



Probabilistic Software Workshop September 29, 2014

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

DNA Mixtures: A Separation Problem



- Multiple people combine their DNA

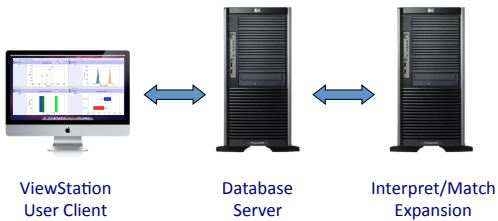


- Laboratory biological separation
extract DNA, amplify, electrophorese



- Computer data separation
infer each person's genotype

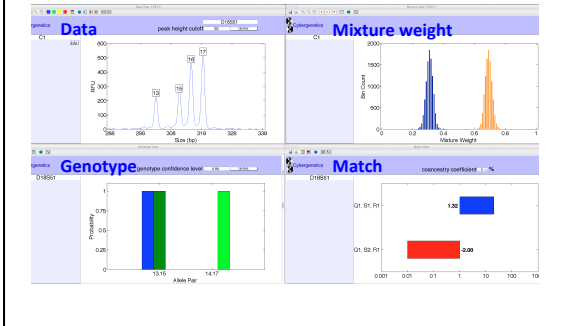
Cybergenetics TrueAllele® Casework



Visual User Interface
VUIer™ Software

Parallel Processing Computers

Visual User Interaction



Development History

- 1999. Version 1
Two hours to write, two seconds to run
Published math, filed patents
- 2004. Refine probability model
Expand hierarchy and variance parameters
Focus: accuracy and robustness
- 2009. Deploy version 25
Continued validation, routine application
Focus: workflow and ease-of-use
- 2014. Growing user community

Design Philosophy

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

Likelihood Ratio

Likelihood ratio (LR) can use separated genotypes

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

Bayes theorem + probability + algebra ...

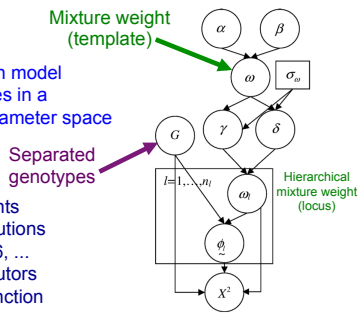
$$= \frac{\sum_x P\{d_x|X=x, \dots\} P\{d_y|Y=x, \dots\} P\{X=x\}}{\sum_{x,y} P\{d_x|X=x, \dots\} P\{d_y|Y=y, \dots\} P\{X=x, Y=y\}}$$

genotype probability: posterior, likelihood & prior

Genotype Inference

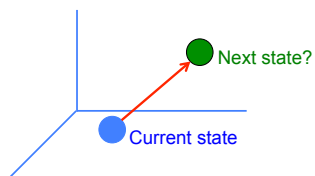
Hierarchical Bayesian model induces a set of forces in a high-dimensional parameter space

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

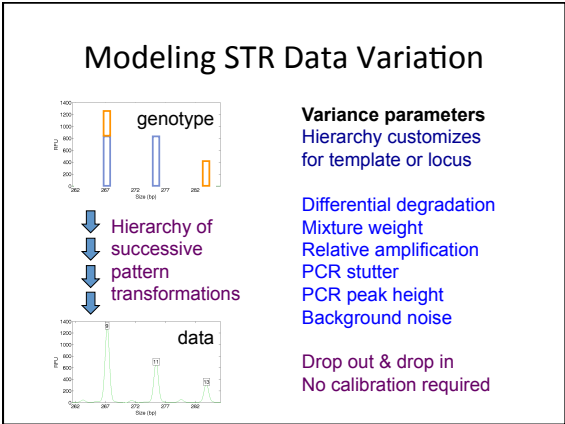


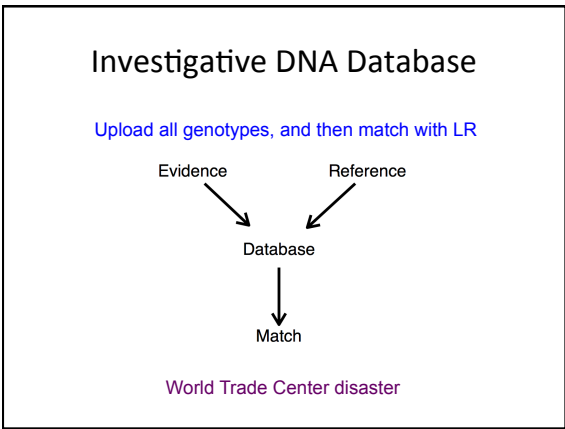
Markov Chain Monte Carlo

Sample from the posterior probability distribution



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





Published Validation Studies

Samples of known composition

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

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, 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.

Greenspoon SA, Schiermeier-Wood L, Jenkins BC. Establishing the limits of TrueAllele® Casework: a validation study. *Journal of Forensic Sciences*. 2015;in press.

Published Validation Studies

Samples from actual casework

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.

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.

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

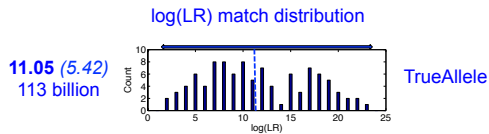
Sensitive

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

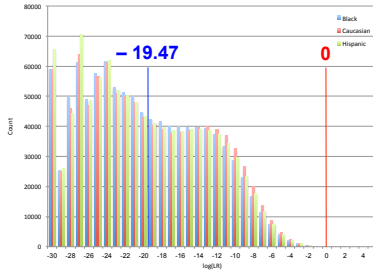
The extent to which interpretation does not misidentify the wrong person

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

log(LR) mismatch distribution



Reproducible

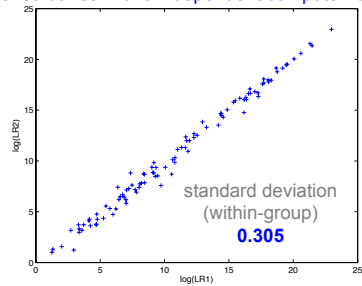
The extent to which interpretation gives the same answer to the same question

MCMC computing has sampling variation

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

TrueAllele Reproducibility

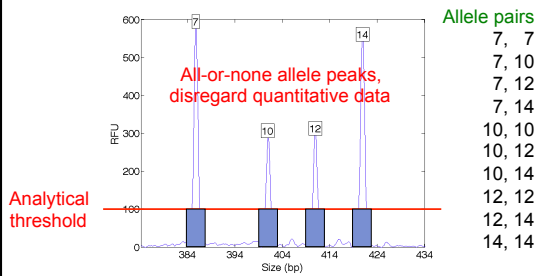
Concordance in two independent computer runs



Manual Inclusion Method

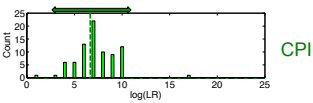
Over threshold, peaks become binary allele events

<https://soundcloud.com/markperlin/threshold>



CPI Information

6.83 (2.22)
6.68 million

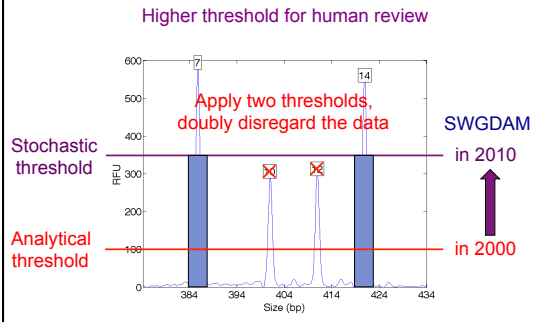


Combined probability of inclusion

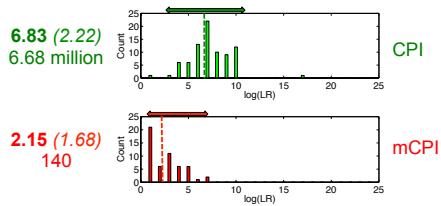
Simplify data, easy procedure,
apply simple formula

$$PI = (p_1 + p_2 + \dots + p_k)^2$$

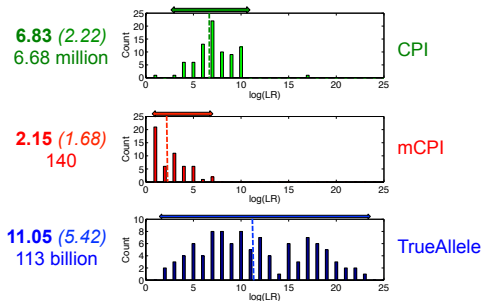
Modified Inclusion Method

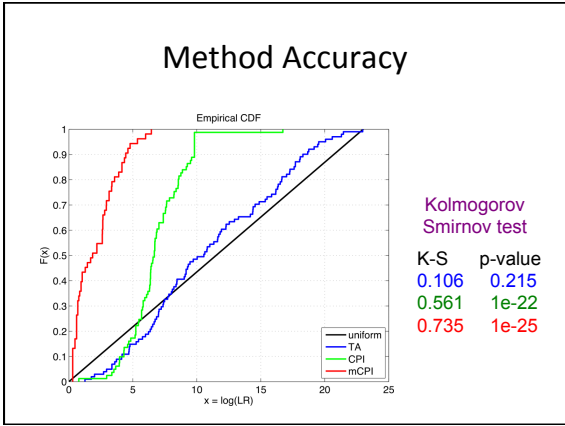


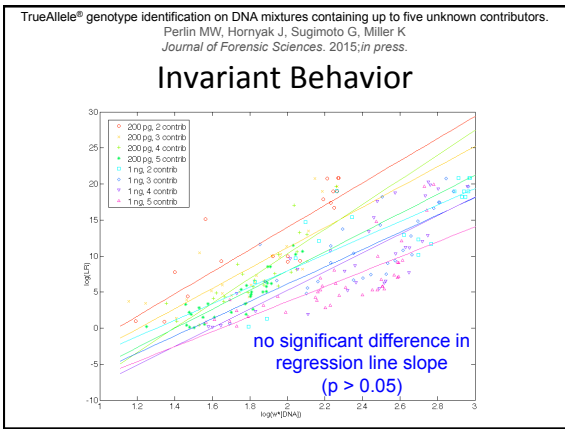
Modified CPI Information

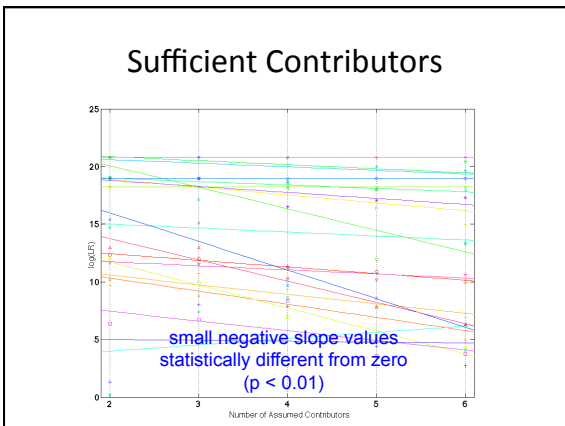


Method Comparison









MIX13: An interlaboratory study on the present state of DNA mixture interpretation in the U.S.
 Coble M, National Institute of Standards and Technology
 5th Annual Prescription for Criminal Justice Forensics, Fordham University School of Law, 2014.

NIST MIX13 Study

MIX13 Case 5 Outcomes with Suspect C (whose genotypes were not present in the mixture)

# Labs	Report Conclusions	Reasons given
6	Exclude Suspect C	detailed genotype checks (D+); TrueAllele negative LR (D+); assumed major/minor and suspects did not fit (D+); 3 labs noted Penia E missing allele 15 (PP16HS)
3	Inconclusive with C only (A & B included)	All these labs used PP16HS
21	Inconclusive for A, B, and C	
70	Include & provide CPI statistics	All over the road...

Range of CPI stats for Caucasian population:
 FBI allele frequencies: **1 in 9** to **1 in 344,000**

An investigation of software programs using "semi-continuous" and "continuous"
 methods for complex DNA mixture interpretation.
 Coble M, Myers S, Klaver J, Kloosterman A
 9th International Conference on Forensic Inference and Statistics, 2014.

Other Comparisons

Quantity	Fraction	Drops	TrueAllele	Continuous	DropOutA	DropOutB
1.00 ng	50%	0	11.89	12.10	9.55	9.68
1.00 ng	10%	2	13.49	11.47	6.75	4.77
0.50 ng	10%	8	8.74	0.48	1.75	-0.14
0.25 ng	10%	14	7.21	0.25	0.57	-1.70

Limited LR methods do not separate out mixed genotypes

$$LR = \frac{P\{\text{data} | H_p\}}{P\{\text{data} | H_d\}} \quad \text{Better: separate the genotypes}$$

Admissibility Hearings

- California
- Louisiana
- Maryland
- New York
- Ohio
- Pennsylvania
- Virginia

- United Kingdom
- Australia

Appellate precedent in Pennsylvania

Genotype Peeling

ISHI workshop-provided three person mixture data

1. Assume nothing, identify major contributor
2. Assume major, identify 1st minor contributor
3. Assume major and 1st minor, identify 2nd minor

Contributor	Ratio	Weight	Assumed Knowns	Cycles	Car Owner	Person 1a	Suspect 2	HH:MM
major	4	0.49		500	9.77			00:12
minor 1	3	0.32	Car Owner	1000		12.58		00:27
minor 2	1	0.19	Car Owner, Person 1a	5000			4.38	01:27

Used in casework to separate up to five related contributors

TrueAllele in Criminal Trials

About 200 case reports filed on DNA evidence

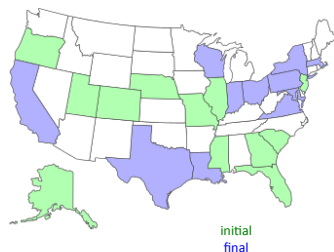
Court testimony:

- state
- federal
- military
- international

Crimes:

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

TrueAllele Case Reports



People of New York v Casey Wilson

Serial rapist in Elmira, New York

- Due to insufficient genetic information, no comparisons were made to the minor contributors of this profile.
- Due to the complexity of the genetic information, no comparisons were made to this profile.

December 11, 2013: crime lab emails data late afternoon
TrueAllele peeling in the evening
preliminary report issued that night

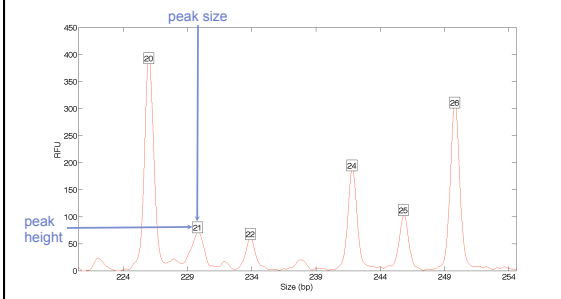
December 19, 2013: Cybergeneics testifies at Grand Jury
September 11, 2014: Cybergeneics testifies at trial

Poster #105

TrueAllele speed for Grand Jury need: same day reporting of complex mixtures

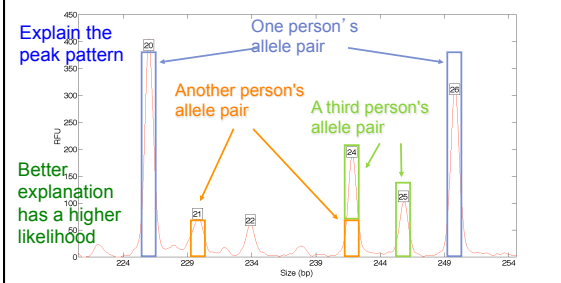
Computers can use all the data

Quantitative peak heights at locus FGA



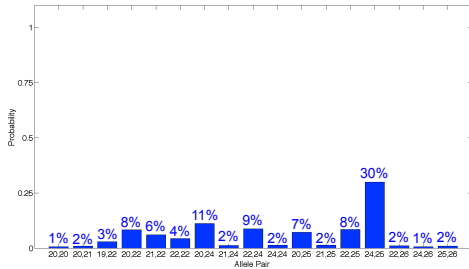
How the computer thinks

Consider every possible genotype solution



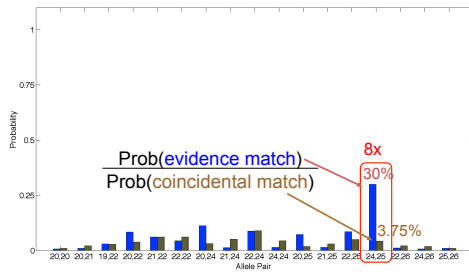
Evidence genotype

Objective genotype determined solely from the DNA data.
Never sees a reference.

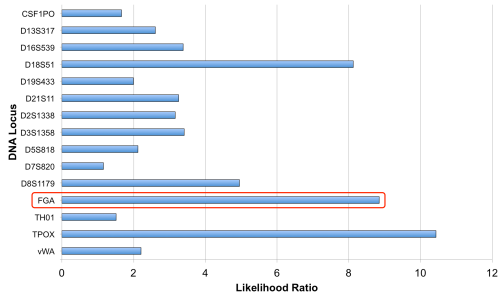


DNA match information

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



Match information at 15 loci



Match statistics

Item	Description	15B	24A	20A
		Victim	Elimination	Defendant
17D-E	Purple knit glove	930 quadrillion	1/2.72	817 thousand
18D-E	Purple knit glove	520 trillion	14.6 thousand	31.3 million

A match between the glove and Casey Wilson is 31.3 million times more probable than coincidence.

September 12, 2014: Casey Wilson convicted on all charges

DNA Mixture Crisis



375 cases/year x 4 years = 1,500 cases
320 M in US / 8 M in VA = 40 factor
1,500 cases x 40 factor = 60,000 inconclusive



1,000 cases/year x 4 years = 4,000 cases
320 M in US / 8 M in NY = 40 factor
4,000 cases x 40 factor = 160,000 inconclusive



+ under reporting of DNA match statistics

DNA evidence data in 100,000 cases
Collected, analyzed & paid for – but unused

Kern County Workflow



Poster #104



Kern County Resolves the DNA Mixture Crisis
By [Name] and [Name]

Introduction
 The DNA mixture crisis (DMC) has become a significant forensic challenge for law enforcement agencies in California and across the United States. This paper discusses the current state of DMC, the challenges it poses for forensic laboratories, and the solutions that have been implemented in Kern County to address this crisis.

Background
 The DMC is a complex issue that arises from the combination of DNA evidence from multiple individuals. This is particularly challenging in cases where the DNA evidence is mixed with that of other individuals, making it difficult to identify the individual(s) responsible for the crime.

Challenges
 The DMC poses several challenges for forensic laboratories, including the need for more sophisticated DNA analysis techniques, the need for more personnel to handle the increased volume of DNA evidence, and the need for more resources to support the analysis of DNA evidence.

Solutions
 Kern County has implemented several solutions to address the DMC, including the use of more sophisticated DNA analysis techniques, the hiring of additional personnel, and the implementation of a new DNA analysis workflow.

Conclusion
 The DMC is a complex issue that requires a multi-faceted approach. Kern County's implementation of a new DNA analysis workflow has been successful in resolving the DMC and ensuring that DNA evidence is analyzed accurately and efficiently.

TrueAllele User Meeting

California
Louisiana
Maryland
Massachusetts
New York
Pennsylvania
South Carolina
Virginia
Australia
Oman
Prosecutors



Bear Mountain Inn, New York
September, 2014

Consistent results on MIX13 data across groups

TrueAllele Cloud



Your cloud, or ours

Interpret and identify
anywhere, anytime

- Crime laboratory
 - Training
 - Validation
 - Spare capacity
 - Rent instead of buy
- Solve unreported cases
- Prosecutors & police
- Defense transparency
- Forensic education

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
