

# DNA Identification as an Information Science

New York State DNA Subcommittee  
May, 2011  
New York, NY

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



Cybergenetics

Cybergenetics © 2003-2011

---

---

---

---

---

---

---

---

## The DNA Crime Solution



### Evidence

- solve a crime
- match DNA evidence to suspect

### Investigation

- DNA database of crimes and criminals
- match evidence to convicted offenders

### Prevention

- all evidence: property crime, sexual assault, ...
- all criminals: felonies, misdemeanors, ...
- DNA databases interrupt criminal careers

---

---

---

---

---

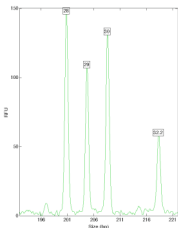
---

---

---

## The DNA Evidence Problem

Most DNA evidence is uncertain,  
producing uncertain genotypes



- mixtures: two or more contributors
- degraded or damaged DNA
- low template DNA amounts

---

---

---

---

---

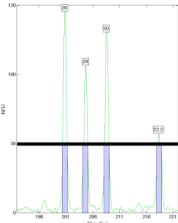
---

---

---

## An Interim Approach

Apply "thresholds" to the quantitative data



- peak height: **variation** vs. **absolute**
- variance scaling: **linear** vs. **constant**
- data: **distribution** vs. **false negative**
- probability: **concentrate** vs. **disperse**
- information: **preserve** vs. **discard**
- identify: **succeed** vs. **fail**
- protect: **better** vs. **worse**

---

---

---

---

---

---

---

---

## Probabilistic Science

Classical Physics

Quantum Mechanics



interim solution (1913):  
electron in an orbital

true solution (1925):  
electron with a probability

---

---

---

---

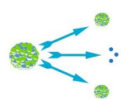
---

---

---

---

## Validation: The Atom Bomb



Splitting the atom



War in Pacific is over

---

---

---

---

---

---

---

---

## Probabilistic Genotypes

Perlin MW. (2003) Simple reporting of complex DNA evidence: automated computer interpretation. Promega's Fourteenth International Symposium on Human Identification; Phoenix, AZ.

Mortera J, Dawid AP, Lauritzen SL (2003) Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology* 63: 191-205.

Cowell RG, Lauritzen SL, Mortera J (2007) Identification and separation of DNA mixtures using peak area information. *Forensic Science International* 166: 28-34.

Cowell RG, Lauritzen SL, Mortera J (2007) A gamma bayesian network for DNA mixture analysis. *Bayesian Analysis* 2.

Curran J (2008) A MCMC method for resolving two person mixtures. *Science & Justice* 48: 168-177.

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

Perlin, MW, MM Legler, CE Spencer, JL Smith, WP Allan, JL Belrose and BW Duceman (2011). Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences*. 56 (November);in press.

Tvedebrink T, Eriksen PS, Mogensen HS, Morling N (2011). Identifying contributors of DNA mixtures by means of quantitative information of STR typing. *J Comput Biol*. 18 (in press).

---

---

---

---

---

---

---

---

---

---

## Information Theory

34 *The Mathematical Theory of Communication*

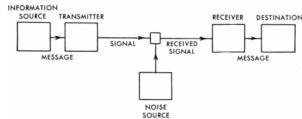


Fig. 1.— Schematic diagram of a general communication system.

Quantify the information in a system

---

---

---

---

---

---

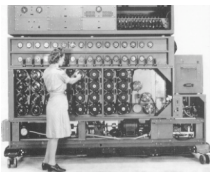
---

---

---

---

## Validation: The Bombe Computer



Breaking the code



War in Europe is over

---

---

---

---

---

---

---

---

---

---

## DNA Likelihood Ratio

Collins, A. and N. E. Morton (1994). Likelihood ratios for DNA identification. *Proc. National Acad. Sci. USA* 91: 6007-6011.

Buckleton J.S., Triggs C.M., Champod C. (2006). An extended likelihood ratio framework for interpreting evidence. *Science and Justice*. 46(2), pp. 69-78.

Gill, P., C. H. Brenner, J. S. Buckleton, A. Carracedo, M. Krawczak, W. R. Mayr, N. Morling, M. Prinz, P. M. Schneider and B. S. Weir (2006). DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160: 90-101.

Buckleton J, Curran J (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Sci Int Genet.* 2(4):343-8.

Perlin, M. W. (2010) Explaining the likelihood ratio in DNA mixture interpretation, in the Proceedings of Promega's Twenty First International Symposium on Human Identification. San Antonio, TX.

Brenner, C. H. (2011) The mythical "exclusion" method for analyzing DNA mixtures – does it make any sense at all? (A111) AAFS 63rd Annual Scientific Meeting, 2011; Chicago, IL. American Academy of Forensic Sciences, p. 79.

Perlin, M. W. (2011) Sherlock Holmes and the DNA likelihood ratio (A142) AAFS 63rd Annual Scientific Meeting; Chicago, IL. American Academy of Forensic Sciences; 2011. p. 95.

---

---

---

---

---

---

---

---

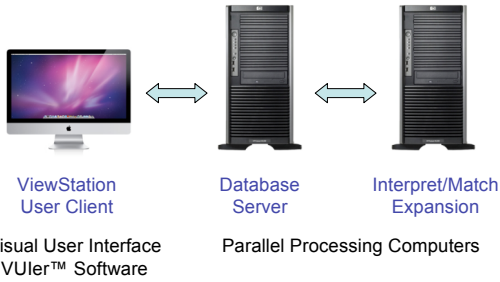
---

---

---

---

## TrueAllele® Casework System




---

---

---

---

---

---

---

---

---

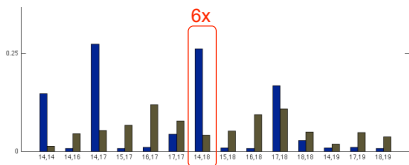
---

---

---

## DNA Identification

- 1 infer genotype  
(up to probability)
- 2 match genotype  
(compute L.R)




---

---

---

---

---

---

---

---

---

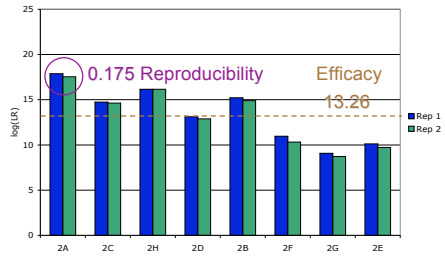
---

---

---

## Validate Reliability

- log(LR) is a standard measure of information
- translates genotype match into a single number



---

---

---

---

---

---

---

---