Peeling Away Uncertainty: A Probabilistic Approach to DNA Mixture Deconvolution

William Allan[1], MS; Hajara Chaudhry[2], MS; John Donahue[3], MA;
Mark Perlin[1], PhD, MD, PhD; Mark Wilson[2], PhD

[1]Cybergenetics Corp, Pennsylvania
[2]George Mason University, Virginia
[3]Beaufort County Sheriff's Office, South Carolina

Cybergenetics
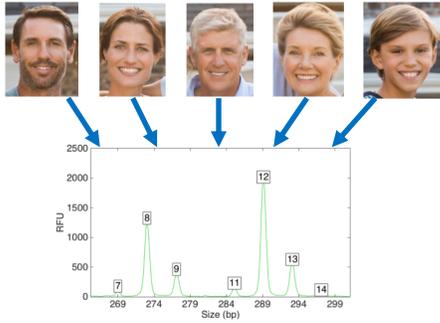
Genotype mixing

Genotype unmixing

Information measurement
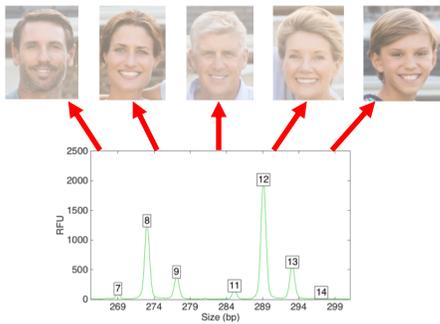
Evidence genotype   Reference genotype
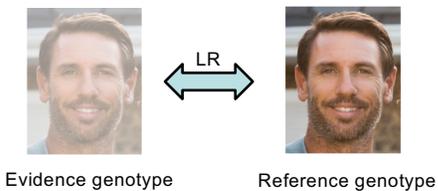
likelihood ratio (LR) = change in genotype probability
made by DNA evidence

log(LR) is a standard information measure

Genotype uncertainty

Diffuse probability
distribution

Reduces match
information
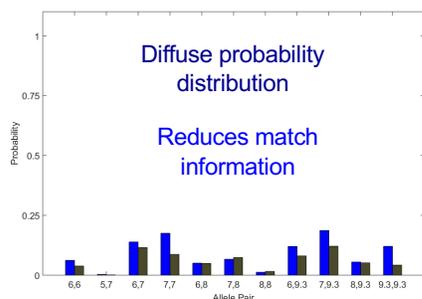
Genotype conditioning

• Assumes a contributor genotype in genotype inference

• Performed iteratively, in subsequent analyses

• Focuses the inference on the other unknown contributor genotypes

Improves genotype separation by
reducing uncertainty

The "peeling" process

Most informative contributor

All unknowns → **Mixture separation** → **Condition with reference**

Zero unknowns

solve for MW

Continue until one contributor remains
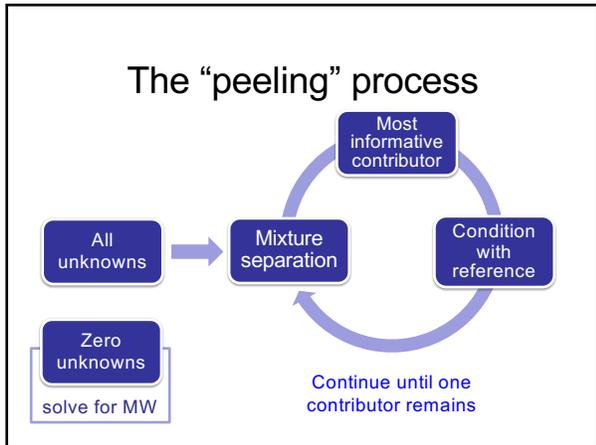
Beaufort data for GMU study

• Dataset
  ➢ 2, 3, 4, and 5 contributor mixtures
  ➢ GlobalFiler PCR Kit
  ➢ AB 3500xl Genetic Analyzer

• TrueAllele®
  ➢ 25k MCMC cycles, in duplicate
  ➢ Conditioned based on highest LR
  ➢ 70 references for comparison

Reducing genotype uncertainty

Genotype View | Match View

Q3, S4, R1 — -0.0945
Q7, S4, R1 — 0.0904
Q13, S4, R1 — 0.2570
Q20, S4, R1 — 0.5315
Q25, S4, R1 — 1.3787

Subject 34 in 5-con mixture at D1S1656
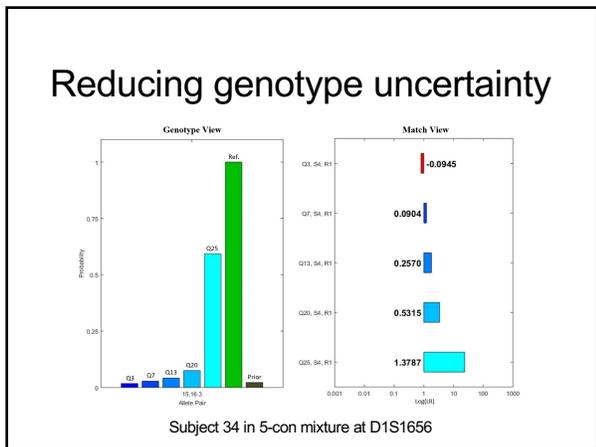
Increasing information yield

5-contributor mixture

Assumed Genotypes	Mixture Weight (%)	Log(LR)				
		58	25	26	30	34
0	33	11.7	1.4	1.3	<1	<1
1	24		5.4	5.4	3.4	4.6
2	20			14.3	7.6	1.8
3	11				13.1	10.6
4	11					17.8

Conditioning improves genotype separation for overlapping contributors, match information doubled overall

---

The Queen v. Robert Xie

On the night of 18 July 2009, in a Sydney suburb, five Lin family members were bludgeoned to death in their North Epping home (Australia).


Blood stain 91

Suspicion fell on the wife's brother-in-law, Robert Xie. Xie had cleaned his garage floor that morning. Police examined the garage floor stains.

FGA locus
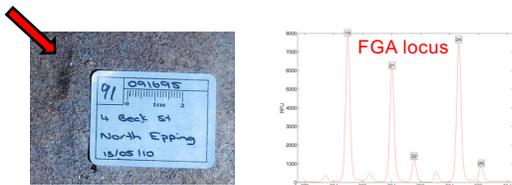
A family question

Forensic question #1: Is there DNA from the Lin family in the garage floor mixture?

Min (45)   Yun Li (43)   Yun Bin (39)

Henry (12)   Terry (9)

A peeling answer

Without peeling, all five relatives look associated. Kinship allele sharing? Or is their DNA truly there?

Item 550, garage floor

Contributors Assumed	Total	142 Terry	127 Henry	108 Min	88 Yun Li	70 Yun Bin
None	5	13.11	6.80	9.04	-0.61	-5.37
Terry	5		6.90	5.27	-16.33	-24.82
Terry, Henry	5			9.37	-15.67	-18.61
Terry, Henry, Min	5				-0.19	-0.11
Terry, Henry, Min, Yun Li	5					3.95

With peeling, four of them are clearly there.

The missing sister

Forensic question #2: Is Brenda Lin's DNA in the garage floor mixture?

Min (45)   Yun Li (43)   Yun Bin (39)

Henry (12)   Terry (9)   Brenda (15) ?
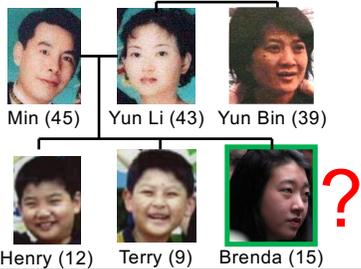
Peeling reveals the truth

Without peeling, Brenda Lin might be present.
Sibling allele sharing? Or is her DNA truly there?

Item 550, garage floor

Contributors Assumed	Total	142 Terry	127 Henry	108 Min	88 Yun Li	70 Yun Bin	63 Brenda
None	5	13.11	6.80	9.04	-0.61	-5.37	3.47
Terry	5		6.90	5.27	-16.33	-24.82	-6.37
Terry, Henry	5			9.37	-15.67	-18.61	-8.26
Terry, Henry, Min	5				-0.19	-0.11	-3.59
Terry, Henry, Min, Yun Li	5					3.95	-5.59
Terry, Henry, Min, Yun Li, Yun Bin	6						-6.68

With peeling, Brenda's DNA clearly isn't there.

Conclusions

- Mixture data contain DNA identification information
- DNA comparisons are made through genotypes
- The likelihood ratio measures match information
- Genotype conditioning can increase information
- Successive "peeling" rounds reveal DNA matches
- Laboratory study demonstrates efficacy
- Criminal case application shows utility

Better use of DNA data can give
more identification information