19 \\ 7.25 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1*1 + 1*1.17 \\ 1*7.19 + 1*7.25 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2.17 \\ 14.44 \end{bmatrix} = \begin{bmatrix} 1.09 \\ 7.22 \end{bmatrix}$$

5. Calculate the fitted area: $b_f = Ax$

$$b_f = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1.09 \\ 7.22 \end{bmatrix} = \begin{bmatrix} 1.09 \\ 1.09 \\ 7.22 \\ 7.22 \end{bmatrix}$$



4-allele example

6. Calculate the error vector in the peak area fit: $e = b - b_f$

$$e = \begin{bmatrix} 1 \\ 1.17 \\ 7.19 \\ 7.25 \end{bmatrix} - \begin{bmatrix} 1.09 \\ 1.09 \\ 7.22 \\ 7.22 \end{bmatrix} = \begin{bmatrix} -0.09 \\ 0.08 \\ -0.03 \\ 0.03 \end{bmatrix}$$

7. Calculate the square of the magnitude of the error vector: $\|e\|^2$

$$\begin{aligned} \text{magsq}_e = \|e\|^2 &= (-0.09)^2 + (0.08)^2 + (-0.03)^2 + (0.03)^2 \\ &= 0.0163 \end{aligned}$$

4-allele example

- Normalize the mass ratio vector and the predicted peak area ratio vector for reporting in the composite profile

$$x_{ls} = \begin{bmatrix} 1.09 \\ 7.22 \end{bmatrix} \xrightarrow{\text{normalize}} \begin{bmatrix} 1 \\ 6.62 \end{bmatrix}$$

x_n : Normalised mass ratio

Post Processing Steps

1. Normalize the mass coefficient vector x into x_n , by dividing the mass coefficients by the smallest one, such that the smallest mass is 1.
2. Report the normalized mass ratio.
3. Do the least square fit for every possible genotype combination for this locus, and calculate the error in the fit for each combination.
4. Rank order the genotype combinations according to the associated magsq_e error, from the smallest to the largest.
5. Eliminate those which result in a negative mass coefficient (unrealistic).
6. For those loci with only 2 alleles, 3 cases will results in fit with 0 error. Record all these. The ones with zero error may not be the correct case.
7. Look for obvious outliers which are not consistent with the mass ratios for the other loci.
8. Special consideration for 2-allele loci, and for those sample cases in which the mass ratio is close to 1:1.



LSD Software Under Development

Screen Shot of the LSD Software Interface Under Development

The screenshot displays the 'Loci Separator' software interface. The window title is 'Loci Separator' and it includes a 'File' menu. The interface is organized into a grid of input fields for different loci. Each locus has a label and a corresponding set of input boxes. The loci and their input boxes are as follows:

Locus	Input Boxes
D3S1385	4
TH01	4
D21S11	4
D18S51	4
D5S818	4
D13S317	4
D7S820	4
D16S539	4
CSF1PO	4
VWA	4
D8S1179	4
TPOX	4
FGA	4

At the bottom right of the interface, there are three buttons: 'Clear', 'Separate', and 'Exit'.

Loci Ratio	P1	P2	Mean Error	Mass
------------	----	----	------------	------

as

D3S1385

1	15, 16	16, 17	0.00	1.0:3.3
2	15	16, 17	0.63	1.0:7.9
3	15, 17	16, 17	1.50	1.0:13.0
4	15, 17	16	2.87	1.0:1.0
5	17	15, 16	6.18	1.0:1.6
6	15, 17	15, 16	15.92	1.0:2.0

TH01

No Information

D21S11

1	29, 30	31.2	0.04	1.0:5.4
2	30, 31.2	29, 31.2	33.09	1.0:1.1
3	29, 30	29, 31.2	47.70	1.0:-2.8
4	29, 30	30, 31.2	52.05	1.0:-2.8
5	30	29, 31.2	60.09	1.0:13.5
6	29	30, 31.2	63.05	1.0:10.4

D18S51

1	15, 17	15, 16	0.09	1.0:6.0
2	17	15, 16	1.17	1.0:15.3
3	16, 17	15, 16	2.14	1.0:48.4
4	16, 17	15	17.29	1.0:1.1
5	16	15, 17	27.47	1.0:1.4
6	16, 17	15, 17	68.09	1.0:1.7

D5S818

1	10, 11	11, 12	0.01	1.0:4.7
2	10	11, 12	0.64	1.0:10.9
3	10, 12	11, 12	1.51	1.0:18.3
4	10, 12	11	7.61	1.0:1.0
5	12	10, 11	12.66	1.0:1.4
6	10, 12	10, 11	32.89	1.0:1.7

D13S317

1	11, 13	9, 12	0.86	1.0:4.9
2	12, 13	9, 11	20.88	1.0:1.3
3	11, 12	9, 13	21.28	1.0:1.2

D7S820

1	10, 11	8	0.00	1.0:7.2
2	8, 11	8, 10	53.71	1.0:1.0
3	10, 11	8, 10	71.97	1.0:-2.5
4	10, 11	8, 11	72.71	1.0:-2.5
5	11	8, 10	93.76	1.0:15.8
6	10	8, 11	94.27	1.0:15.2

D16S539

No Information

CSF1PO

No Information

VWA

1	14, 19	14, 20	0.84	1.0:2.9
2	19	14, 20	3.34	1.0:10.3
3	19, 20	14	4.15	1.0:1.3
4	19, 20	14, 20	4.28	1.0:-27.0
5	20	14, 19	14.94	1.0:1.9
6	19, 20	14, 19	29.13	1.0:4.4

D8S1179

1	13, 14	11, 12	0.06	1.0:7.1
2	12, 14	11, 13	37.96	1.0:1.0
3	12, 13	11, 14	37.96	1.0:1.0

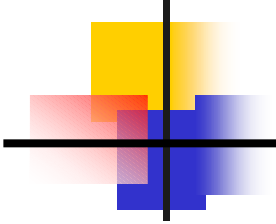
TPOX

No Information

FGA

1	24, 25	21	0.00	1.0:8.2
2	21, 24	21, 25	69.98	1.0:1.0
3	24, 25	21, 25	90.63	1.0:-2.4
4	24, 25	21, 24	90.81	1.0:-2.4
5	24	21, 25	119.95	1.0:17.5
6	25	21, 24	120.08	1.0:17.4

LSD Software
Output



LSD Software Demonstration