Forensics
Paternity cases, complex identification cases

Thore Egeland[(1)]

(1) Norwegian University of Life Sciences

thore.egeland@nmbu.no

Contents

- What is forensics?
  - Principles for evaluation of evidence
- Practical evaluation of evidence;
  - Hypotheses
  - Likelihood Ratio (LR)
  - Assumptions. Interpretation
- Complications;
  - Mutations
  - Complex pedigrees: Large, inbred
  - ...
- Part II: Alternatives to LR, interpretation;
  - Introducing prior information like: we may have *some* information on say age
  - Exclusion power

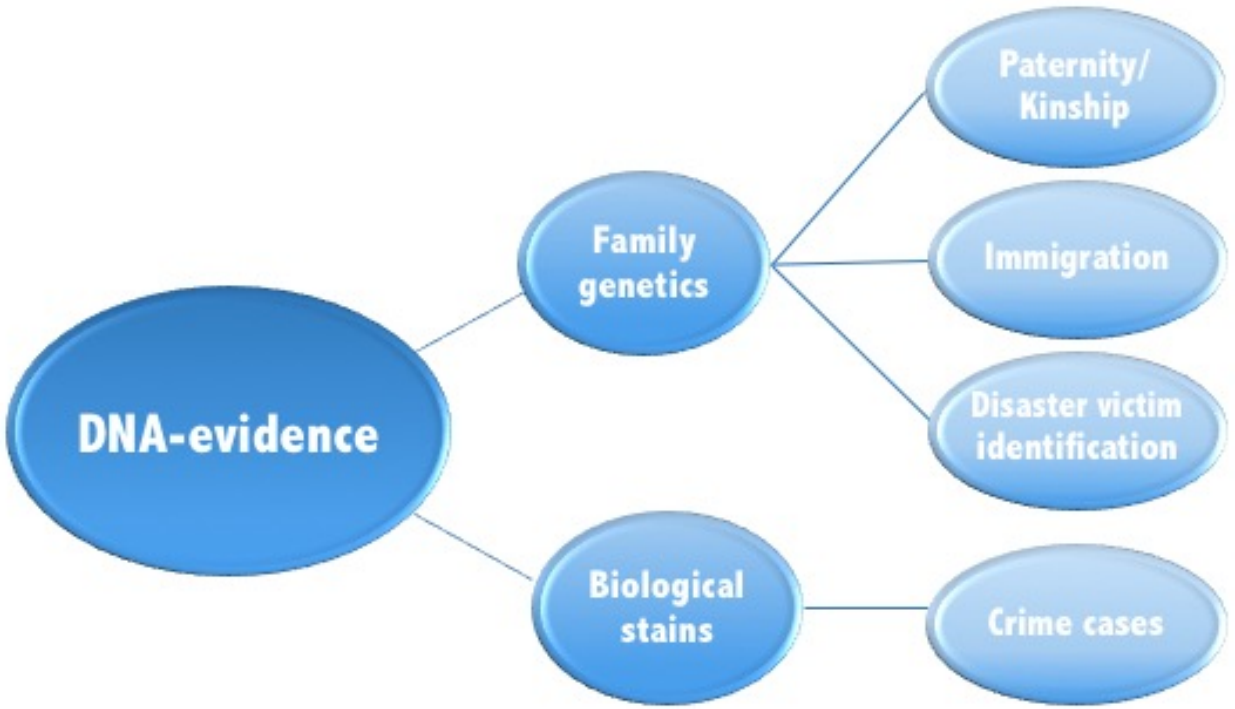
Different legal systems

- Forensics: the application of science in legal settings.
- Different legal systems, traditions, have implications for the role of the *forensic expert*:
  - Adversarial. US, UK, other English speaking countries;
    - "battle of experts"
  - Inquisitorial. Large parts of mainland Europe:
    - "unbiased, independent expert opinion"

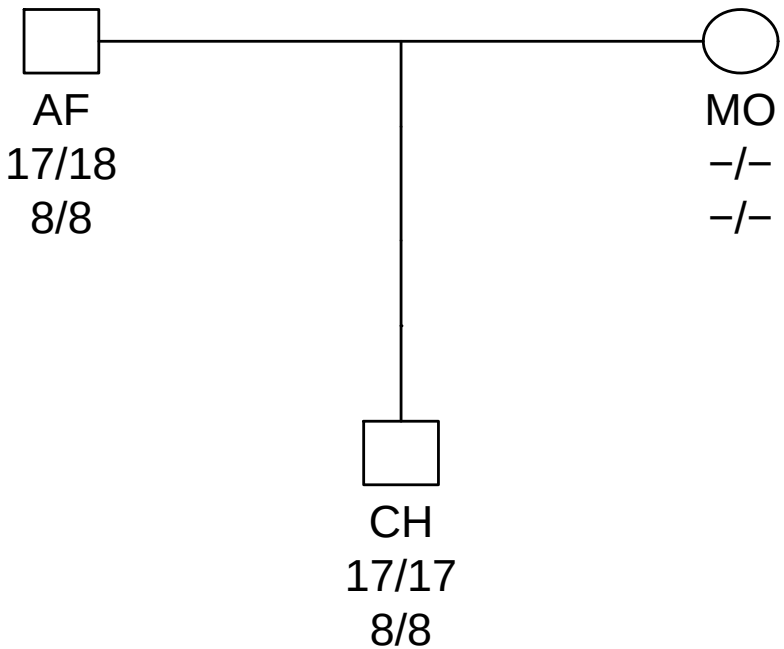
Principles for evaluation of evidence

1. To evaluate the uncertainty of any given proposition it is necessary to consider at least one alternative proposition.
2. Scientific interpretation is based on questions of the following kind: What is the probability of the data given the proposition?
3. Scientific evidence is conditioned not only by the competing propositions, but also by the framework of circumstances within which they are to be evaluated.

Overview of forensic genetics



Hypotheses



- H_1: AF biological father of CH.
- H_2: AF and CH unrelated.
- Notation. Sometimes:
- $H_1 = H_P$:
  "prosecution hypothesis",
- $H_2 = H_D$:
  "defence hypothesis".

Likelihood Ratio (LR)

Definition of the LR

$$LR_{H_1,H_2}(E) = \frac{P(E \mid H_1)}{P(E \mid H_2)},$$

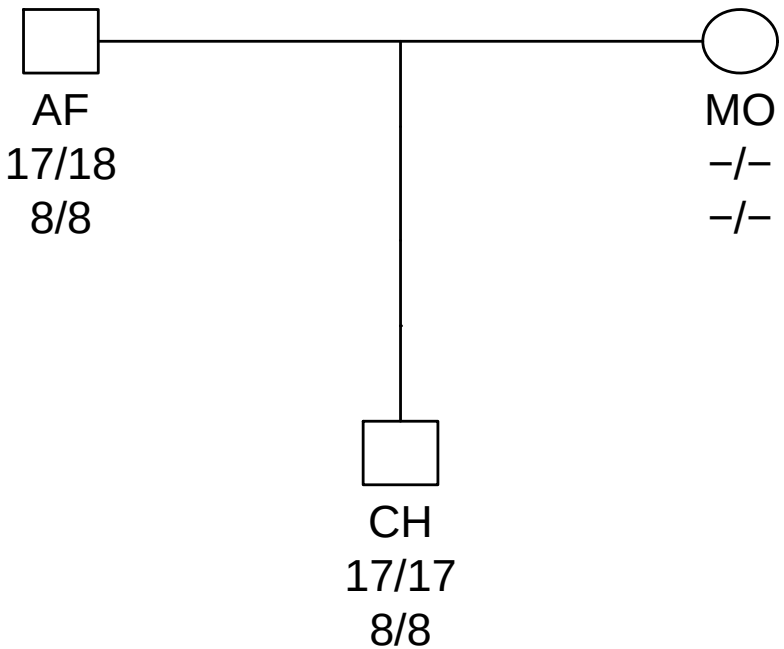
depending on

- The hypotheses H_1, H_2 under consideration
- The data E that we are considering

Meaning of the LR

- $P(E \mid H)$ is the probability to get E, if hypothesis H is true
- It is also called the likelihood of the hypothesis, given the evidence E
- The LR says how much better the explanation for E offered by H_1 is, compared to the explanation offered by H_2.
- The individual likelihoods $P(E \mid H_i)$ do not allow for any inference considered on their own: the issue is not to predict the evidence (as $P(E \mid H)$ does) but to see which mechanism explains it better
- Special LR-s: PI (paternity index), SI (sib index),...

Likelihood Ratio. Example



$$LR = \frac{P(E \mid H_1)}{P(E \mid H_2)} = \cdots = \frac{P(g_{CH} \mid g_{AF})}{P(g_{CH})}$$

$$LR_1 = \frac{\frac{1}{2}p_{17}}{p_{17}^2} = \frac{1}{2 \times 0.204} = 2.45$$

$$LR_2 = \frac{p_8}{p_8^2} = \frac{1}{0.554} = 1.81.$$

Multiplying LR–s

Recall that for events A and B

$$P(A \cap B) = P(A)P(B)$$

if A and B are independent. Similarly

$$LR = LR_1 \times LR_2 = 2.45 \times 1.81 = 4.4.$$

if markers are independent.

- The independence assumption holds if markers are unlinked (not always needed) and in *linkage equilibrium:*

Linkage equilibrium

- Locus 1 with allele frequencies p_a
- Locus 2 with allele frequencies q_a

- Haplotype frequencies H_{ab}

- If $H_{ab} - p_a q_b = 0$: "linkage equilibrium" (LE). Otherwise Linkage Disequilibrium (LD).

- This is a *statistical* property
- It does not depend on the loci themselves, e.g., loci may be in LE in a single population but not in a composed population
- Is a property similar to Hardy-Weinberg equilibrium: a statistical property, following from Mendelian segregation. LE is asymptotically reached (LD diminishes per generation) in a homogeneous infinite population if recombination is possible.

Example: Haplotype frequencies

loc1	loc2	freq1	freq2	$P(hap \mid LE)$	Count	$P(hap \mid Count)$
A	B	0.2	0.3	0.2*0.3=0.06	10	10/100=0.10
A	b	0.2	0.7	0.2*0.7=0.14	15	15/100=0.15
a	B	0.8	0.3	0.8*0.3=0.24	25	25/100=0.25
a	b	0.8	0.7	0.8*0.7=0.56	50	50/100=0.50
tot				1.00	100	1.00

Table 1: LE and count based haplotype frequency estimates

Likelihood Ratio. Software

- `Familias`, `http://familias.no`. R version not maintained
- `forrel`. *This course*
- `DNA-View`, ...

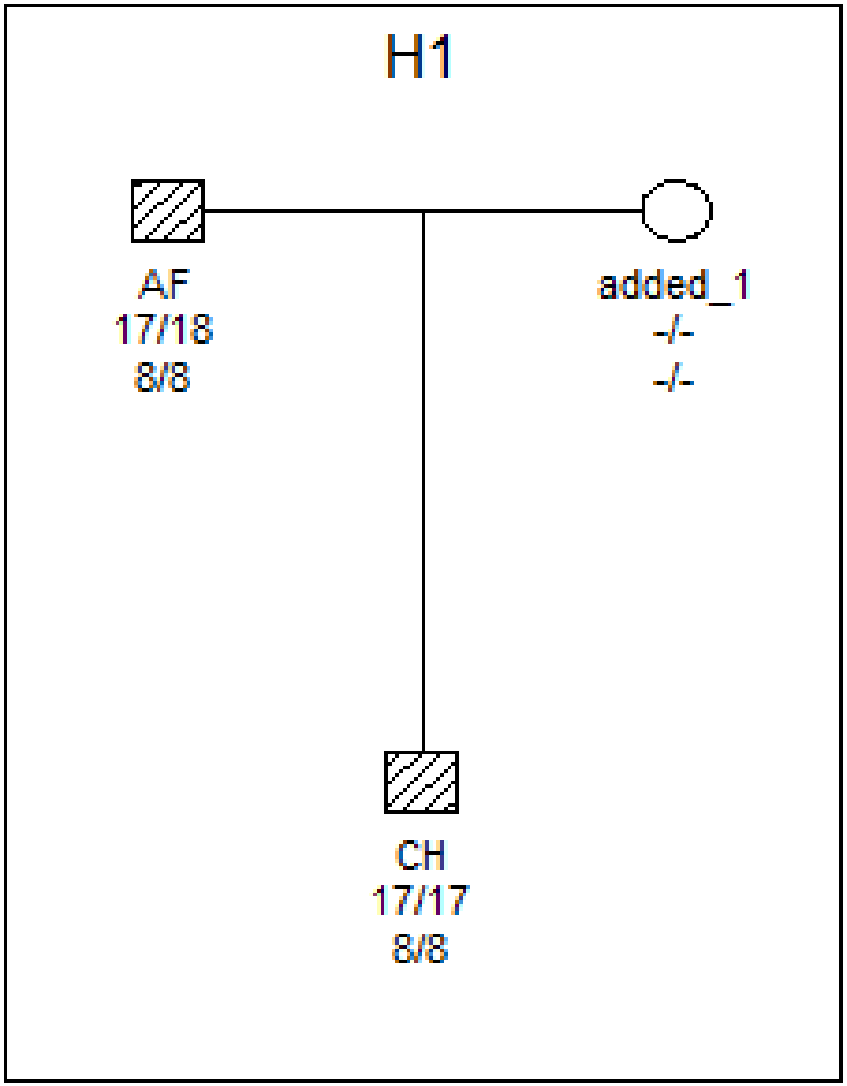
forrel

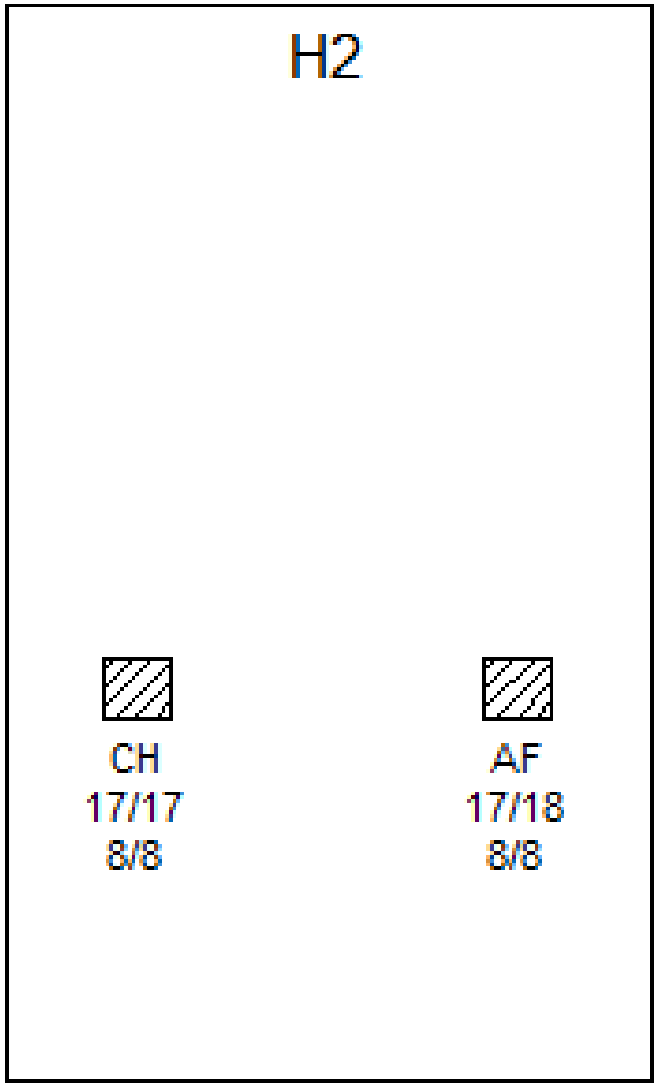
- Input can be entered manually or from files, see exercises
- For simplicity, in the lecture, we convert `.fam` - files
- Basic functions: `readFam`, `plotPedList`, `kinshipLR`

Step 1: Input and plot

```
library(forrel)
peds = readFam("Demo2markers.fam", verbose = FALSE)
plotPedList(peds, marker=1:2, shaded = typedMembers,
                frametitles = c("H1","H2"))
# http://familias.name/norbisRelatedness/Demo2markers.fam
```

Likelihood Ratio. Plot





Step 2: Calculation

```
res = kinshipLR(peds, ref = 2)
res # main output
unclass(res) # all output
```

```
Total LR:
    H1: father H2: not father
      4.423845       1.000000
> unclass(res)
$LRtotal
    H1: father H2: not father
      4.423845       1.000000

$LRperMarker
         H1: father H2: not father
D3S1358   2.450403              1
TPOX      1.805354              1
```

Advantages of R: Generality

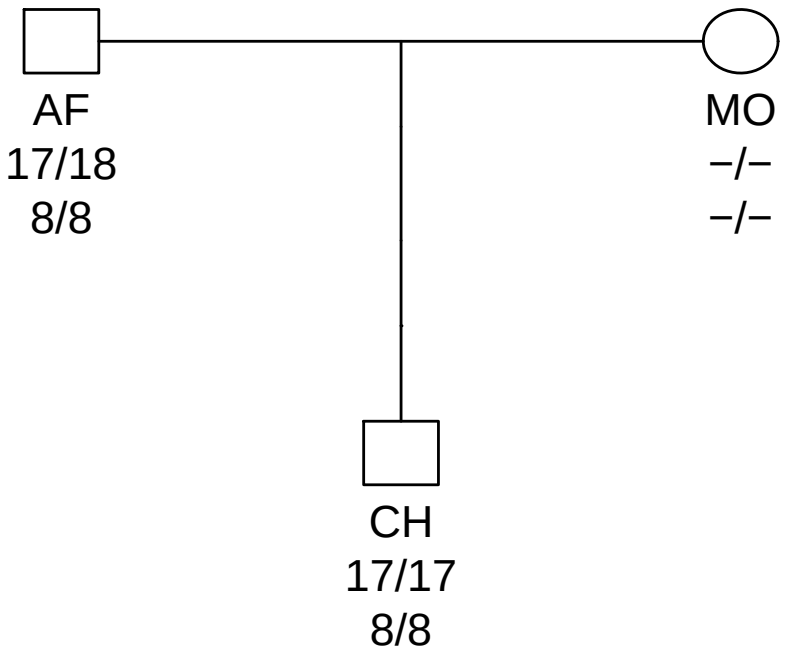
Assume "AF" is the child of first cousins:

```
H1 = peds[[1]]
H2 = peds[[2]]
founderInbreeding(H1,"AF") = 1/16
founderInbreeding(H2[[2]],"AF") = 1/16
kinshipLR(peds, ref = 2) #Same LR in *this* case
```

Inbreeding does not change LR in *this case* since

$$LR = \frac{P(g_{CH} \mid g_{AF})}{P(g_{CH})}$$

Step 3: Interpretation and assumptions



- Interpretation LR=4.4: The data is 4.4 times more likely if AF is assumed to be the father compared to the unrelated alternative.
- Assumptions
  - Hardy–Weinberg Equilibrium (HWE).
  - Independent markers.
  - No artefacts: no mutation, no silent alleles, no drop–out/in, no error; discussed later)

One Verbal Scale for LR

LR	Expert guidance*
1	… do not support one proposition over the other
2 - 10	weak support
10 - 100	moderate support
100 - 1000	moderately strong support
1000 - 10000	strong support
10000 - 1 million	very strong support
Over 1 million	extremely strong support

*ENFSI Guideline for Evaluative Reporting in Forensic Science

Real case. Output from Familias

Compare DNA

System	LR	Child	Alleged father
D3S1358	2.46675184	17, 18	17, 17
TH01	1.194605231	6, 9	6, 7
D21S11	1.095934095	29, 30	28, 29
D18S51	2.153261166	14, 16	16, 17
PENTA_E	0	7, 11	10, 16
D5S818	1.406126529	12, 12	12, 13
D13S317	4.041610583	8, 8	8, 11
D7S820	1.433569585	9, 10	9, 13
D16S539	8.312297405	13, 14	11, 14
CSF1PO	2.024678178	10, 10	10, 11
PENTA_D	11.98925175	8, 11	8, 13
VWA	5.565000184	19, 19	17, 19
D8S1179	9.650567455	13, 16	11, 16
TPOX	1.78765206	8, 8	8, 8
FGA	2.956393798	21, 22	21, 21
D12S391	2.183521522	19, 22	19, 23
D1S1656	3.333333333	14, 16	14, 15
D2S1338	3.147059638	18, 20	18, 23
D22S1045	26.74815224	12, 12	12, 15
D2S441	1.445947587	10, 13	10, 15
D19S433	3.343765883	12, 15	12, 14

Total LR: 0 Save Close

Beyond standard cases: Complicating factors

- Mutations.
- Complex pedigrees: Large, inbred.
- Deviations from HWE. *Theta corrrection*.
- Inbred founders. founderInbreeding.
- Silent alleles: Homozygote or silent allele?
- Artefacts: Drop–out, drop–in, genotyping error.

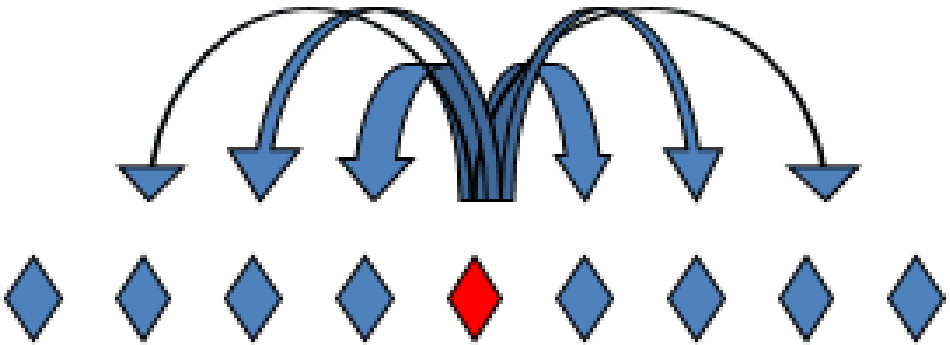
Mutation. Motivation

Marker	CH	AF	LR	LR(mut)
D3S1358	17/17	17/18	2.45	2.45
TPOX	8/8	8/8	1.81	1.80
D6S474	16/17	14/15	0.000	0.001
...	...	...	...	...
D19S433	12/15	12/14	3.34	3.34
Total			0	25070642

Mutation:

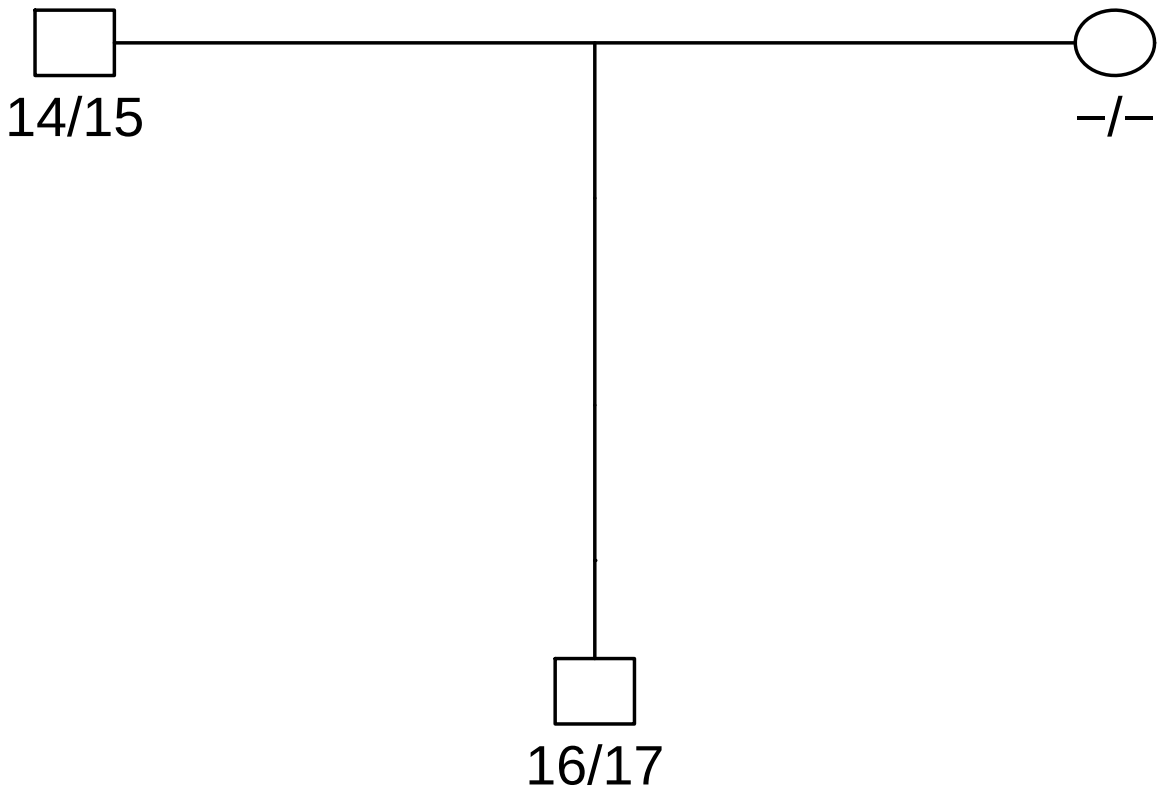
- Observed if parent and child share no alleles.
- Other examples? Mendelian inconsistencies.
- Mutation models interesting also in population genetics
- The forensic community is well positioned to study mutations.

Mutation: Biology

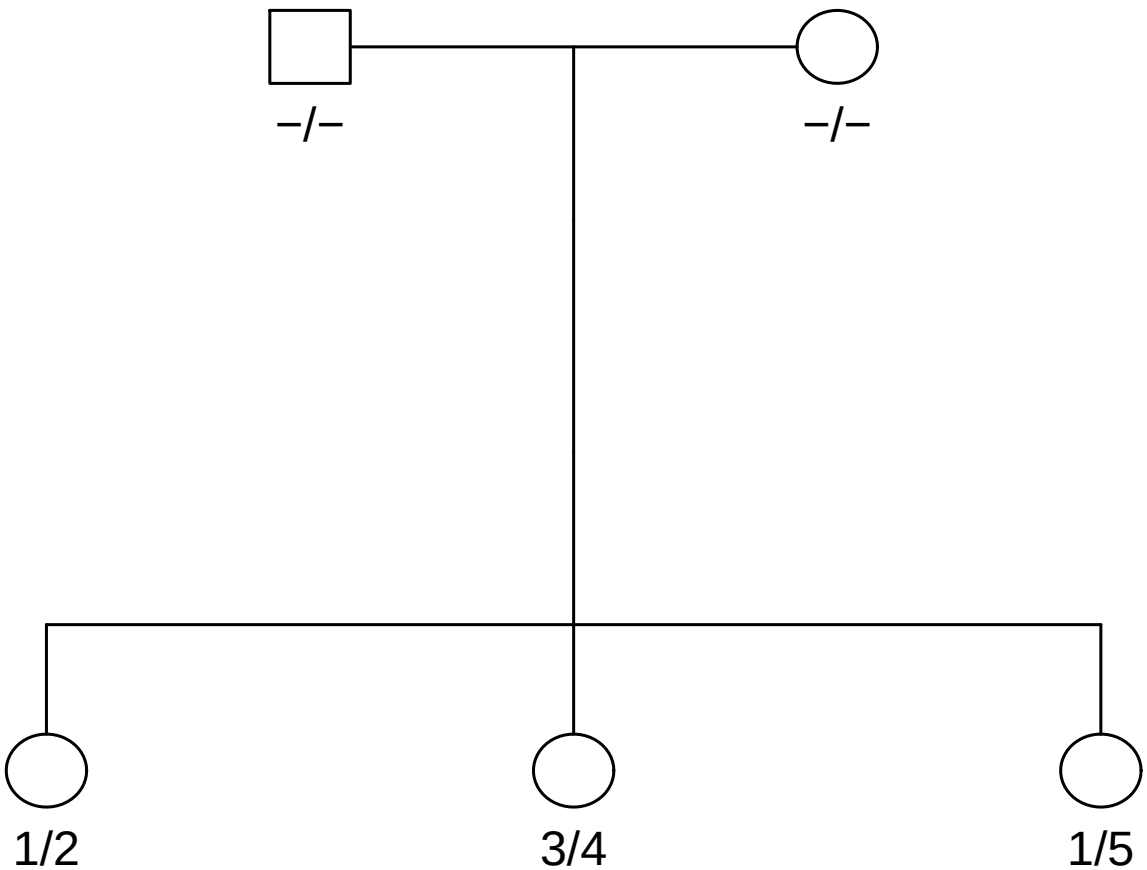


- Mutation rates higher in males.
- Short mutations more likely: One step mutation more likely than two steps and so on.
- Mutation rates:
  `http://www.cstl.nist.gov/strbase/mutation.htm`

Standard example



Non-standard example



The mutation matrix specifies the model

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} & \dots & m_{1n} \\ m_{21} & m_{22} & m_{23} & \dots & m_{2n} \\ m_{31} & m_{32} & m_{33} & \dots & m_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ m_{n1} & m_{n2} & m_{n3} & \dots & m_{nn} \end{bmatrix}$$

$m_{ij} =$ allele i transmitted as j

Mutation models in pedmut

- custom. Completely general, see exercise.
- equal. Simplest.
- proportional. Favoured by mathematicians, not used much.
- stepwise. Favoured by forensic case workers,
- onestep. Favoured by population geneticists.

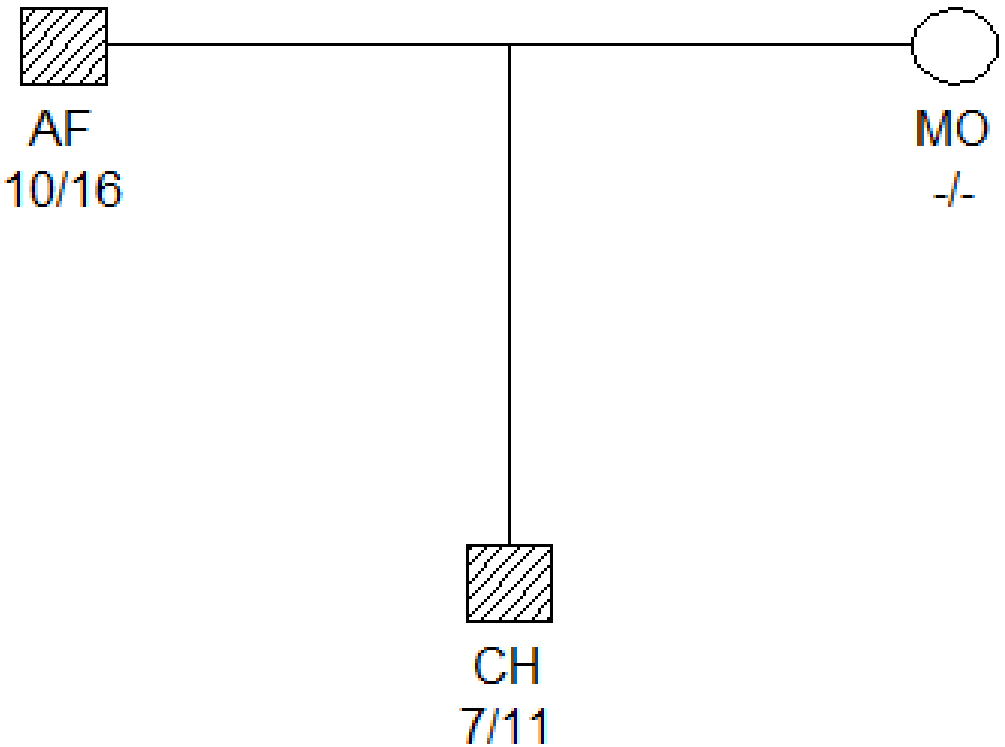
Stepwise mutation model

	14	15	16	17
14	0.995000	0.00450	0.00045	0.000045
15	0.002380	0.99500	0.00238	0.000238
16	0.000238	0.00238	0.99500	0.002380
17	0.000045	0.00045	0.00450	0.995000

Paternity case with mutation: R version

```
> peds = readFam("Solutions2_9.fam", verbose = FALSE)
> kinshipLR(peds, ref = 2)$LRperMarker
         H1: father H2: not father
D3S1358    2.466752               1
TH01       1.194605               1
D21S11     1.095934               1
D18S51     2.153261               1
PENTA_E    0.000000               1
D5S818     1.406127               1
D13S317    4.041611               1
D7S820     1.433570               1
D16S539    8.312297               1
CSF1PO     2.024678               1
PENTA_D   11.989252               1
VWA        5.565000               1
D8S1179    9.650567               1
TPOX       1.787652               1
FGA        2.956394               1
D12S391    2.183522               1
D1S1656    3.333333               1
D2S1338    3.147060               1
D22S1045  26.748152               1
D2S441     1.445948               1
D19S433    3.343766               1
```

Paternity case with mutation: plot



pedmut

Below the "proportional" model is introduced for "PENTA_E":

```
peds = readFam("Solutions2_9.fam", verbose = FALSE)
H1 = peds[[1]]
H2 = peds[[2]]
mutmod(H1,marker="PENTA_E")=list("prop",rate = 0.005)
mutmod(H2,marker="PENTA_E")=list("prop",rate = 0.005)
kinshipLR(peds, ref = 2)
Total LR:
      H1       H2
53573994        1
```

Stationarity. Intuitively

p

Name	Frequency
1	0.5
2	0.5

M

Alle...	1	2
1	0.99	0.01
2	0.01	0.99

$$\text{Freq of '1' next generation} = \overset{1 \cdot \ 1 \to 1}{0.5} \cdot 0.99 + \overset{2 \cdot \ 2 \to 1}{0.5} \cdot 0.01 = 0.5.$$

Freq of '2' next generation = 1 – 0.5 =0.5.

Frequencies do not change, i.e., remain 0.5 and 0.5 and so the model (M, p) is stationary.

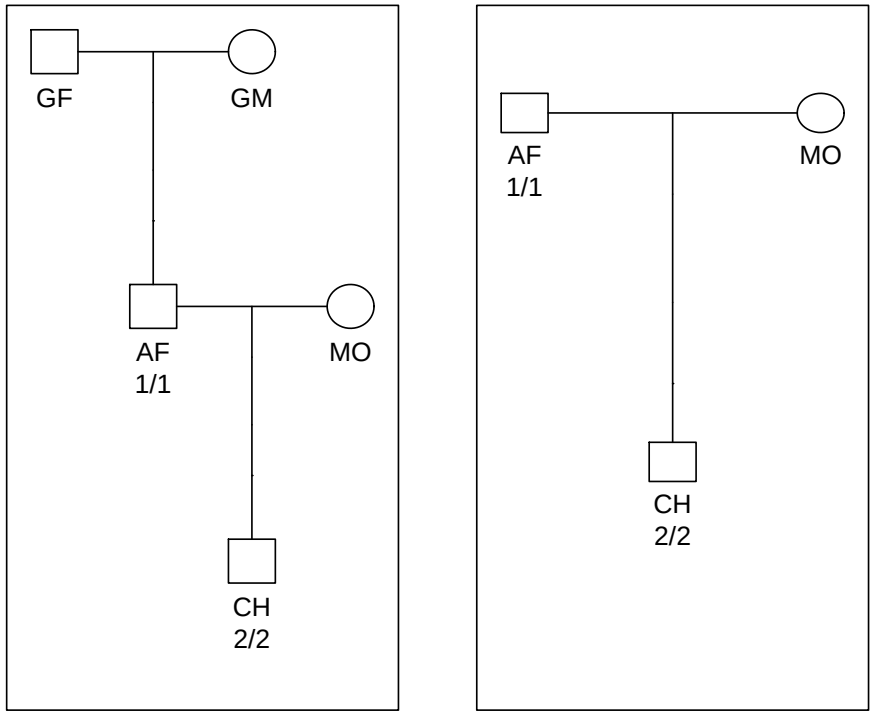
Stationarity*. Mathematics

- $M = \{m_{ij}\}$ mutation matrix.
- v row vector with probabilities for the n alleles.
- Probability distribution for the alleles after one generation: vM.
- After k generations: vM^k.
- If p is the vector of population probabilities for the alleles, a mutation model (M, p) is stationary if and only if $pM = p$.

Problems with non stationary models

- LR changes if irrelevant (untyped) persons are added.
- Frequencies change between generations; arbitrary choice required.
- Differences can be inflated, cause confusion.

Example with non–stationary mutation model



$$LR = 1.010025$$

Problems also with stationary models

Example

$$M = \begin{pmatrix} 1-2R & \frac{3}{2}R & \frac{1}{2}R \\ \frac{1}{2}R & 1-R & \frac{1}{2}R \\ \frac{1}{2}R & \frac{1}{2}R & 1-R \end{pmatrix}$$

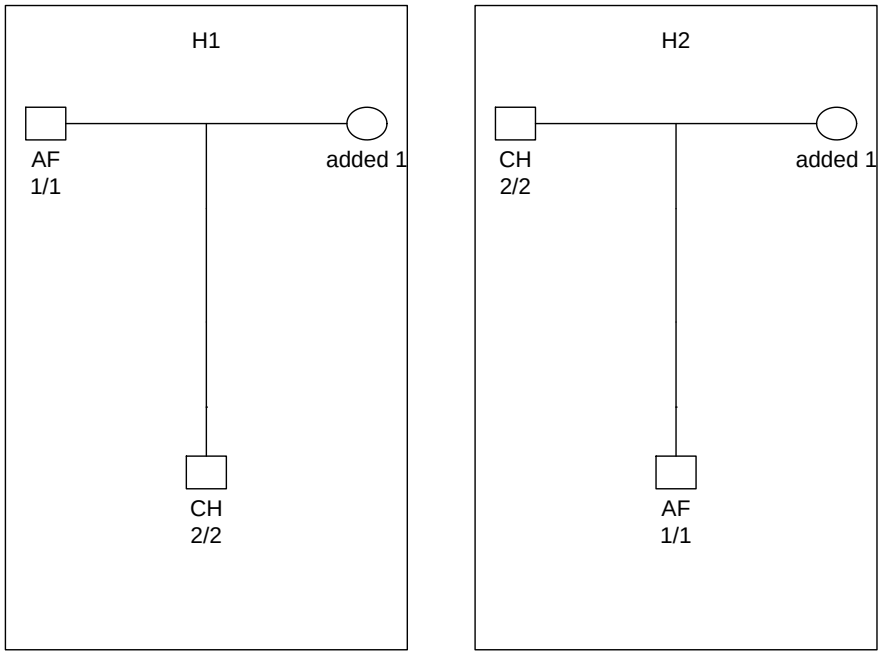
Stationary distribution $p = (3/15, 7/15, 5/15),\ 0 \leq R < 0.5.$

$$PI(AF = 1/1, CH = 2/2 \mid \text{AF father of CH}) = \frac{m_{12}}{p_2} = \frac{45R}{14},$$

$$PI(AF = 1/1, CH = 2/2 \mid \text{CH father of AF}) = \frac{m_{21}}{p_1} = \frac{35R}{14}.$$

Equality if and only if the mutation model is *balanced*, then:
$p_1 m_{12} = p_2 m_{21}$.

Example with non–balanced mutation model



$$LR = 0.66$$

Model	Parameters	Stationary?[1]	Lumpable?[2]	Reversible?[3]	Biologically OK?[4]
Equal	Mutation **Rate**	NO	YES	NO	NO
Proportional	Expected (E) mutation **Rate**[5]	YES	YES	YES	NO
Stepwise	Mutation **Rate** and **Range**[6]	NO	NO	NO	For integer alleles
Stepwise (stationary)	Mutation **Rate**[6] and **Range**	YES	NO	NO	For integer alleles
Extended stepwise[8]	Mut **Rate**[6], **Range, Rate 2**[7]	NO	NO	NO	YES[8]