DNA Mixture Interpretation:
Where did we come from? What are we doing?
Where are we going?

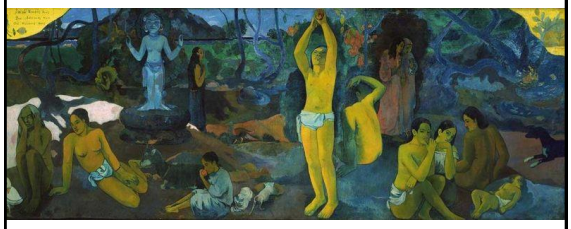
Michael Coble, PhD
NIST

Official Disclaimer

The opinions and assertions contained herein are solely those of the author and are not to be construed as official or as views of the U.S. Department of Commerce, U.S. Department of Justice, or the U.S. Department of Defense.

Commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the U.S. Department of Commerce, U.S. Department of Justice, or the U.S. Department of Defense nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Where Do We Come From? What Are We? Where Are We Going?



Paul Gauguin, 1897

http://en.wikipedia.org/wiki/File:Woher_kommen_wir_Wer_sind_wir_Wohin_gehen_wir.jpg

Where are we going?
(2015 -)

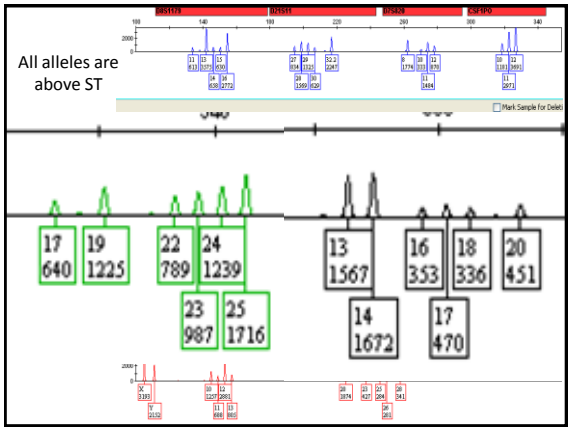
ASCLD Trivia Time!!!

Three Questions

- What were the last words of Julius Caesar before he died?
- Et tu, Brute? Then fall Caesar!
- What is the capital of Bangladesh?
- Dhaka

Three Questions

- How many people are in this mixture?



Do you have any uncertainty in your answer?

Whatever way uncertainty is approached, probability is the *only* sound way to think about it.

-Dennis Lindley

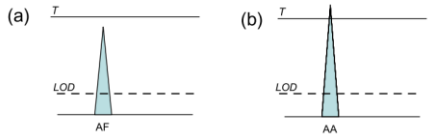
Handling Complex Mixtures

- Stochastic thresholds are necessary in combina… with CPI statistics
  - but a stochastic threshold may not hold much meaning for >2 person mixtures (due to potential allele sharing)
- Most labs are not adequately equipped to cope with complex mixtures
  - Extrapolating validation studies from simple mixtures will not be enough to create appropriate interpretation SOPs

David Balding (UK professor of statistical genetics): "LTDNA cases are coming to court **with limited abilities for sound interpretation.**" (Rome, April 2012 meeting)

"Falling off the Cliff Effect"

- If T = an arbitrary level (e.g., 150 rfu), an allele of 149 rfu is subject to a different set of guidelines compared with one that is 150 rfu even though they differ by just 1 rfu (Fig. 1).



Gill and Buckleton *JFS* 55: 265-268 (2010)

Gill and Buckleton *JFS*
55: 265-268 (2010)

- "The purpose of the ISFG DNA commission document was to provide a way forward to demonstrate the use of ***probabilistic models to circumvent the requirement for a threshold*** and to safeguard the legitimate interests of defendants."

What should we do with data below our Stochastic Threshold?

- Continue to use RMNE (CPI, CPE) (not optimal)
- Use the Binary LR with 2p (not optimal)
- Probabilistic Models

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…).

R. v Garside and Bates

- James Garside was accused of hiring Richard Bates to kill his estranged wife, Marilyn Garside.
- Marilyn was visiting her mother when someone knocked on the door. Marilyn answered and was stabbed to death.
- A profile from the crime scene stain gave a low-level DNA profile of the perpetrator.

Summary

Locus	Mrs Garside	Bates	CSP: minor component
D3	16,16	13,16	13
VWA	15,17	16,16	16
D16	11,12	11,12	–
D2	20,20	**19**,22	22
D8	12,13	8,13	8
D21	30,32.2	30,31.2	31.2
D18	14,14	**12,15**	–
D19	12,14	12,15	15
THO1	9.3,9.3	7,7	7
FGA	23,25	21,21	21

Three alleles from Bates were not present in the evidence

Court case

- The Crown expert dropped the D18 locus (gave a LR = 1) from the statistical results and used “2p” for D2 to give an overall odds for Bates of 1 in 610,000.
- David Balding argued for the defense that dropping loci is not conservative.

Balding and Buckleton (2009)

Forensic Science International: Genetics 4 (2009) 1–10

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

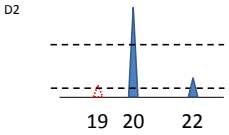
Interpreting low template DNA profiles

David J. Balding [a,*], John Buckleton [b]

[a] Department of Epidemiology and Public Health, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK
[b] ESR Private Bag 92021, Auckland, New Zealand

Present the "Drop model" for interpreting LT-DNA profiles

Drop Model



V = 20, 20
S = 19, 22

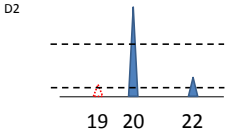
Pr(Drop-out) = 0.05
Pr(Drop-in) = 0.01

$P(E|H_1)$ = Pr(no Drop-out at 22) Pr(Drop-out at 19) Pr(No Drop-in)

= 0.95 0.05 0.99

= 0.047

Drop Model



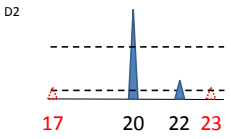
V = 20, 20
S = 19, 22

Pr(Drop-out) = 0.05
Pr(Drop-in) = 0.01

$$\frac{P(E|H_1)}{P(E|H_2)} = \frac{0.047}{}$$

The defense can now argue that someone else in the population unrelated to Bates was the true perpetrator!

Drop Model



V = 20, 20
UC = 17, 23

Pr(Drop-out) = 0.05
Pr(Drop-in) = 0.01

$P(E|H_2)$

Pr(Drop-out at 17)	Pr(Drop-out at 23)	Pr(Drop-in at 22)
0.05	0.05	0.01

= **0.000025** x **$2pq_{17,23}$** **(0.027)** = **0.000000675**

Summary

- Using "2p" for D2 gave a LR = 11. This is non-conservative compared to the probabilistic approach where a Pr(D) was incorporated into the calculation, the LR = 2.8
- The use of a probabilistic approach uses all of the information in the profile.
- The final LR in favor of the Hp was ≈ 400,000.

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)

Semi-continuous methods

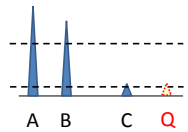
- Use a Pr(DO) and LRs
- Speed of analysis – "relatively fast" depending on the mixture.
- The methods do not make full use of data - only the alleles present.

What should we do with data below our Stochastic Threshold?

- Continue to use RMNE (CPI, CPE) (not optimal)
- Use the Binary LR with 2p (not optimal)
- Semi-continuous methods with a LR (Drop models)
- Fully continuous methods with LR

Continuous Models

- Mathematical modeling of "molecular biology" of the profile (mix ratio, PHR (Hb), stutter, etc…) to find optimal genotypes, giving **WEIGHT** to the results.



Probable Genotypes
AC – 40%
BC – 25%
CC – 20%
CQ – 15%

Some Continuous Model Examples

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

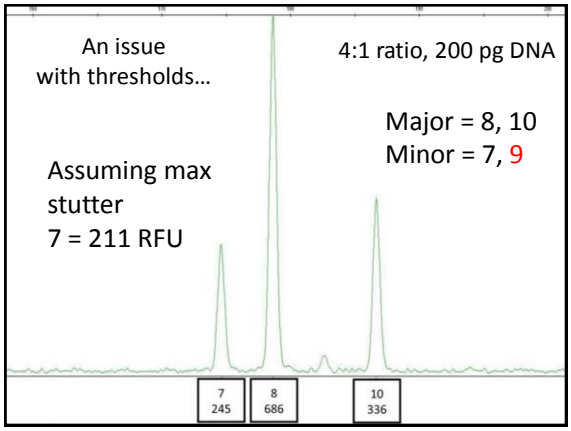
Weights may be determined by performing simulations of the data (Markov Chain Monte Carlo - MCMC).

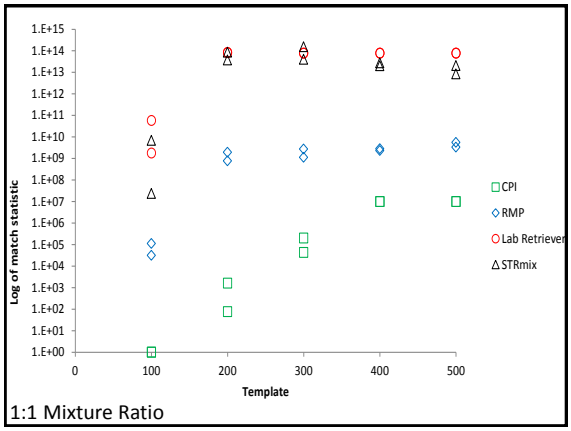
Fully continuous methods

- Use a Pr(DO) and LRs
- Speed of analysis – can vary
- Attempts to use all of the data

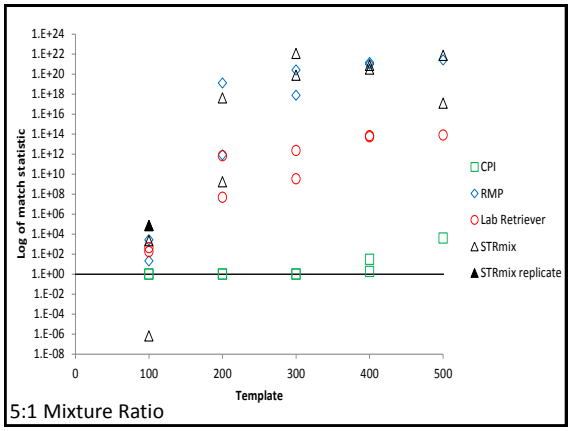
Advantages of Probabilistic Models

- Bille et al. *Electrophoresis*
- Used two samples with low allele sharing (10 markers – 4 alleles, 5 markers – 3 alleles). 2 PCR amplifications.
- 1:1, 2:1, 3:1, 4:1, and 5:1
- 500, 400, 300, 200, 100 pg input DNA
- CPI, RMP (2p), Lab Retriever, STRmix



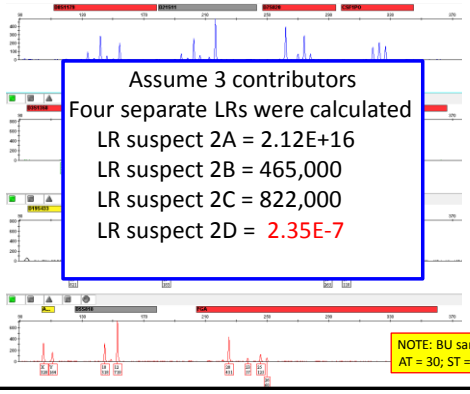


1:1 Mixture Ratio



5:1 Mixture Ratio

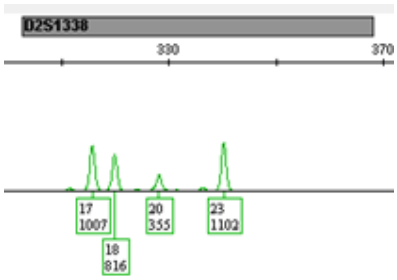
Case 02 – IDFiler



Case 05

"Couldn't help but note the need for mix deconvolution software tools for case 05"

(a) Deconvolution as 3p mixture



GENOTYPE PROBABILITY DISTRIBUTION
D2S1338

[23,23]	[18,20]	[17,17]	3.1435619345160034E-4
[17,20]	[18,23]	[17,17]	0.013779123510160775
[18,20]	[18,23]	[17,17]	0.0025562385293281887
[20,20]	[18,23]	[17,17]	3.5330685147076245E-4
[20,23]	[18,23]	[17,17]	0.09463609425559072
[18,18]	[20,23]	[17,17]	1.3307620480543583E-4
[18,20]	[23,23]	[17,17]	2.2360678716779012E-4
[23,23]	[17,20]	[17,18]	1.860003332375718E-5
[18,20]	[17,23]	[17,18]	0.011194437871043312
[20,23]	[17,23]	[17,18]	0.0022887419156283734
[17,20]	[18,23]	[17,18]	0.0025434216996429106
[20,23]	[18,23]	[17,18]	4.670327695074493E-5
[17,18]	[20,23]	[17,18]	5.917624047373503E-5
[17,23]	[20,23]	[17,18]	2.2663906150796565E-5
[18,23]	[20,23]	[17,18]	1.1378843915710276E-5
[17,20]	[23,23]	[17,18]	4.217987388864801E-4
[18,20]	[23,23]	[17,18]	3.3877069097404465E-4

[20,23]	[23,23]	[17,18]	5.689421957855138E-6
[23,23]	[17,20]	[18,18]	2.0319364135196922E-6
[17,20]	[17,23]	[18,18]	0.012963941881616883
[18,20]	[17,23]	[18,18]	0.004694742192596937
[20,20]	[17,23]	[18,18]	5.3705642437335 86E-5
[20,23]	[17,23]	[18,18]	0.07806365212748431
[17,17]	[20,23]	[18,18]	4.020108042748191E-5
[17,23]	[20,23]	[18,18]	2.1225920381228785E-5
[17,20]	[23,23]	[18,18]	7.19930701590131E-5
[18,23]	[17,18]	[17,20]	2.5571138250140126E-5
[23,23]	[17,18]	[17,20]	0.01846580047822405
[23,23]	[18,18]	[17,20]	1.8006082679805273E-5
[18,18]	[17,23]	[17,20]	0.012773252464348113
[18,23]	[17,23]	[17,20]	0.0017543426388726942
[17,17]	[18,23]	[17,20]	0.0153325545285766 84
[17,18]	[18,23]	[17,20]	0.0027471155099582077
[17,20]	[18,23]	[17,20]	3.0322743401755406E-5
[17,23]	[18,23]	[17,20]	0.07381384148832777
[18,23]	[18,23]	[17,20]	2.3154696946062092E-4
[17,18]	[23,23]	[17,20]	1.6071054002930366E-4
[18,18]	[23,23]	[17,20]	1.5758448400850412E-4
[18,23]	[17,17]	[18,20]	1.1128759434046313E-5

[23,23]	[17,18]	[18,20]	0.006032975514541006
[17,18]	[17,23]	[18,20]	0.00808063598984532
[18,18]	[17,23]	[18,20]	0.006747373096974322
[18,20]	[17,23]	[18,20]	2.341415959578845E-5
[17,23]	[17,23]	[18,20]	0.0024700844253949536
[18,23]	[17,23]	[18,20]	0.07586022028066357
[23,23]	[17,23]	[18,20]	5.9645148876854965E-5
[17,17]	[18,23]	[18,20]	0.0031901401692259164
[17,23]	[18,23]	[18,20]	2.420817782507153E-4
[17,17]	[23,23]	[18,20]	8.84705114446474E-4
[17,18]	[23,23]	[18,20]	4.2464344986540767E-4
[18,18]	[17,23]	[20,20]	3.0313365233693005E-4
[17,17]	[18,23]	[20,20]	2.3336008195268463E-4
[17,23]	[18,23]	[20,20]	1.5036329460045721E-5
[18,20]	[17,18]	[17,23]	1.0003379266558484E-5
[20,23]	[17,18]	[17,23]	0.013604189415236835
[18,23]	[17,20]	[17,23]	3.238594037548309E-5
[17,23]	[18,20]	[17,23]	8.277796343077145E-5
[18,20]	[17,23]	[17,23]	0.0015531184128138287
[17,20]	[18,23]	[17,23]	0.07913420127236732
[18,20]	[18,23]	[17,23]	3.572144214967619E-4
[20,23]	[18,23]	[17,23]	3.1072996846747295E-5
[18,20]	[17,17]	[18,23]	2.3320377915164468E-5
[20,23]	[17,17]	[18,23]	5.867607151040711E-5

[17,20]	[17,18]	[18,23]	5.501858596607167E-6
[20,23]	[17,18]	[18,23]	0.00291845464045823
[17,23]	[17,20]	[18,23]	6.170834585058266E-5
[18,23]	[17,20]	[18,23]	8.221527334702755E-5
[17,20]	[17,23]	[18,23]	0.0026107882068911402
[18,20]	[17,23]	[18,23]	0.07008842674666035
[20,20]	[17,23]	[18,23]	1.2254139601534142E-5
[20,23]	[17,23]	[18,23]	3.1291820768203256E-5
[17,20]	[18,23]	[18,23]	2.1735467512619107E-4
[17,17]	[20,23]	[18,23]	3.532443303503465E-6
[18,18]	[17,17]	[20,23]	1.0253463748222446E-5
[18,23]	[17,17]	[20,23]	4.689084031199296-6
[17,18]	[17,18]	[20,23]	1.4192294334429856-5
[17,23]	[17,18]	[20,23]	0.017445580835275917
[18,23]	[17,18]	[20,23]	0.0036050545102892 6
[20,23]	[17,18]	[20,23]	5.6269008374391474E-6
[23,23]	[17,18]	[20,23]	0.07682076351417463
[17,18]	[17,23]	[20,23]	0.0022137165711291844
[18,18]	[17,23]	[20,23]	0.10066731917876008
[18,23]	[17,23]	[20,23]	8.346569575534735E-5
[17,17]	[18,23]	[20,23]	0.07069522548085776
[17,18]	[18,23]	[20,23]	1.3617100026602736E-4
[17,23]	[18,23]	[20,23]	2.654021561658798E-5
[17,20]	[17,18]	[23,23]	0.020532436113574616

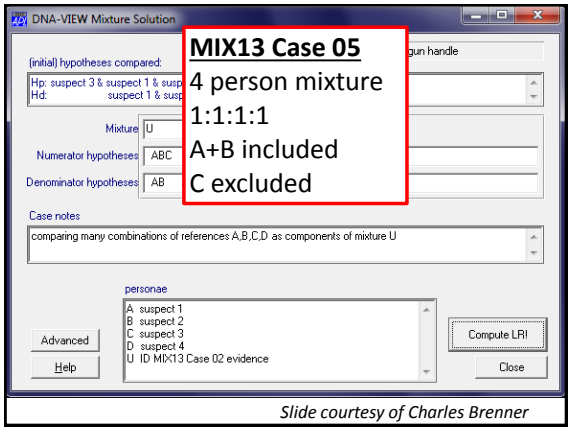
[18,20]	[17,18]	[23,23]	0.006249798760143661
[20,20]	[17,18]	[23,23]	1.1363213635606278E-4
[20,23]	[17,18]	[23,23]	0.07546711873604803
[18,18]	[17,20]	[23,23]	1.5942885706077583E-5
[18,23]	[17,20]	[23,23]	4.0638728270393845E-6
[17,17]	[18,20]	[23,23]	2.1663568224140716E-5

93 Possible **Genotype** Combinations

```
Number of known contributors under Hp: 3
, Suspect_05A_ref.csv, Suspect_05B_ref.csv, Suspect_05C_ref.csv
Number of known contributors under Hd: 0


Locus 1(D8S1179): Pr(E|Hp) = 0.0, Pr(E|Hd) = 8.0E-5, LR = 0.0
Locus 2(D21S11): Pr(E|Hp) = 0.01875, Pr(E|Hd) = 3.0E-5, LR = 603.86846
Locus 3(D7S820): Pr(E|Hp) = 0.0, Pr(E|Hd) = 4.6E-4, LR = 0.0
Locus 4(CSF1PO): Pr(E|Hp) = 3.4E-4, Pr(E|Hd) = 0.00108, LR = 0.31407
Locus 5(D3S1358): Pr(E|Hp) = 0.19148, Pr(E|Hd) = 8.6E-4, LR = 222.85716
Locus 6(TH01): Pr(E|Hp) = 0.0, Pr(E|Hd) = 9.8E-4, LR = 0.0
Locus 7(D13S317): Pr(E|Hp) = 0.0, Pr(E|Hd) = 4.5E-4, LR = 0.0
Locus 8(D16S539): Pr(E|Hp) = 0.19217, Pr(E|Hd) = 0.00211, LR = 91.11187
Locus 9(D2S1338): Pr(E|Hp) = 0.0, Pr(E|Hd) = 5.0E-5, LR = 0.0
Locus 10(D19S433): Pr(E|Hp) = 0.0, Pr(E|Hd) = 5.3E-4, LR = 0.0
Locus 11(vWA): Pr(E|Hp) = 0.0, Pr(E|Hd) = 4.7E-4, LR = 0.0
Locus 12(TPOX): Pr(E|Hp) = 3.3E-4, Pr(E|Hd) = 0.00526, LR = 0.06341
Locus 13(D18S51): Pr(E|Hp) = 0.01176, Pr(E|Hd) = 7.0E-5, LR = 176.38643
Locus 14(D5S818): Pr(E|Hp) = 4.9E-4, Pr(E|Hd) = 0.00243, LR = 0.19993
Locus 15(FGA): Pr(E|Hp) = 0.0, Pr(E|Hd) = 1.8E-4, LR = 0.0
LR total = 0.0
```

LR Total = 0.0



Slide courtesy of Charles Brenner

Summary of the Issues

- We need to move away from the interpretation of mixtures from an "allele-centric" point of view.
- Methods to incorporate probability will be necessary as we make this transition to interpret low-level profiles with drop-out.
- "Just as logic is reasoning applied to truth and falsity, probability is reasoning with uncertainty"

-Dennis Lindley

Summary of the Issues

- The LR is only logical approach to evaluate complex, low-level mixture evidence. Probabilistic genotyping software can overcome ***many (noticed I didn't say "ALL")*** of the limitations we are facing today.
- Software solutions will be helpful – but it's also important that we understand the limitations of these programs.

Concluding Thoughts

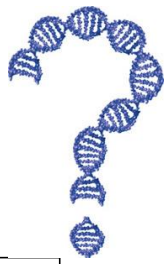
- Despite the improvements in protocols and interpretation guidelines since 2010, mixture interpretation tends to be all over the place.
- Some of this is a consequence of using a statistical approach that is inappropriate for complex mixture interpretation – CPI is often being used as a substitute for interpretation.
- Software solutions can greatly assist in the statistical evaluation and removal of bias.

Acknowledgments

National Institute of Justice and NIST Law Enforcement Standards Office

Dr. Charlotte Word
Dr. Robin Cotton
Dr. John Butler

Dan Katz



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