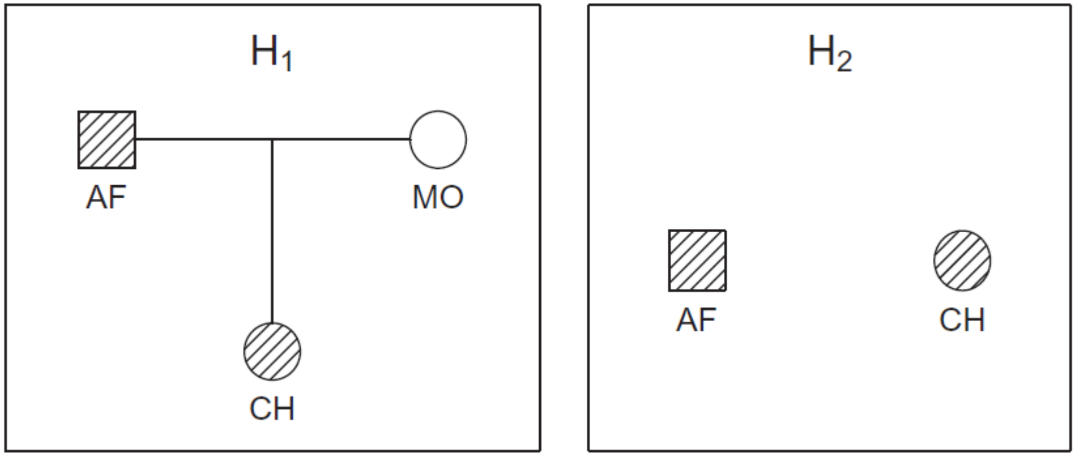
Workshop Alcalá de Henares 2026

Forensic kinship testing



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

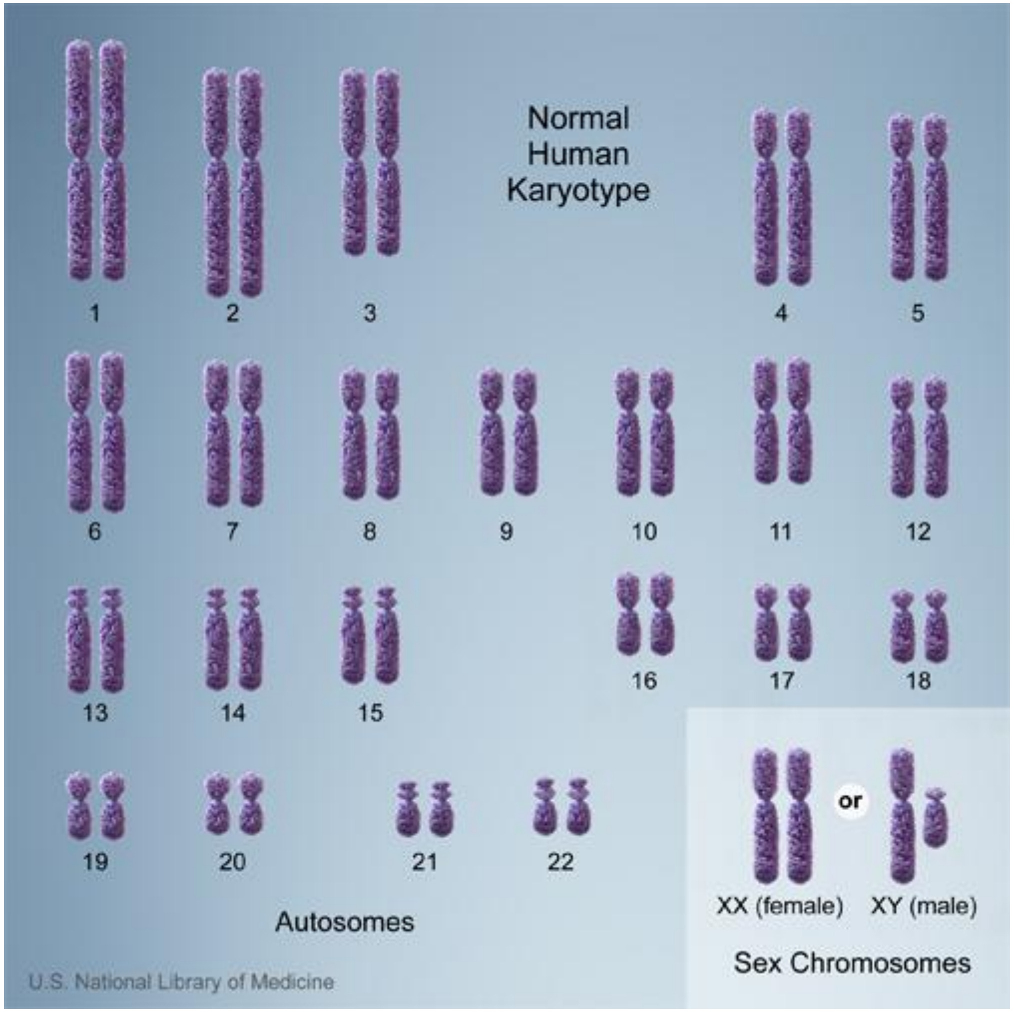
Motivating examples

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

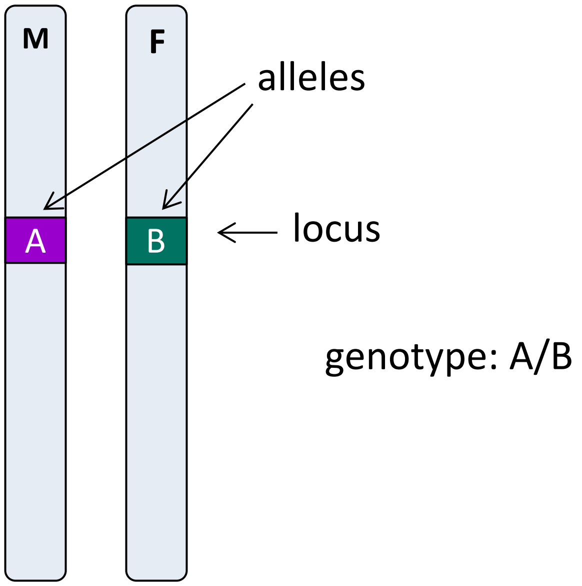
- We distinguish between
  - *kinship testing*, current topic, where a specific set of alternatives are compared, and
  - *relatedness inference* aiming to find the most probable relationship without restrictions

Genetics terminology

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



Locus, allele, genotype



Homologous chromosomes

- **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

$$P(\overset{\text{data}}{\text{genotypes}} \mid \text{pedigree}, \overbrace{\text{inheritance model, allele freqs, ...}}^{\Theta})$$

- Often referred to as a *pedigree likelihood*:

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

Software for pedigree likelihoods

- Familias
  - GUI for forensic applications
  - Elston-Stewart. Mutations, theta correction, ++
  - More shortly

- KLINK
  - Next lecture
  - Input: Familias file (.fam)

- MERLIN
  - command line program, not developed for forensics
  - Lander-Green. Not mutations, not theta correction
  - gold standard for cases with dense SNP markers (but not too large pedigrees)

- R/pedsuite
  - Elston-Stewart
  - mutations, theta correction, ++

Three principles of interpretation

1. It is necessary to consider at least one alternative proposition
2. Interpretation is based on questions like:
   «What is the probability of the evidence given the proposition»
3. We condition on proposition and known circumstances

- These principles lead to the likelihood ratio (LR):

The Likelihood Ratio

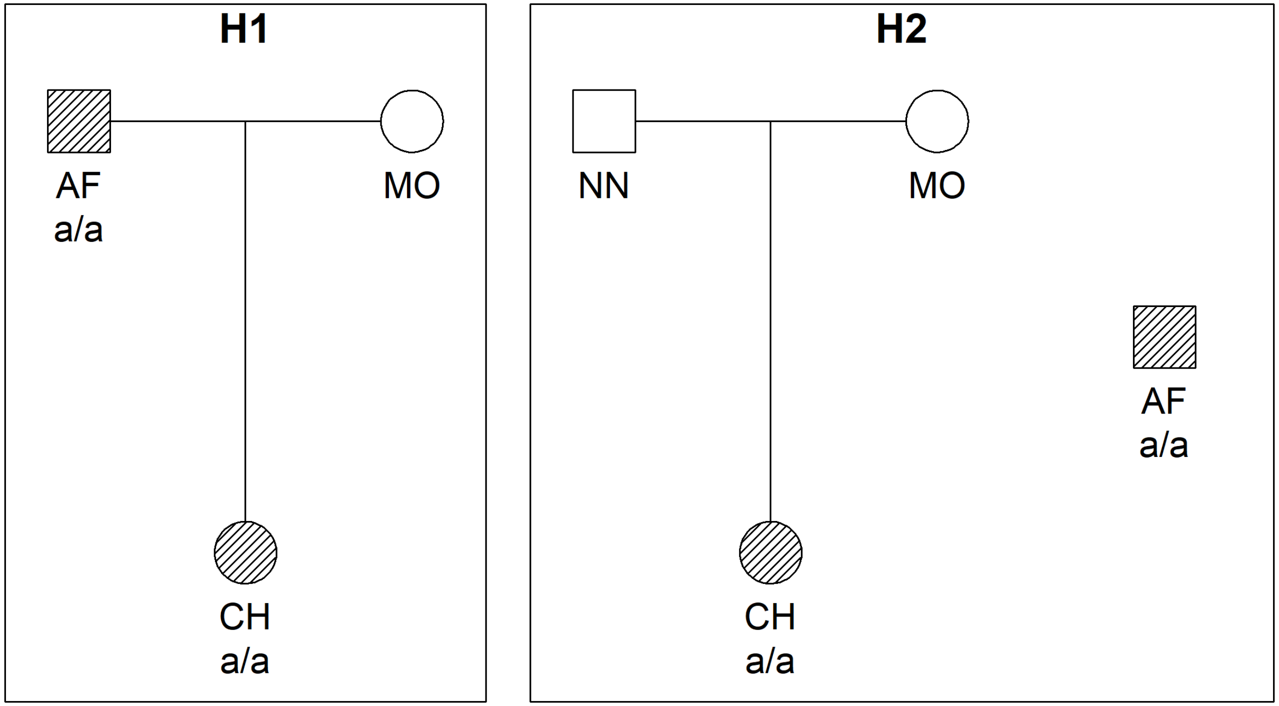
- H_1: The individuals are related according to some pedigree $\mathcal{P}_1$.
- H_2: The individuals are related according to a different pedigree $\mathcal{P}_2$.

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

- 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

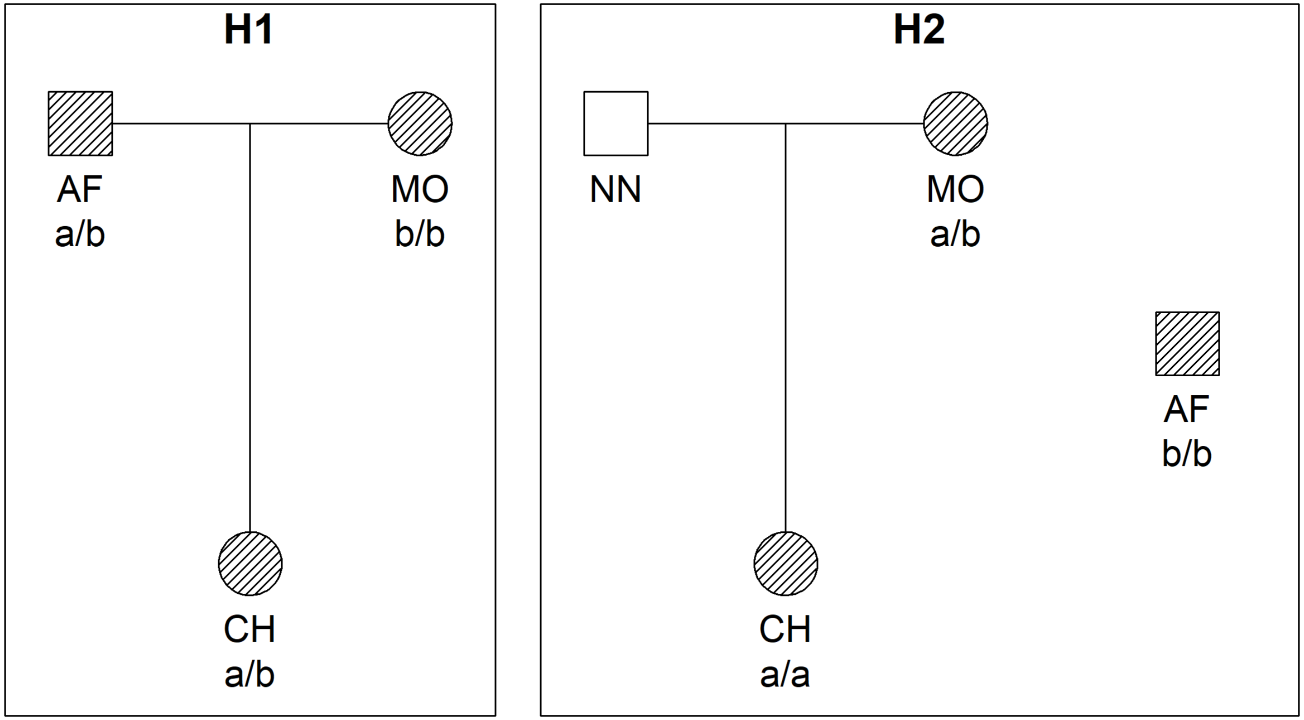
- Default assumptions:
  - ✓ Hardy Weinberg Equilibrium
  - ✓ No mutations
  - ✓ No artefacts (drop out, drop in, genotyping error)
  - ✓ Independence between markers

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



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

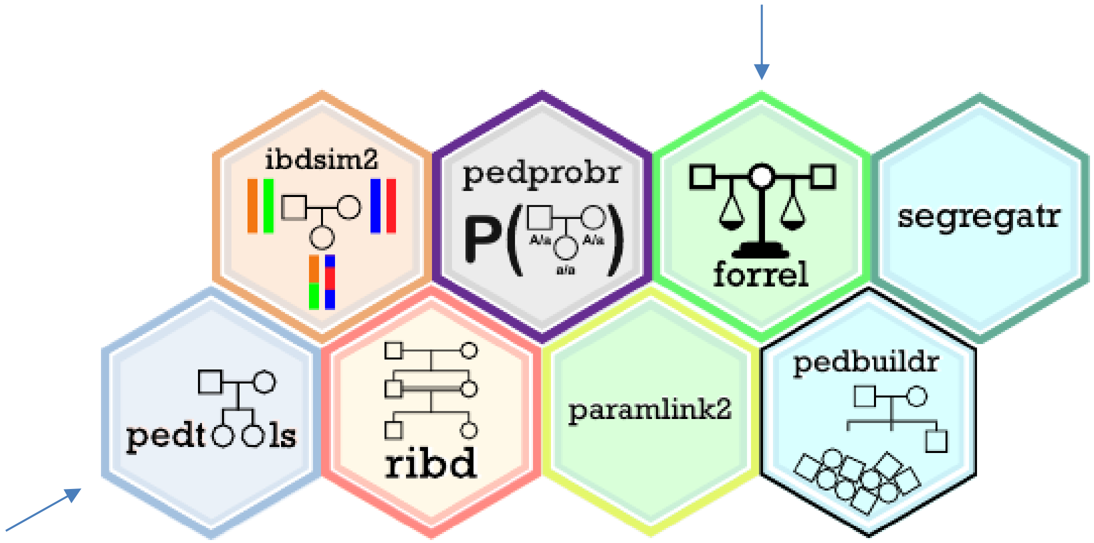
Combined LR

- 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

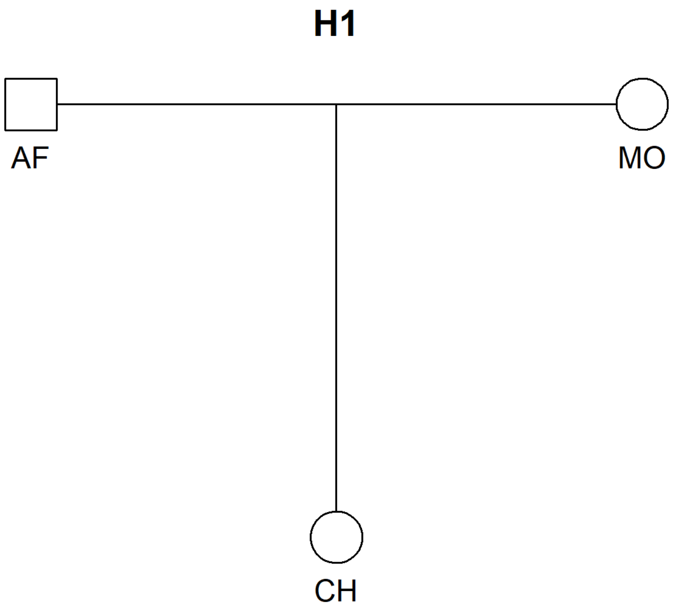
Kinship testing in R with the pedsuite

- 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

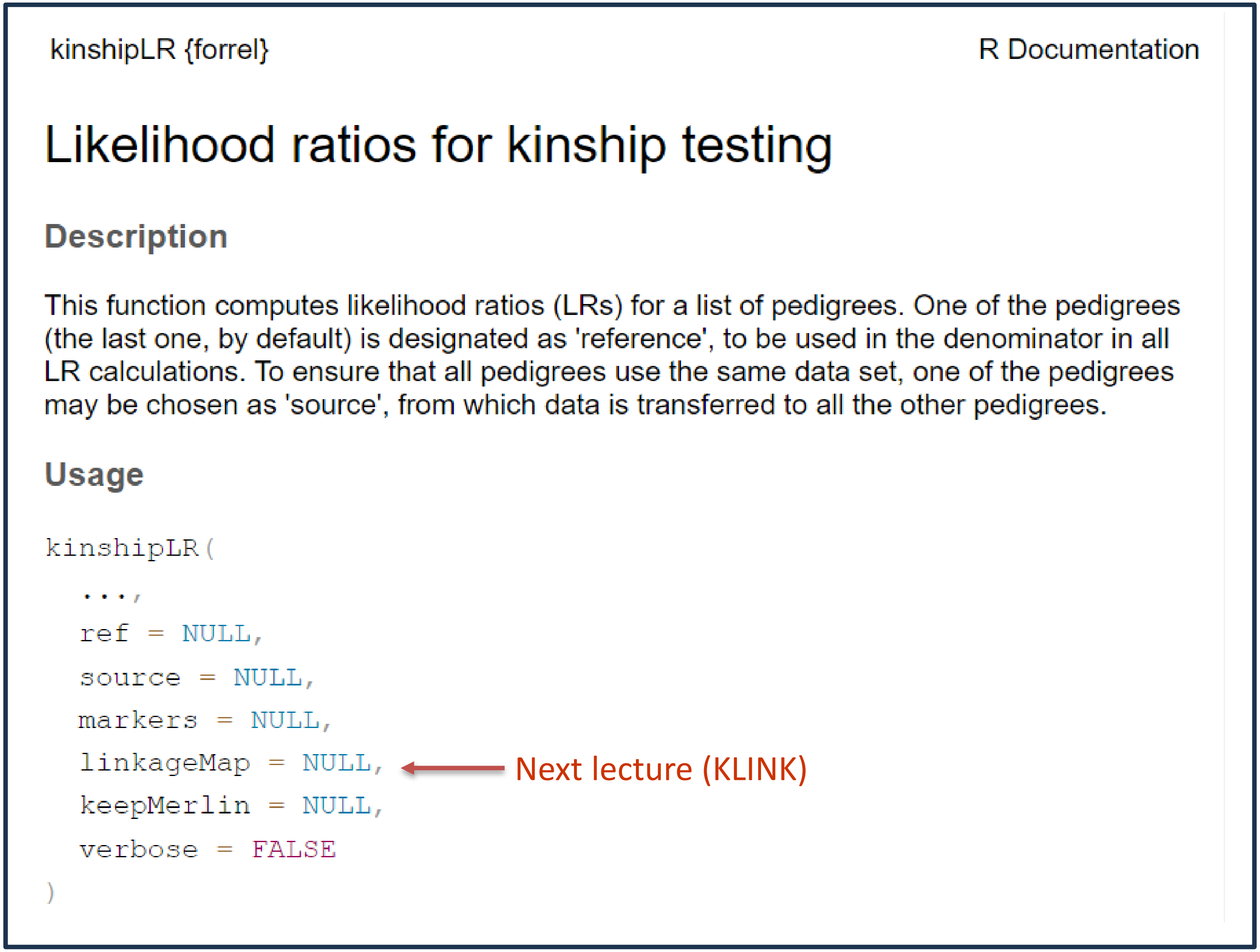
```
> library(pedsuite)
> H1 = nuclearPed(fa = "AF", mo = "MO", child = "CH", sex = 2)
> plot(H1, title = "H1")
```



- Some further R code will be introduced in the exercises

LR calculations

```
> ?kinshipLR
```



kinshipLR {forrel} R Documentation

Likelihood ratios for kinship testing

Description

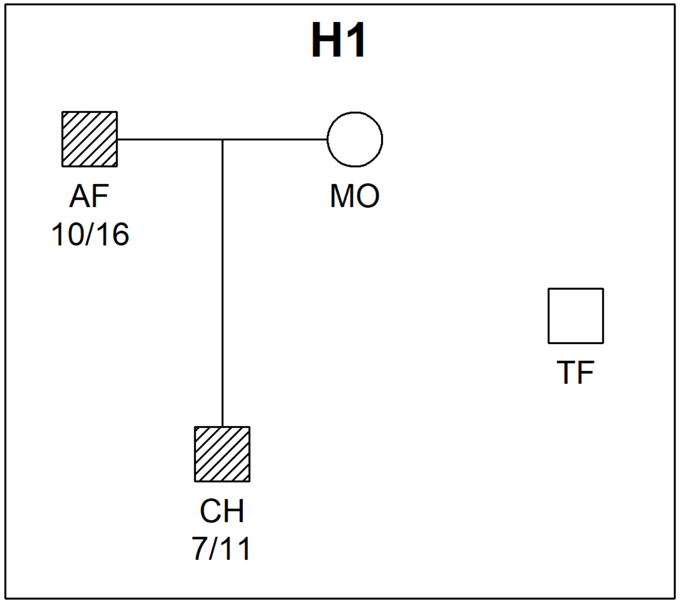
This function computes likelihood ratios (LRs) for a list of pedigrees. One of the pedigrees (the last one, by default) is designated as 'reference', to be used in the denominator in all LR calculations. To ensure that all pedigrees use the same data set, one of the pedigrees may be chosen as 'source',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

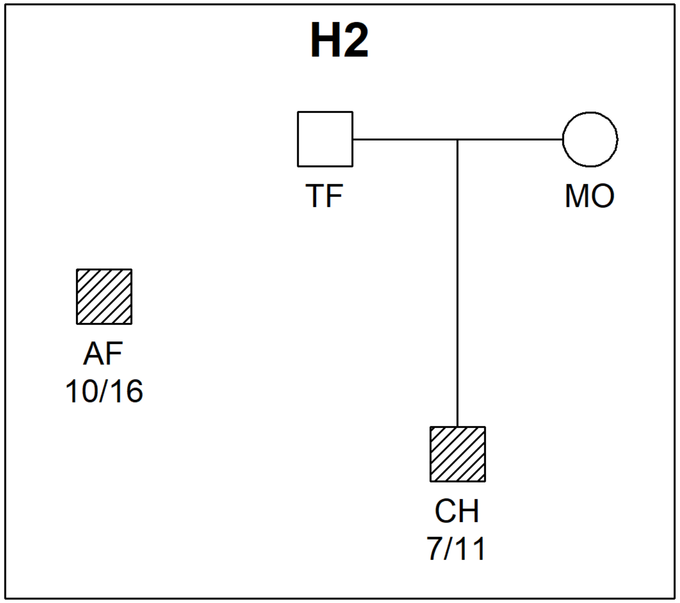
Usage

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

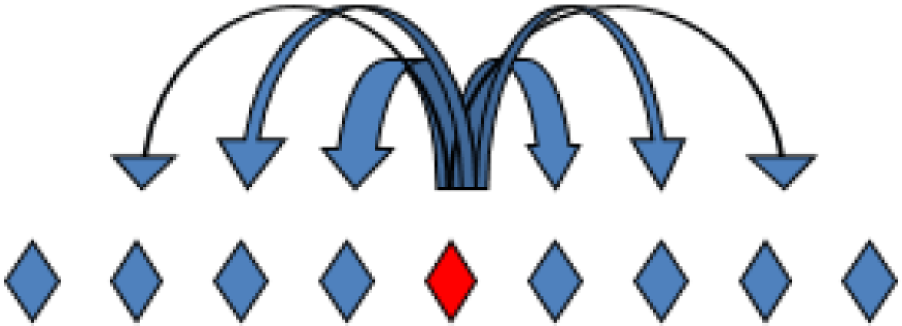
Next lecture (KLINK)

Mutation?



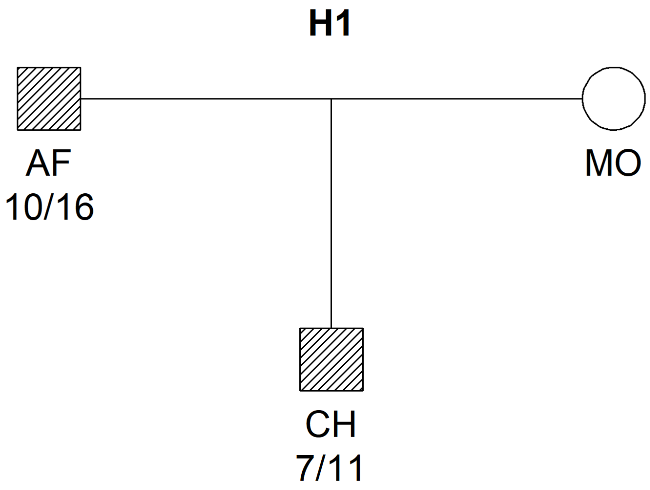


Mutations. Models



- 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

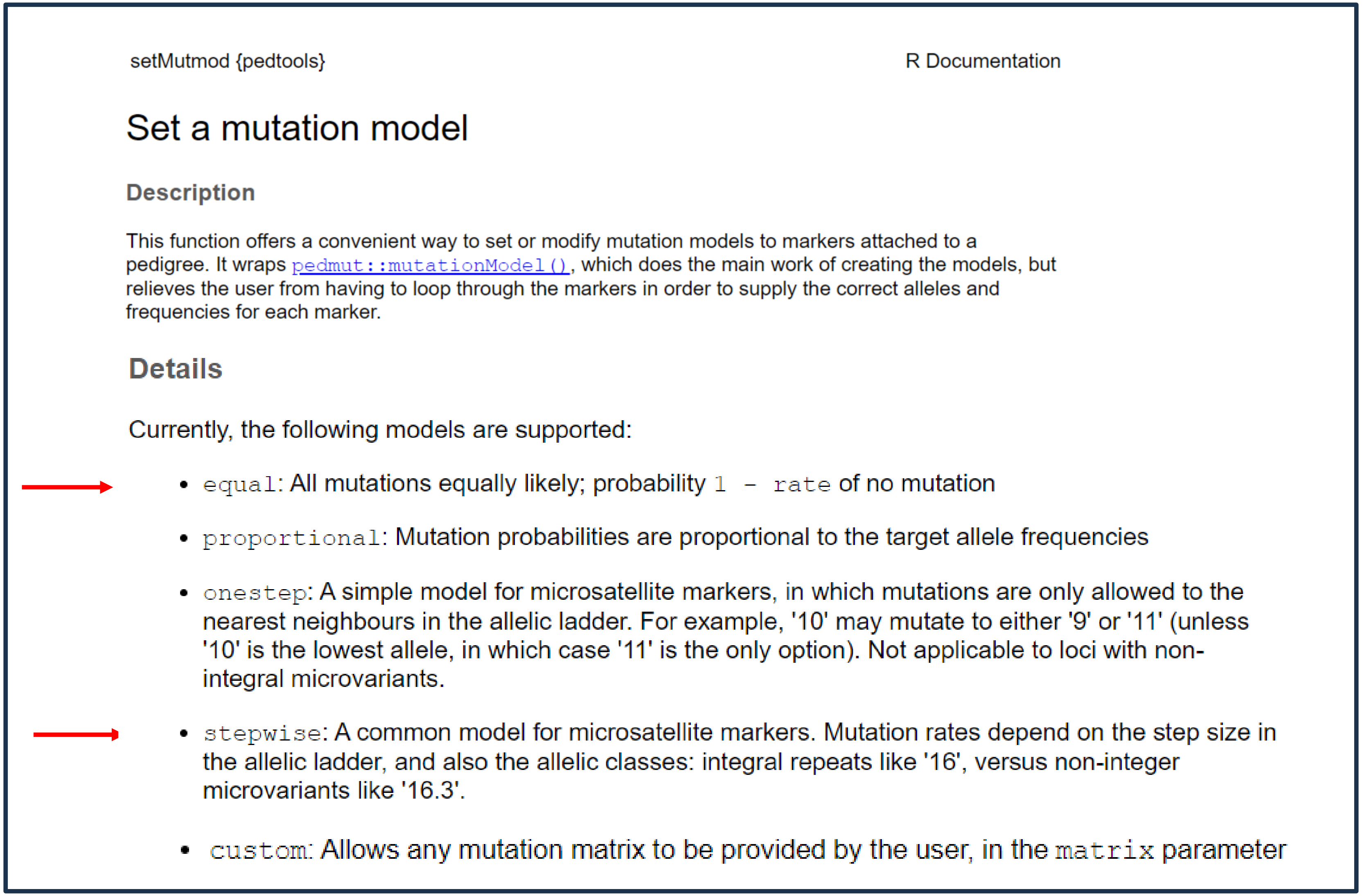


Strategies for handling mutations

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

Mutation models

Exercise



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

Next: Introduction to Familias

https://familias.name/tutorial/familias_tutorial_english.pdf

https://familias.name/tutorial/familias_tutorial_spanish.pdf

Familias tutorial