A Match Likelihood Ratio for DNA Comparison

American Academy of Forensic Sciences
February, 2009

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

Cybergenetics

Cybergenetics © 2003-2009

Uncertain Genotype

Q probability distribution q(x)

R r(x)

S s(x)

Match Likelihood Ratio

Q

Pr(Q=S)

S

Pr(Q=R)

R

$$LR = \frac{Pr(Q=S)}{Pr(Q=R)}$$

Interpreting DNA Evidence

A. Obtain DNA data
B. Infer genotype
   1. Data
   2. Model
   3. Compare
   4. Probability
C. Likelihood ratio

Genotype Inference

1. Data
   - evidence
   - victim
2. Model
   - genotype candidate
   - generate pattern
3. Compare
   - likelihood function (bell curve)
   - product rule for data
4. Probability
   - genotype probability distribution
   - genotype probability = $\frac{\text{genotype likelihood*}}{\text{sum of all genotype likelihoods*}}$

Different Methods: Data Used

Data Used	inclusion	subtraction	addition
victim profile	NO	YES	YES
original data	NO	NO	YES

Statistical Inference View

inclusion method vs. likelihood ratio approach

"often robs the items of any probative value" - B. Weir

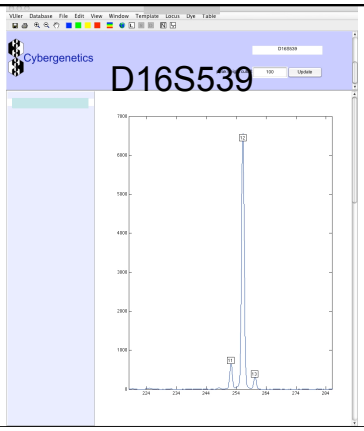
"usually discards a lot of information compared to the correct likelihood ratio approach" - C. Brenner

"does not use as much of the information included in the data as the LR approach but, conceptually, they are equivalent" - M. Krawczak

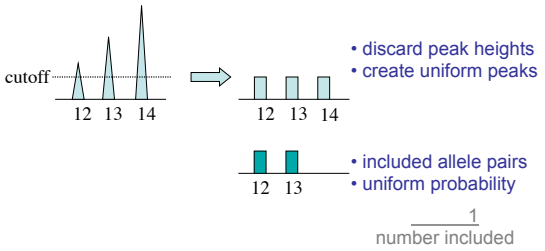
"Recommendation 1: The likelihood ratio is the preferred approach to mixture interpretation." - DNA commission of the International Society of Forensic Genetics

Mixture Case

- DNA from under victim's fingernails
- two contributors to DNA mixture
- 93.3% victim & 6.7% unknown
- 2 ng DNA in 50 ul
- ProfilerPlus + Cofiler STR analysis
- three different mixture interpretations
  1. inclusion
  2. subtraction
  3. addition



Inclusion Method



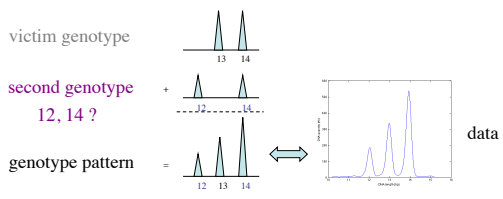
- discard peak heights
- create uniform peaks
- included allele pairs
- uniform probability

$$\frac{1}{\text{number included}}$$

'Inclusion' LR at D16S539

		Genotype Probability Distributions			Match Probabilities	
		Q	**R**	**S**	**Pr(Q=S)**	**Pr(Q=R)**
allele pair		q(x)	r(x)	s(x)	q(x)*s(x)	q(x)*r(x)
11	11	0.16667	0.07413			0.01236
11	12	0.16667	0.18466			0.03078
11	13	0.16667	0.08896	1	0.16667	0.01483
12	12	0.16667	0.11499			0.01917
12	13	0.16667	0.11080			0.01847
13	13	0.16667	0.02669			0.00445
				Pr(Q=S)	0.16667	
				Pr(Q=R)		0.10004
				Likelihood Ratio		1.666

Addition Method

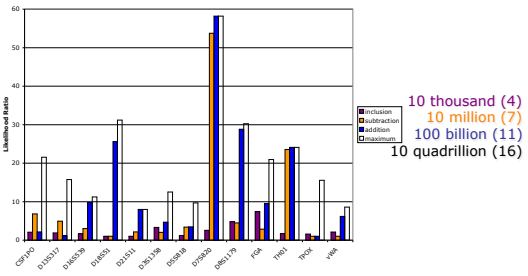


- similar pattern, high likelihood
- dissimilar pattern, low likelihood

'Addition' LR at D16S539

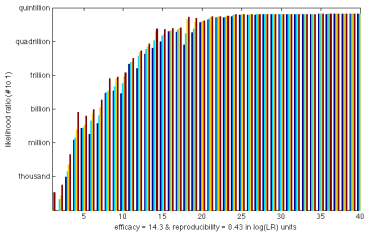
		Genotype Probability Distributions			Match Probabilities	
		Q	**R**	**S**	**Pr(Q=S)**	**Pr(Q=R)**
allele pair		q(x)	r(x)	s(x)	q(x)*s(x)	q(x)*r(x)
11	11		0.07413			
11	12		0.18466			
11	13	0.722	0.08896	1	0.72200	0.06423
12	12		0.11499			
12	13	0.028	0.11080			0.00310
13	13	0.250	0.02669			0.00667
				Pr(Q=S)	0.72200	
				Pr(Q=R)		0.07400
				Likelihood Ratio		9.756

Information

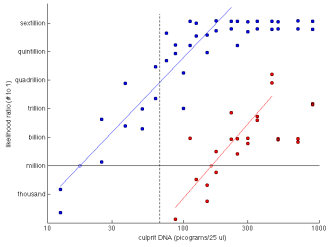


10 thousand (4)
10 million (7)
100 billion (11)
10 quadrillion (16)

Validation



Calibration



Bibliography

- Quantitative STR Peak Information
- Genotype Probability Distributions
- Computer Interpretation of STR Data
- Statistical Modeling and Computation
- Likelihood Ratio Literature
- Mixture Interpretation Admissibility
- Computer Systems for Quantitative DNA Mixture Deconvolution
- TrueAllele Casework Publications

Conclusions: MLR

- a useful tool for determining identification information with uncertain genotypes
- works well on forensic mixture cases
- enables quantitative validation, calibration, and comparison of genotype inference methods
- based on generally accepted scientific principles