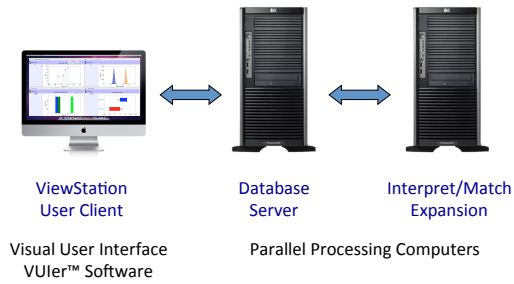




Probabilistic Software Workshop September 29, 2014

TrueAllele® Casework
Mark W. Perlin, PhD, MD, PhD

Cybergenetics TrueAllele® Casework



Design Philosophy

- Use all the data (peak heights, replicates)
- Objective, no examination bias (no suspect)
- One architecture: evidentiary & investigative

- Model STR parameters & variation
- Infer genotypes, then match them
- Likelihood ratio (LR) match statistic

Basic Features

- Visual user interface
- Fast and easy to use
- Flexible workflow (batch, case, confirm)
- Greater productivity with fewer samples

- Consistent and accurate answers
- LR number can include or exclude
- Easy to explain results

Basic Capabilities

- Client (many users)
- Server (central database, parallel computing)
- Solves many problems at once (10 to 100)

- Fast on most mixtures (1-2 hours)
- Thorough on challenging mixtures
- Preserves identification information

Intended Application

- Low-template & degraded DNA
- Mixtures (any number of contributors)
- Kinship & paternity

- Investigative database
- Non-suspect CODIS search
- Familial search
- Disaster victim identification

Input Files

Data

- Original sequencer data file (.fsa, .hid)
- GeneMapper® ID Peak Table Export (.txt)

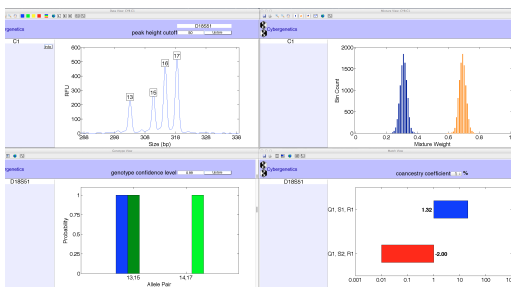
Genotype

- Reference profile (.txt)
- Kinship pedigree file (.txt)

Population

- Many populations included (FBI, NIST, country, state, ...)
- Customizable allele frequency database (.txt)

Visual User Interaction



Likelihood Ratio

Likelihood ratio (LR) requires genotype probability

$$LR = \frac{O(H | \text{data})}{O(H)}$$

Bayes theorem + probability + algebra ...

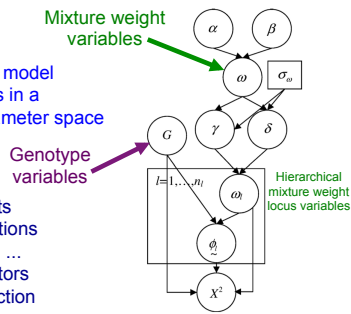
$$= \frac{\sum_x P\{d_x|X=x, \dots\} P\{d_y|Y=x, \dots\} P\{X=x\}}{\sum_{x,y} P\{d_x|X=x, \dots\} P\{d_y|Y=y, \dots\} P\{X=x, Y=y\}}$$

genotype probability: posterior, likelihood & prior

Genotype Inference

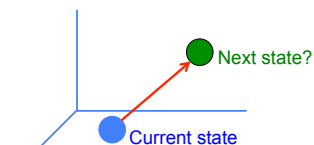
Hierarchical Bayesian model induces a set of forces in a high-dimensional parameter space

- small DNA amounts
- degraded contributions
- K = 1, 2, 3, 4, 5, 6, ...
- unknown contributors
- joint likelihood function



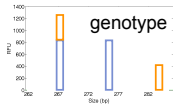
Markov chain Monte Carlo

Sample from the posterior probability distribution

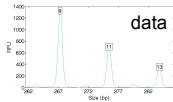


$$\text{Transition probability} = \frac{P\{\text{Next state}\}}{P\{\text{Current state}\}}$$

Modeling STR Data Variation



Hierarchy of successive pattern transformations



Variance parameters

Hierarchical
(e.g., customized for DNA template or locus)

Differential degradation
Mixture weight
Relative amplification
PCR stutter
PCR peak height
Background noise

Types of DNA Profiles

- Simple mixtures (e.g., 2-3 contributors)
- Low-template DNA mixtures
- Low minor contributors (e.g., 5%-15%)
- Differentially degraded mixtures
- Multiple amplifications, jointly analyzed
- Many contributors (e.g., 4, 5, 6, ...)

STR Kits

- PowerPlex® 16
- PowerPlex® 21
- PowerPlex® Fusion
- Profiler®, COfiler® & Profiler Plus®
- Identifier® & Identifier® Plus
- GlobalFiler™
- SGM plus®
- MiniFiler™
- IDplex

Genetic Analyzers

- ABI 310
- ABI 3100
- ABI 3100-Avant
- ABI 3130
- ABI 3130xl
- ABI 3500
- ABI 3500xl
- ABI 3700
- ABI 3730

Published Validation Studies

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

Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW. Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences*. 2011;56(6):1430-47.

Ballantyne J, Hanson EK, Perlin MW. DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information. *Science & Justice*. 2013;53(2):103-14.

Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele® Casework validation study. *Journal of Forensic Sciences*. 2013;58(6):1458-66.

Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S. TrueAllele® Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLoS ONE*. 2014;(9)3:e92837.

Perlin MW, Hornyak J, Sugimoto G, Miller K. TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors. *Journal of Forensic Sciences*. 2015;in press.

TrueAllele Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S. *PLoS ONE* (2014) 9(3): e92837

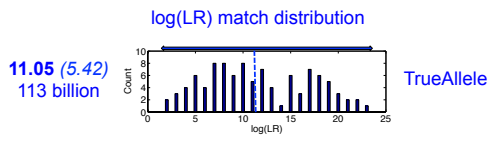
Sensitive

The extent to which interpretation identifies the correct person

True DNA mixture inclusions

101 reported genotype matches
82 with DNA statistic over a million

TrueAllele Sensitivity



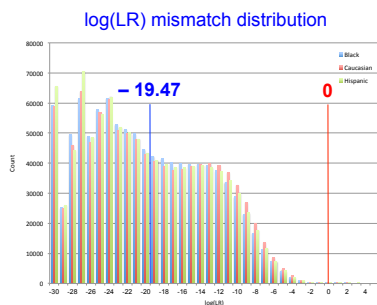
Specific

The extent to which interpretation does not misidentify the wrong person

True exclusions, without false inclusions

101 matching genotypes x 10,000 random references
x 3 ethnic populations,
for over 1,000,000 nonmatching comparisons

TrueAllele Specificity



Reproducible

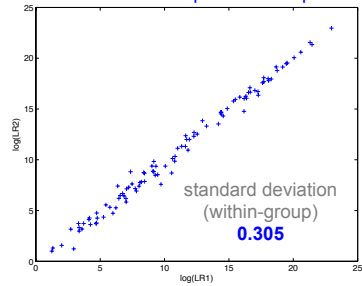
The extent to which interpretation gives the same answer to the same question

MCMC computing has sampling variation

duplicate computer runs
on 101 matching genotypes
measure log(LR) variation

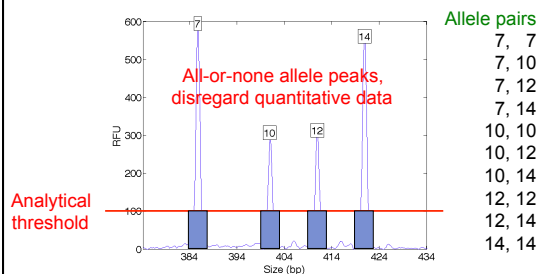
TrueAllele Reproducibility

Concordance in two independent computer runs



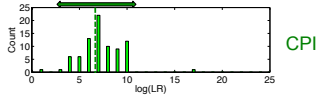
Manual Inclusion Method

Over threshold, peaks become binary allele events



CPI Information

6.83 (2.22)
6.68 million

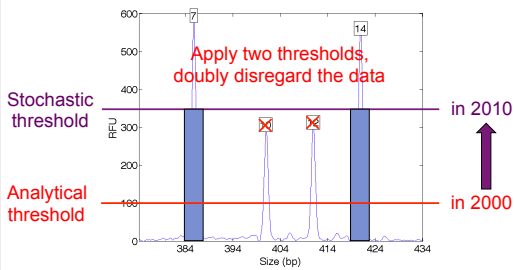


Combined probability of inclusion

Simplify data, easy procedure,
apply simple formula
 $PI = (p_1 + p_2 + \dots + p_k)^2$

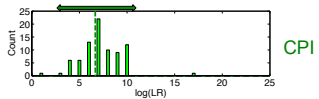
Modified Inclusion Method

Higher threshold for human review

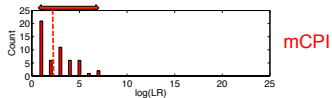


Modified CPI Information

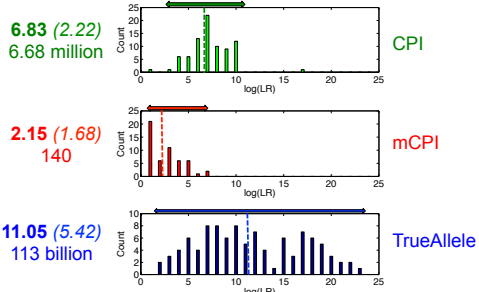
6.83 (2.22)
6.68 million



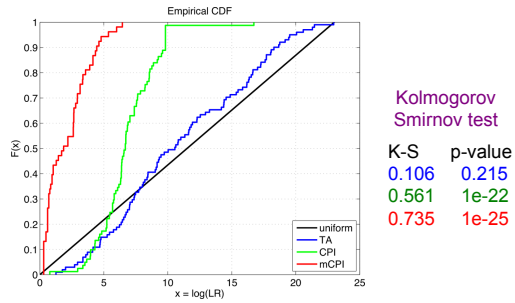
2.15 (1.68)
140



Method Comparison

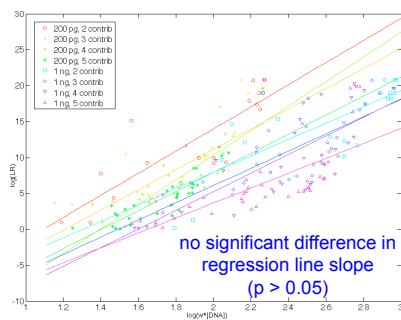


Method Accuracy

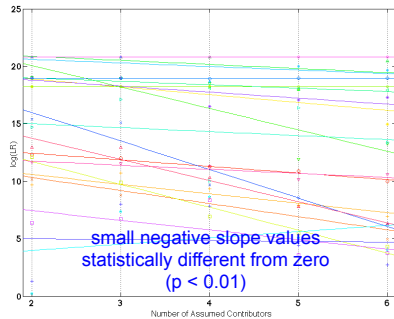


Perlin MW, Hornyak J, Sugimoto G, Miller K.
TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors.
Journal of Forensic Sciences. 2015;in press.

Invariant Behavior



Sufficient Contributors



Training Requirements

- Science & Software
 - Lectures and reading
 - Three day hands-on instruction
- Operator Training
 - Problem solving on challenging data
 - Grading, examinations & certification
- Reporting & Testifying (Optional)

User Support

- User education and training
- Software manuals and procedures
- Validation assistance and service
- Admissibility, reporting & testifying

- Cybergenetics website
- YouTube TrueAllele channel
- Software, hardware & networking
- On-site and remote (by phone or Internet)

Computer Hardware

System Requirements (For VUIer™ Client Software)

| | <i>Windows</i> | <i>Mac</i> |
|------------------|-----------------|--|
| Operating System | Windows XP | Mac OS X v10.6 Snow Leopard |
| | Windows 7 | Mac OS X v10.7 Lion |
| | | Mac OS X v10.8 Mountain Lion Mac OS X v10.9 Mavericks |
| Processor | At least 1 GHz | At least 1 GHz Intel (Core 2 duo) |
| Memory | At least 256 MB | At least 1 GB |

Future Updates

- 1999. Version 1
- 2004. Model refinement
- 2009. Version 25, deployment
- 2014. Cloud computing

Interpret and identify
anywhere, anytime



Your cloud, or ours

Cybergenetics Experience

| | |
|-------------------------------|-----------------|
| Invented math & algorithms | 20 years |
| Developed computer systems | 15 years |
| Support users and workflow | 10 laboratories |
| Used routinely in casework | 3 labs |
| Validate system reliability | 20 studies |
| Educate the community | 50 talks |
| Train & certify analysts | 200 students |
| Go to court for admissibility | 5 hearings |
| Testify about LR results | 20 trials |
| Educate lawyers and laymen | 1,000 people |
| Make the ideas understandable | 200 reports |

Admissibility Hearings

- California
- Pennsylvania
- Virginia
- United Kingdom
- Australia

Appellate precedent in Pennsylvania

TrueAllele in Criminal Trials

About 200 case reports filed on DNA evidence

Court testimony:

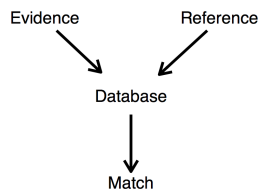
- state
- federal
- military
- international

Crimes:

- armed robbery
- child abduction
- child molestation
- murder
- rape
- terrorism
- weapons

Investigative DNA Database

Infer genotypes, and then match with LR



World Trade Center disaster

Further Information

<http://www.cybgen.com/information>



- Courses
- Newsletters
- Newsroom
- Patents
- Presentations
- Publications
- Webinars

<http://www.youtube.com/user/TrueAllele>
TrueAllele YouTube channel



Cybergenetics



perlin@cybgen.com
