TrueAllele® Genetic Calculator: Implementation in the NYSP Crime Laboratory

NYS DNA Subcommittee
May 19, 2010

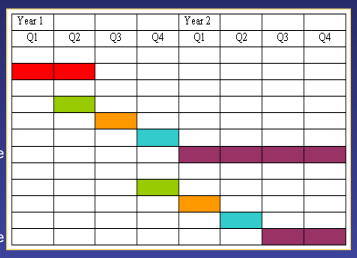
Barry Duceman, Ph.D
New York State Police
Forensic Investigation Center

Implementation Timeline

Establish core group

Identify first case type
Side-by-side comparison
Assessment of comparison
Deploy TrueAllele case type

Identify second case type
Side-by-side comparison
Assessment of comparison
Deploy TrueAllele case type



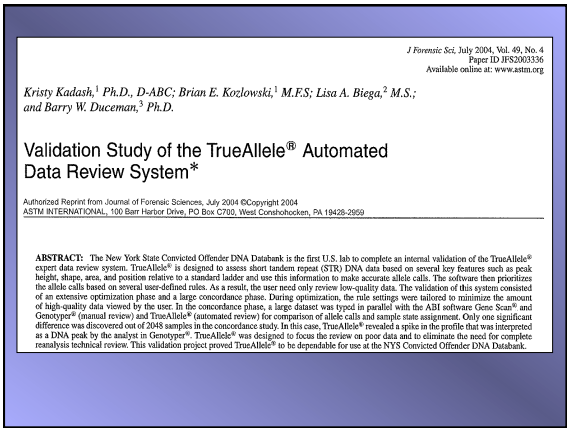
Deployment Metrics

Concordance
Productivity
Information Content
Objectivity
Ease of Use

Software Concept

A software tool that will provide…

- Data quality assessment



J Forensic Sci, July 2004, Vol. 49, No. 4
Paper ID JFS2003336
Available online at: www.astm.org

Kristy Kadash,[1] Ph.D., D-ABC; Brian E. Kozlowski,[1] M.F.S; Lisa A. Biega,[2] M.S.; and Barry W. Duceman,[3] Ph.D.

Validation Study of the TrueAllele® Automated Data Review System*

Authorized Reprint from Journal of Forensic Sciences, July 2004 ©Copyright 2004
ASTM INTERNATIONAL, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959

ABSTRACT: The New York State Convicted Offender DNA Databank is the first U.S. lab to complete an internal validation of the TrueAllele® expert data review system. TrueAllele® is designed to assess short tandem repeat (STR) DNA data based on several key features such as peak height, shape, area, and position relative to a standard ladder and use this information to make accurate allele calls. The software then prioritizes the allele calls based on several user-defined rules. As a result, the user need only review low-quality data. The validation of this system consisted of an extensive optimization phase and a large concordance phase. During optimization, the rule settings were tailored to minimize the amount of high-quality data viewed by the user. In the concordance phase, a large dataset was typed in parallel with the ABI software GeneScan® and Genotyper® (manual review) and TrueAllele® (automated review) for comparison of allele calls and sample state assignment. Only one significant difference was discovered out of 2048 samples in the concordance study. In this case, TrueAllele® revealed a spike in the profile that was interpreted as a DNA peak by the analyst in Genotyper®. TrueAllele® was designed to focus the review on poor data and to eliminate the need for complete reanalysis technical review. This validation project proved TrueAllele® to be dependable for use at the NYS Convicted Offender DNA Databank.

Software Concept

A software tool that will provide…

- Data quality assessment
- Unattended parallel processing of multiple data inquiries.
- Access to the relevant data at any point in the analysis.
- Challenge to the software responses with alternative scenarios.
- Resolves mixtures
- Maximum information from the data

Standard 3.2.2.

• • • The laboratory must establish alternative criteria (e.g., quantitation values or use of a probabilistic genotype approach) for addressing potential stochastic amplification. The criteria must be supported by empirical data and internal validation and must be documented in the standard operating procedures.

– SWGDAM Interpretation Guideline for Autosomal STR Typing

Software Concept

A software tool that will provide…

- Data quality assessment
- Unattended parallel processing of multiple data inquiries.
- Access to the relevant data at any point in the analysis.
- Challenge to the software responses with alternative scenarios.
- Resolves mixtures
- Maximum information from the data
- Objective genotype inference for STR DNA patterns (including perpetrators contributing to mixtures).

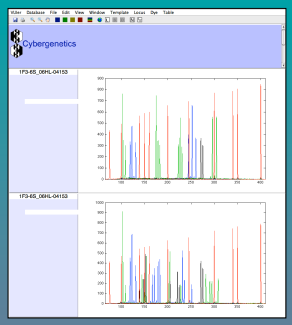
TrueAllele® Process Steps

1. Data
2. Request
3. Solve
4. Review
5. Explain
6. Report

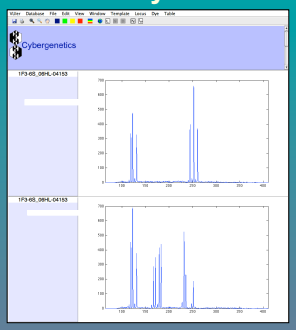
Easy Mixture Case

- 50:50 Mixture Data
- One Unknown Contributor
- Victim Reference
- Suspect Reference

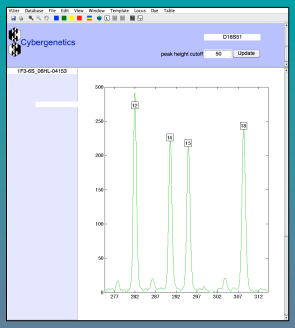
1. STR Evidence Data



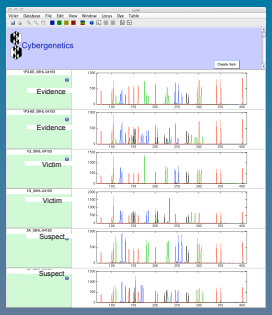
5-FAM Dye View



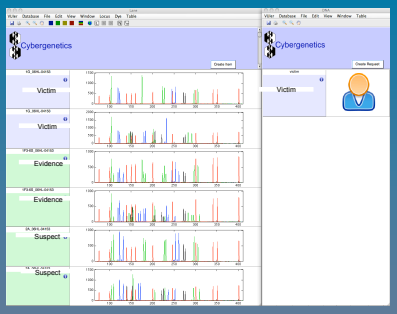
D18S51 Locus



2. Request



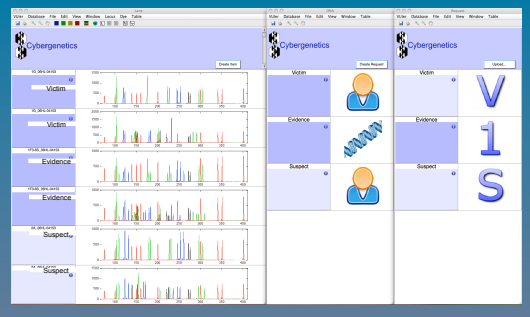
Create Victim Item



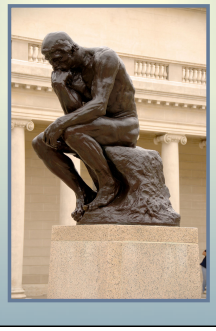
Create Victim Request

Create Evidence Request

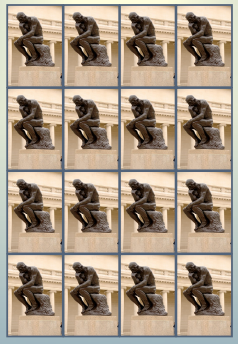
Interpretation Requests



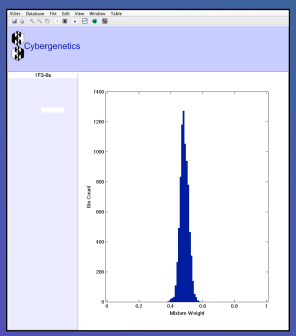
3. Solving with TrueAllele



Parallel Interpretation

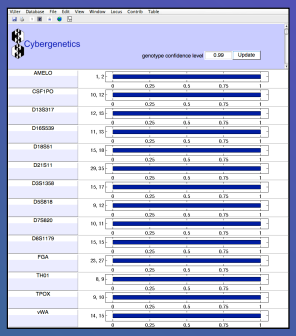


4. Review: Mixture Weight

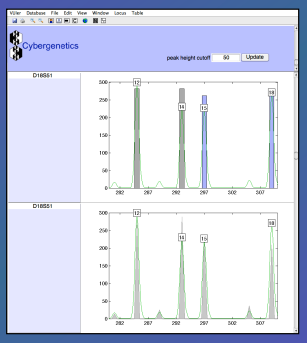


49%

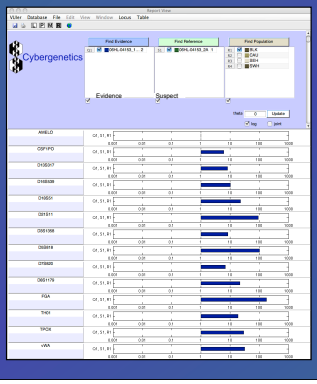
Definite Genotype



5. Explain Reasoning



6. Report: Likelihood Ratio



log(LR)
17.65

log(CLR)
13.52

TrueAllele LR Calculation

File Signature Statement Summary Calculation

The LR calculation assumes one unknown contributor in the evidence with one known contributor reference relative to a African-American human population.
The likelihood ratio for the identification hypothesis is 456 quadrillion.

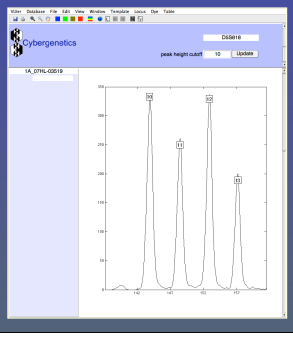
The joint LR is approximately 456 quadrillion.
The log(LR) information is 17.65.

locus	allele pair	Q	R	S	LR	log(LR)
CSF1PO	10, 12	1	0.1621	1	6.167	0.790
D13S317	12, 13	1	0.1209	1	8.272	0.918
D16S539	11, 13	1	0.0967	1	10.339	1.014
D18S51	15, 18	1	0.0443	1	22.589	1.354
D21S11	29, 35	1	0.0107	1	93.586	1.971
D3S1358	15, 17	1	0.1173	1	8.522	0.931
D5S818	9, 12	1	0.0099	1	100.882	2.004
D7S820	10, 11	1	0.1443	1	6.929	0.841
D8S1179	15, 15	1	0.0458	1	21.850	1.339
FGA	23, 27	1	0.0056	1	178.837	2.252
TH01	8, 9	1	0.0537	1	18.616	1.270
TPOX	9, 10	1	0.0338	1	29.587	1.471
VWA	14, 15	1	0.0313	1	31.919	1.504

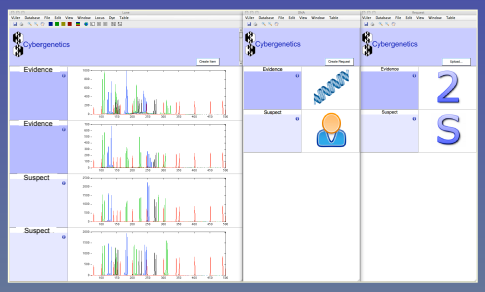
Interesting Mixture Case

- 65:35 Mixture Data
- Two Unknown Contributors
- No Victim Reference
- Suspect Reference

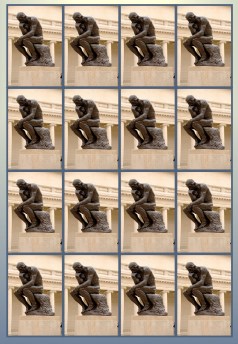
D5S818 Locus Data



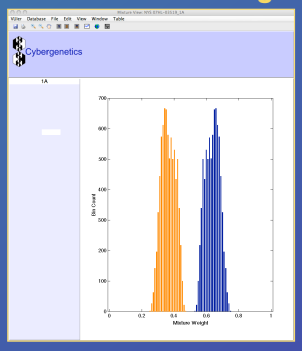
Interpretation Requests



Solving with TrueAllele

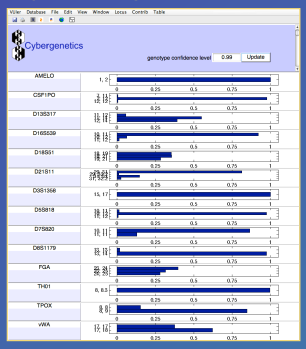


Review: Mixture Weight

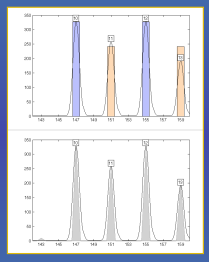


64%

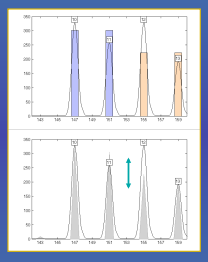
Genotype: Major Contributor



Better Fit's More Likely It

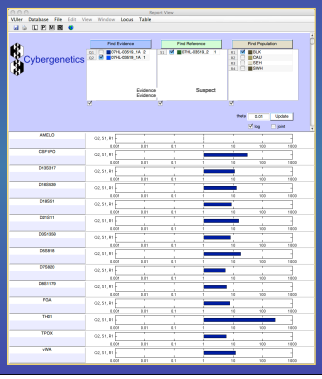


10, 12



10, 11

Likelihood Ratio

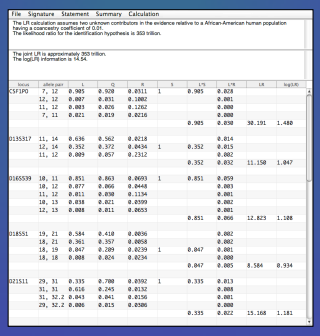


log(LR)
14.54

log(CPI)
7.01

theta = 1%

Report: LR Calculation



File Signature Statement Summary Calculation

The LR calculation assumes two unknown contributors in the evidence relative to a African-American human population having a coincidence of 0.01.
The likelihood ratio for the identification hypothesis is 303 trillion.

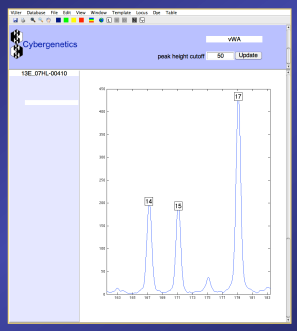
The joint LR is approximately 303 trillion.
The log(LR) information is 14.54.

Locus	Allele pair	L	Q	R	S	L*S	L*R	LR	log(LR)
CSF1PO	7, 12	0.905	0.920	0.0311	1	0.905	0.028		
	12, 12	0.007	0.031	0.1002			0.001		
	11, 12	0.003	0.026	0.1262			0.000		
	7, 11	0.021	0.019	0.0216			0.000		
						0.905	0.030	30.191	1.480
D13S317	11, 14	0.636	0.562	0.0218			0.014		
	12, 14	0.352	0.372	0.0434	1	0.352	0.015		
	11, 12	0.009	0.057	0.2312			0.002		
						0.352	0.032	11.150	1.047
D16S539	10, 11	0.851	0.863	0.0693	1	0.851	0.059		
	10, 12	0.077	0.066	0.0448			0.003		
	11, 12	0.011	0.030	0.1154			0.001		
	10, 13	0.038	0.021	0.0399			0.002		
	12, 13	0.008	0.011	0.0653			0.001		
						0.851	0.066	12.823	1.108
D18S51	19, 21	0.584	0.410	0.0036			0.002		
	18, 21	0.361	0.357	0.0058			0.002		
	18, 19	0.047	0.209	0.0239	1	0.047	0.001		
	18, 18	0.008	0.024	0.0234			0.000		
						0.047	0.005	8.584	0.934
D21S11	29, 31	0.335	0.700	0.0392	1	0.335	0.013		
	31, 31	0.616	0.245	0.0132			0.008		
	31, 32.2	0.043	0.041	0.0156			0.001		
	29, 32.2	0.006	0.015	0.0506			0.000		
						0.335	0.022	15.168	1.181

Hard Mixture Case

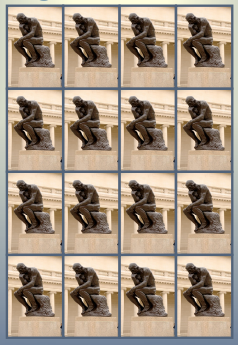
- 60:40 Mixture Data
- Two Unknown Contributors
- No Victim Reference
- Two Suspect References

vWA Locus Data

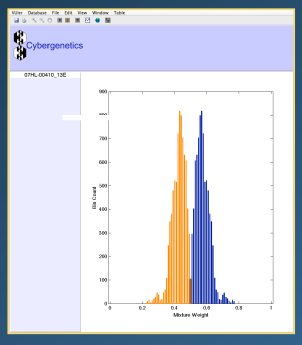


Interpretation Requests

Solving with TrueAllele

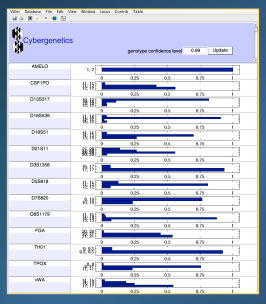


Review: Mixture Weight

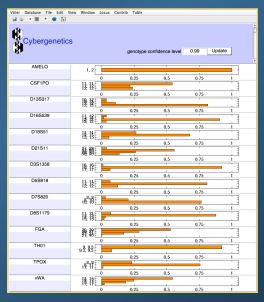


58%
42%

Genotypes

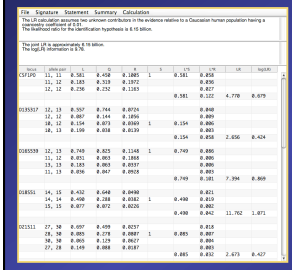


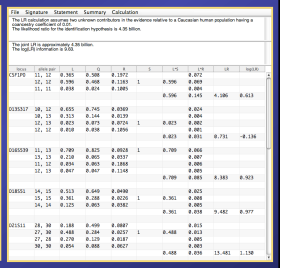
Major



Minor

Reports: Likelihood Ratio





Suspect 1
9.78

theta = 1%
log(LR)

Suspect 2
9.63

Human Review: No Match Statistic

The STR DNA mixture profile from the blood stained swab of the trigger of the rifle (evidence item) is consistent with DNA from suspect 1 admixed with DNA from suspect 2.

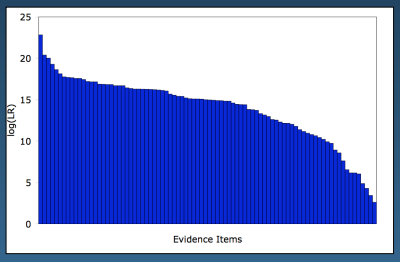
STATUS: EXPERT SYSTEM FUNCTIONS

Without human intervention, the software

1. Identifies peaks, assigns allele designations
2. Performs lab-defined quality checks
5. Assigns mixture weights
6. Perform mixture deconvolution
7. Matches DNA patterns
8. Assigns statistics
9. Prepares reports

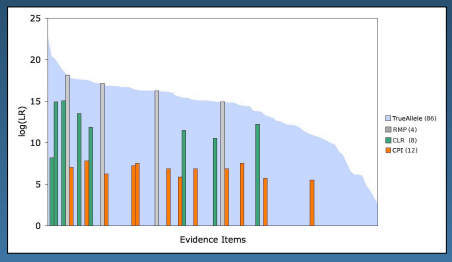
TrueAllele® 2 Validation
NYS DNA Databank and Cybergenetics

TrueAllele Information: 86 Match Stats (100%)



Preserves all the identification information

Human Review Information: 24 Match Stats (28%)



Preserves 20% of the identification information