

New York State Police TrueAllele® Casework Developmental Validation

New York State DNA Subcommittee
March, 2010



Cybergenetics



Cybergenetics © 2003-2010

Information Gain (LR)

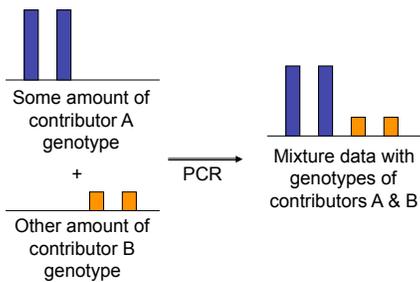
identification hypothesis:
the suspect contributed to the evidence

$$\text{information gain (likelihood ratio)} = \frac{\text{Odds(hypothesis | data)}}{\text{Odds(hypothesis)}}$$

↑ data
after
before

Additive information units: log(LR)
Order of magnitude, powers of ten

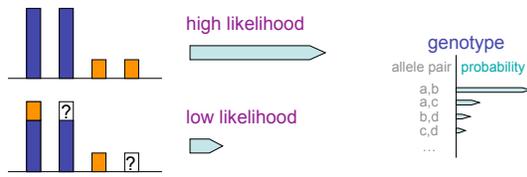
DNA Mixture Data



Quantitative Mixture Interpretation

Step 1: infer genotype

- consider every possible allele pair
- compare pattern with DNA data
- Rule: *better fit's more likely it*



Information Gain (LR)

Step 2: match genotypes

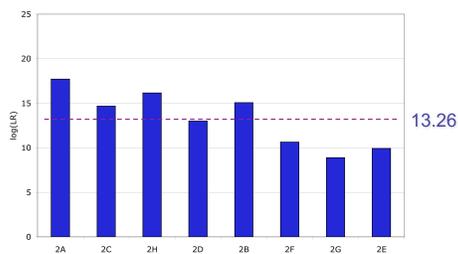
At the suspect's genotype allele pair, what is the locus *information gain*?

$$\text{information gain (likelihood ratio)} = \frac{\text{Prob}(\text{allele pair} \mid \text{data})}{\text{Prob}(\text{allele pair})}$$

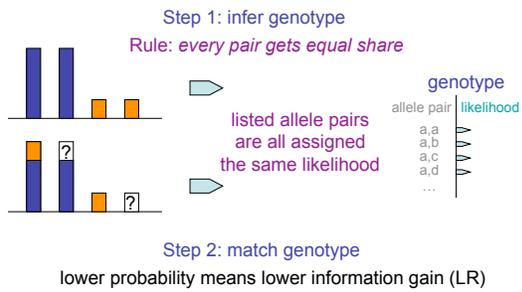
↑ data
after
before (population)

Computer objectivity:
(Step 1) infer evidence genotype from data
(Step 2) compare genotype with suspect

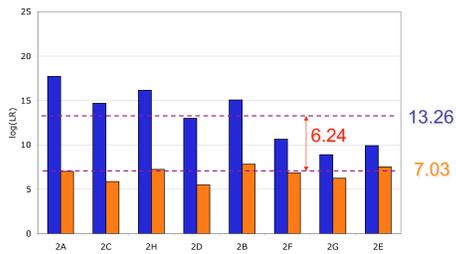
Efficacy (2 unknown)



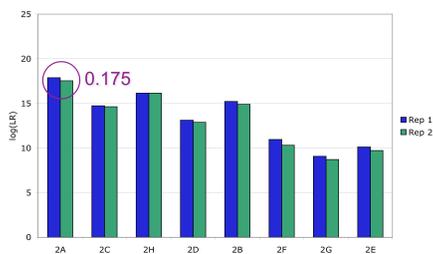
Qualitative Manual Review

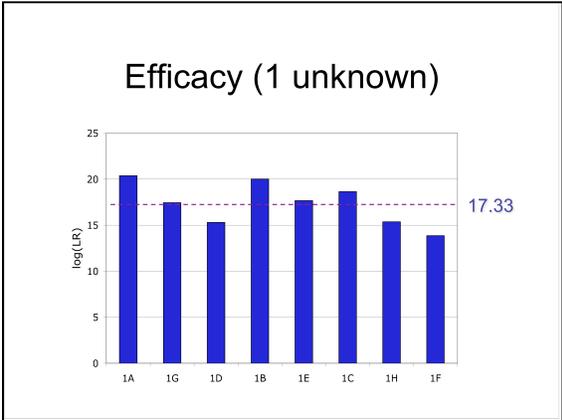


Improvement

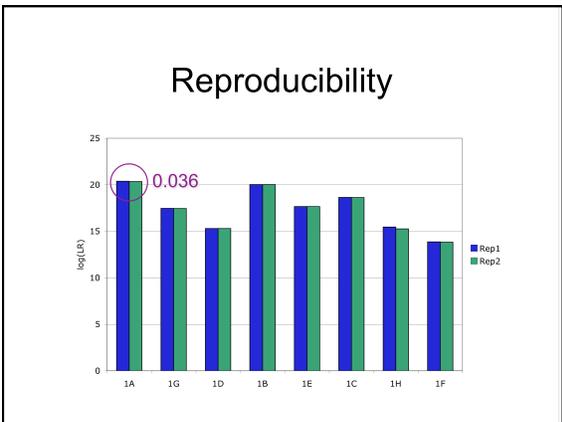


Reproducibility









Comparison

interpretation method	two unknown (without victim)	one unknown (with victim)
quantitative computer	13.26 (0.175) (ten trillion)	17.33 (0.036) (hundred quadrillion)
qualitative human	7.03 (ten million)	12.66 (fifty trillion)
improvement	6.24 (one million)	4.67 (fifty thousand)

Summary

- **information gain** (LR) is a universal DNA metric
- **efficacy**: computer extracts useful information
- **improvement**: computer mixture interpretation is more informative than human review
with victim 50,000x - without victim 1,000,000x
- **reproducibility**: tenths of a log(LR) unit
- **objectivity**: "parallel unmasking", infer then match
- **productivity**: lab gives statistic for 1 of 3 items
- **utility**: science, investigation and evidence

Commonwealth vs. Foley

Apr 2006: Blairsville Dentist John Yelenic murdered

Nov 2007: Trooper Kevin Foley charged with crime



Feb 2008: Defense questions 13,000 DNA match score

DNA Evidence

- DNA from under victim's fingernails (Q83)
- two contributors to DNA mixture
- 93.3% victim & 6.7% unknown
- 1,000 pg DNA in 25 ul
- STR analysis with ProfilerPlus®, Cofiler®
- know victim contributor genotype (K53)
- TrueAllele® computer interpretation (using genotype addition method)
infer unknown contributor genotype
- only after having inferred unknown, compare with suspect genotype (K2)

Three DNA Match Statistics

ScoreMethod
13 thousandinclusion
23 millionsubtraction
189 billionaddition

- Why are there different match results?
- How do mixture interpretation methods differ?
- What should we present in court?

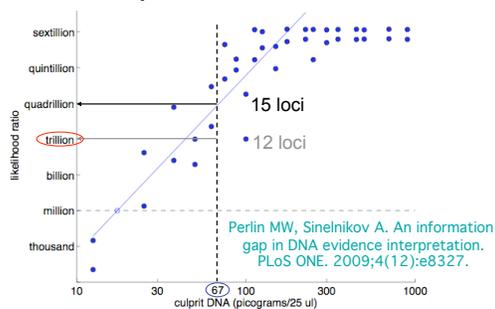
Different Interpretation Methods

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

Frye: General Acceptance in the Relevant Community

- 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

Expected Result

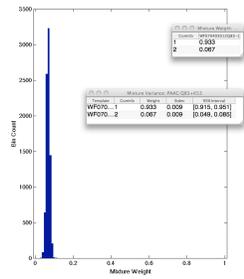


Expert Testimony

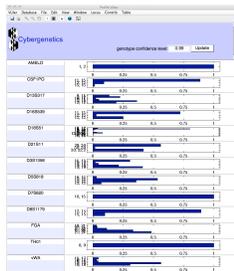
Dr. Perlin explained to the jury why these apparently different results were expected by DNA science. "The less informative methods ignored some of the data," said Dr. Perlin, "while the TrueAllele computation considered all of the available DNA data."

"A scientist may look at the same slide using the naked eye, a magnifying glass, or a microscope," analogized Dr. Perlin. "A computer that considers all the data is a more powerful DNA microscope."

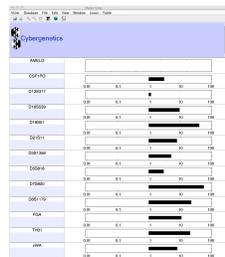
Mixture Weight



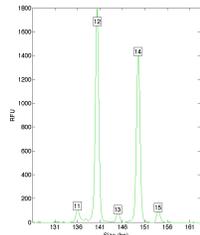
Inferred Genotype



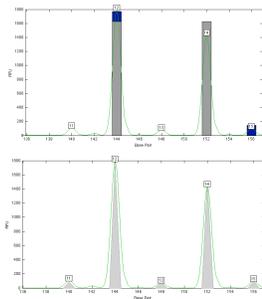
log(LR) Match Information



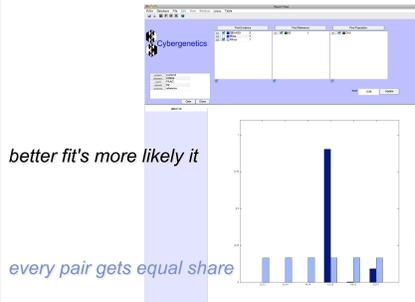
Locus D8S1179 Data



Explain D8S1179 Genotype



Likelihood Comparison



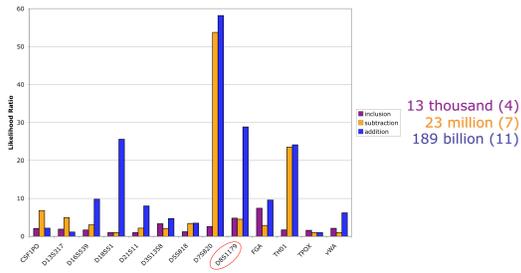
Generate Report

locus	allele pair	Q	R	S	LR	log10(LR)
C57J40	12, 13	0.891	0.0518	1	1.755	0.244
D1S5117	8, 11	0.136	0.0685	1	1.990	0.299
D16S539	11, 13	0.722	0.0928	1	7.775	0.891
D18S51	12, 13	0.803	0.0354	1	22.683	1.356
D21S11	29, 30	0.561	0.0877	1	6.388	0.805
DSS1958	15, 18	0.213	0.0859	1	2.338	0.465
D5S818	12, 13	0.358	0.1077	1	3.324	0.522
D7S820	10, 13	1	0.0226	1	44.188	1.645
D8S1179	12, 15	0.895	0.0365	1	24.325	1.390
FGA	21, 24	0.483	0.0314	1	9.388	0.973
TH01	8, 9	1	0.0450	1	22.201	1.346
VWA	17, 18	0.562	0.1199	1	4.689	0.671

Locus information gain is genotype probability ratio:
LR = after/before

Joint information is the sum of the locus information

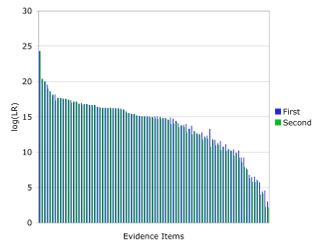
More Data In, More Information Out



Case Observations

- objective review never saw suspect
- easy to testify about in court
- understandable to judge and jury
- have precedent: admitted, testified
- preserve match information in data
- rapid response to attorney
- multiple match scores presented
- all information to the triers of fact – nothing was withheld from the jury
- this should be standard practice

Efficacy & Reproducibility



Productivity

case classification	TrueAllele log(LR) information	log(LR) standard deviation	human review match score success rate
simple N = 35	16.3 (ten quadrillion)	0.10	49%
intermediate N = 20	13.1 (ten trillion)	0.26	25%
complex N = 33	11.9 (one trillion)	0.44	21%
