Forensics applications I: Principles and methods

Thore Egeland[(1),(2)]

(1) Norwegian University of Life Sciences,
(2) Department of Forensic Sciences, Oslo University Hospital

NORBIS, 2022, thore.egeland@nmbu.no

Contents

- What is forensics?
  - Principles for evaluation of evidence
- Practical evaluation of evidence
  - Hypotheses
  - Likelihood Ratio (LR)
  - Assumptions. Interpretation
  - Mutations
  - ...
- Part II: Alternatives to LR, applications
  - Introducing prior information like: we may have *some* information on say age
  - Exclusion power
  - Disaster Victim Identification

Different legal systems

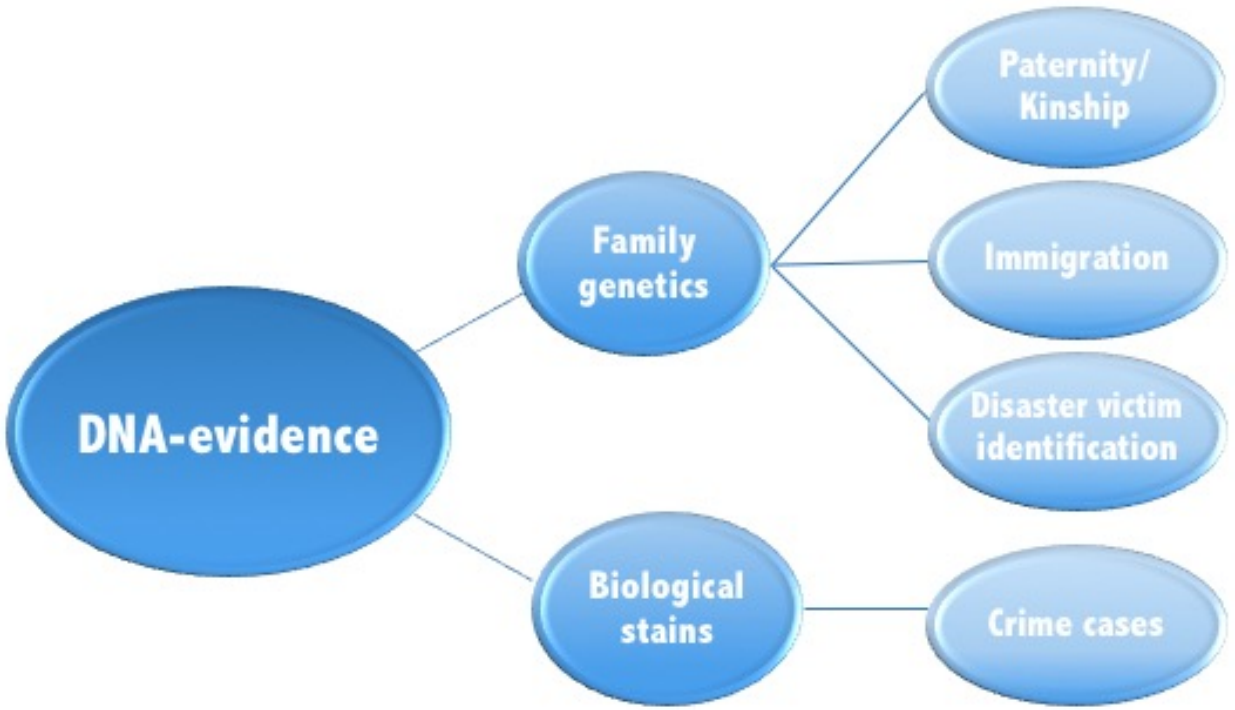
- Forensics: the application of science in legal settings.
- Different legal systems, traditions, have implications for the role of the *forensic expert*:
  - Adversarial. US, UK, other English speaking countries;
    - "battle of experts"
  - Inquisitorial. Large parts of mainland Europe:
    - "unbiased,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

“These are my principles.
If you don't like them I have others”.
Groucho Marx

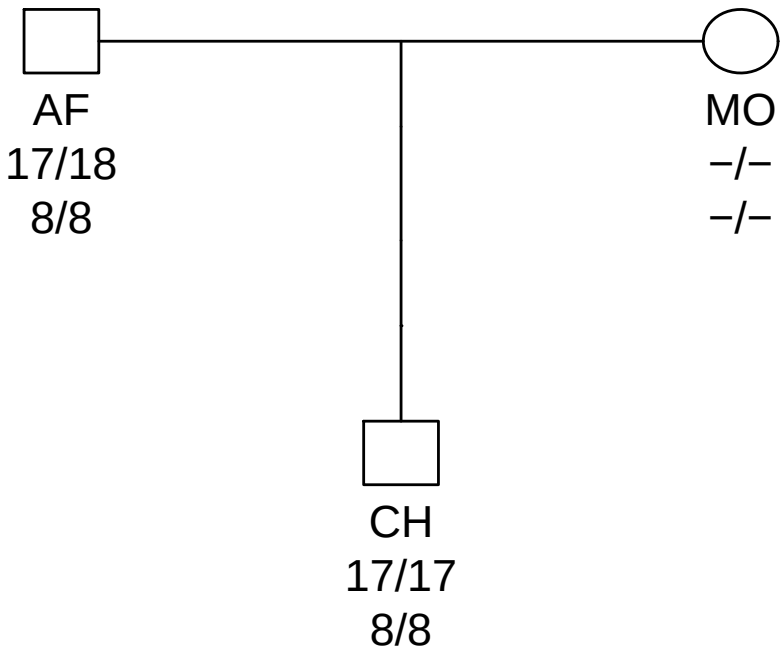
Principles for evaluation of evidence

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

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

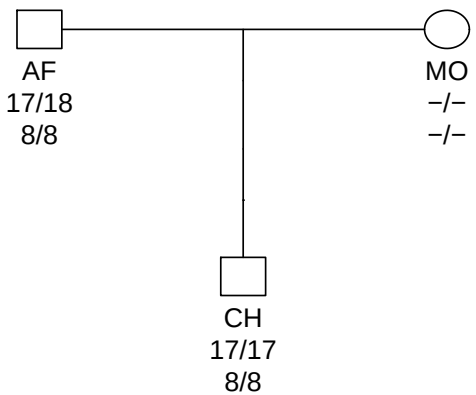
depending on

- 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



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

$$LR_1 = \frac{2p_{17}p_{18}\,\frac{1}{2}p_{17}}{2p_{17}p_{18}\,p_{17}^2} = \frac{1}{2p_{17}} = \frac{1}{2 \times 0.204} = 2.45$$

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

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* and in *linkage equilibrium*

Linkage equilibrium (skip in presentation?)

- 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.

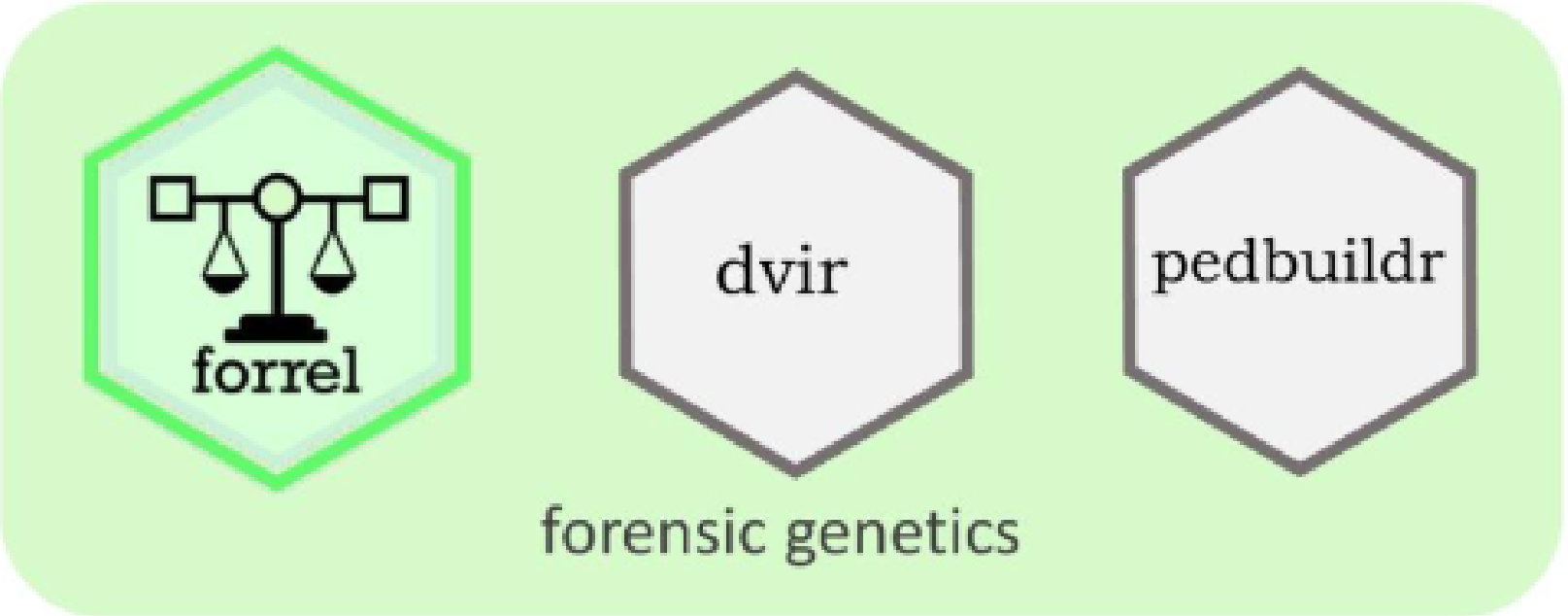
Example: Haplotype frequencies (skip in presentation?)

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
- `DNA-View`, ...
- `forrel`.



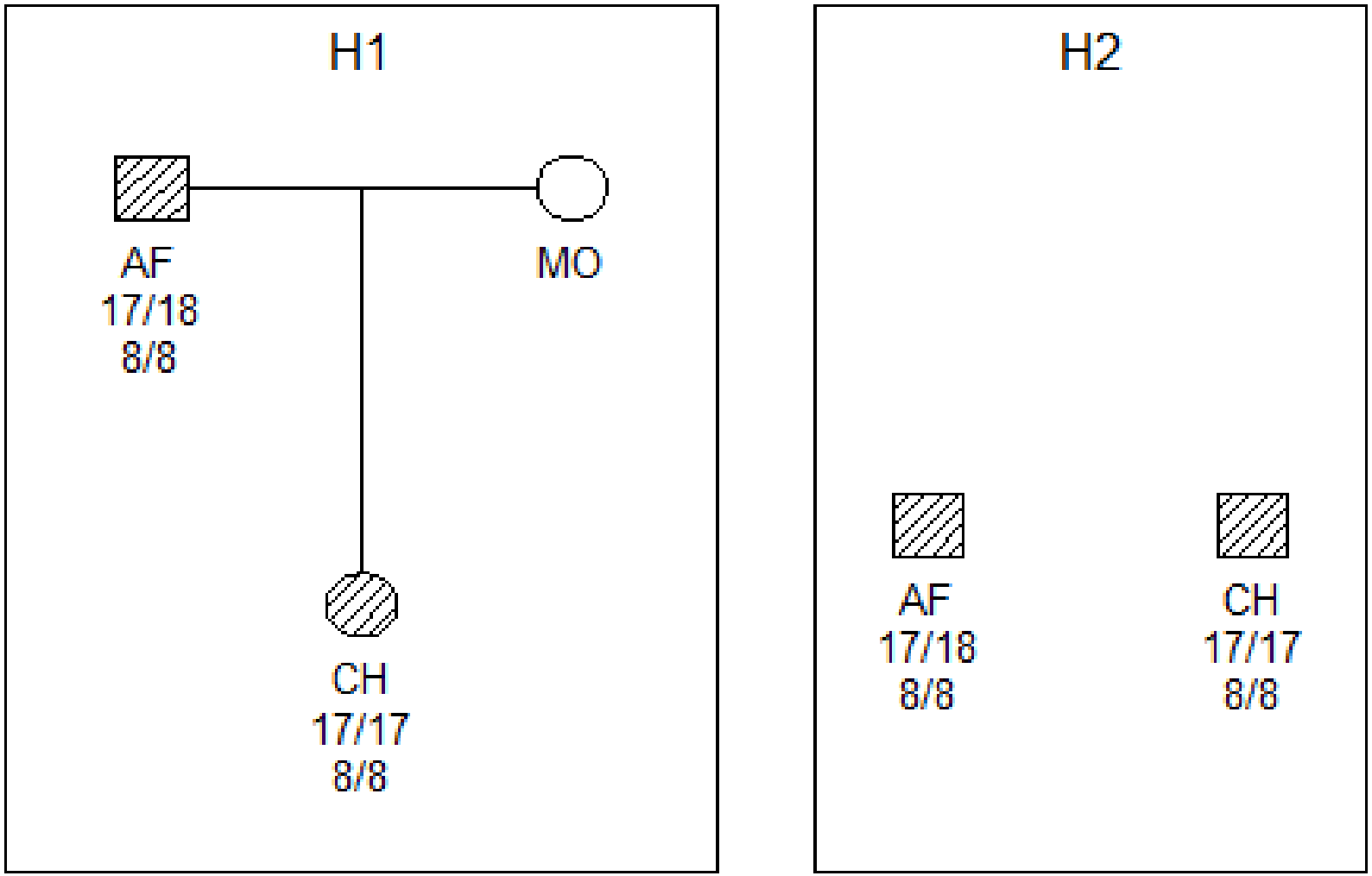
- Main function: `kinshipLR`

Step 1: Input and plot

```
library(forrel)
H1 = nuclearPed(fa = "AF", mo = "MO", child = "CH",
                sex = 2)
H2 = list(singleton("AF"), singleton("CH"))

afr1 = c("17" = 0.204, "18" = 0.140,
         rest = 1 - 0.204 -0.140)
H1 = addMarker(H1, AF = "17/18", CH = "17/17",
               afreq = afr1)
afr2 = c("8" = 0.554, rest = 1 - 0.554)
H1 = addMarker(H1, AF = "8/8", CH = "8/8",
               afreq = afr2)
plotPedList(list(H1, H2), titles = c("H1", "H2"),
            marker = 1:2, source = 1,
            hatched = typedMembers)
```

Paternity case. Plot



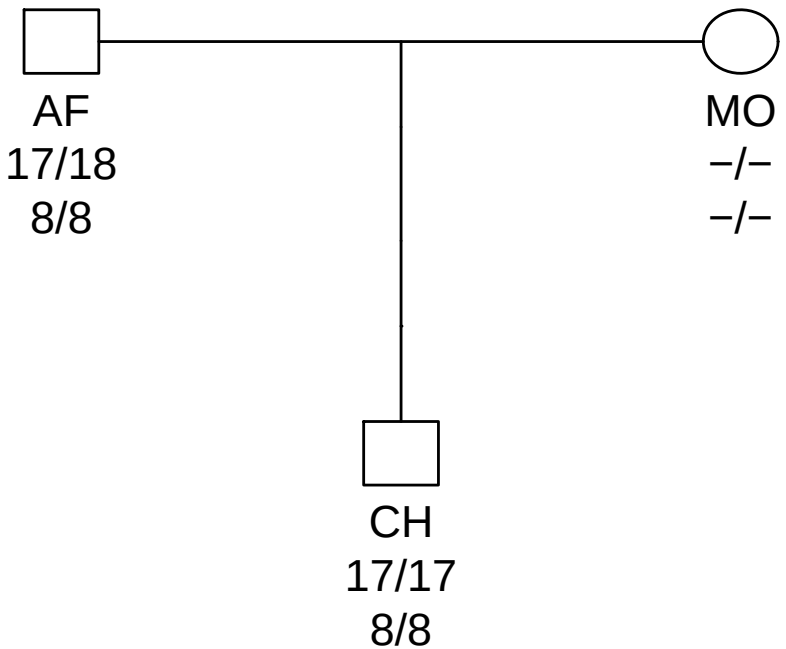
Step 2: Calculation

```
res = kinshipLR(H1, H2, 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
```

Step 3: Interpretation and assumptions



- Interpretation of 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 genotyping error.

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

Beyond standard cases: Complicating factors

- Pairwise relationships
- 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.

Alternative formulation of hypotheses: p-values?

Forensic formulation

- H_1: AF biological father of CH.
- H_2: AF and CH unrelated.

Forensic practice: Claim H_1 if LR > T (= 10,000, say)

Parametric reformulation:

- H_1: $\kappa = (0, 1, 0)$
- H_2: $\kappa = (1, 0, 0)$

Generalisation: consider all (non-inbred) alternatives:

- H_1: $\kappa = (0, 1, 0)$
- H_2: $\kappa \neq (0, 1, 0)$

Standard practice: Reject H_1 if p-value < α (= 0.05)

Pairwise relationshps

- A single marker:
  - Genotypes G_1 and G_2 observed in the two individuals
  - Idea for computing L(κ): *Condition on IBD status 0, 1 or 2*

$$L(\kappa) = P(G_1, G_2 \mid \kappa) = P(G_1, G_2 \mid UN)\,\kappa_0 + P(G_1, G_2 \mid PO)\,\kappa_1 + P(G_1, G_2 \mid MZ)\,\kappa_2$$

UN = unrelated
PO = parent/offspr
MZ = monozygotic

- With several independent markers:

$$L(k) = \prod L_i(k)$$

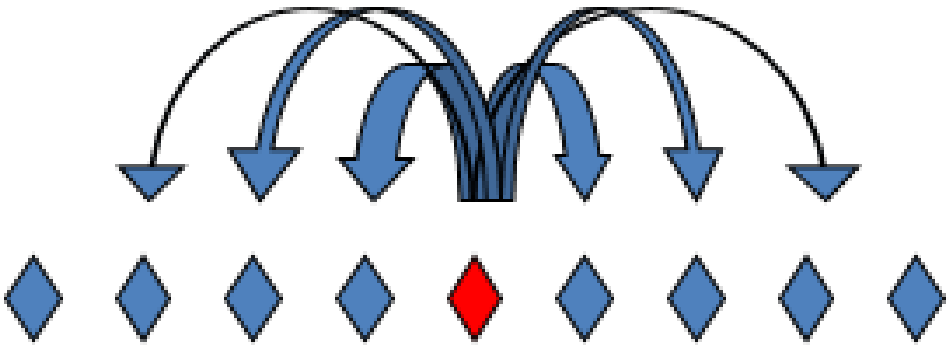
Exclusion or mutation?

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

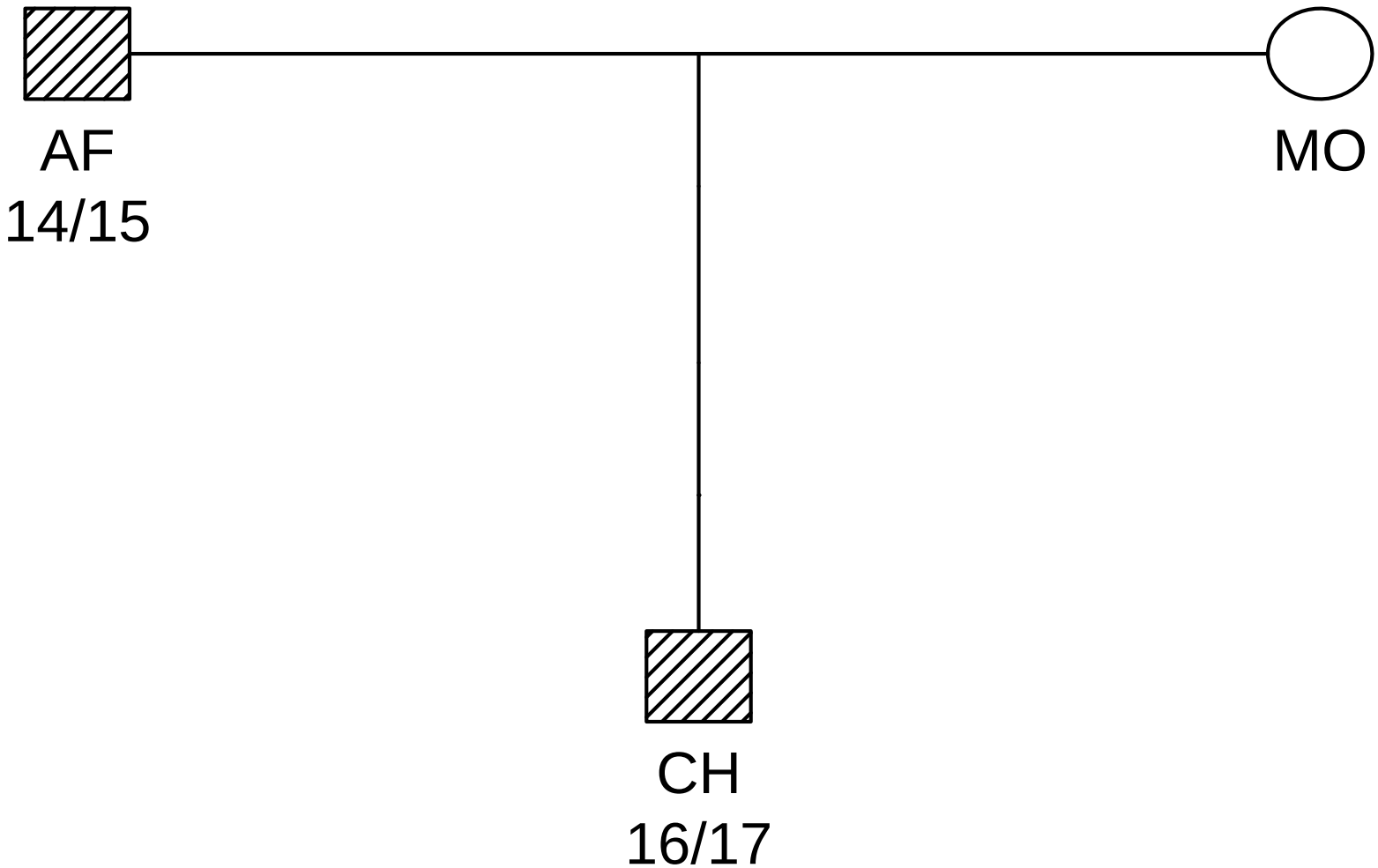
- Observed if parent and child share no alleles.
- Other examples? Mendelian inconsistencies.
- Mutation models interesting also in other applications.
- The forensic community is well positioned to study mutations.

Mutation: Biology

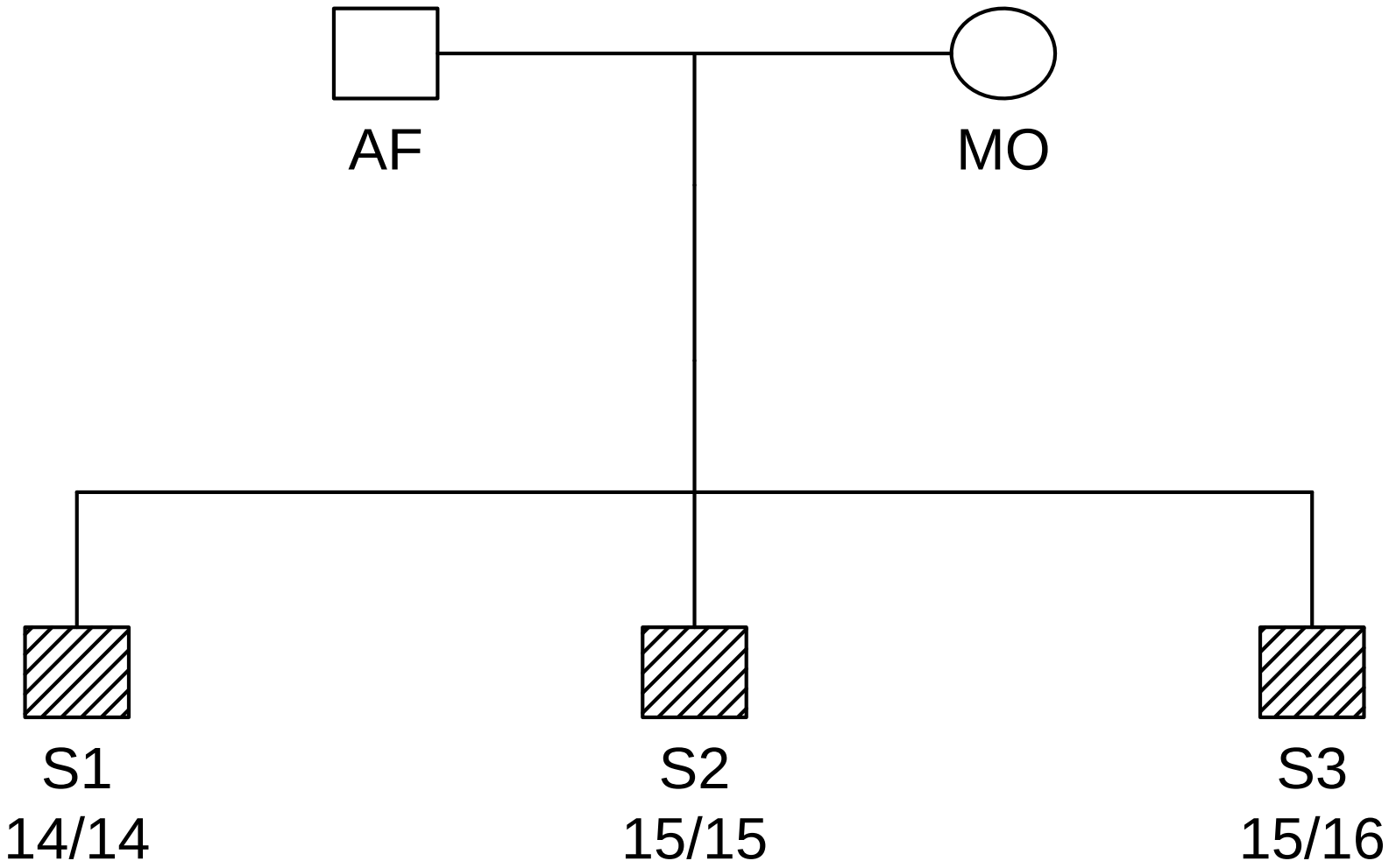


- Mutation rates higher in males.
- Short mutations more likely: One step mutation more likely than two steps and so on.
- Mutation rates:
  `http://www.cstl.nist.gov/strbase/mutation.htm`

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.
- equal. Simplest.
- proportional. Favoured by mathematicians, not used much.
- stepwise. Favoured by forensic case workers,
- onestep. Favoured by population geneticists.

Equal mutation model

```
library(pedmut)
mutationModel("eq", alleles = 14:17, rate = 0.003)
```

	14	15	16	17
14	0.997	0.001	0.001	0.001
15	0.001	0.997	0.001	0.001
16	0.001	0.001	0.997	0.001
17	0.001	0.001	0.001	0.997

$$P(\text{14 transmitted as 14, 15, 16, or 17})$$
$$= 0.997 + 0.001 + 0.001 + 0.001 = 1$$

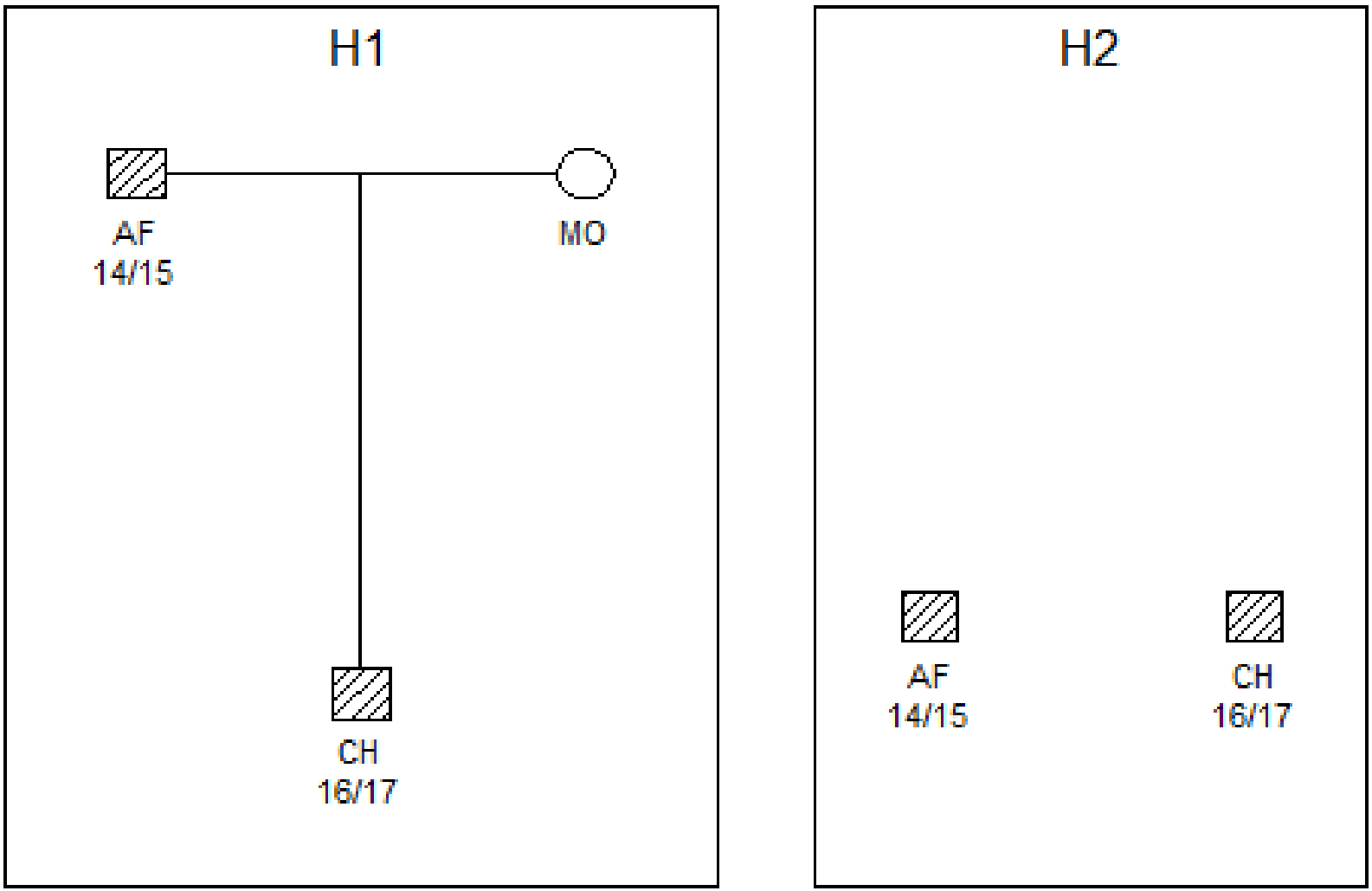
Stepwise mutation model

```
mutationModel("stepwise", alleles = 14:17, rate = 0.005,
              rate2 = 0, range = 0.1)
```

	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

```
?mutationMatrix  #Help page
```

Paternity case with mutation: plot



Paternity case with mutation: calculation

```
> plotPedList(list(H1, H2), source = 1, titles = c("H1", "H2"),
+             marker = 1,hatched = typedMembers)
> H1 = nuclearPed(father = "AF", child = "CH", mother ="MO")
> H2 = list(singleton("AF"), singleton("CH"))
> freq = c("14" = 0.25, "15" = 0.25, "16" = 0.25, "17" = 0.25)
> H1 = addMarker(H1, AF = "14/15", CH = "16/17", afreq = freq)
> H1 = setMutationModel(H1, model = "eq", rate = 0.003)
> kinshipLR(H1, H2, source = 1, ref = 2)
H1:H2 H2:H2
0.004 1.000
```

Summary

- General principles for evidence evaluation
- Likelihood ratio. Interpretation and assumptions
- Complications. Mutations