

INTERPRETING DNA PROFILE EVIDENCE IN COMPLEX DISPUTED PATERNITY CASES: BAYESIAN NETWORKS TO THE RESCUE

Philip Dawid
 University College London
 Julia Mortera & Paola Vicard
 Università Roma Tre

FORENSIC USES FOR DNA PROFILES

- Murder/Rape/...: Is **A** the culprit?
- Paternity: Is **A** the father of **B**?
- Immigration: Is **A** the mother of **B**? How are **A** and **B** related?
- Disasters: 9/11, tsunami, Romanovs,...

DNA Profile

- From blood, saliva, semen, hair root, ...
- Can be amplified from a single cell
- Record *genotypes* for 12–20 DNA *markers*
 - unlinked (different chromosomes)

A typical DNA profile

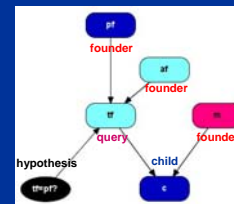
Marker	Genotype
FGA	20/24
FES	8/11
TH01	7/9
VWA	15/18
D3S1358	15
TPOX	8/10
CSF1PO	11/12
D5S818	12
D13S317	11/13
D7S820	8/9
D16S539	12/13
D2S1338	24/25
D8S1179	12
D21S11	30/33.2
D18S51	14/22
D19S433	14/14.2

D7S820

D7S820 is one of the 13 core CODIS STR genetic loci. This DNA is found on human chromosome 7. The DNA sequence of a representative allele of this locus is shown below. The tetrameric repeat sequence of D7S280 is **GATA**. Different alleles of this locus have from 6 to 15 tandem repeats of the **GATA** sequence.

```
001 AATTTTGTATTTTTTTAGAGACGGGGTTCCACATGTGGTCAGGCTGACTATGGAGT
061 TATTTAAGGTTAATATATATAAAAGGGTATGATAGAACACTTGTCTAGTTTAGAACGAA
121 CTAACGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT
181 TGATAGTTTTTTTTATCTGACTAAATAGTCTATAGTAAACATTTAATTACCAATATTG
241 GTGCAATTCGTGCAATGAGGATAAATGTGGAAATCGTTATAATCTTAAAGAATATATTC
301 CCTCTGAGTTTTTGATACCTCAGATTTTAAAGGCC
```

Disputed Paternity



We have DNA data **D** from a disputed child **c**, its mother **m** and the putative father **pf**

If **pf** is not the true father **tf**, this is a “random” alternative father **af**

Building blocks: **founder**, **child**, **query**

Disputed Paternity

LIKELIHOOD RATIO

$$LR_D = \frac{\text{Prob}(D|P)}{\text{Prob}(D|\bar{P})}$$

$$= \frac{\text{Prob}(c|m, pf, P)}{\text{Prob}(c|m, \bar{P})}$$

(Essen-Möller 1938)

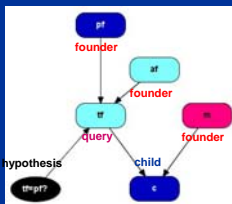
We have DNA data D from a disputed child c , its mother m and the putative father pf

If pf is not the true father tf , this is a "random" alternative father af

MISSING DNA DATA

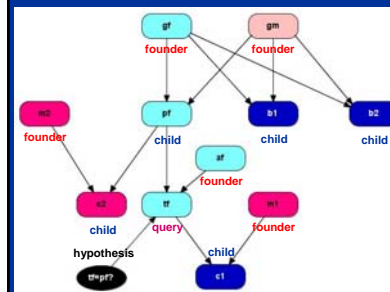
- What if we can not obtain DNA from the suspect ? (or other relevant individual?)
- Sometimes we can obtain indirect information by DNA profiling of **relatives**
- But analysis is complex and subtle...

Disputed Paternity Case



Building blocks: **founder, child, query**

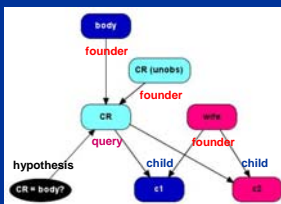
Complex Paternity Case



We have DNA from a disputed child $c1$ and its mother $m1$ but not from the putative father pf . We do have DNA from $c2$ an undisputed child of pf , and from her mother $m2$ as well as from two undisputed full brothers $b1$ and $b2$ of pf .

Building blocks: **founder, child, query**

Criminal Identification Case



Building blocks: **founder, child, query**

A **body** has been found, burnt beyond recognition, but there is reason to believe it might be that of a missing criminal **CR**. DNA is available from the **wife** of **CR**, and from two children $c1$ and $c2$ of **CR** and **wife**

Object-Oriented Bayesian Network

HUGIN 6

- Each building block (**founder / child / query**) in a pedigree can be an **INSTANCE** of a generic **CLASS** network — which can itself have further structure
- The pedigree is built up using simple mouse clicks to insert new nodes/instances and connect them up
- Genotype data are entered and propagated using simple mouse clicks

Under the microscope...

- Each **CLASS** is itself a Bayesian Network, with internal structure
- **Recursive**: can contain instances of further class networks
- Communication via *input* and *output* nodes

Single-marker analysis

(multiply LR's across markers)

12	.0003
13	.0018
14	.1009
15	.1004
16	.1949
17	.2834
18	.2162
19	.0866
20	.0137
21	.0015
22	.0003

Marker vWA

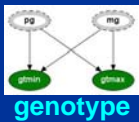
(Austro-German population allele frequencies)

Lowest Level Building Blocks

gene

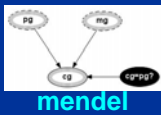
STR MARKER

having associated repertory of alleles together with their frequencies



GENOTYPE

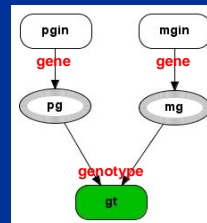
consisting of maximum and minimum of paternal and maternal genes



MENDELIAN SEGREGATION

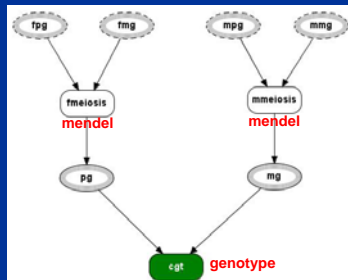
Child's gene copies paternal or maternal gene, according to outcome of fair coin flip

founder



FOUNDER INDIVIDUAL represented by a pair of genes *pgin* and *mgin* (instances of *gene*) sampled independently from population distribution, and combined in instance *gt* of *genotype*

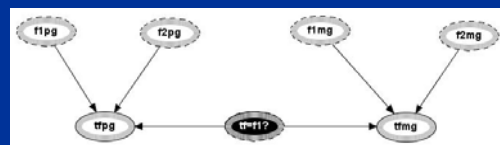
child



CHILD INDIVIDUAL

paternal [maternal] gene selected by instances *fmeiosis* [*mmeiosis*] of *mendel* from father's [mother's] two genes, and combined in instance *cgt* of *genotype*

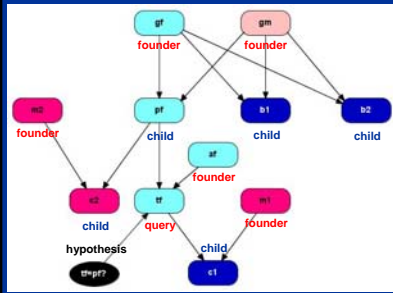
query



QUERY INDIVIDUAL

Choice of true father's paternal gene *tfpg* [maternal gene *mfpg*] as either that of *f1* or that of *f2*, according as *tf=f1?* is *true* or *false*.

Complex Paternity Case

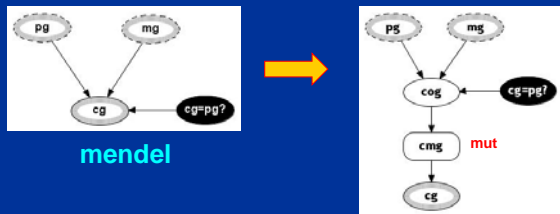


- Measurements for 12 DNA markers on all 6 individuals
- Enter data, "propagate" through system
- Overall Likelihood Ratio in favour of paternity: **1300**

MORE COMPLEX DNA CASES

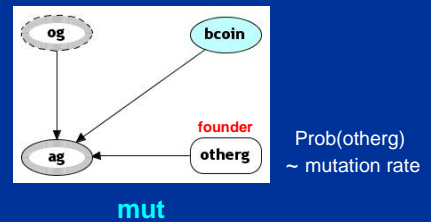
- Mutation
- Silent/missed alleles,...
- Mixed crime stains
 - rape
 - scuffle
- Multiple perpetrators and stains
- Database search
- Contamination, laboratory errors
 - ...

MUTATION



+ appropriate network **mut** to describe mutation process

e.g. proportional mutation:

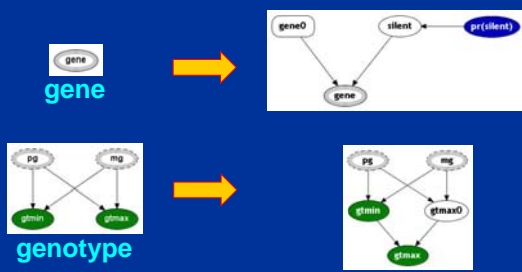


– or build other, more realistic, models

SILENT ALLELES

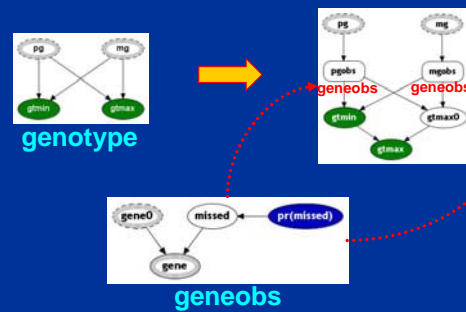
unobserved + *inherited*
e.g. 5 = 5/5 or 5/s

Code by additional allele (99)



MISSED ALLELES

unobserved + *non-inherited*



COMBINATION

- Can **combine** any or all of above features (and others), by using all appropriate subnetworks
- Can use any desired pedigree network
– *no visible difference at top level*
- Simply enter data (and desired parameter-values) and propagate...

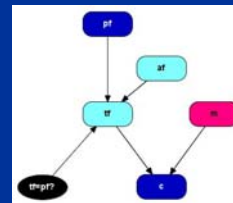
Effect of accounting for silent allele

- Simple paternity testing
- Paternity testing with additional measured individuals

Marker **vWA**
(Austro-German population allele frequencies)

12	.0003
13	.0018
14	.1009
15	.1004
16	.1949
17	.2834
18	.2162
19	.0866
20	.0137
21	.0015
22	.0003

Simple paternity testing – allowing for silent alleles



Paternal incompatibility

mgt = 12/20 pfgt = 13 cgt = 12

with mutation ~ 0.005

pr(silent)	LR	LR
0	0	3.8
0.000015	26	30
0.0001	125	127
0.001	203	203

$p_{12} = 0.0003$ – rare allele

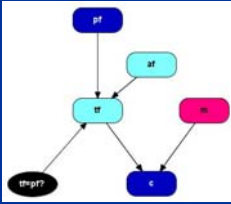
Maternal incompatibility

mgt = 16 pfgt = 18 cgt = 18

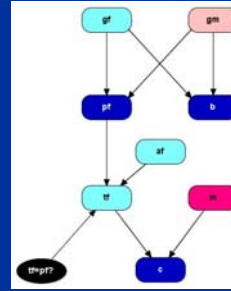
The mother must have passed a silent allele to the child
– who must have inherited allele 18 from his father

pr(silent)	LR
0	Impossible
0.000015	4.6
0.0001	4.6
0.001	4.6

Paternity testing



Paternity testing with brother too



Consider *additional* information carried by the brother's data B :

$$LR_B = \frac{\Pr(B | D, P)}{\Pr(B | D, \bar{P})}$$

where D denotes data on triplet (pf, c, m)

Overall likelihood ratio is

$$LR_{overall} = LR_D \times LR_B$$

Incompatible triplet

mgt = 12/15 pfgt = 14 cgt = 12

$B =$

		16/20	12/14	14	22*
$p(\text{silent})$	LR_D	LR_B	LR_B	LR_B	LR_B
0	0	1	0.55	1	3334
0.000015	0.5	1	0.55	1.00	1595
0.0001	2.5	1	0.55	1.00	404
0.001	7.5	1	0.55	1.00	46

$p_{12} = .0003$

$p_{22} = .0003$

*Maximum $LR_{overall}$ is 1027, at $p(\text{silent}) = 0.0000642$

Compatible triplet

mgt = 12/15 pfgt = 13 cgt = 12/13

$B =$

		13	13/16	21/22	22
$p(\text{silent})$	LR_D	LR_B	LR_B	LR_B	LR_B
0	556	1	1	1	1
0.000015	551	1	1.00	1	0.51
0.0001	528	1	1.02	1	0.52
0.001	410	1	1.11	1	0.61

Extensions

- Estimation of mutation rates from paternity data
- Peak area data
 - mixtures
 - contamination
 - low copy number

REFERENCES

- Dawid, A. P., Mortera, J., Pascali, V. L. and van Boxel, D. W. (2002). Probabilistic expert systems for forensic inference from genetic markers. *Scand. J. Statist.* **29**, 577–595.
- Dawid, A. P. (2003). An object-oriented Bayesian network for estimating mutation rates. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, January 3–6 2003, Key West, Florida, edited by Christopher M. Bishop and Brendan J. Frey. ISBN 0-9727358-0-1. Online at <http://nips.cc/papers/2003/>.
- Mortera, J., Dawid, A. P. and Lauritzen, S. L. (2003). Probabilistic expert systems for DNA mixture profiling. *Theor. Pop. Biol.* **63**, 191–205.
- Cowell, R. G., Lauritzen, S. L. and Mortera, J. (2006). Identification and separation of DNA mixtures using peak area information. *Forensic Science International* (to appear).
- Dawid, A. P., Mortera, J. and Vicard, P. (2007). Object-oriented Bayesian networks for complex forensic DNA profiling problems. *Forensic Science International: Genetics* **1** (to appear).