The Science of Quantitative DNA Mixture Interpretation

SWGDAM Meeting
January, 2011
Fredericksburg, VA

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

Cybergenetics

TrueAllele® Casework

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

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

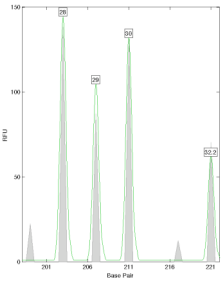
- created in 1999, now in version 25
- used on 100,000 evidence samples
- available as product, service or both

Question 1

What advancements have been made in your casework software that would make the software eligible to be a casework expert system?

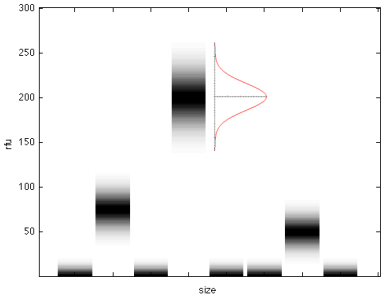
Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. J Forensic Sci. 2001;46(6):1372-7.

Model Quantitative Data



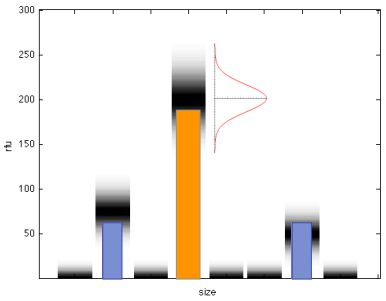
MW Perlin, A Sinelnikov. An information gap in DNA evidence interpretation. PLoS ONE, 2009.

Model 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



Perlin MW. Explaining the likelihood ratio in DNA mixture interpretation.
Promega's Twenty First International Symposium on Human Identification, 2010; San Antonio, TX.

Report Likelihood Ratio

How much more does
the suspect match the evidence
than some random person?

$$\text{information gain in DNA match} = \frac{\text{Prob(evidence match)}}{\text{Prob(coincidental match)}}$$

A match between the suspect and the evidence
is a billion times more probable than
a coincidental match.

Question 2

Are there issues that are still being
addressed prior to releasing to
laboratories for this consideration?

TrueAllele® System

ViewStation
User Client

Database
Server

Interpret/Match
Expansion

Visual User Interface
VUIer™ Software

Parallel Processing Computers

Cybergenetics Support

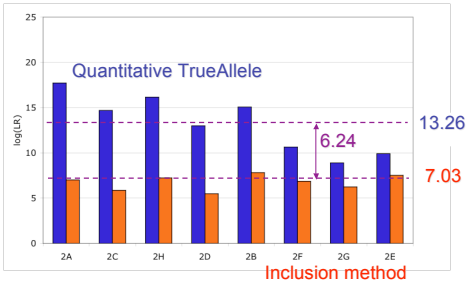
- process planning
- science education
- software training
- user documentation
- data assessment
- validation services
- ongoing meetings
- project management
- workflow integration
- testifying support

Question 3

Why can it be relied upon as a casework expert system?

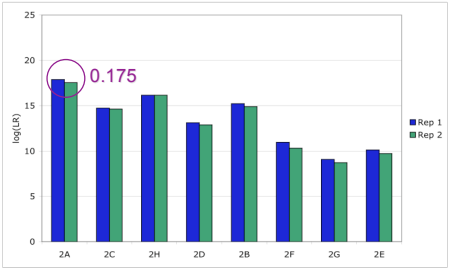
MW Perlin, MM Legler, CE Spencer, JL Smith, WP Allan, JL Belrose, BW Duceman. Validating TrueAllele DNA Mixture Interpretation. Journal of Forensic Sciences, 2011.

Validated: Information Efficacy



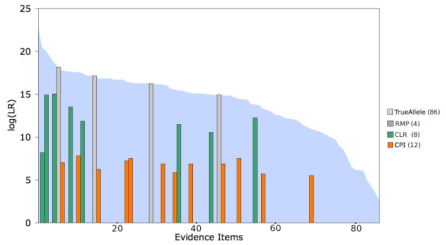
MW Perlin, MM Legler, CE Spencer, JL Smith, WP Allan, JL Belrose, BW Duceman. Validating TrueAllele DNA Mixture Interpretation. Journal of Forensic Sciences, 2011.

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.

Preserves DNA Information



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

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

Courtroom Admissibility

Commonwealth v. Foley

Score	Method
13 thousand	inclusion
23 million	obligate allele
189 billion	TrueAllele

- probability modeling preserves information
- peak threshold discards information

Reliable Weight of Evidence

When labs cannot report a match score, Cybergenetics can run TrueAllele to infer (and testify to) a likelihood ratio.

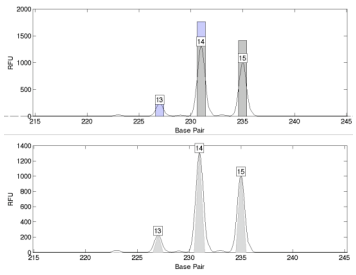
United States
- Maryland
- Pennsylvania
- South Carolina

International
- England
- Northern Ireland
- The Netherlands

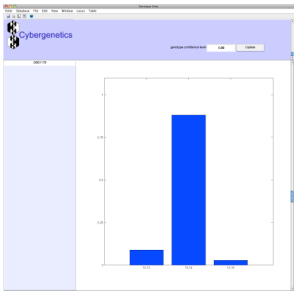
Question 4

Specifically, how are mixtures addressed?

Consider All Possible Explanations

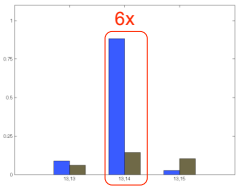


Genotype Probability Distribution

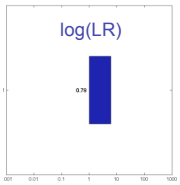


Perlin MW. Explaining the likelihood ratio in DNA mixture interpretation.
Promega's Twenty First International Symposium on Human Identification, 2010; San Antonio, TX.

Information: Likelihood Ratio



posterior to prior ratio
of genotype probability

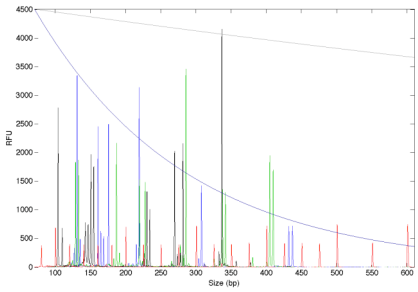


evidence to coincidence ratio
of DNA match probability

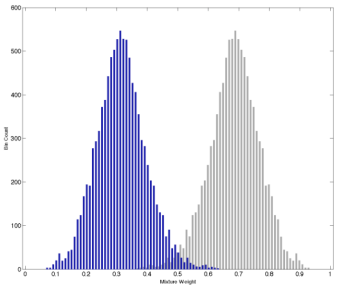
Question 5

How are peak height differences addressed in terms of degradation vs. mixture?

Degradation Parameter



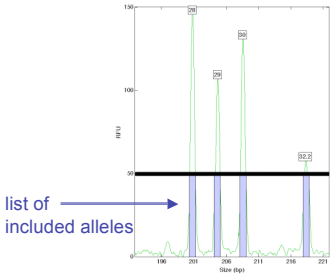
Mixture Weight Parameter



Question A

Why don't thresholds work?

Wrong Data, Wrong Model

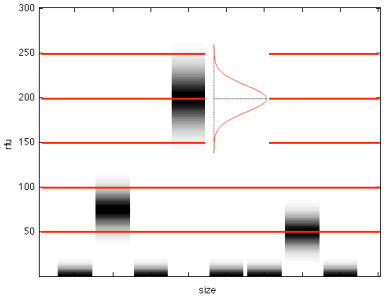


list of included alleles

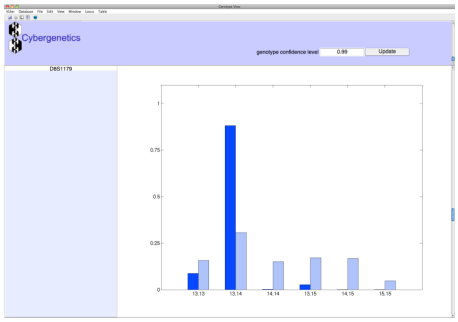
Over threshold, peaks are treated as allele events.

Under threshold, alleles do not exist.

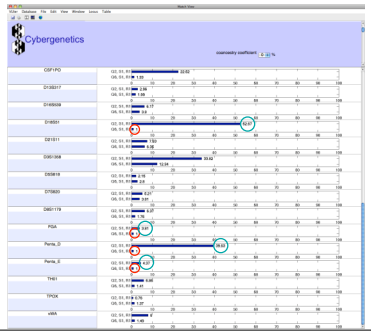
Wrong Peak Uncertainty Model



Disperse Genotype Probability



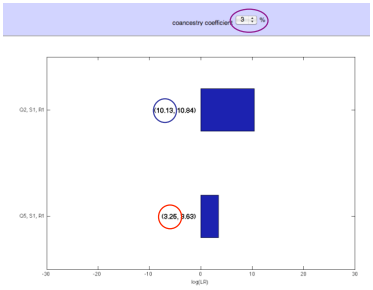
Failure to Preserve Information



TrueAllele
9.46 trillion

Threshold
43 thousand

Defense: Coancestry, Confidence

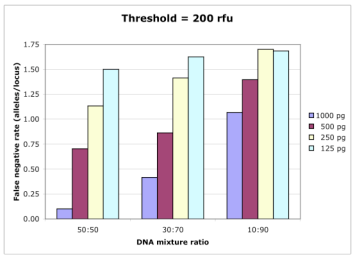


TrueAllele
13 billion

Threshold
1,778

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

Missed Allele Error > 100%



CODIS Information Loss

- CODIS "allele" approach loses information
- TrueAllele Database:
  stores & matches probabilistic genotypes
- LR preserves identification information

  - evidence vs. convicted offender
  - disaster victim identification (WTC)
  - finding missing people
  - automated familial search
  - customizable to each state's statutes

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.

Question B

How can I learn more about
scientific DNA mixture interpretation
that uses all the quantitative data?

Cybergenetics Resources

The science of quantitative DNA mixture interpretation

www.cybgen.com/information

- Courses
  for scientists and lawyers
- Presentations
  handouts, movies, transcripts
- Publications
  abstracts, manuscripts

Cybergenetics TrueAllele Services

If you see it, we can solve it™

1. Have some interesting test case mixtures?
   Send the data to Cybergenetics, and we will report back TrueAllele results via webinar.

2. Have an important case that needs an answer?
   Inform your police or prosecutor about TrueAllele.

3. Not ready to replace human review in your lab?
   Complement your own work on challenging cases. TrueAllele interpretation, reporting and testifying.

Conclusions

- quantitative DNA mixture interpretion preserves identification information
- thresholds discard information
- validated, courtroom-tested system
- TrueAllele is available today
- (A) lab-wide system implementation
- (B) forensic intepretation services can complement existing methods

Cybergenetics

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