

Investigative DNA Databases that Preserve Identification Information

American Academy of Forensic Sciences
February, 2012
Atlanta, GA

Mark W Perlin, PhD, MD, PhD
Cybergenetics, Pittsburgh, PA



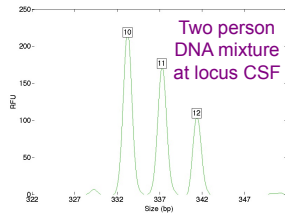
Cybergenetics

Cybergenetics © 2003-2012

DNA Uncertainty

Genetic uncertainty
in forensic DNA data

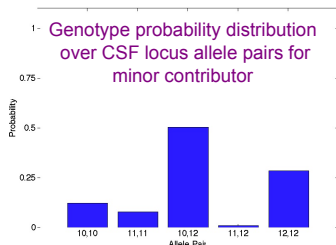
- DNA mixtures
- low DNA amounts
- kinship relations
- stochastic effects



Probabilistic Genotype

Genotype uncertainty expressed through probability

Genotype probability distribution
over CSF locus allele pairs for
minor contributor

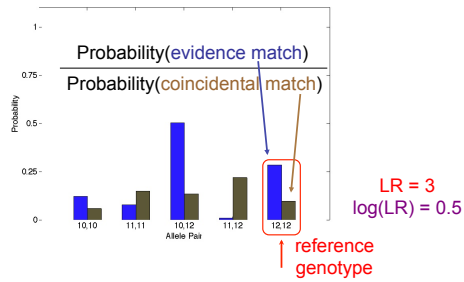


- Laplace (1812)
- Mendel (1865)
- SWGDAM (2010)
- ANSI/NIST (2011)

MW Perlin, MM Legler, CE Spencer, JL Smith, WP Allan, JL Belrose, BW Duceaman.
Validating TrueAllele® DNA mixture interpretation. Journal of Forensic Sciences, 2011.

Identification Information

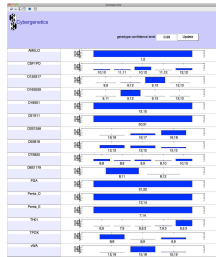
Likelihood ratio (LR) DNA match statistic



Investigative DNA Database

probabilistic genotypes

reference genotypes



LR match statistic
10^{17.5}

AMELO	1	2
CSF1PO	12	12
D13S317	9	13
D16S539	9	12
D18S51	13	15
D21S11	30	31
D3S1358	16	17
D5S818	12	12
D7S820	10	10
D8S1179	8	11
FGA	21	22
Penta_D	12	14
Penta_E	7	14
TH01	9	9.3
TPOX	8	8
vWA	15	18

Investigative Applications

Information-preserving DNA database

- *evidence-to-suspect*, solve cold cases
- *evidence-to-evidence*, connect serial crime
- *evidence-to-kinship*, find missing people
- *kinship-to-reference*, conduct familial search
- *remains-to-missing*, identify disaster victims

Perlin MW. Identifying human remains using TrueAllele® technology. In: Okoye MI, Wecht CH, editors. Forensic Investigation and Management of Mass Disasters. Lawyers & Judges Publishing; 2007.

Study Data: 40 Mixtures

5 weights

Set 1		ng amplified			
Mixture ratio		1	0.5	0.25	0.125
90%	10:0	A1	A2	A3	A4
70%	9:1	B1	B2	B3	B4
50%	7:3	C1	C2	C3	C4
30%	5:5	D1	D2	D3	D4
10%	3:7	E1	E2	E3	E4
	1:9	F1	F2	F3	F4
	0:10	G1	G2	G3	G4

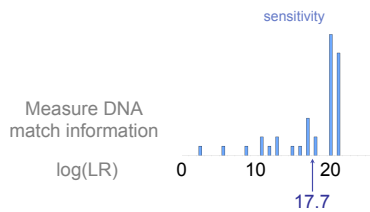
1, 1/2, 1/4, 1/8 ng
4 dilutions

2 pairs

Set 2		ng amplified			
Mixture ratio		1	0.5	0.25	0.125
10:0	H1	H2	H3	H4	
9:1	I1	I2	I3	I4	
7:3	J1	J2	J3	J4	
5:5	K1	K2	K3	K4	
3:7	L1	L2	L3	L4	
1:9	M1	M2	M3	M4	
0:10	N1	N2	N3	N4	

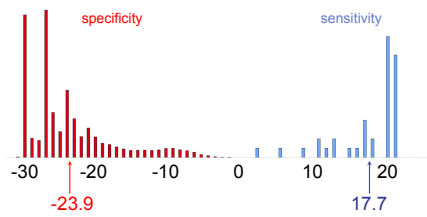
Sensitivity

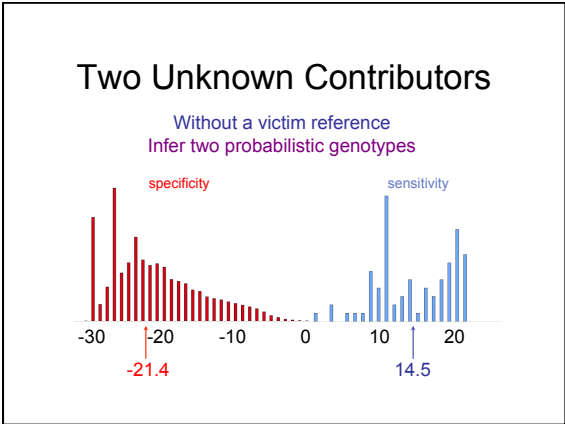
Two contributors, victim known
Infer one probabilistic genotype

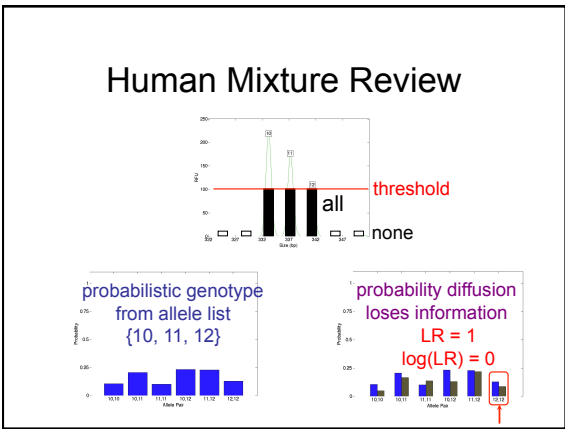


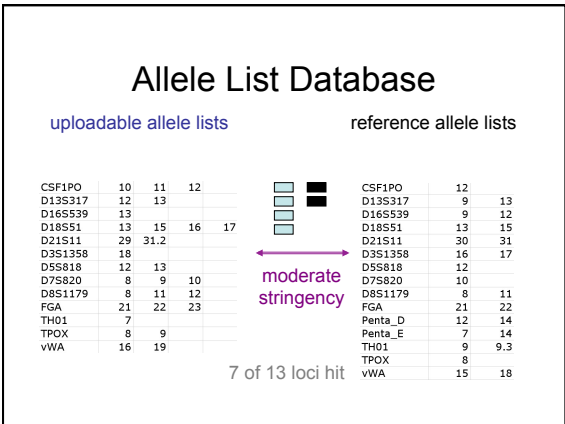
Specificity

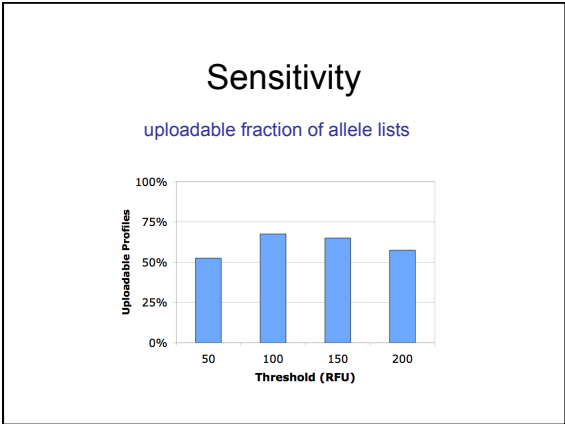
Compare with 1,000 random genotypes

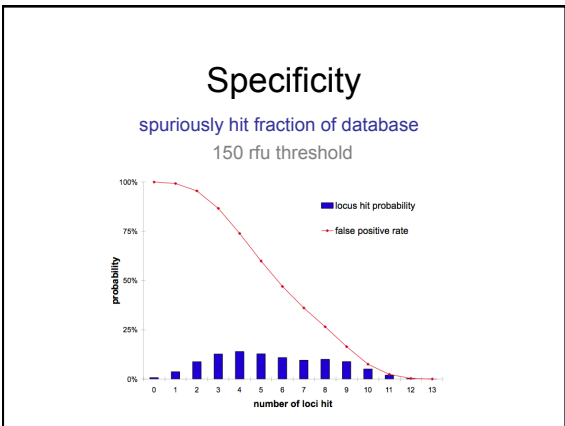












Information Comparison

Type of DNA database	Sensitivity	Specificity
Probabilistic genotype	LR average is about a quadrillion	False positive rate < 0.01%
Allele list (moderate stringency)	Upload fails about 1/3 of the time	Hits 5%-25% of the DNA database

The Information Age

Saves

- work
- time
- money
- evidence
- lives

*Cybergenetics: Meredith Clarke, Matt Legler and Duquesne Univ interns
National Institute of Science and Technology: Margaret Kline
National Institute of Justice: Award Number 2001-IJ-CX-K003*



Cybergenetics

www.cybgen.com/information

perlin@cybgen.com
