Probabilistic Software Workshop

September 29, 2014

Probabilistic Genotyping – An Overview

Michael D. Coble, Ph.D.

Product Disclaimer

- **I will mention software programs and STR kit names and information, but I am in no way attempting to endorse any specific products.**
- **NIST Disclaimer**: Certain commercial equipment, instruments, software programs, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.
- **Points of view are mine** and do not necessarily represent the official position of the National Institute of Standards and Technology or the U.S. Department of Justice. **Our group receives or has received funding from the FBI Laboratory and the National Institute of Justice.**

Two Parts to Mixture Interpretation

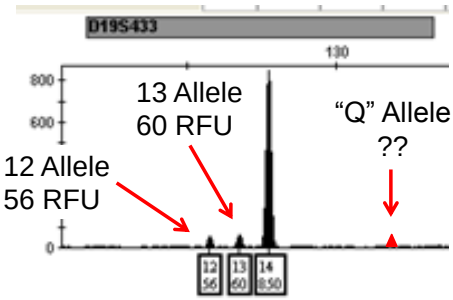
- Determination of alleles present in the evidence and **deconvolution of mixture components** where possible
  - Many times through comparison to victim and suspect profiles
- **Providing some kind of statistical answer** regarding the weight of the evidence
  - There are multiple approaches and philosophies

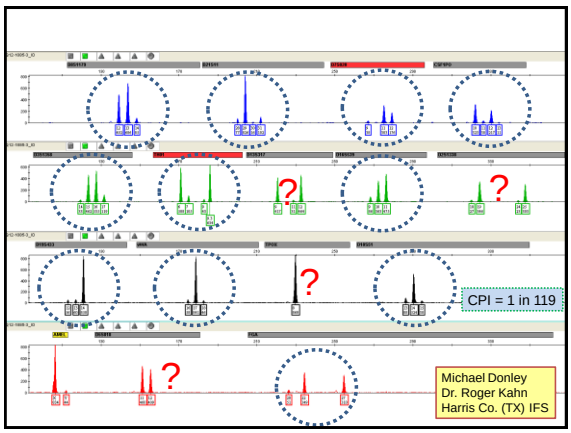
Statistical Approaches with Mixtures

See Ladd *et al.* (2001) *Croat Med J.* 42:244-246

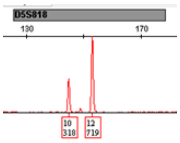
"Exclusionary" Approach	"Inferred Genotype" Approach
Random Man Not Excluded (RMNE)	**Random Match Probability (RMP)**
Combined Prob. of Inclusion (CPI)	**(mRMP)**
Combined Prob. of Exclusion (CPE)	**Likelihood Ratio (LR)**

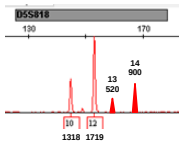
Challenging Mixtures





A cautionary note on using CPI when drop-out is possible





$PI = (f_{10} + f_{12})^2$

$PI = (0.051 + 0.361)^2$

$PI = 0.169$ or **1 in 5.92**

$PI = (f_{10} + f_{12} + f_{13} + f_{14})^2$

$PI = (0.051+0.361+0.384+0.141)^2$

$PI = 0.878$ or **1 in 1.14**

Drop-out ***inflates*** your statistics for CPI (not conservative!)

2013 JFS Article

J Forensic Sci, March 2013, Vol. 58, No. 2
doi: 10.1111/1556-4029.12067
Available online at: onlinelibrary.wiley.com

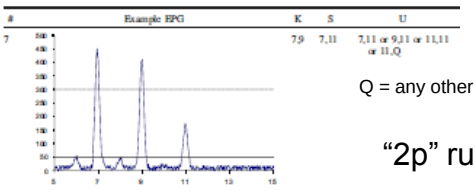
TECHNICAL NOTE

CRIMINALISTICS

Todd Bille,[1] M.Sc.; Jo-Anne Bright,[2] M.Sc.; and John Buckleton,[2] Ph.D.

Application of Random Match Probability Calculations to Mixed STR Profiles

When data is below ST



Q = any other allele

"2p" rule

CPI = drop the locus

mRMP = "2p"

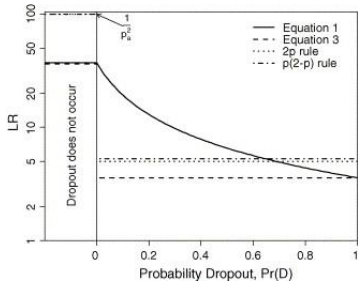
The “2p” Rule

- “This rule arose during the VNTR era. At that time many smaller alleles “ran off the end of the gel” and were not visualised.”

  - Buckleton and Triggs (2006)

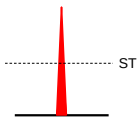
Is the 2*p* rule always conservative?”

The “2p” Rule



Stain = AA

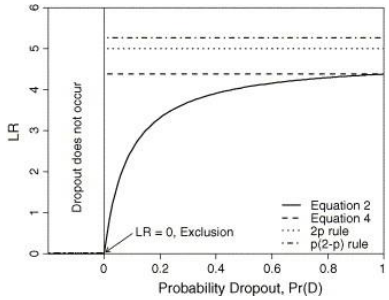
Suspect = AA



LR = 100

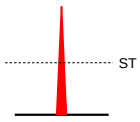
f(a) = 0.10 $1/p^2 = 100$ 1/2p = 5

The “2p” Rule



Stain = AA

Suspect = AB



Exclusion

f(a) = 0.10 1/2p = 5

Likelihood Ratio

Likelihood Ratios in Forensic DNA Work

- We evaluate the evidence (*E*) relative to alternative pairs of hypotheses
- Usually these hypotheses are formulated as follows:
  - The probability of the evidence if the crime stain originated with the suspect or Pr(*E*|*S*)
  - The probability of the evidence if the crime stain originated from an unknown, unrelated individual or Pr(*E*|*U*)

$$LR = \frac{\Pr(E \mid S)}{\Pr(E \mid U)}$$

The numerator

The denominator

Slide information from Peter Gill

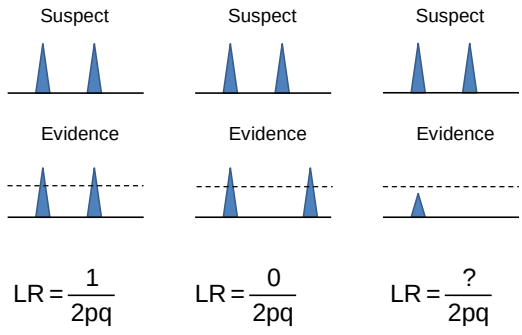
Likelihood Ratio (LR)

- Provides ability to express and evaluate both the prosecution hypothesis, H_p (the suspect is the perpetrator) and the defense hypothesis, H_d (an unknown individual with a matching profile is the perpetrator)

$$LR = \frac{H_p}{H_d}$$

- **The numerator, H_p, is usually 1** – since in theory the prosecution would only prosecute the suspect if they are 100% certain he/she is the perpetrator
- The denominator, $\mathbf{H_d}$, is typically the profile frequency in a particular population (based on individual allele frequencies and assuming HWE) – i.e., **the random match probability**

Slide information from Peter Gill



The Binary LR approach

"2p"

Forensic Science International: Genetics 6 (2012) 679–688

DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods

P. Gill [a,b,*], L. Gusmão [c], H. Haned [d], W.R. Mayr [e], N. Morling [f], W. Parson [g], L. Prieto [h], M. Prinz [i], H. Schneider [j], P.M. Schneider [k], B.S. Weir [l]

Summary of recommendations of the ISFG DNA commission

- (1) Probabilistic methods following the 'basic model' described here can be used to evaluate the evidential weight of DNA results considering drop-out and/or drop-in.
- (2) Estimates of drop-out and drop-in probabilities should be based on validation studies that are representative of the method used.

Summary of recommendations of the ISFG DNA commission

- (3) The weight of the evidence should be expressed following likelihood ratio principles.
- (4) The use of appropriate software is highly recommended to avoid hand-calculation errors.

Probabilistic Approaches

- "Semi-Continuous" or "Fully Continuous"
- Semi-Continuous – information is determined from the alleles present – peak heights are not considered.
- Fully Continuous – incorporation of biological parameters (PHR [Hb], Mx ratio, Stutter percentage, etc…).

Some Semi-Continuous Examples

- LR mix (Haned and Gill)
- Balding (likeLTD - R program)
- FST (NYOCME, Mitchell *et al.*)
- Kelly *et al.* (University of Auckland, ESR)
- Lab Retriever (Lohmueller, Rudin and Inman)
- Armed Expert (NicheVision)
- Puch-Solis *et al.* (LiRa and LiRaHT)
- GenoProof Mixture (Qualitype)

Some Continuous Model Examples

- TrueAllele Casework (Cybergenetics)
- STRmix (ESR [NZ] and Australian collaboration)
- DNA-View Mixture Solution (Charles Brenner)
- DNAmixtures (Graversen 2013a,b) – open source, but requires HUGIN.

Weights may be determined by performing simulations of the data (Markov Chain Monte Carlo - MCMC).

Goals of this Workshop

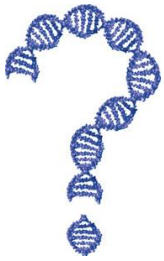
- To develop a greater understanding of the software systems presented.
- To foster discussions about training, validation, and scientific support of the software systems.
- To interact and ask questions of the software developers.

Acknowledgments

National Institute of Justice and NIST OLES

Charlotte Word

Carol Bingham
Lotte Downey



Contact info:
mcoble@nist.gov
+1-301-975-4330