DNA Mixture De-Convolution by Binomial Sampling of Individual Cells

Jack Ballantyne and Mark Perlin

International Conference on Inference and Statistics, July 20 2011, Seattle, WA

Scope of Talk

- Introduction to Mixtures
  - Binomial sampling hypothesis
- Laser Capture Technology (LCM)
  - Hardware
  - Outline of Method
  - Efficacy of profile recovery
- Proof of Principle
  - Experimental strategy (50:50 mixture as model)
  - Quantitative Computer Interpretation
  - Information gain
    - Different weight ratios obtained from same sample
    - Joint likelihood
- Conclusions

Introduction to Mixtures

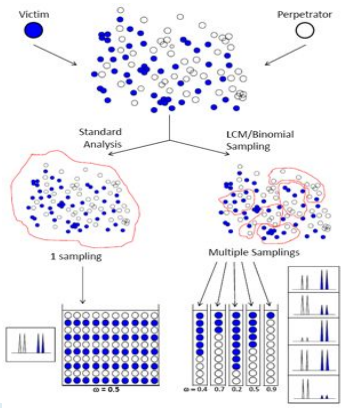
STR Interpretation Guidelines

-SWGDAM (2000)

3. Interpretation of Results

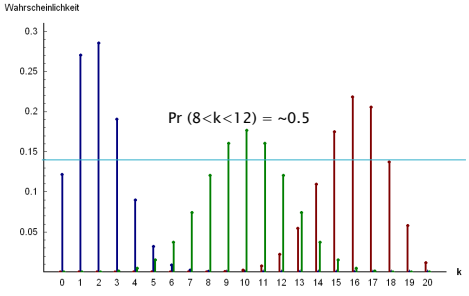
- 3.1.1. *Single Contributor*
  - when the observed number of alleles at each locus and the signal intensity ratios of alleles at a locus are consistent with a profile from a single contributor
  - all loci should be evaluated in making this determination
- 3.1.2. *Mixtures With Major/Minor Contributors*
  - if there is a distinct contrast in signal intensities among the alleles. The difference is evaluated on a case-by-case context. All loci should be evaluated in making this determination
- 3.1.3. *Mixtures With a Known Contributor(s)*
  - when one of the contributors (e.g., the victim) is known, the genetic profile of the unknown contributor may be inferred.
  - This can be accomplished by subtracting the contribution of the known donor from the mixed profile
- 3.1.4. *Mixtures With Indistinguishable Contributors*
  - When major or minor contributors cannot be distinguished because of similarity in signal intensities or the presence of shared or masked alleles, individuals may still be included or excluded as possible contributors

Standard versus Binomial Sampling



Binomial distribution for $n = 20$

($p = 0.1$ (blue), $p = 0.5$ (green) and $p = 0.8$ (red))



adapted from http://en.wikipedia.org/wiki/Binomial_distribution

Hypothesis

- Multiple samplings of individual cells (e.g. 10 x 20) from a mixture will result in an information gain compared to a single sampling
  - More likely that a beneficial (ie more informative) weight ratio will be obtained (ie w = 0.2 is better than w = 0.5)
  - Joint likelihood function (combining data from the different samplings) will further increase information

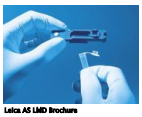
Laser Capture Technology

Laser Capture Micro-dissection (LCM)

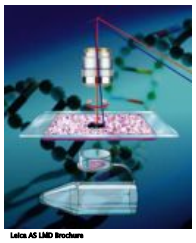
- Technology that permits the recovery and isolation of single cells or groups of cells from various samples
- First developed to separate cancer cells from normal tissues
  - 1996 – Conceived by NIH, developed by Arcturus
- Gained interest in the forensic community within the past few years
  - Separation of sperm and epithelial cells in sexual assault cases

Our LCM: "Leica LMD"

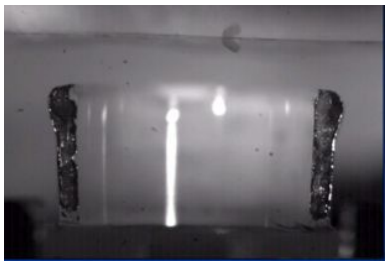
- Uses a UV laser pulse to excise selected cells from a membrane slide
- Cut areas then fall by gravity into a collection tube


Leica AS LMD Brochure


Leica AS LMD Brochure

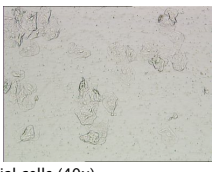

Leica AS LMD Brochure

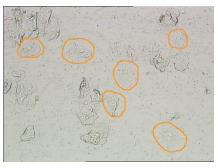
Cell Collection

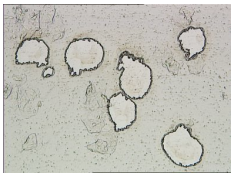


Laser-cut cell falls by gravity into the cap of a 0.2ml PCR tube located directly below the stage

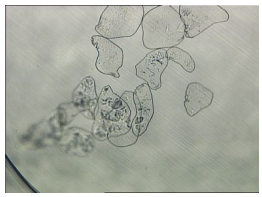
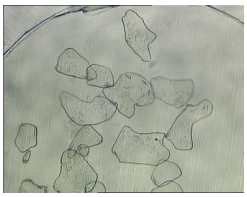
Epithelial Cell Collection


Epithelial cells (40x)


Marked for collection


After collection

Cells in Tube Cap After Collection

LCM method

- Sample preparation
  - Buccal cell suspensions prepared
    - Cell count – hemocytometer
  - Appropriate volume of each cell suspension mixed to create desired mixture ratio
  - Applied to Leica PEN membrane slide
    - Heat fixed; no staining
- Sample Collection (400x magnification)
  - Semi-automated
    - No manual manipulation of slide after placed in holder
    - Objectives and stage controlled by joystick
- Cell lysis to release DNA
  - 10 ml in cap of 0.2ml PCR tube
    - Lysis in thermocycler – 75ºC 15 min, 95ºC 5 min

Efficacy of Profile Recovery with LCM

- 10 donors (buccal swabs)
  - 5 female, 5 male
  - Varying number of cells collected
    - 1, 2, 3, 4, 5, 10, 20
      - 10 replicates for each cell number
        - both biological and technical replicates
- Samples amplified with Identifiler (ABI)
  - Increase to 34 cycles (LCN)
  - Separation and detection on CE ABI Prism 3130

LCM Success Rate – Epithelial Cells

Number of Micro-dissected Cells	Avg. % Partial/Full Profile	Avg. % Full Profiles	Avg. % No Profile
1	74 (± 5%)	4 (± 2%)	26 (± 5%)
2	69 (± 5%)	19 (± 8%)	30 (± 5%)
3	81 (± 3%)	20 (± 5%)	20 (± 5%)
4	81 (± 3%)	39 (± 6%)	19 (± 3%)
5	81 (± 3%)	46 (± 7%)	18 (± 4%)
10	90 (± 3%)	66 (± 6%)	11 (± 3%)
20	99 (± 1%)	80 (± 6%)	1 (± 1%)

*Std error values listed in parentheses next to each percentage

**Average % profile recovery determined by averaging the success rates (number of profiles recovered from 10 replicates) of the 10 individual donors.

Good profile recovery obtained with 20-cell samples
(work is underway to improve profile recovery with 1 cell which would preclude the need to perform mixture analysis)

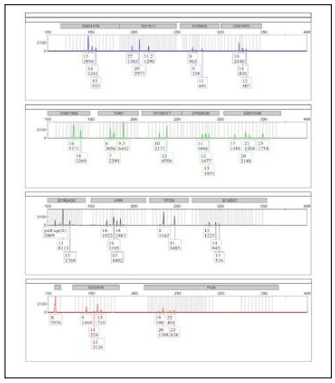
Proof of Principle

Binomial Sampling of 2 Person Mixtures

Proof of Concept: Experimental Design

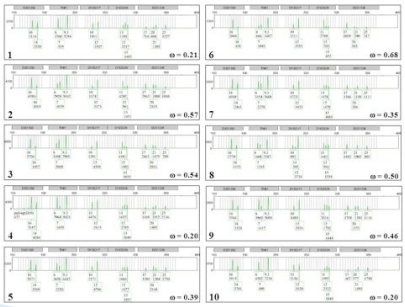
- Two 50:50 mixture samples prepared
  - Male-Female and Female-Female
- 20-cell samples collection from each mixture
  - 10 separate samplings performed (i.e. 10 x20)
  - Direct lysis and Identifiler (34 cycle) amplification
- DNA profiles obtained (electropherograms)
- Analysis
  - Quantitative Computer interpretation (TrueAllele)
    - Inference of Genotypes
    - Attach Statistical Weight

Mixture Profile Example (20 cells putative 1:1 mix F1:F2)



Female–Female (1:1)
Identifiler, 34 cycles

Mixture Profiles Recovered from 10 Separate 20-cell samples



VIC channel shown

Quantitative Computer Interpretation

DNA Identification

Questions

- Does binomial sampling diverge from 50:50?
- Does this increase identification information?
- How else can we increase information yield?

Information Solution

Requirements

- Infer mixture weights
- Infer genotypes
- Compute log(LR) match information
- Represent uncertainty using probability

MW Perlin, MM Legler, CE Spencer, JL Smith, WP Allan, JL Belrose, BW Duceman. Validating TrueAllele® DNA mixture interpretation. Journal of Forensic Sciences, 2011.

TrueAllele® Casework Information

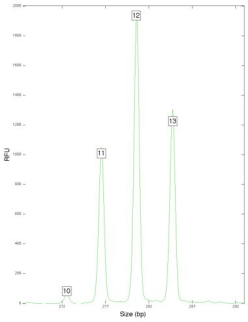
Forensic inference via computational statistics

- Hierarchical probability accounts for all relevant variables
  (genotype, mixture weight, DNA quantity, PCR stutter, preferential amplification, degradation, peak variation, data combination, 2, 3, 4, 5+ contributors, ...)
- Markov chain Monte Carlo statistical search
  (1,000+ dimensional parameter space)
- Extensively validated, courtroom proven, government approved
- Rape, murder, burglary, kinship, mass disaster, database, etc.

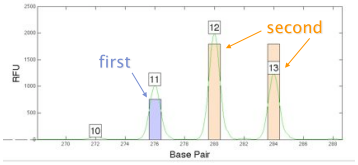
STR Data

- Two contributors
- 50:50 ratio
- Binomial sampling
- LCM 20 cells
- Identifiler STR kit
- LCN 34 PCR cycles

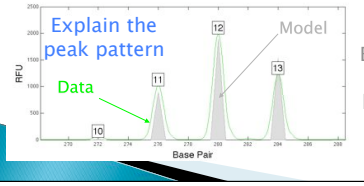
D16S539



Likelihood Function



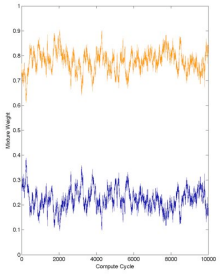
Consider every possible genotype solution



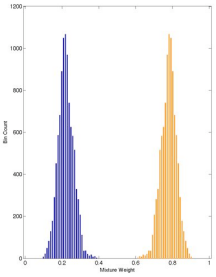
Better explanation gives a higher probability

Mixture Weight

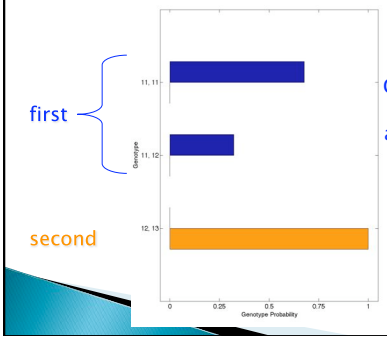
Markov chain



Probability



Genotype

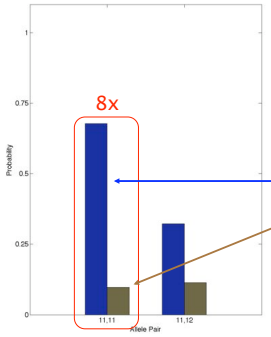


Genotype uncertainty described by allele pair probability

Objective inference never sees suspect

MW Perlin. Explaining the likelihood ratio in DNA mixture interpretation. Proceedings of Promega's 21st International Symposium on Human Identification, 2010.

Information Gain: LR



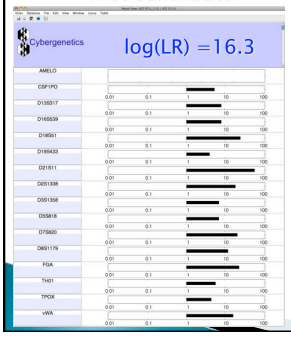
How much more does the suspect match the evidence than a random person?

$$\frac{\text{Probability(evidence match)}}{\text{Probability(coincidental match)}}$$

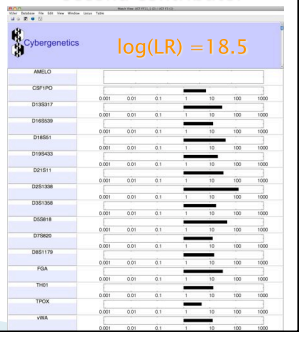
LR = 8

Joint Likelihood Ratio

first contributor

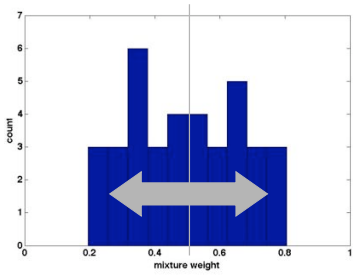


second contributor

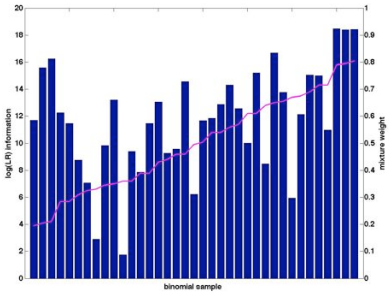


Binomial Sampling: Mixture Weight

Diverges from 50:50 ratio



Identification Information

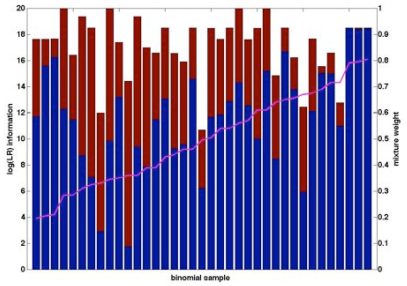


Combining Data: Joint Likelihood

Hypothesize model

Explain all the data

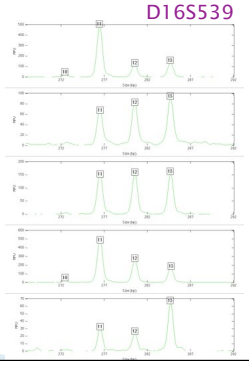
Binomial Pairs: Information Gain



Low level STR Data

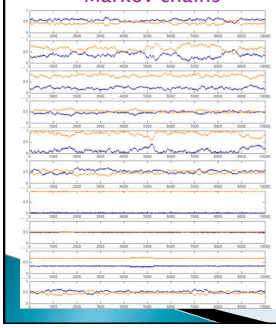
- Two contributors
- 50:50 ratio
- Binomial sampling
- LCM 20 cells
- Identifiler STR kit
- 28 PCR cycles
- low peak height
- pattern variation

D16S539

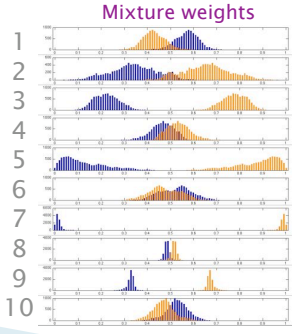


Combine Ten Samples

Markov chains

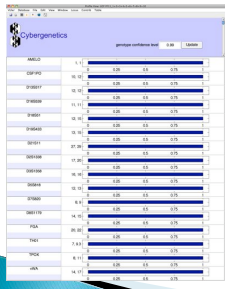


Mixture weights

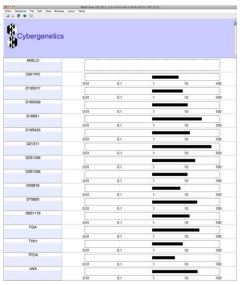


Unique Genotype, Exact Match

100% probability



full log(LR) information



Observations

- Binomial sampling separates DNA mixtures
- Probabilistic genotypes extract much information
- TrueAllele quantifies mixtures and information
- Statistical data and interpretation are useful

www.cybgen.com/information
perlin@cybgen.com

Conclusions

- Binomial sampling of cells by LCM from a two person admixture yields sub-populations of cells with different weight ratios
- Genotype inference is facilitated by binomial sampling
- Information gain due to
  - Multiple LRs obtained from the same mixture
  - Joint likelihood function

Library of Medicine (IHM)

"There are no facts, only interpretations"

Nietzsche

Acknowledgements:

•NCFS/UCF: Dr. Erin Hanson

•Cybergenetics: Cara Lincoln, Jessica Smith

•Funding – NIJ/FBI