ISFG-GHEP Online School 2024
October 7,14,21,28

Kinship and pedigree analysis: Methods and applications

Teachers

Magnus Dehli Vigeland, PhD

Thore Egeland, PhD

Department of Forensic Sciences,

Oslo University Hospital, Norway

Schedule

The course runs each Monday of October 2024, from 16 to 20 (CEST). The following schedule is tentative:

Oct 7: Theory of relatedness

- 16:00–17:00 Lecture: **Introduction to pedigrees, QuickPed and R** (MDV)
- 17:00–17:45 Exercises
- 17:45–18:00 *Break*
- 18:00–19:00 Lecture: **Measures of relatedness** (MDV)
- 19:00–19:45 Exercises
- 19:45–20:00 Wrap-up

Oct 14: Kinship testing

- 16:00–17:00 Lecture: **Introduction to forensic kinship testing** (TE)
- 17:00–17:45 Exercises
- 17:45–18:00 *Break*
- 18:00–19:00 Lecture: **Kinship testing with Familias** (TE)
- 19:00–19:45 Exercises
- 19:45–20:00 Wrap-up

Oct 21: Relatedness inference

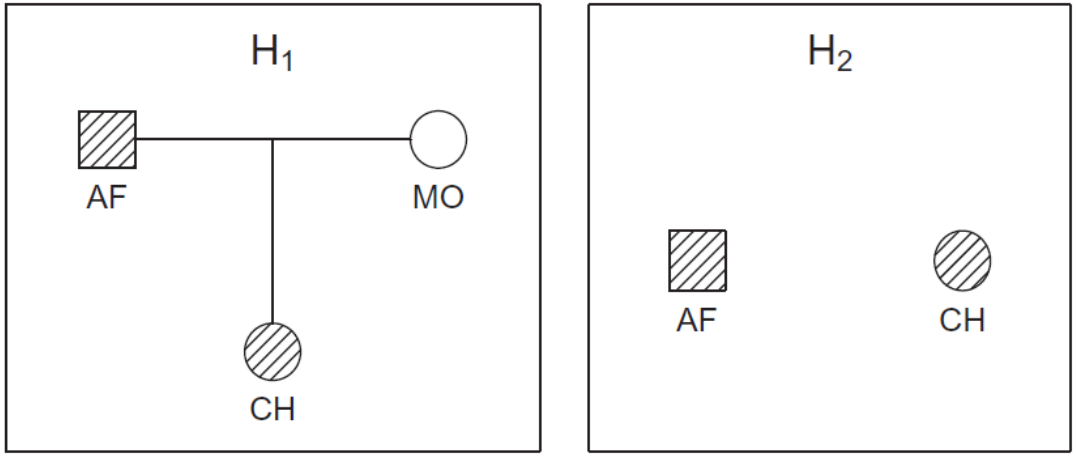
- 16:00–17:00 Lecture: **Realised relatedness: Why are some siblings more alike than others?** (MDV)
- 17:00–17:45 Exercises
- 17:45–18:00 *Break*
- 18:00–19:00 Lecture: **Pedigree reconstruction** (MDV)
- 19:00–19:45 Exercises
- 19:45–20:00 Wrap-up

Oct 28: Disaster victim identification

- 16:00–17:00 Lecture: **DNA-based disaster victim identification** (TE)
- 17:00–17:45 Exercises
- 17:45–18:00 *Break*
- 18:00–19:00 Lecture: **Practical DVI with Diviana** (MDV)
- 19:00–19:45 Exercises
- 19:45–20:00 Wrap-up

Home page
https://magnusdv.github.io/pedsuite/articles/web_only/course-ghep2024.html

Lecture 3:
Introduction to forensic kinship testing



Thore Egeland
Norwegian University of Life Sciences &
Department of Forensic Medicine, Norway

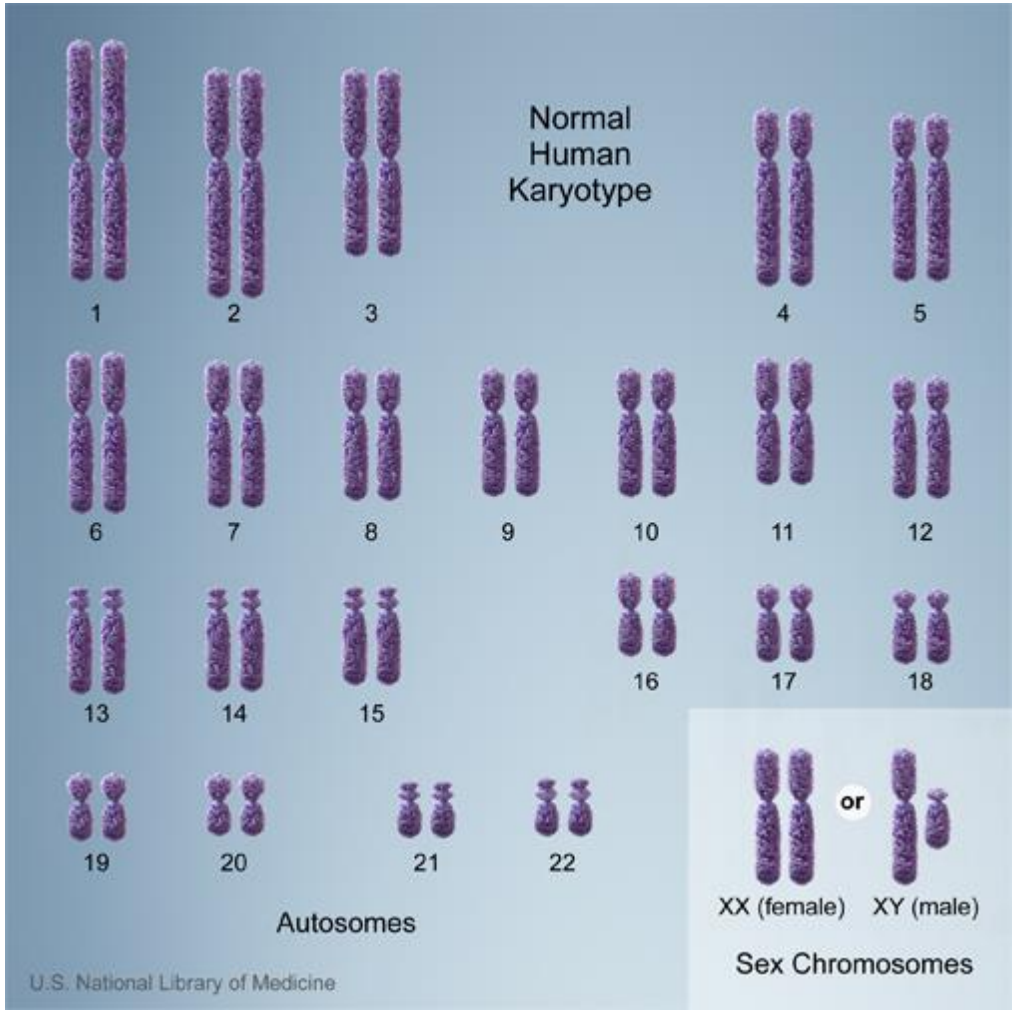
Motivating examples

- Kinship testing
  - Close (paternity) or distant (second cousins)
  - Disaster victim identification (DVI)
  - Pedigree reconstruction
  - …

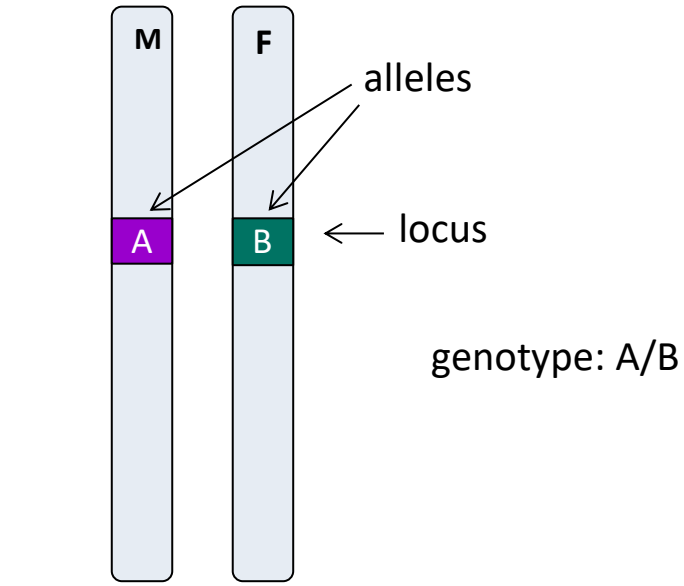
- We distinguish between
  - *kinship testing*, current topic, where a specific set of alternatives are compared, and
  - *relatedness inference* aiming to find the most probable relationship without restrictions

Genetics terminology

- Locus
- Allele
- Genotype
- Genetic markers
  - SNPs
  - microsatellites



Locus, allele, genotype



- **LOCUS** = a specific place in the genome
- **ALLELE** = any of the alternative forms of a locus
- **GENOTYPE** = the set (usually: pair) of alleles carried at a given locus

Genetic markers

- Small parts of the genome which ...
  - have known position
  - vary in the population
  - are easy to genotype

- SNPs (single nucleotide polymorphisms)
  - two alleles
  - usual requirement: MAF > 1% = minor allele frequency
  - very common in the genome (millions!)
  - used in medical genetics +++

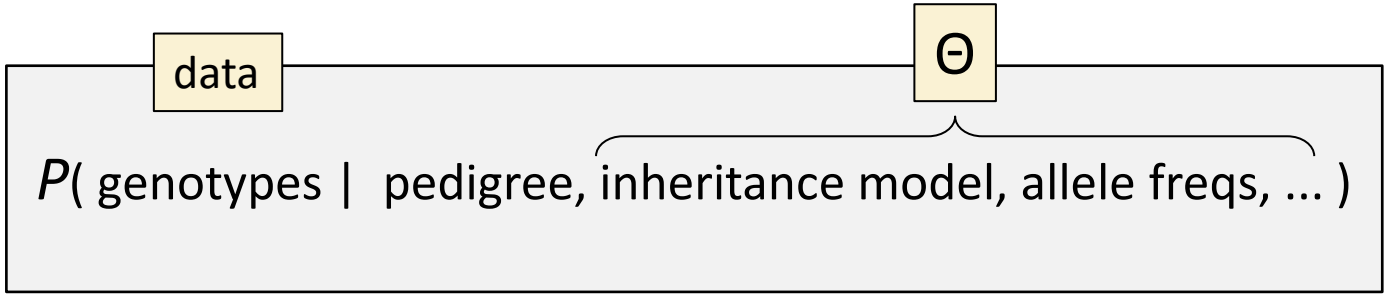
...CCGTTATATGGGC...
...CCGTTAGATGGGC...
...CCGTTATATGGGC...
...CCGTTATATGGGC...
...CCGTTAGATGGGC...

- STRs (short tandem repeats)
  - consecutive repeats of typically 2-5 bases
  - multiallelic: typically 5 - 50 alleles
  - allele names: # repeats
  - used in forensics

...ACG TTAG TTAG TTAG TTAG AAC..
...ACG TTAG TTAG AAC..
...ACG TTAG TTAG TTAG TTAG TTAG AAC..

Pedigree likelihoods

- Many applications involve probabilities of the following form



- Often referred to as a *pedigree likelihood*:

$$L(\text{ pedigree} \mid \text{data}) = P(\text{ data} \mid \text{pedigree}, \Theta)$$

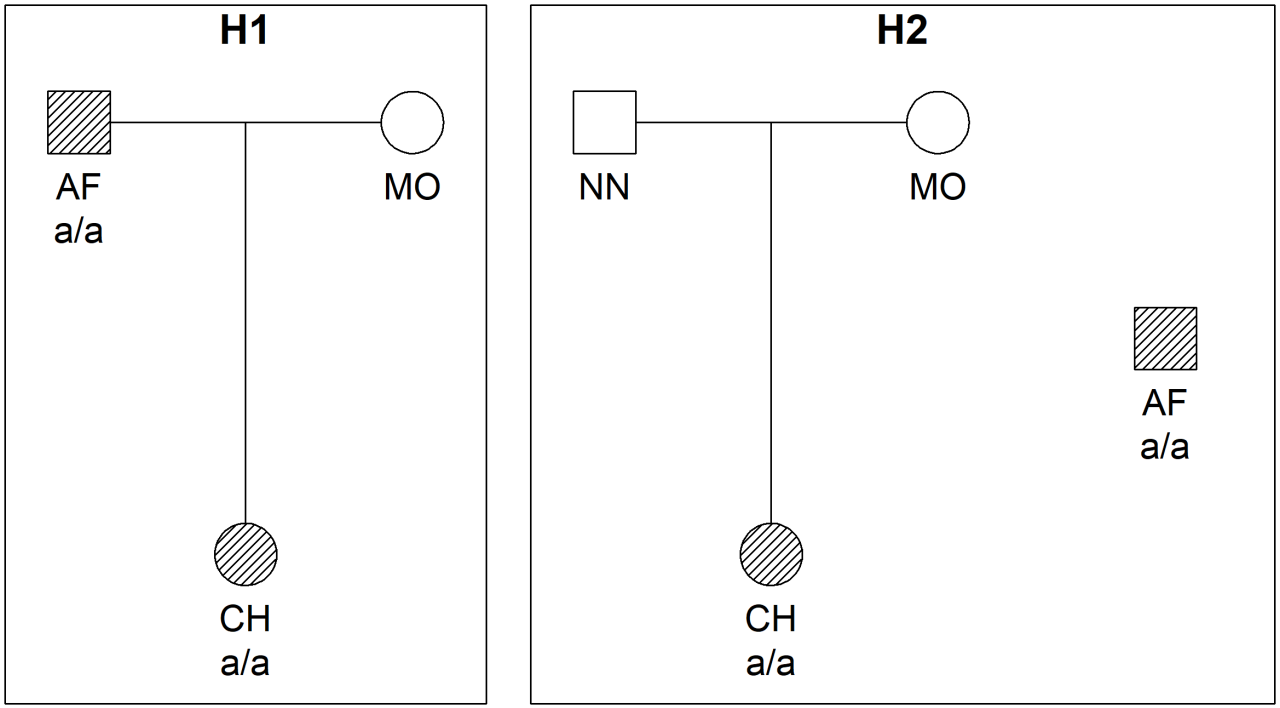
The Likelihood Ratio (LR)

- H_1: The individuals are related according to some pedigree $\mathcal{P}_1$.
- H_2: The individuals are related according to a different pedigree $\mathcal{P}_2$.

$$\mathrm{LR} = \frac{P(\text{data} \mid H_1, \Theta)}{P(\text{data} \mid H_2, \Theta)}.$$

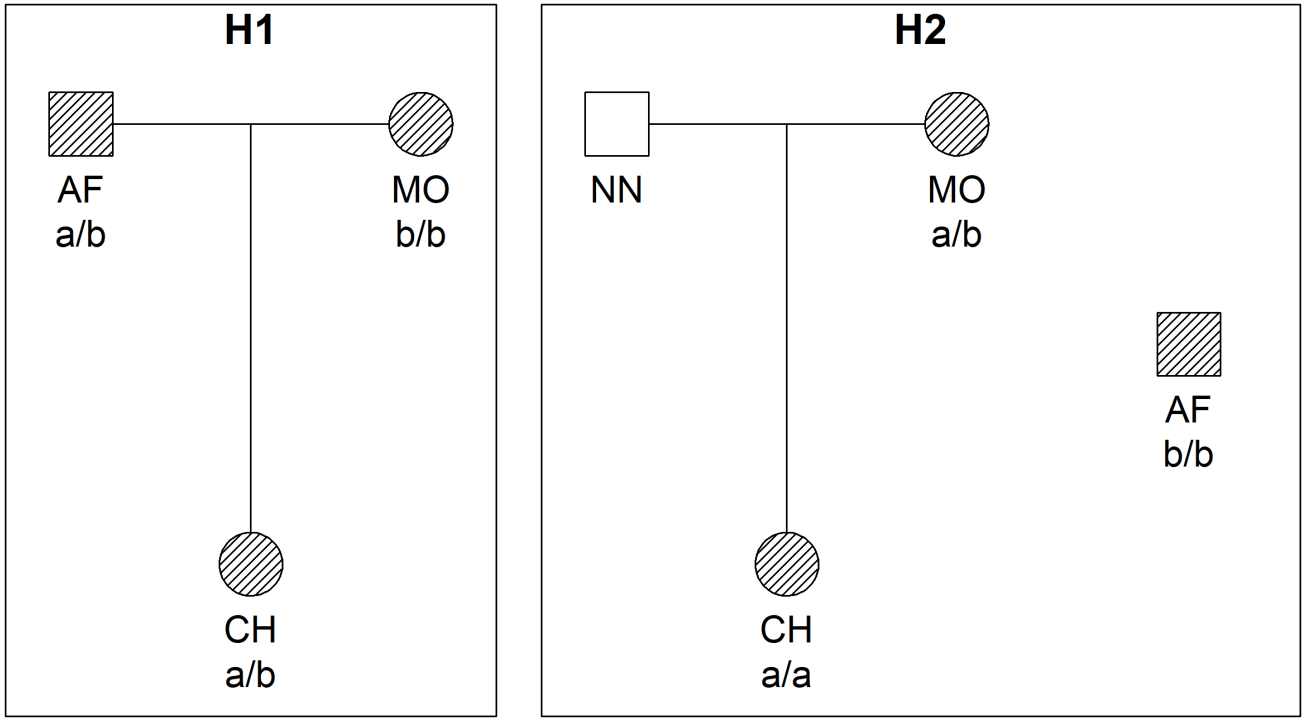
- data: available genotypes
- Θ : fixed model parameters common to both hypotheses
- Interpretation:
  - The LR measures how well H_1 explains the data compared to H_2
- Default assumptions:
  - ✓ Hardy Weinberg Equilibrium
  - ✓ No mutations
  - ✓ No artefacts (drop out, drop in, genotyping error)
  - ✓ Independence between markers

Example 1: Paternity case

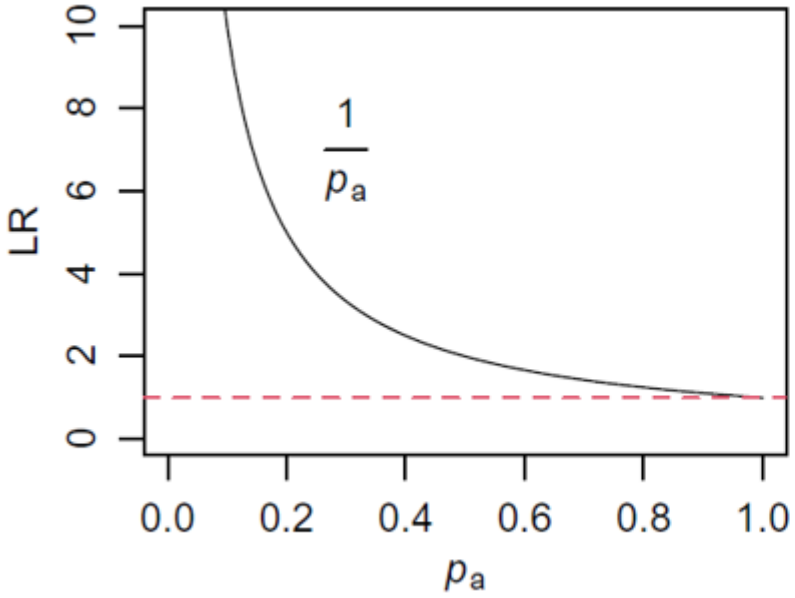


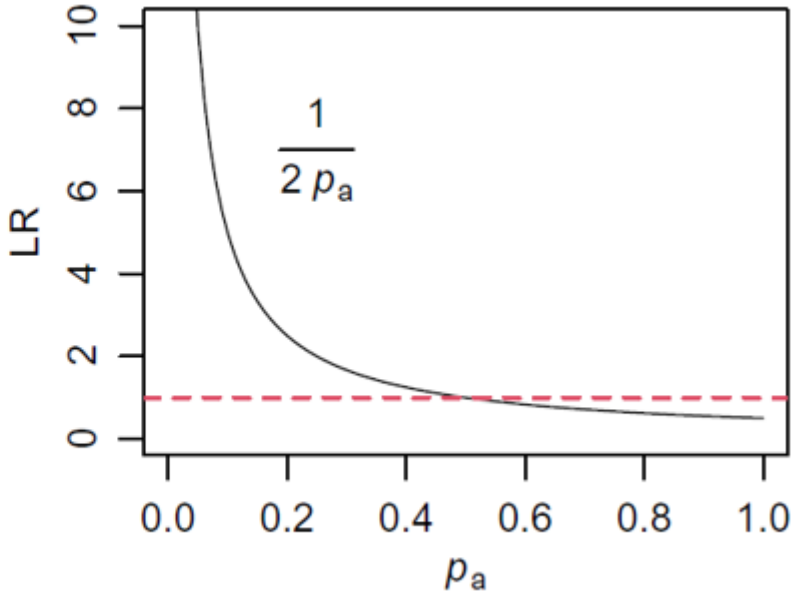
$$LR_1 = \frac{P(\mathrm{AF} = a/a, \mathrm{CH} = a/a \mid H_1)}{P(\mathrm{AF} = a/a, \mathrm{CH} = a/a \mid H_2)} = \frac{p_a^2 \cdot p_a}{p_a^2 \cdot p_a^2} = \frac{1}{p_a}.$$

Mother genotyped



$$LR_2 = \frac{P(\mathrm{AF} = a/b, \mathrm{MO} = b/b, \mathrm{CH} = a/b \mid H_1)}{P(\mathrm{AF} = a/b, \mathrm{MO} = b/b, \mathrm{CH} = a/b \mid H_2)} = \frac{2p_a p_b \cdot p_b^2 \cdot \frac{1}{2}}{2p_a p_b \cdot p_b^2 \cdot p_a} = \frac{1}{2p_a}.$$





- Observe
  - ✓ LR < 1 if $p_a > 0.5$ in right panel! Why?

Combined LR

- Assume $p_a = 0.05$ for both markers:
  - $LR_1 = \frac{1}{p_a} = 20$
  - $LR_2 = \frac{1}{2p_a} = 10$
- Assuming independence:
  - $LR = LR_1 \cdot LR_2 = 20 \cdot 10 = 200$

- Interpretation:
  The data is 200 times more likely if we assume H_1 to be true rather than H_2

Software for pedigree likelihoods

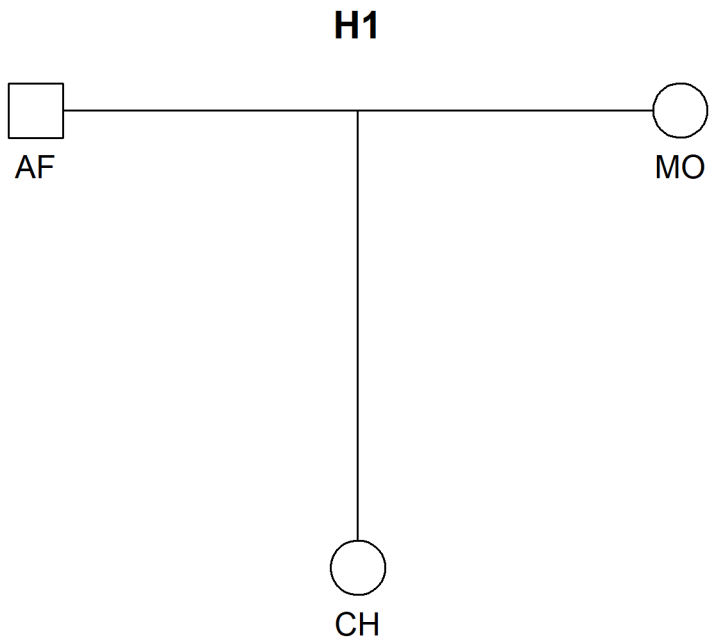
- MERLIN
  - command line program
  - Lander-Green
  - gold standard for cases with dense SNP markers (but not too large pedigrees)
  - used by FamLink & pedsuite to handle linked markers
  - not mutations, not theta correction
- Familias
  - GUI for forensic applications
  - Elston-Stewart
  - mutations, theta correction, ++
- R/pedsuite
  - Elston-Stewart
  - mutations, theta correction, ++
- Main software today: Familias. But
  - the pedsuite is introduced first and interaction with Familias is emphasised

Kinship testing in R with the pedsuite

- Create pedigrees representing the hypotheses.
- Attach the given genotype data to one of the pedigrees.
- Invoke the function kinshipLR() to calculate LRs.

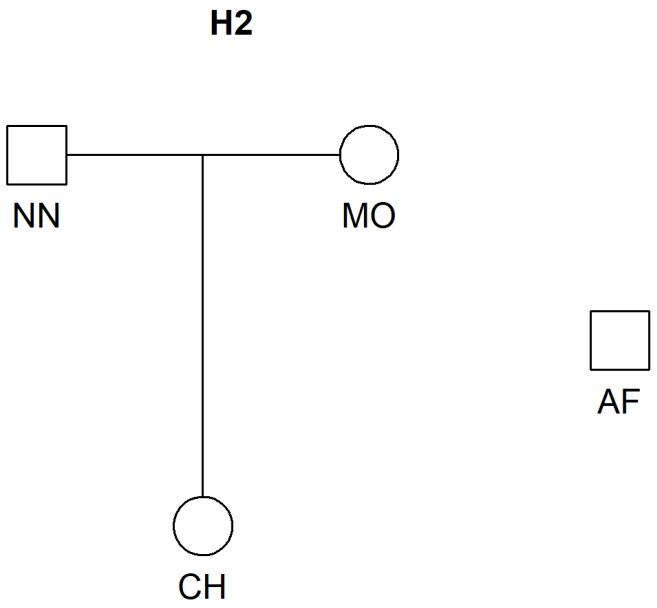
Create pedigrees. H1

```
> library(pedsuite)
> H1 = nuclearPed(fa = "AF", mo = "MO", child = "CH", sex = 2)
> plot(H1, title = "H1")
```



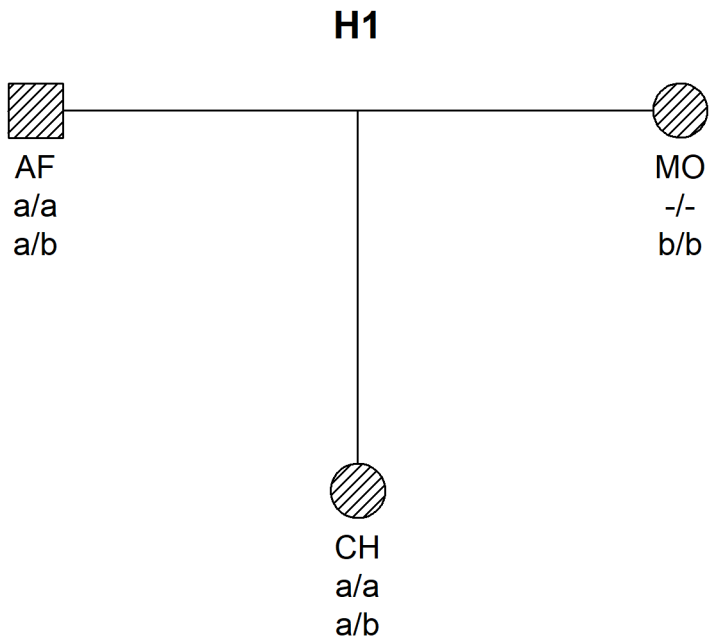
Create pedigrees. H2

```
> H2 = list(nuclearPed(fa = "NN", mo = "MO", child = "CH", sex = 2),
>           singleton("AF"))
> plotPedList(H2)
```



Attach genotype data to one of the pedigrees

```
> afr = c(a = 0.05, b = 0.95)
> H1 = addMarker(H1, AF = "a/a", CH = "a/a", afreq = afr)
> H1 = addMarker(H1, AF = "a/b", MO = "b/b", CH = "a/b",
>                afreq = afr)
> plot(H1, marker = 1:2, hatched = typedMembers)
```



LR calculations

```
> ?kinshipLR
```

kinshipLR {forrel} R Documentation

Likelihood ratios for kinship testing

Description

This function computes likelihood ratios (LRs) for a list of pedigrees. One of the pedigrees (the last one, by default) is designated as 'reference', to be used in the denominator in all LR calculations. To ensure that all pedigrees use the same data set, one of the pedigrees may be chosen as 'source', from which data is transferred to all the other pedigrees.

Usage

```
kinshipLR(
  ...,
  ref = NULL,
  source = NULL,
  markers = NULL,
  linkageMap = NULL,
  keepMerlin = NULL,
  verbose = FALSE
)
```

Not discussed

Invoke the function kinshipLR() to calculate LRs

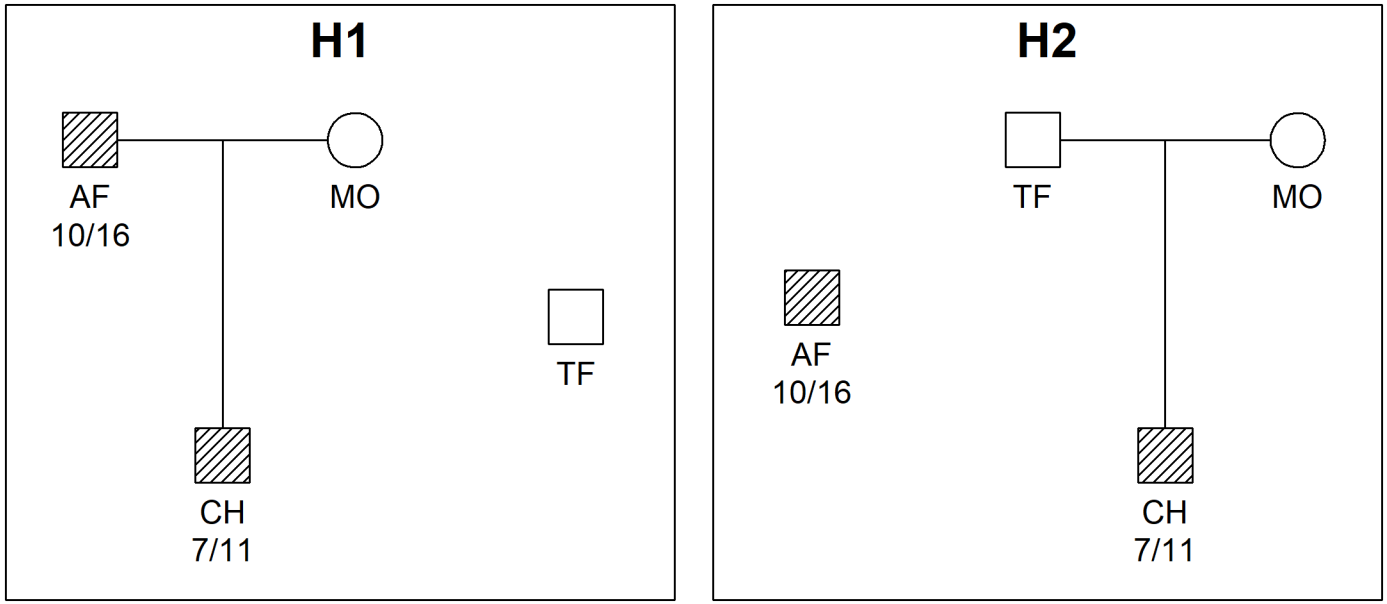
```
> lr = kinshipLR(H1, H2, source = 1)
```

```
H1:H2 H2:H2
  200     1
```

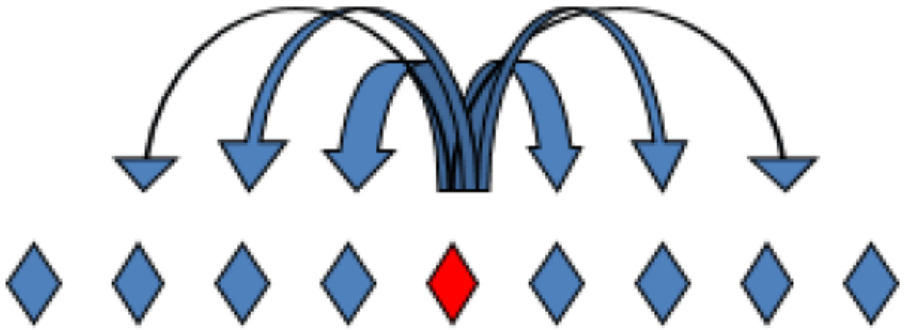
```
> lr$LRperMarker
```

```
    H1:H2 H2:H2
<1>    20     1
<2>    10     1
```

Mutation?

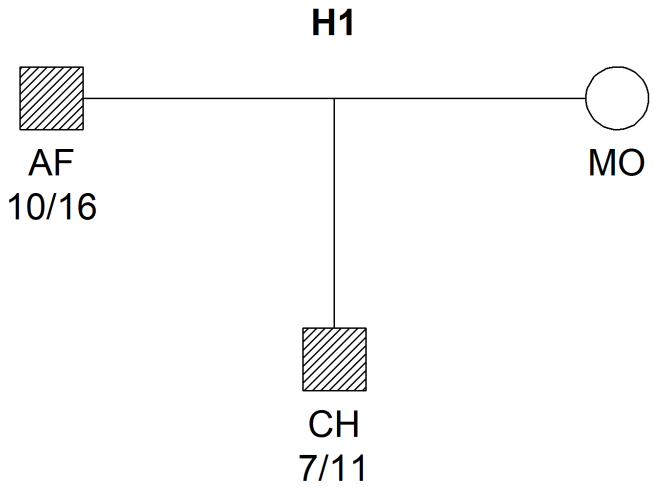


Mutations. Models



- 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`

Dealing with mutations

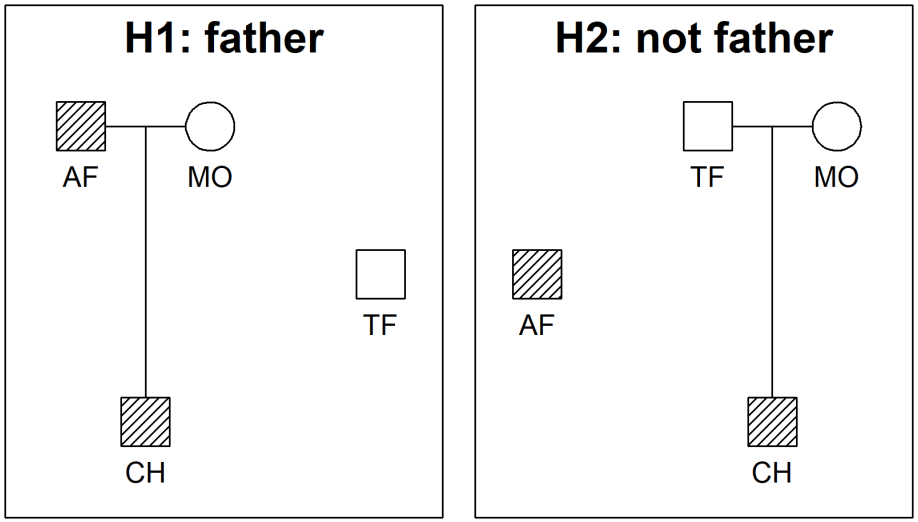


Strategies for handling mutations

- Exclude inconsistent markers from the analysis. Not recommended
- Apply mutation modelling only to inconsistent markers
- Apply mutation modelling to *all* markers. Recommended

Read data and compute LR

```
> ?readFam
```



Read data from Familias file, plot and find LR:

```
> filename = "http://familias.name/norbisRelatedness/paternityCase.fam"
> dat = readFam(filename)
> plotPedList(dat, hatched = typedMembers)
> lr1 = kinshipLR(dat)
> lr1
```

```
H1:H2 H2:H2
    0     1
```

Inspect each marker

```
> lr1$LRperMarker
```


	H1:H2
D3S1358	2.466752
TH01	1.194605
D21S11	1.095934
D18S51	2.153261
PENTA_E	0.000000
D5S818	1.406127
D13S317	4.041611
D7S820	1.433570
D16S539	8.312297
CSF1PO	2.024678
PENTA_D	11.989252
VWA	5.565000
D8S1179	9.650567
TPOX	1.787652
FGA	2.956394
D12S391	2.183522
D1S1656	3.333333
D2S1338	3.147060
D22S1045	26.748152
D2S441	1.445948
D19S433	3.343766

Mutation models

```
> ?setMutmod
```

setMutmod {pedtools} R Documentation

Set a mutation model

Description

This function offers a convenient way to set or modify mutation models to markers attached to a pedigree. It wraps `pedmut::mutationModel()`, which does the main work of creating the models, but relieves the user from having to loop through the markers in order to supply the correct alleles and frequencies for each marker.

Details

Currently, the following models are supported:

- `equal`: All mutations equally likely; probability `1 - rate` of no mutation
- `proportional`: Mutation probabilities are proportional to the target allele frequencies
- `onestep`: A simple model for microsatellite markers, in which mutations are only allowed to the nearest neighbours in the allelic ladder. For example, '10' may mutate to either '9' or '11' (unless '10' is the lowest allele, in which case '11' is the only option). Not applicable to loci with non-integral microvariants.
- `stepwise`: A common model for microsatellite markers. Mutation rates depend on the step size in the allelic ladder, and also the allelic classes: integral repeats like '16', versus non-integer microvariants like '16.3'.
- `custom`: Allows any mutation matrix to be provided by the user, in the `matrix` parameter

Recompute with mutation model

```
> H1 = dat$H1
> H2 = dat$H2
> H2 = setMutmod(H2, model = "proportional", rate = 0.001)
> lr2 = kinshipLR(H1, H2, ref = 2, source = 2)
```

```
   H1:H2    H2:H2
107132.1      1.0
```

A closer look at the impact of mutation

	lrNoMut	lrMut	ratio
D3S1358	2.46675	2.46673	0.99999
TH01	1.19461	1.19460	1.00000
D21S11	1.09593	1.09593	1.00000
D18S51	2.15326	2.15325	0.99999
PENTA_E	0.00000	0.00001	Inf
D5S818	1.40613	1.40612	1.00000
D13S317	4.04161	4.04157	0.99999
D7S820	1.43357	1.43356	1.00000
D16S539	8.31230	8.31220	0.99999
CSF1PO	2.02468	2.02466	0.99999
PENTA_D	11.98925	11.98912	0.99999
VWA	5.56500	5.56494	0.99999
D8S1179	9.65057	9.65046	0.99999
TPOX	1.78765	1.78764	0.99999
FGA	2.95639	2.95637	0.99999
D12S391	2.18352	2.18351	0.99999
D1S1656	3.33333	3.33331	0.99999
D2S1338	3.14706	3.14703	0.99999
D22S1045	26.74815	26.74780	0.99999
D2S441	1.44595	1.44594	1.00000
D19S433	3.34377	3.34374	0.99999

A Relationship Riddle. *Exercise*

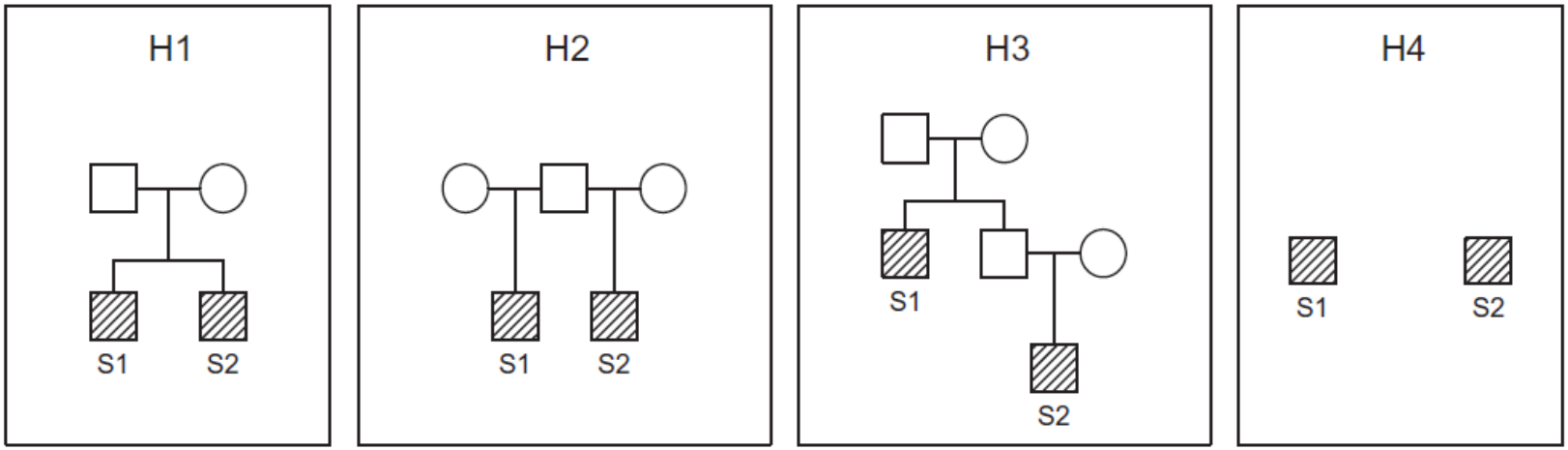


Fig. 6.4 A relationship riddle: Four hypothesised relationships between S1 and S2.

- H_1: Full brothers
- H_2: Half-brothers
- H_3: Uncle and nephew
- H_4: Unrelated