Taming Uncertainty in Forensic DNA Evidence

ENFSI Meeting
April, 2011
Brussels, Belgium

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

Cybergenetics

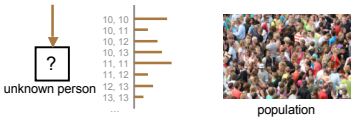
Uncertainty

- people (and the law) want *certainty*
- scientific data are *uncertain*

- *probability* describes uncertainty
- *information* is a change in probability

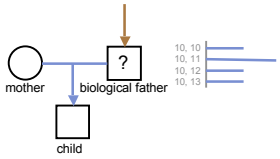
- DNA identification is an information science
- goal is *maximal objective information*

Genotype



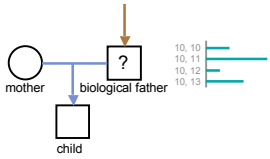
Prior probability

Genotype



Likelihood function

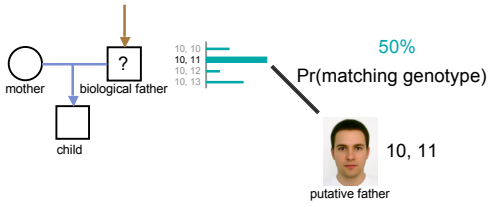
Genotype



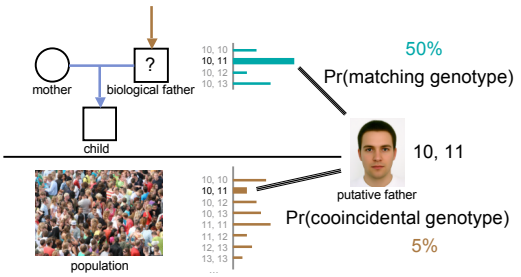
Posterior probability

Objective Bayesian inference
posterior = prior x likelihood

Match



Match



MW Perlin. Explaining the likelihood ratio in DNA mixture interpretation. Proceedings of Promega's 21st International Symposium on Human Identification, 2010.

Information

$$LR = \frac{\text{Pr(matching genotype)}}{\text{Pr(cooincidental genotype)}}$$

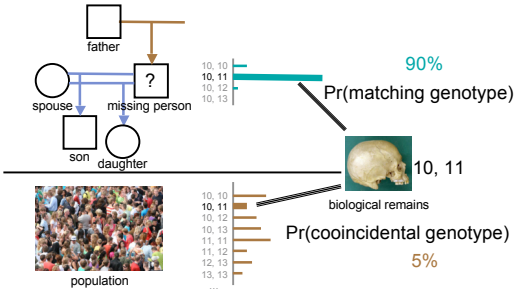
Likelihood ratio

$$= \frac{50\%}{5\%} = 10$$

log(LR) is a standard measure of information

$\log_{10}(10) = 1$ information unit

Kinship



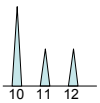
Phenotype

evidence

reality

Phenotype

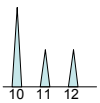
evidence → Lab → STR data



reality observation

Phenotype

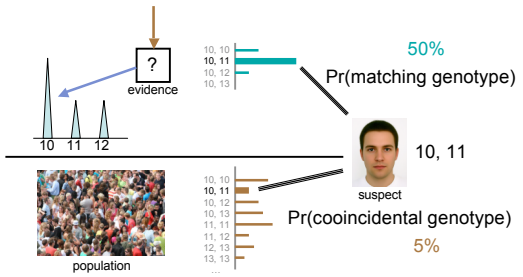
evidence → Lab → STR data → Infer → genotype



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

reality observation model

Phenotype

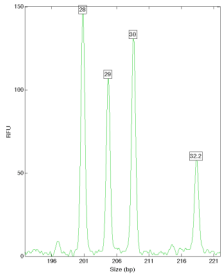


Computer Interpretation

- quantitative computer interpretation
- statistical search of probability model
- preserve all identification information
- objectively infer genotype, then match

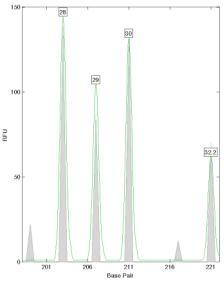
- any number of mixture contributors
- stutter, imbalance, degraded DNA
- calculate uncertainty of every peak

Quantitative Data



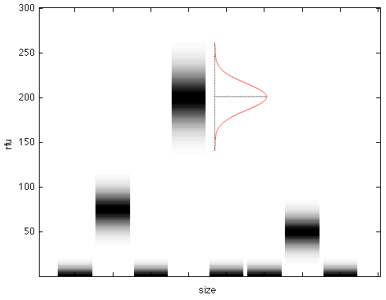
Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. Journal of Forensic Sciences, 2001.

Quantitative Interpretation



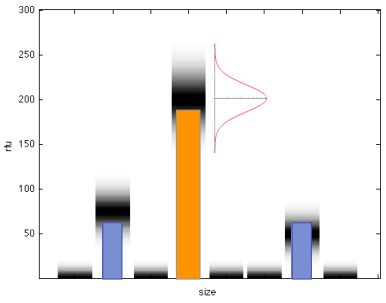
MW Perlin, A Sinelnikov. An information gap in DNA evidence interpretation. PLoS ONE, 2009.

Calculate Peak Uncertainty

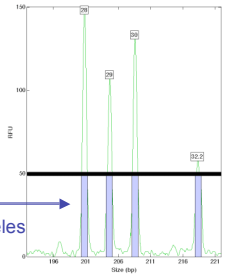


MW Perlin, MM Legler, CE Spencer, JL Smith, WP Allan, JL Belrose, BW Duceman. Validating TrueAllele DNA mixture interpretation. Journal of Forensic Sciences, 2011.

Infer Accurate Genotype



Qualitative Thresholds



list of included alleles

Over threshold, peaks are treated as allele events.

Under threshold, alleles do not exist.

Qualitative Method Variation

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	87,500	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)

The DNA Investigator™ Newsletter, 2009
Same Data, More Information – Murder, Match and DNA

Fingernail: 7% Mixture

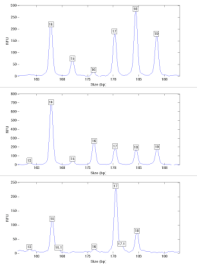
Commonwealth v. Foley

Score	Method
13 thousand	inclusion
23 million	use victim
189 billion	quantitative

- probability modeling preserves information
- peak threshold discards information

More Data, More Information

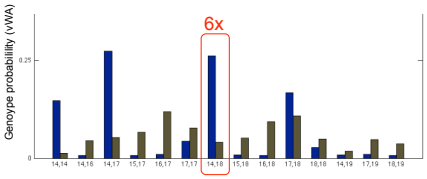
vWA locus data



- low template mixture
- three DNA contributors
- triplicate amplification
- post-PCR enhancement

- no match score found
- computer interpretation
- joint likelihood function

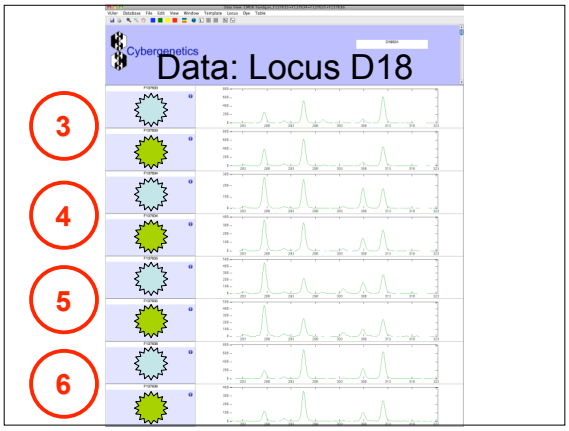
Information

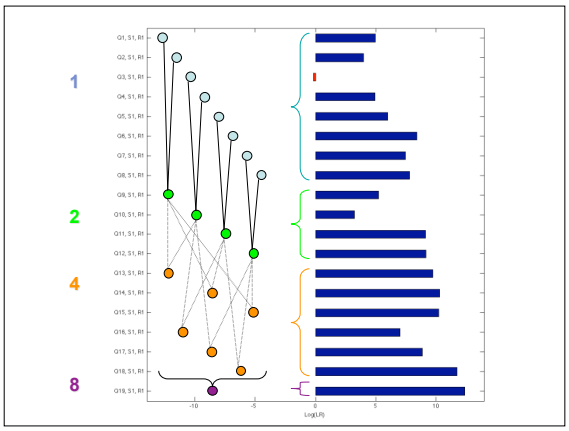


Assume theta = 1%, and three contributors.
A match between the suspect and the evidence
is 3,620,000 times more probable than coincidence.

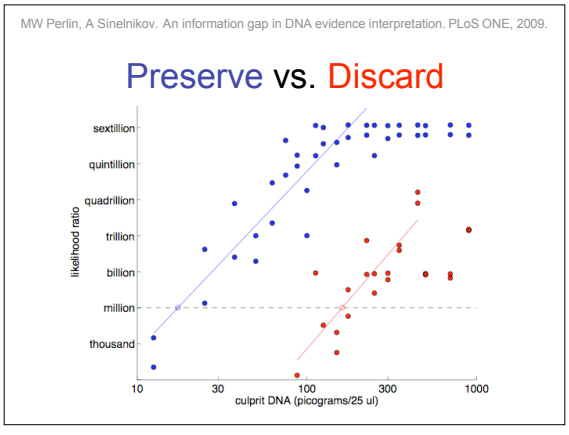
Perlin MW, Greenhalgh M. Scientific combination of DNA evidence: a handgun mixture in eight parts. Twentieth International Symposium on the Forensic Sciences of the Australian and New Zealand Forensic Science Society, Sydney, Australia. 2010.

More Data, More Information



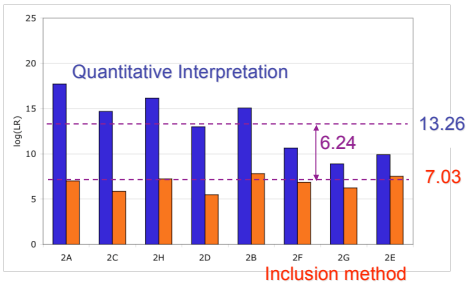


MW Perlin, A Sinelnikov. An information gap in DNA evidence interpretation. PLoS ONE, 2009.



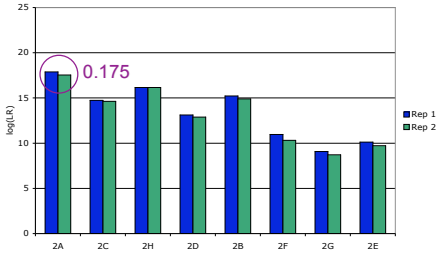
MW Perlin, MM Legler, CE Spencer, JL Smith, WP Allan, JL Belrose, BW Duceman. Validating TrueAllele DNA mixture interpretation. Journal of Forensic Sciences, 2011.

Preserve vs. Discard



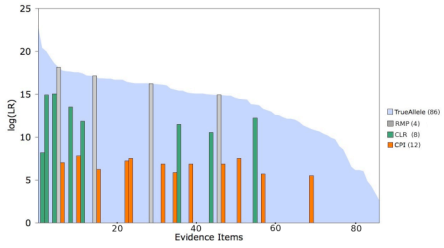
MW Perlin, MM Legler, CE Spencer, JL Smith, WP Allan, JL Belrose, BW Duceman. Validating TrueAllele DNA mixture interpretation. Journal of Forensic Sciences, 2011.

Validated Reproducibility



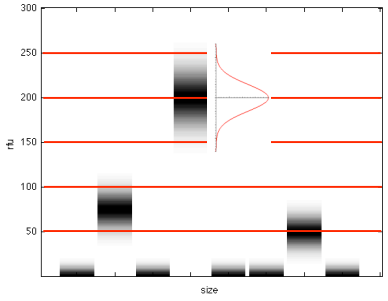
Perlin MW, Duceman BW. Profiles in productivity: greater yield at lower cost with computer DNA interpretation. Twentieth International Symposium on the Forensic Sciences of the Australian and New Zealand Forensic Science Society, Sydney, Australia. 2010.

Preserve vs. Discard

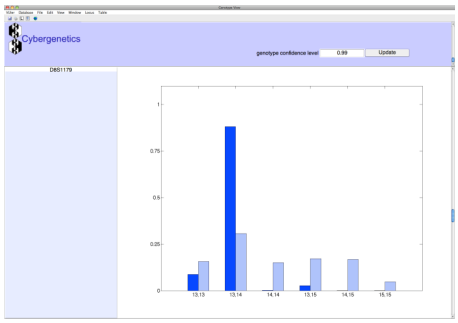


- quantitative interpretation preserves information - every time
- peak threshold discards information - 70% of the time

Data Peak Height is a Random Variable with Mean and Variance

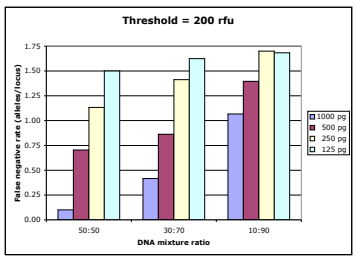


Accurate vs. Dispersed Genotype Probability



Perlin MW. Reliable interpretation of stochastic DNA evidence. Canadian Society of Forensic Sciences 57th Annual Meeting; Toronto, ON. 2010.

Missed Allele Error > 100%



MW Perlin, JB Kadane, RW Cotton. Match likelihood ratio for uncertain genotypes. Law, Probability and Risk, 2009.

Investigative DNA Database

- *genotype probability* representation (not alleles)
- fully *preserves* DNA identification *information*
- enables *LR calculation* with every match
  - connect crimes to criminals
  - disaster victim identification (WTC)
  - find missing people
  - automatic familial search
  - combat terrorism through DNA

The DNA Investigator™ Newsletter, 2011
DNA Intelligence and Forensic Failure – What you don't know can kill you

Societal Consequences

- much informative DNA evidence goes unused
- lost in "interim" qualitative interpretation methods, thresholds applied for analyst convenience
- DNA databases (e.g., CODIS) lose information, using inclusion-based "alleles", instead of ISFG's preferred LR-based "genotype" probability

Inaccurate, uninformative DNA interpretation:
- DNA analysts lose genotypes
- prosecutors lose criminal cases
- databases lose investigative leads
- innocent law abiding citizens lose lives

M Perlin; P Gill, J Buckleton, B Budowle, A van Daal. Low template DNA controversy. Twentieth International Symposium on the Forensic Sciences of the Australian and New Zealand Forensic Science Society, Sydney, Australia. 2010.

International Consensus

1. DNA data is continuous, and has random variation
2. Thresholds do not work for low template DNA
3. Mathematical models can account for random variation

4. The 21st century might be a good time to move away from potentially biased human review of low level (or almost any) DNA data to some sort of objective computer interpretation that can infer genotypes up to probability, without ever looking at suspects, that gives some (possibly uninformative) objective answer.

Strengthening Forensic Science

- objective, reliable, consistent interpretation
- preserve all available information, from the scene of crime to court
- probabilistic genotypes
- modern statistics and computation
- solid mathematical foundation
- platinum standard for preserving DNA information
- move beyond less informative interim methods

Cybergenetics

http://www.cybgen.com/information
perlin@cybgen.com

Learning More

The science of quantitative DNA mixture interpretation

www.cybgen.com/information

- Newsletters
  gentle introduction to ideas
- Courses
  for scientists and lawyers
- Presentations
  handouts, movies, transcripts
- Publications
  abstracts, manuscripts