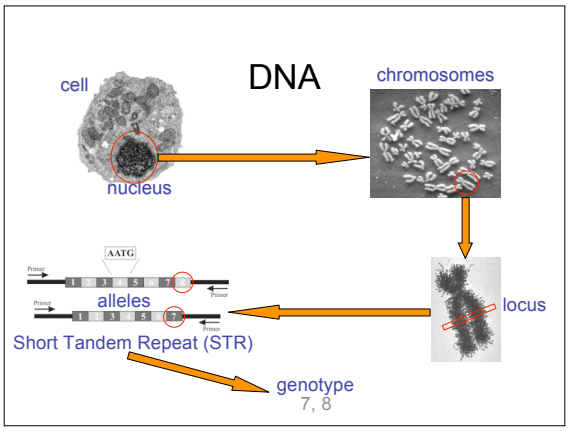
DNA Identification Science: The Search for Truth

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

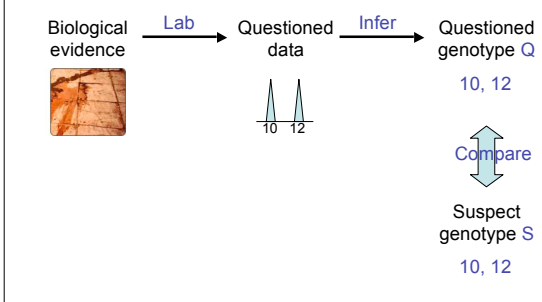
Summer Research Symposium
Duquesne University
July, 2010

Cybergenetics

Cybergenetics © 2003-2010

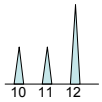


Identification



Uncertainty

Biological evidence → Lab → Questioned data → Infer → Questioned genotype Q



10, 12 @ 50%
11, 12 @ 30%
12, 12 @ 20%

Compare

Suspect genotype S

10, 12

Science

Prof Jevons, 1874

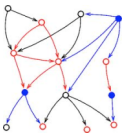
Scientific method

1. Propose hypothesis
2. Examine consequences
3. Compare with data

Prof Feynman, 1965

Agree: hypothesis likely
Sort of agree: sort of likely
Disagree: unlikely

Search for Truth



Computers, 1985

Rev Bayes, 1765

Bayes Theorem

Our belief in a hypothesis after we have seen data *is proportional to* how well that hypothesis explains the data *times* our initial belief.

All hypotheses must be considered.
Need computers to do this properly.

Find the probability of causes by examining effects.

Information Gain (LR)

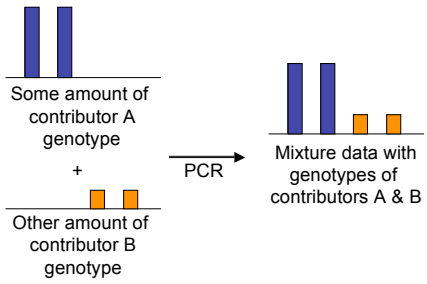
identification hypothesis:
the suspect contributed to the evidence

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

after ↑ data
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

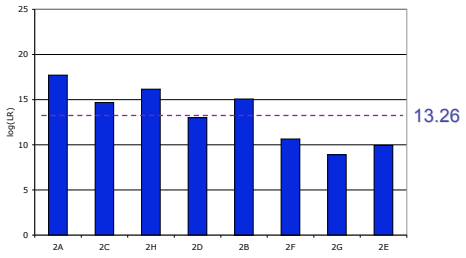
At the suspect's genotype allele pair,
what is the locus *information gain*?

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

after ↑ data
before (population)

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

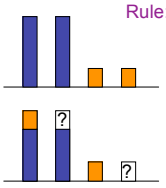
Efficacy (2 unknown)



Qualitative Manual Review

Step 1: infer genotype

Rule: *every pair gets equal share*



listed allele pairs
are all assigned
the same likelihood

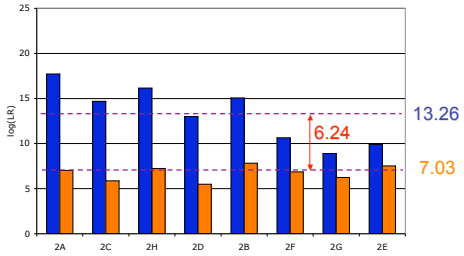
genotype

allele pair	likelihood
a,a	
a,b	
a,c	
a,d	
…	

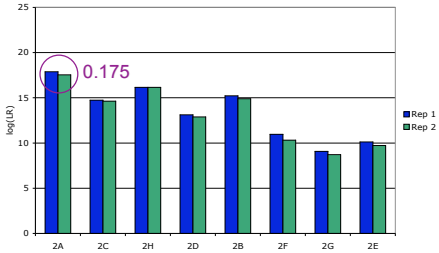
Step 2: match genotype

lower probability means lower information gain (LR)

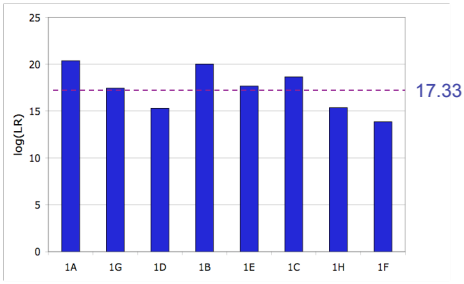
Improvement



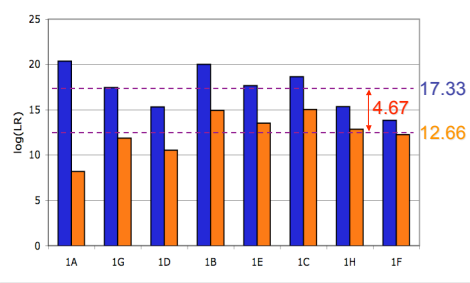
Reproducibility



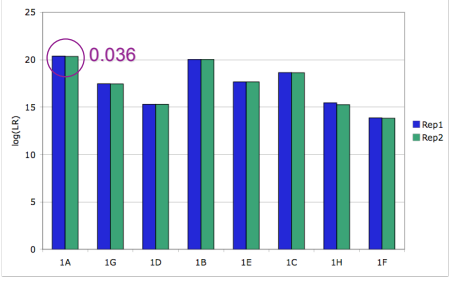
Efficacy (1 unknown)



Improvement



Reproducibility



Validation Summary

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 (five trillion)
improvement	6.24 (one million)	4.67 (fifty thousand)

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

Score	Method
13 thousand	inclusion
23 million	subtraction
189 billion	addition

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

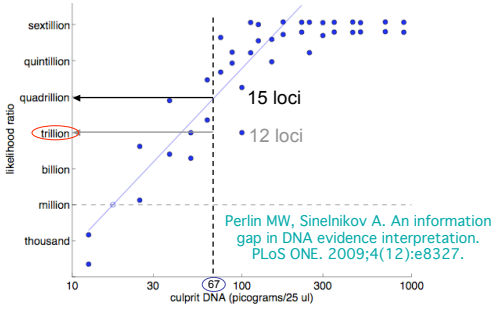
Different Interpretation Methods

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



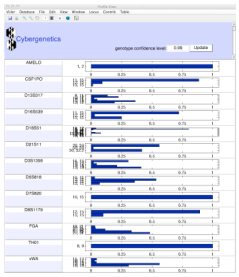
Perlin MW, Sinelnikov A. An information gap in DNA evidence interpretation. PLoS ONE. 2009;4(12):e8327.

Expert Testimony

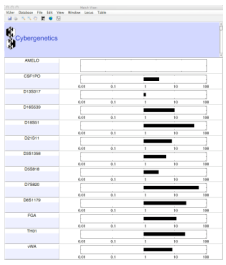
Dr. Perlin explained to the jury why these apparently different results were expected by DNA science. "The less informative methods ignored some of the data," said Dr. Perlin, "while the TrueAllele computation considered all of the available DNA data."

"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."

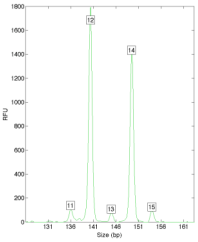
Inferred Genotype



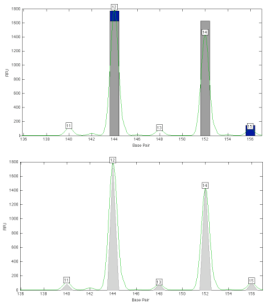
log(LR) Match Information



Locus D8S1179 Data



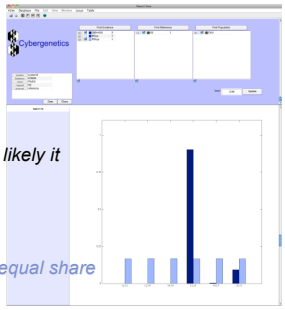
Explain D8S1179 Genotype



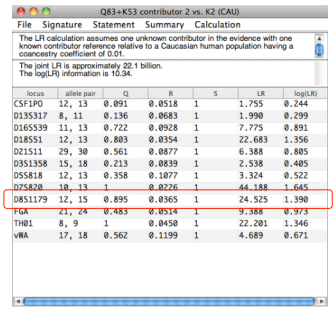
Likelihood Comparison

better fit's more likely it

every pair gets equal share



Generate Report

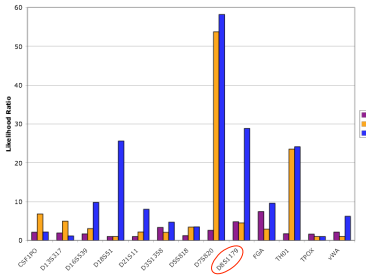


locus	allele pair	Q	R	S	LR	log(LR)
CSF1PO	12, 13	0.091	0.0518	1	1.755	0.244
D13S317	8, 11	0.136	0.0683	1	1.990	0.299
D16S539	11, 13	0.722	0.0928	1	7.775	0.891
D18S51	12, 13	0.803	0.0354	1	22.683	1.356
D21S11	29, 30	0.561	0.0877	1	6.388	0.805
D3S1358	15, 18	0.213	0.0839	1	2.538	0.405
D5S818	12, 13	0.358	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	24.525	1.390
FGA	21, 24	0.483	0.0514	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



13 thousand (4)
23 million (7)
189 billion (11)

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

Indiana Gazette.com
Indiana, Pennsylvania
IN PRINT DAILY. ONLINE ALWAYS.

HOME SUBSCRIBERS MARKETPLACE NEWS OBITUARIES SPORTS BUSINESS MULTIMEDIA O
FYI JOURNALS

ARCHIVES > NEWS

One Verdict

Print | E-mail | Comment (No comments posted.) | Rate | Text Size

Jury convicts trooper of dentist slaying

Published: Thursday, March 19, 2009 12:46 AM EDT

An Indiana County Court jury this evening convicted state trooper Kevin Foley of first degree murder in the April 13, 2006, slashing death of Blairsville dentist John Yelenic.

"John Yelenic provided the most eloquent and poignant evidence in this case," said the prosecutor, senior deputy attorney general Anthony Krastek. "He managed to reach out and scratch his assailant," capturing the murderer's DNA under his fingernails.

The DNA Investigator Newsletter. Same Data, More Information - Murder, Match and DNA, Cybergenetics, 2009.
www.cybgen.com/information/newsletters/CybgenNews1.pdf

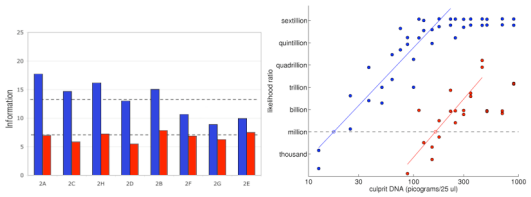
Public Safety

- DNA databases of criminal offenders
- police investigation: DNA database hits
- prevent crime by catching criminals
- could prevent 100,000 stranger rapes
- ensure conviction of the guilty
- avoid implicating the innocent

DNA public policy assumes that crime labs preserve DNA identification information

Information Loss

Discarding DNA identification informaton
instead of preserving it with science



No Information Incentive

Crime lab evaluation
- NIJ funding based on other metrics
- ASCLD/ISO accreditation standards
- FBI/SWGDAM method guidelines

Actual metrics and incentives
- eliminating backlogs
- passing audits
- testifying comfort

DNA identification information is not assessed

No Information Consistency

National Institute of Standards and Technology
Two Contributor Mixture Data, Known Victim

Some Differences in Reporting Statistics

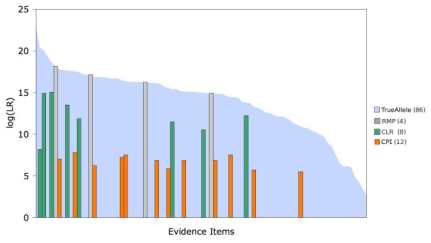
		Case1		
LabID	Kits Used	Caucasians	African-Americans	Hispanics
90	ProPlus/Cofiler	1.18E+15	2.13E+14	3.09E+15
34	ProPlus/Cofiler	2.40E+11	7.80E+09	9.80E+10
33	ProPlus/Cofiler	2.94E+08	1.12E+08	1.74E+09
6	ProPlus/Cofiler	40,000,000	3,500,000	280,000,000
9	ProPlus/Cofiler	1.14E+07	1.97E+07	1.54E+08
79	ProPlus/Cofiler	930,000	87,700	1,350,000
16	ProPlus/Cofiler	434,600	31,710	399,100

Remember that these labs are interpreting the same MIX05 electropherograms

213 trillion (14)

31 thousand (4)

Failure to Identify



Scientist Task

- practice identification science, not forensic art
- teach public about scientific methods
- learn more mathematics (probability theory)
- research accurate and objective methodology

Acknowledgements

Northeast Regional Forensics Institute
Jamie Belrose

Cybergenetics
William Allan
Meredith Clarke
Matthew Legler
Jessica Smith
Cara Spencer

New York State Police
Barry Duceman
Melissa Lee
Shannon Morris
Elizabeth Staude

Boston University
Robin Cotton

Carnegie Mellon
Jay Kadane

Cybergenetics