

Lecture 10

Case study:

The missing grandchildren of Argentina

Magnus Dehli Vigeland

Statistical methods in genetic relatedness and pedigree analysis

NORBIS course, 6th – 10th of January 2020, Oslo

Project

- **Evaluating the statistical power of DNA-based identification, exemplified by «The missing grandchildren of Argentina»**

Daniel Kling, Thore Egeland, Mariana H. Piñero, Magnus D. Vigeland



Argentina 1976 - 1983

- Military dictatorship
- *Dirty war* against left-wing guerrillas
- Opponents killed or disappeared
 - counts: 20,000 - 30,000



- 500 children abducted
 - kidnapped with their parents, or born in captivity
 - parents killed
 - raised by police or military families.





1982: Falklands war against UK

1983: Democracy restored

1985: First trial against dictatorship leaders: life imprisonment sentence

1986: Final point law

1987: Due obedience law

← amnesty laws

Exceptions: Theft, rape and child abduction

The missing grandchildren

- *Grandmothers of Plaza de Mayo*
 - formed in 1977
 - weekly marches ever since
- 1984: First grandchild recovered
 - HLA typing + blood groups
- 1989: National genetic data bank
- 2020: 130 reunifications so far
 - ~80 of these by BNDG





Press conference April 2017: Grandchild no. 122 recovered



Genetics of family reunification

- DNA-based evidence

- autosomal markers
- mtDNA
- Y chromosome

Forensic markers:

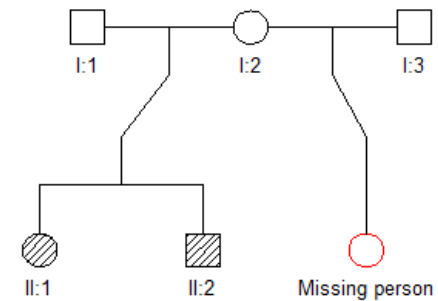
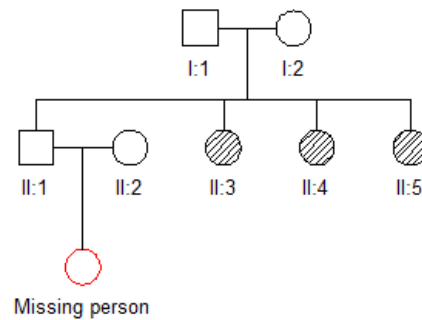
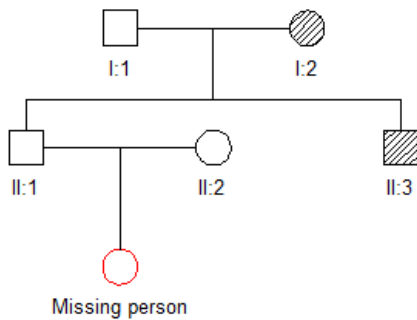
- standard kits, 15 - 24 STRs
- up to 50 alleles
- unlinked (mostly)

- Simplest when

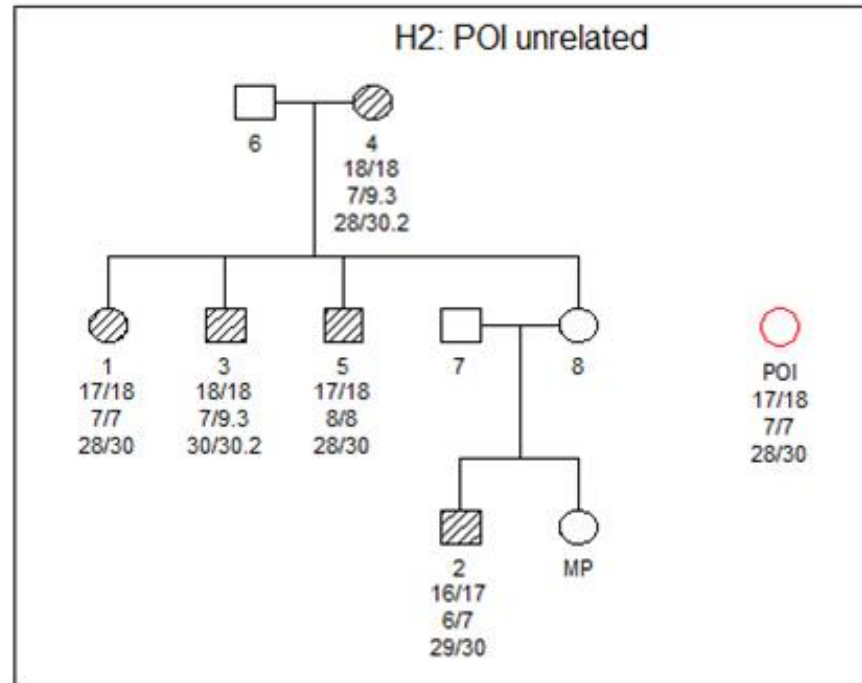
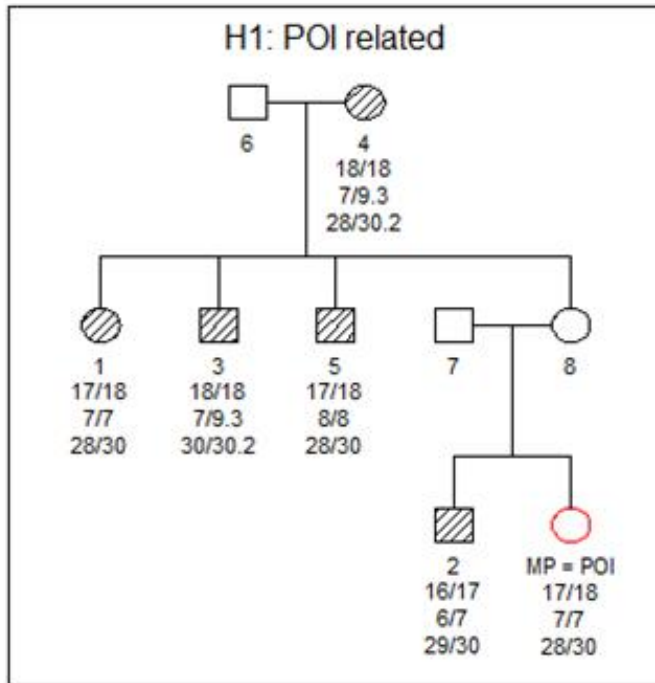
- available DNA from the missing person
- available DNA from parents

Similar to standard paternity cases

- Argentina: Parents usually dead/missing



Genetics of family reunification



POI = person of interest
MP = missing person

$$LR = \frac{P(\text{marker data} \mid MP = POI)}{P(\text{marker data} \mid POI \text{ unrelated})}$$

Positive match if $LR > 10\,000$

Genetics of family reunification

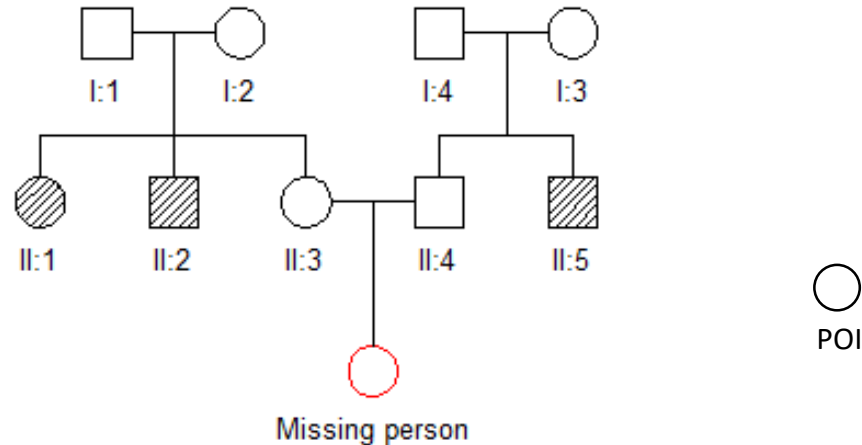


- Software: Familias
 - Original publication: Egeland, Mostad et al, 2000
 - Now maintained by Daniel Kling
- Used in all identifications by BNDG

Currently in BNDG

- 350 families
- 10 000 POIs

Statistical power of reunification - part 1



- Do we enough data to give a positive match if $POI = MP$?
- **Inclusion power** (or *exceedance probability*):

$$IP_{10000} = P(LR > 10000 \mid POI = MP)$$

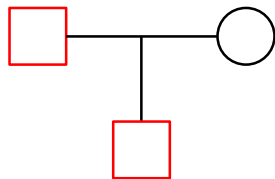
- Can be computed by simulation!
 - Unconditional → average for all pedigrees of this type
 - Conditional → probability for this particular case

In forrel:
> missingPersonIP()

Unconditional simulation

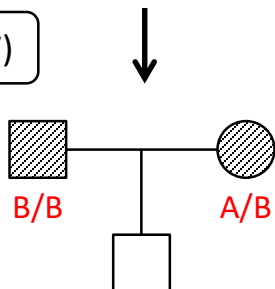
SNP: $p_A = p_B = 0.5$

A/A	A/B	B/B
0.25	0.5	0.25

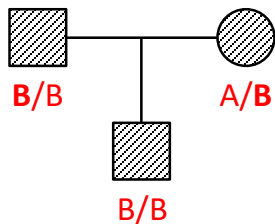


A/A	A/B	B/B
0.25	0.5	0.25

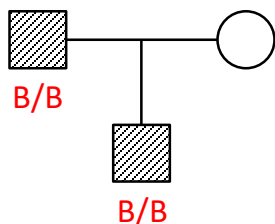
1. Founders (HW)



2. Gene dropping

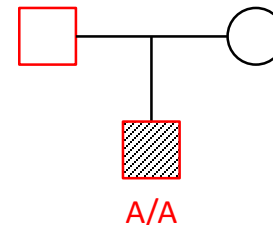


3. Remove untyped



Conditional simulation

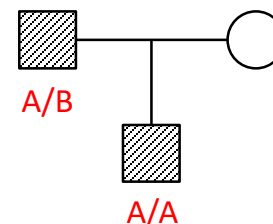
SNP: $p_A = p_B = 0.5$



1. Compute conditional distribution in the father

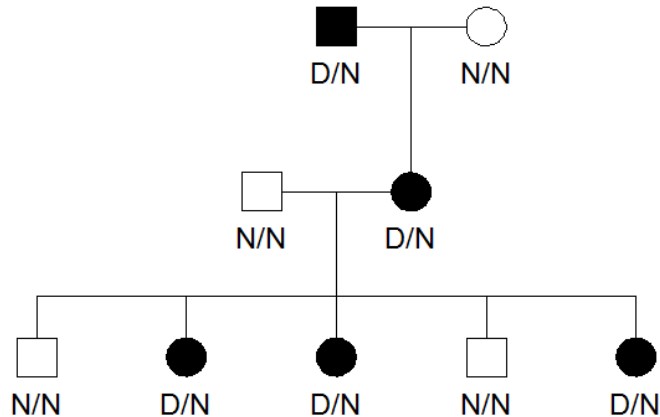
A/A	A/B	B/B
0.5	0.5	0

2. Sample from this



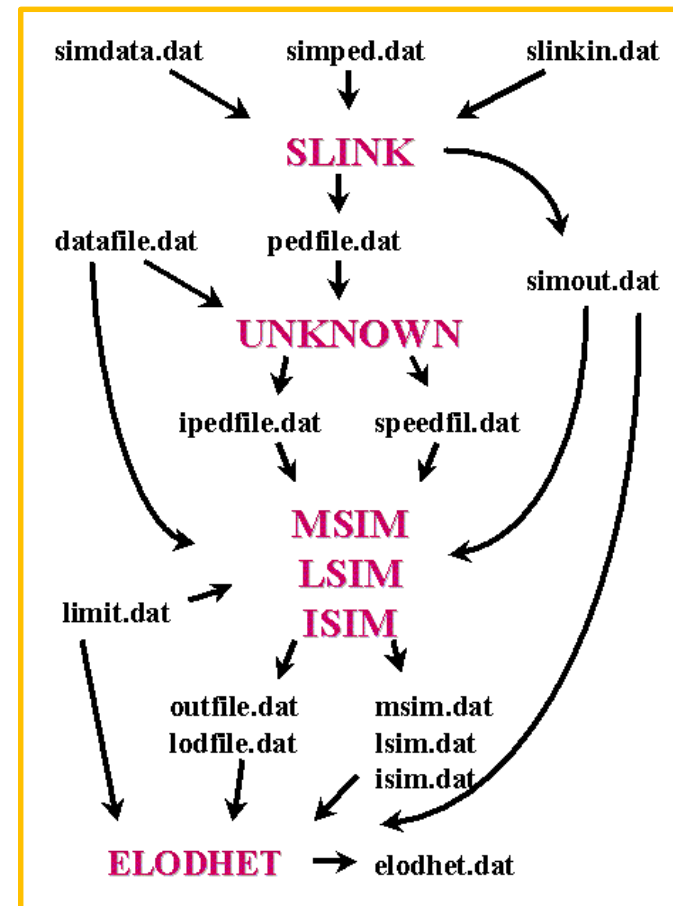
MUCH harder than unconditional

Conditional sims was first done in linkage analysis



- Power analysis for linkage requires simulation
 - conditional on disease genotype
 - conditional on the distance to disease locus
- Weeks, Ott, Lathrop (1990)
 - SLINK: a general simulation program for linkage analysis

Not for the faint of heart...



Conditional simulation: Strategies

1. Brute force

- Sequential sampling
- One marker at a time
- One individual at a time

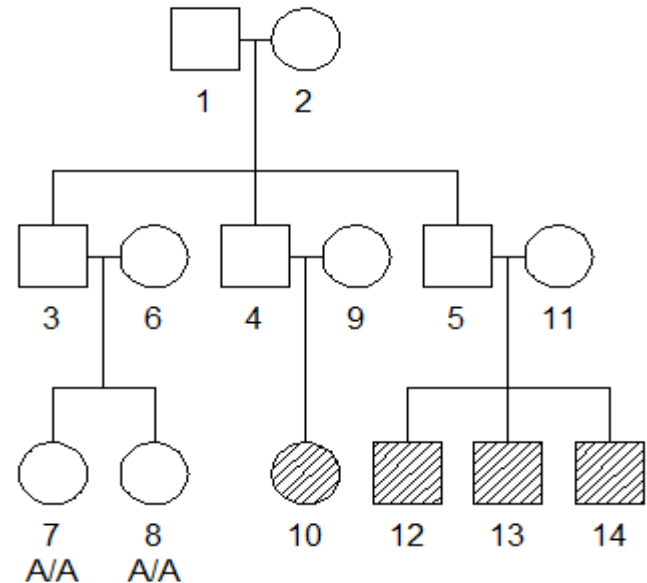
2. Founders + gene dropping

- Founders: Sample alleles from HW
- Rest: Mendelian coin tosses

3. Sample from the *joint* genotype distribution

- Efficient when few alleles

Usually fastest: Combining 2 and 3.



```
> markerSim(x, N=2, partial=m, avail=c(10,12:14))
```

```
Simulation strategy:
```

```
Pre-computed joint distribution: 1.
```

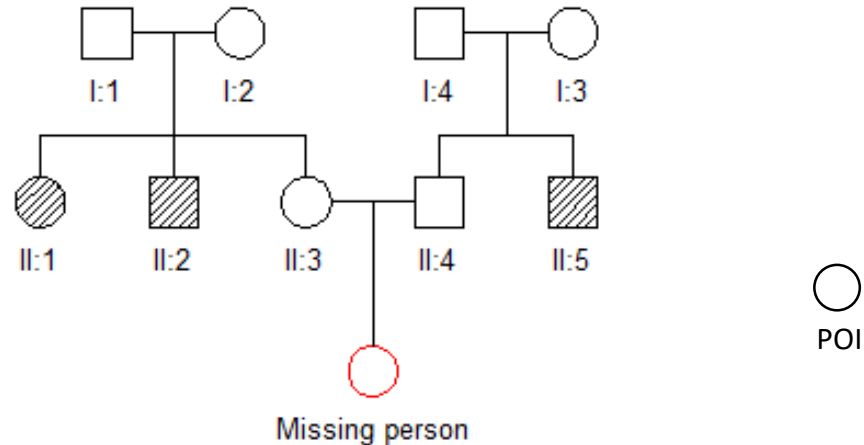
```
Brute force conditional simulation: 2.
```

```
Hardy-Weinberg sampling (founders): 9 and 11.
```

```
Simple gene dropping: 4, 10, 5, 12, 13 and 14.
```

```
Required likelihood computations: 9
```

Statistical power of reunification - part 2



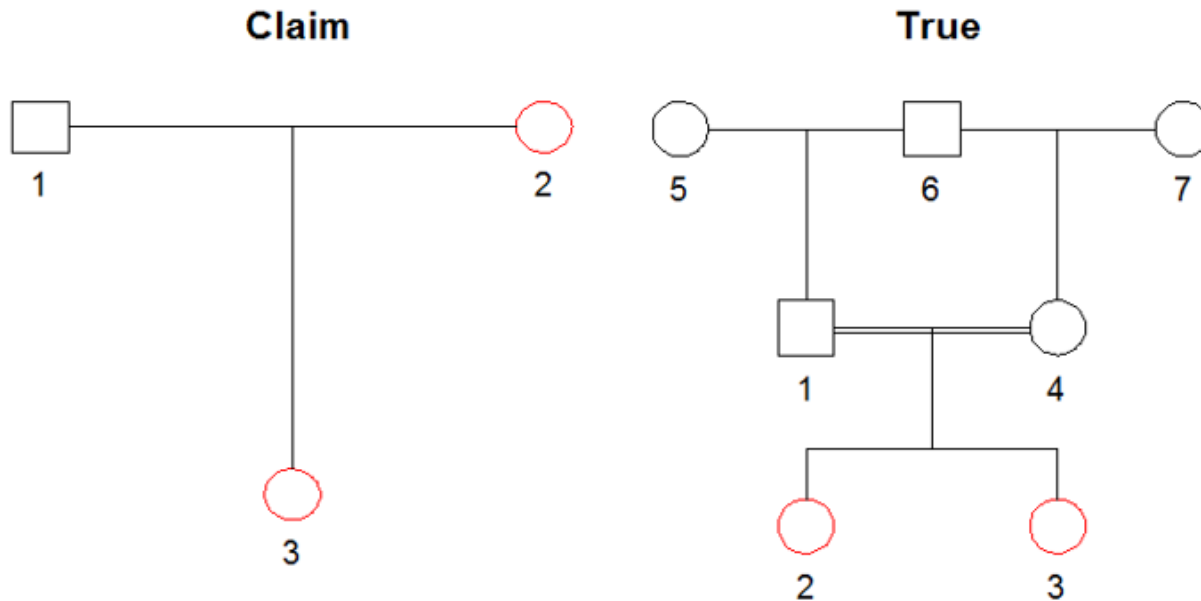
- Do we enough data to *exclude* an unrelated POI?
- **Exclusion power**

$$PE = P(\text{data incompat. with ped} \mid \text{POI unrelated})$$

- Can be computed exactly.
(Egeland, Pinto, Vigeland, 2014).

In forrel:
> exclusionPower()
> missingPersonEP()

The exclusion power formula



$$\begin{aligned}
 PE &= P(\text{data incompat with } Claim \mid True) \\
 &= \sum_{\substack{(g_2, g_3) \\ \text{impossible} \\ \text{in Claim}}} P(g_2, g_3 \mid True)
 \end{aligned}$$

Back to Argentina ...

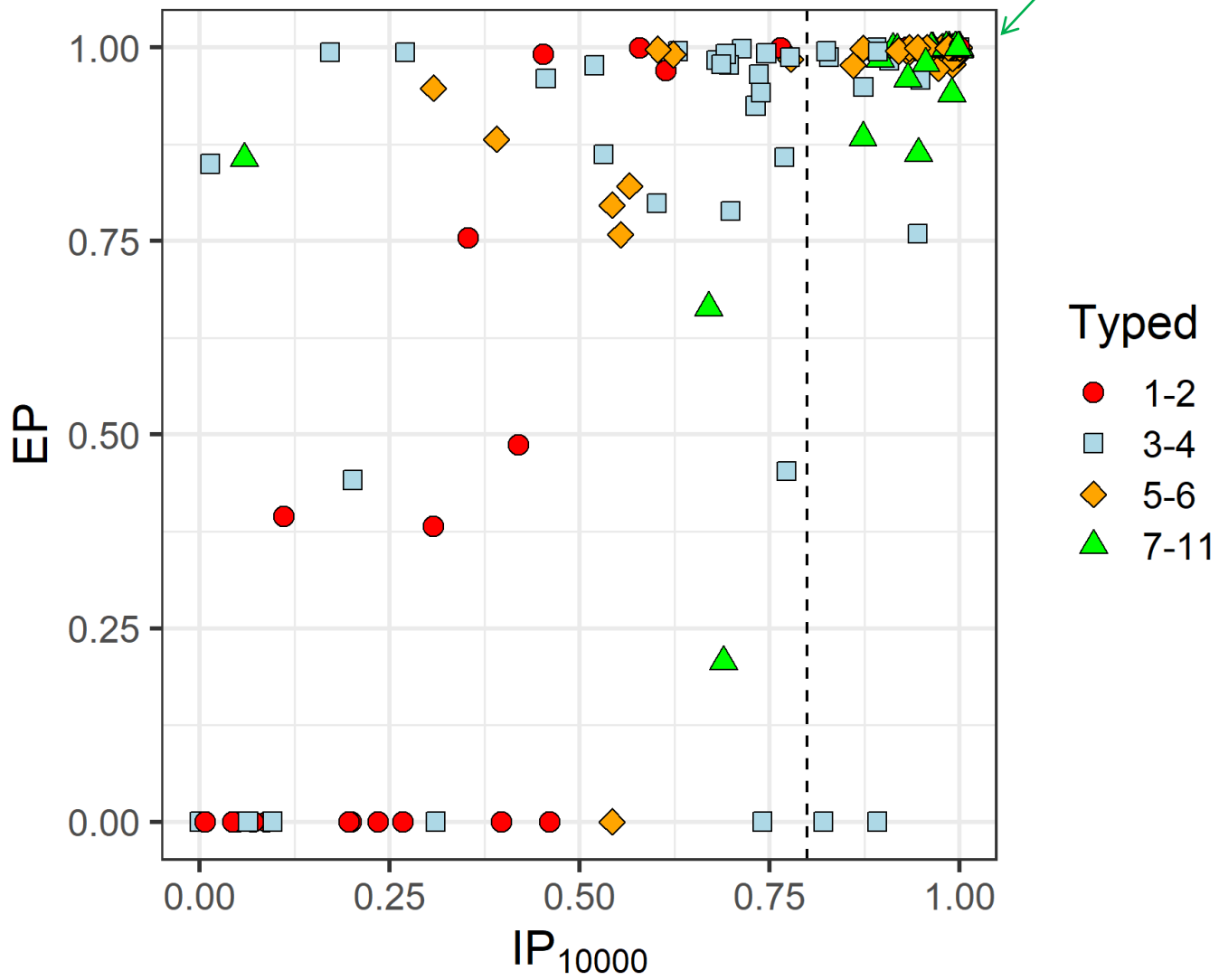
- Power evaluation of ~200 families in the BNDG database
 - most of them unsolved

Typed	Families	Parent(s) typed	2 nd degree only
1	11	5	5
2	20	5	15
3	29	7	20
4	41	6	33
5	29	7	14
6	26	5	14
7	14	0	8
8	14	0	9
9	5	0	3
10+	7	0	5
Total	196	35	126

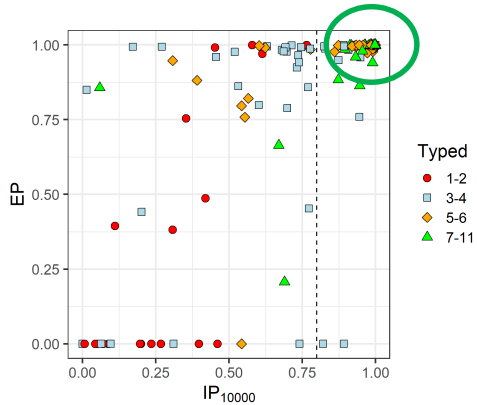
Missing	Total	mtDNA	Y
Male	27	27	25
Female	17	17	-
Unknown	152	142	108
Total	196	186	133

- For each family: Compute IP_{10000} and EP

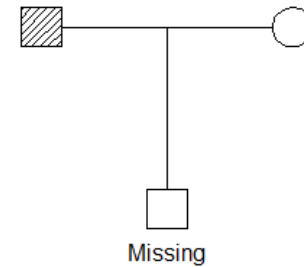
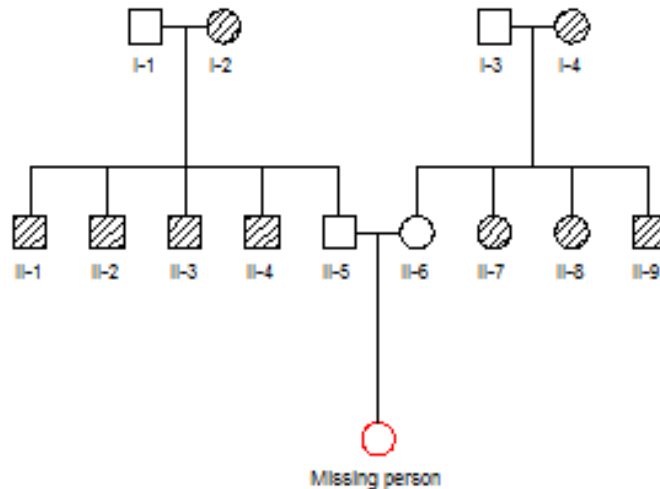
Results



Excellent power and PE

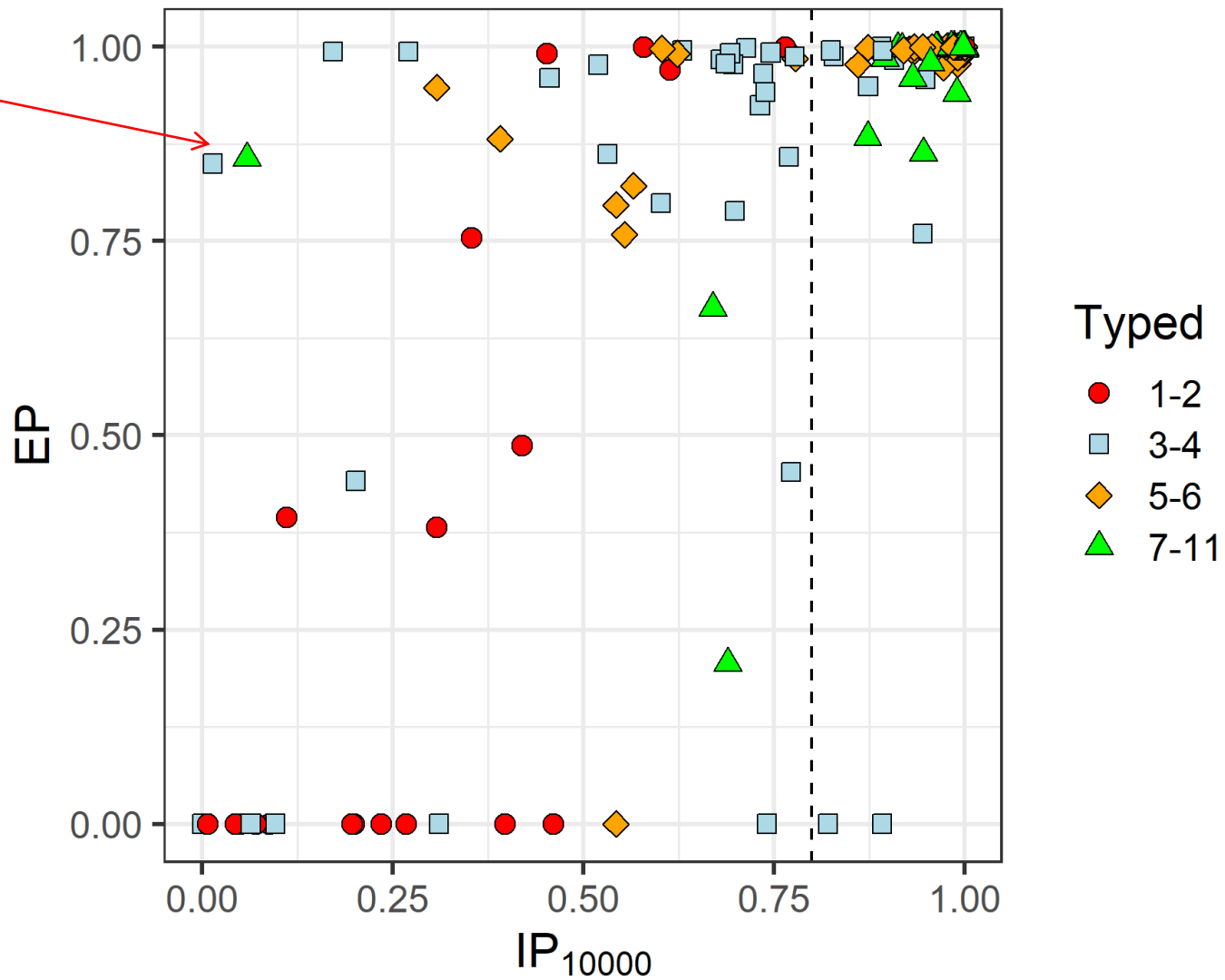


- 68 families with $PE > 99\%$ and $E_{10000} > 99\%$
- Includes all (except 1) of the 31 cases with parental data

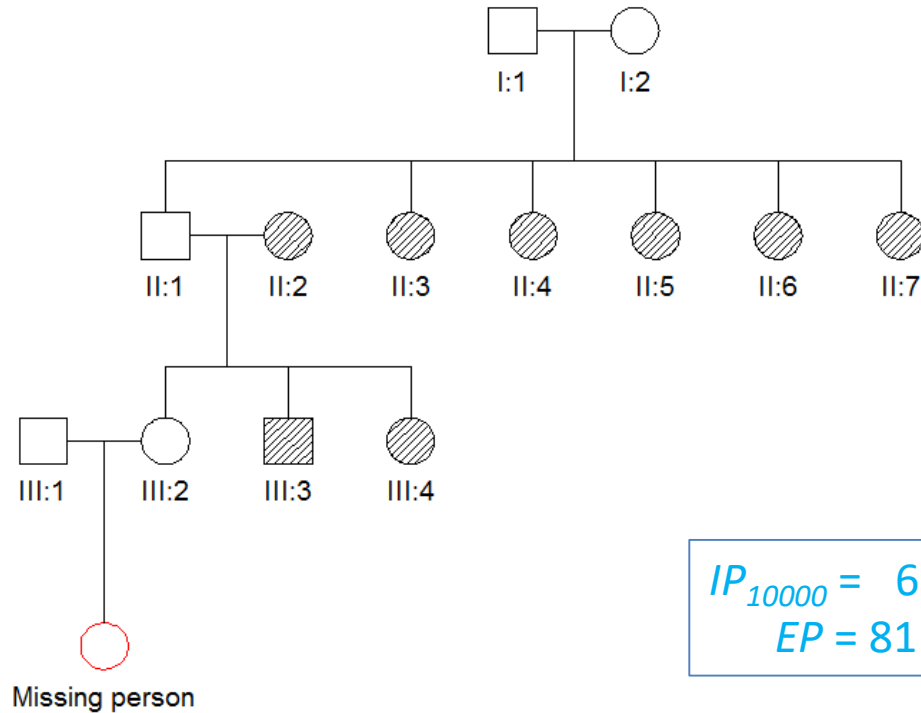
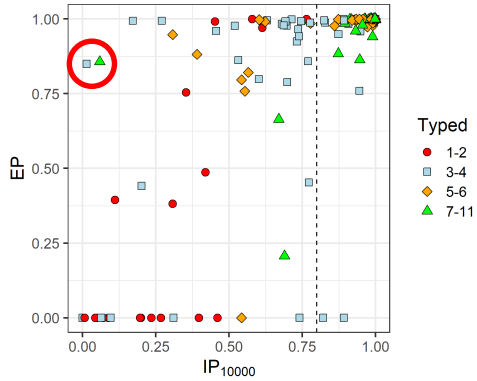


$IP_{10000} = 100\%$
 $EP = 100\%$

Results

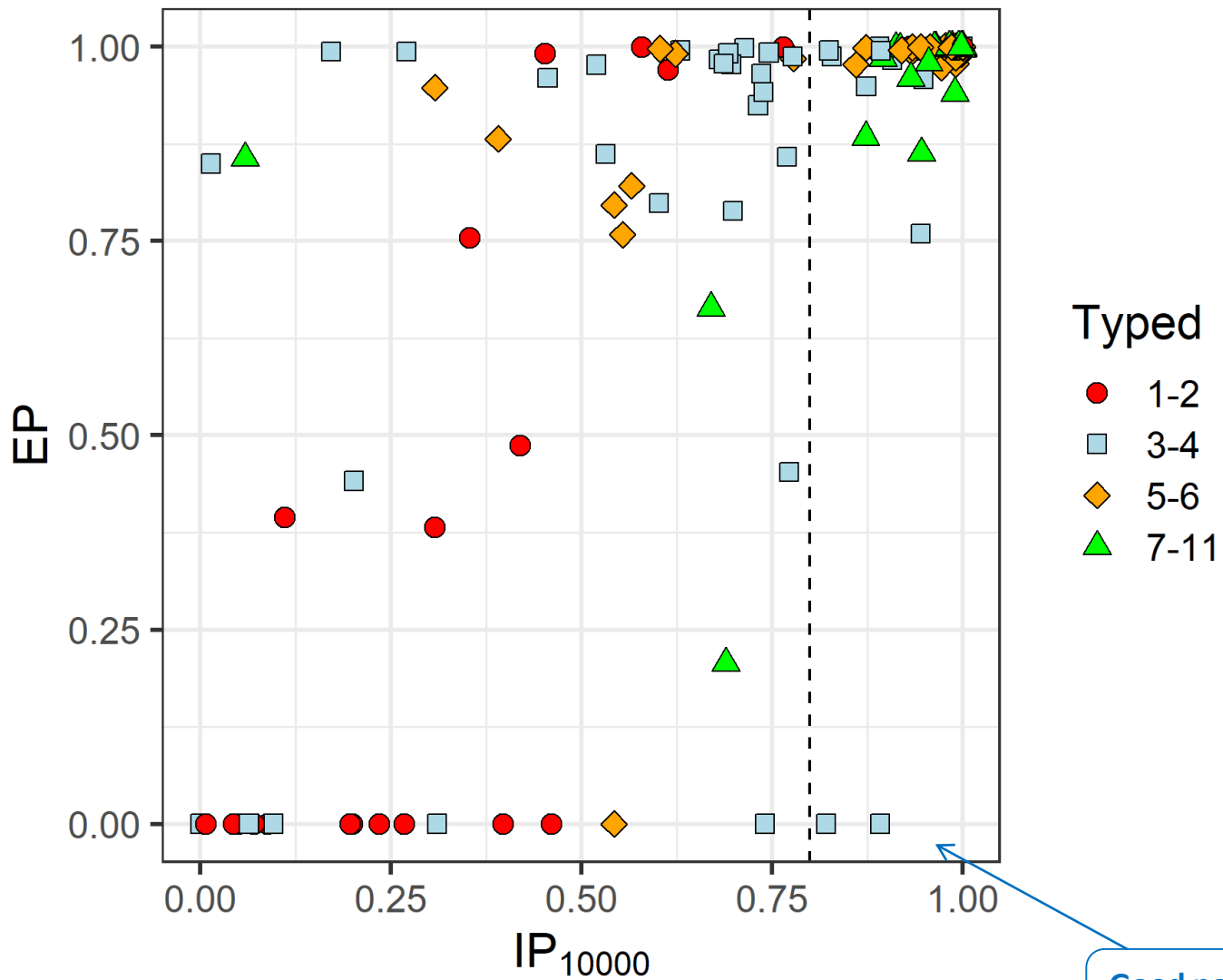


Low power despite many typed



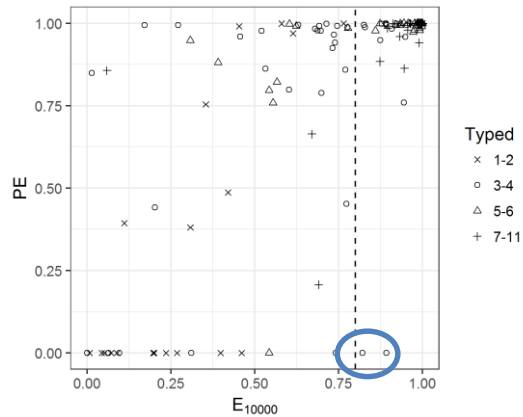
$IP_{10000} = 6\%$
 $EP = 81\%$

Results

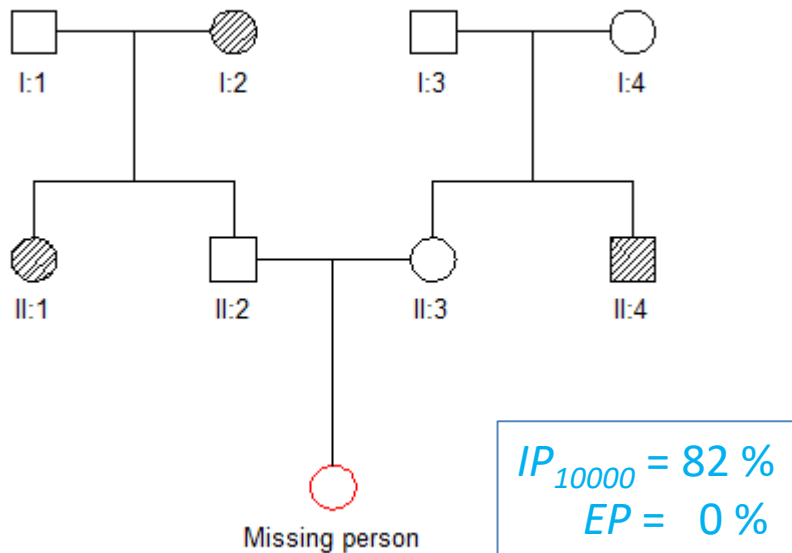


Good power, but not for exclusion

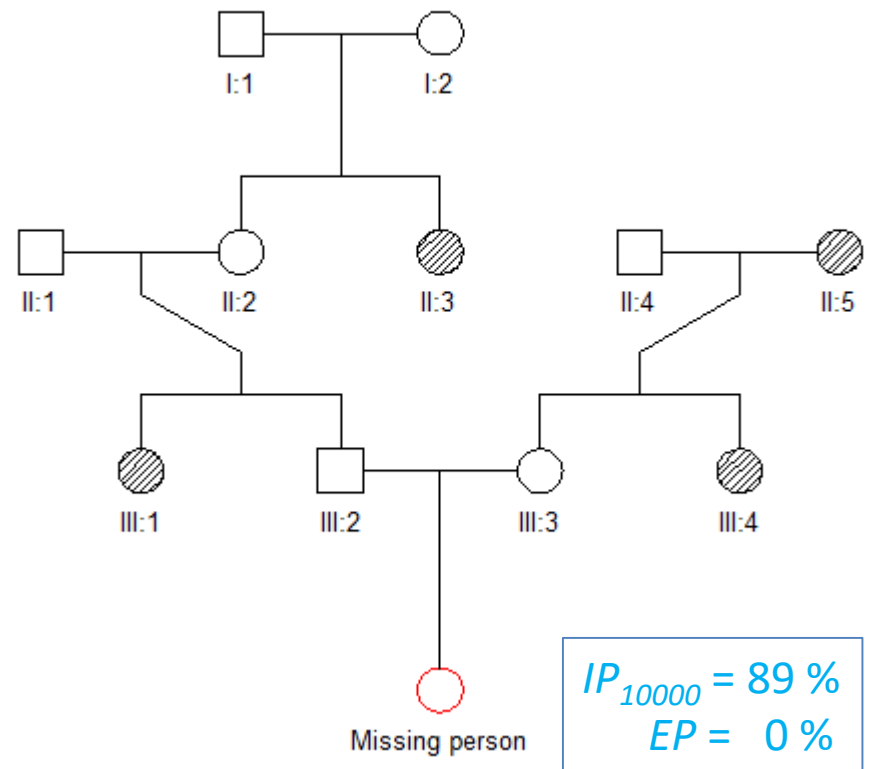
Good power, but exclusion impossible



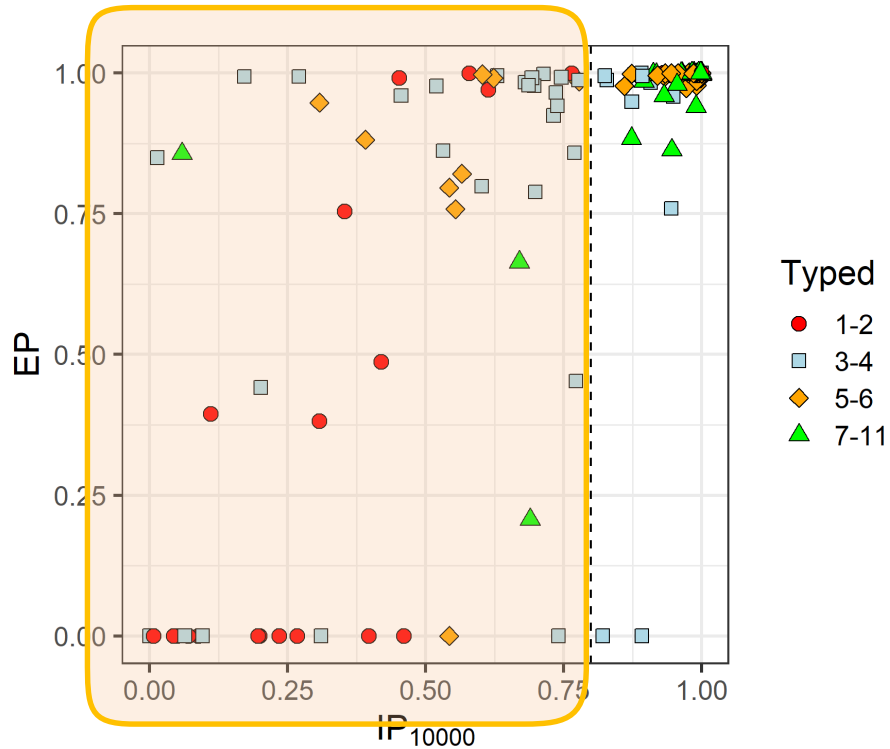
E184



AF107



Overall



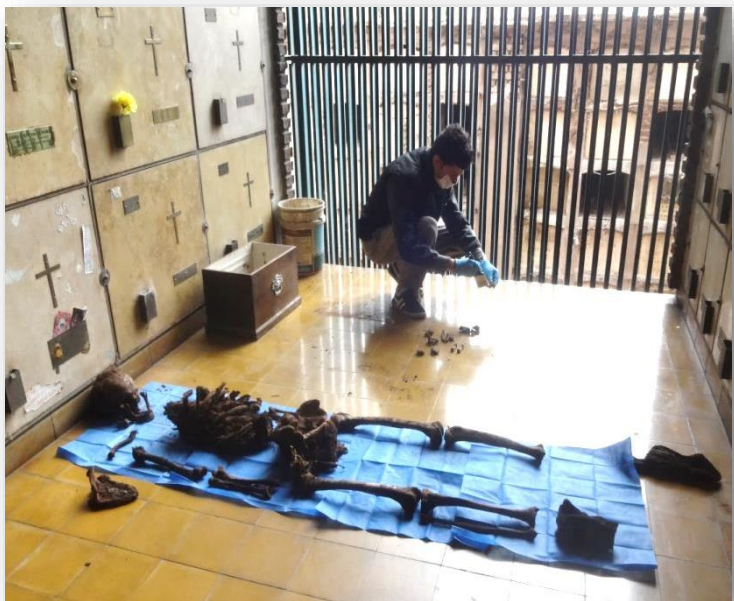
- 34% of the unsolved families had poor power

- Reasons:
 - few markers
 - few typed relatives

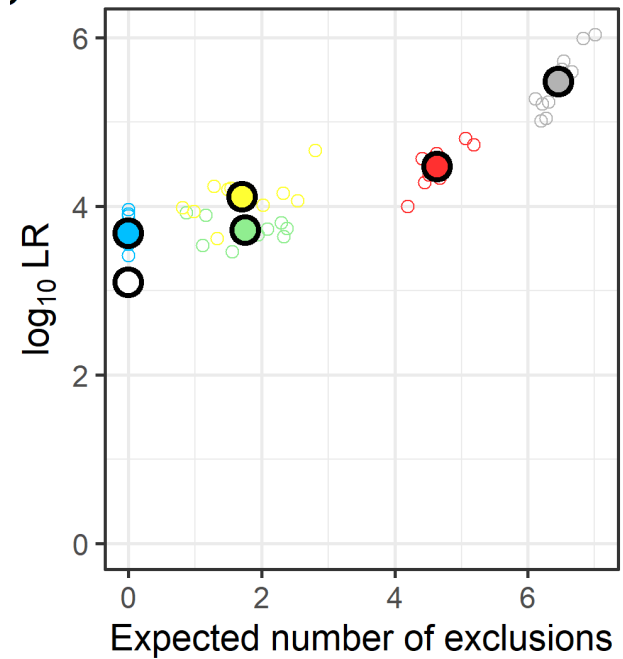
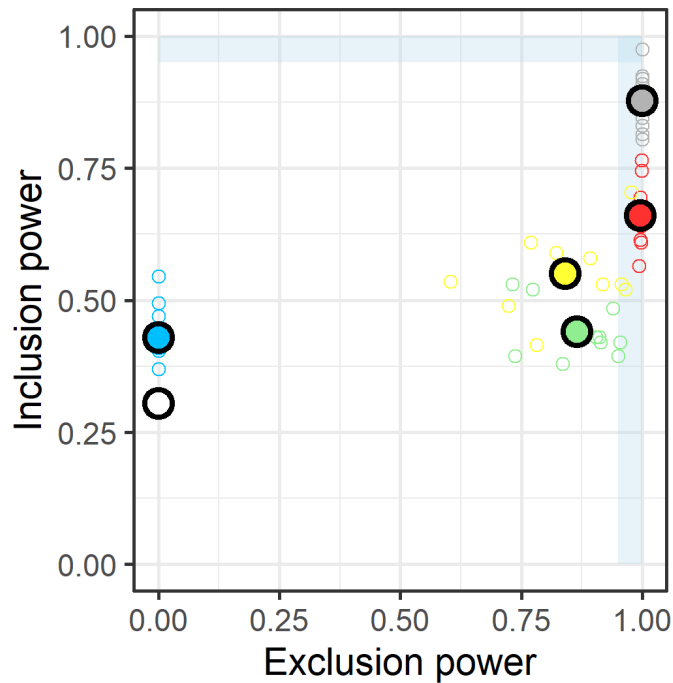
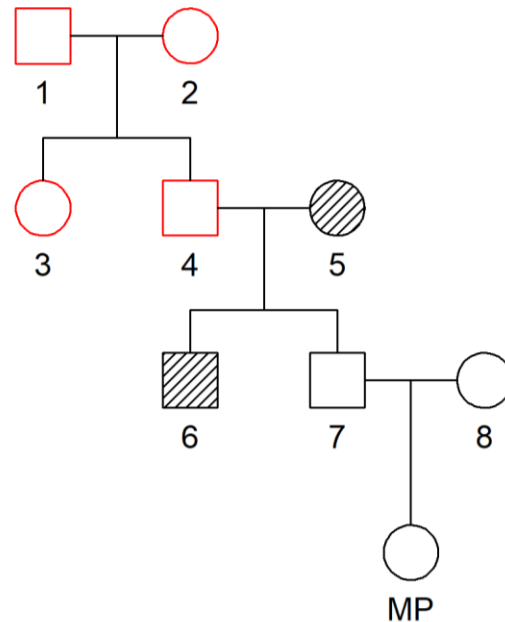
Ongoing actions:

- retyping 1000 individuals
- exhumation of 100 (!) informative relatives

Forensic Anthropology Unit at BNDG



Who should you choose?

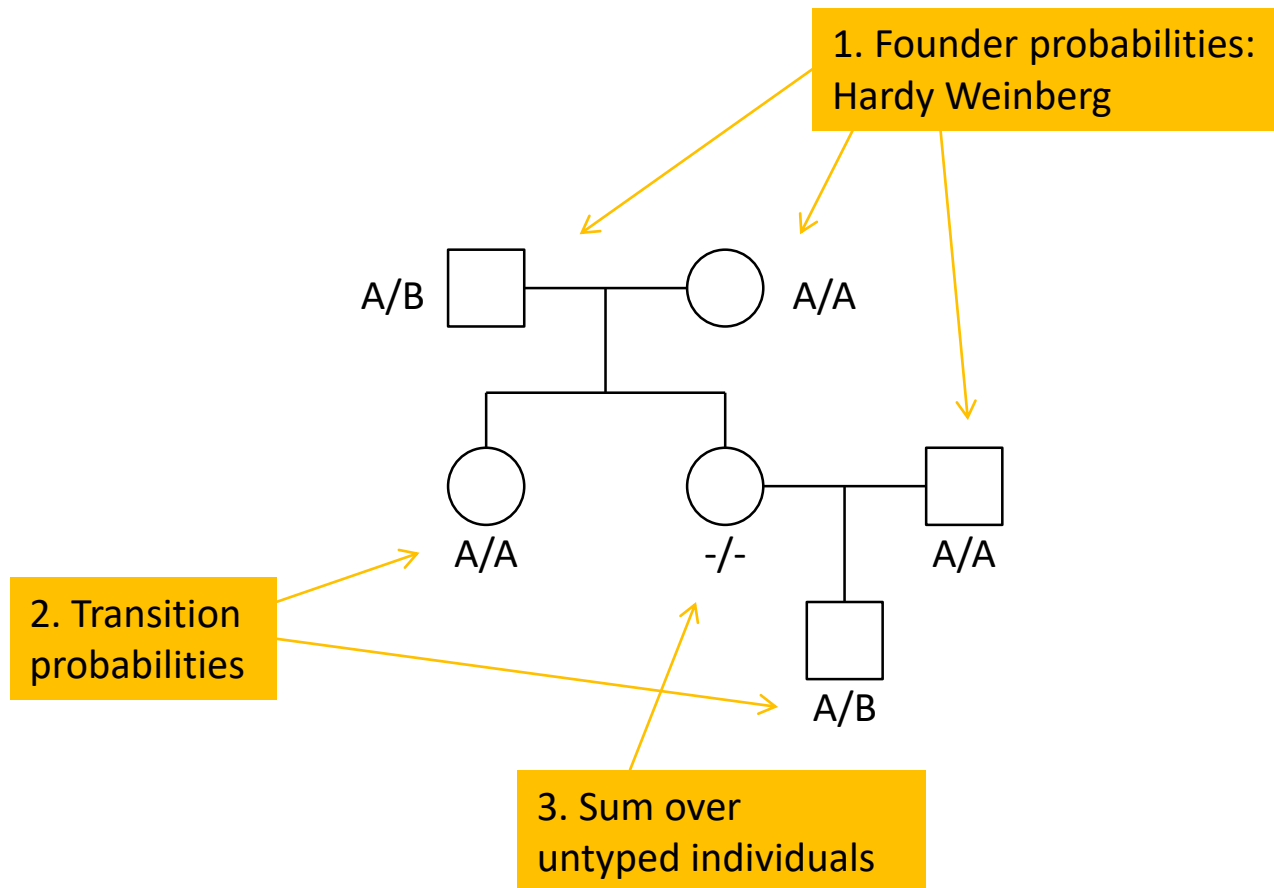


- Coffee break!
- Next: Summary ...

- Monday 1

Pedigrees, genetics and probabilities

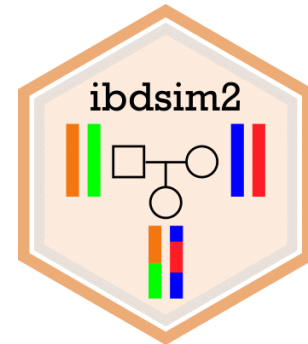
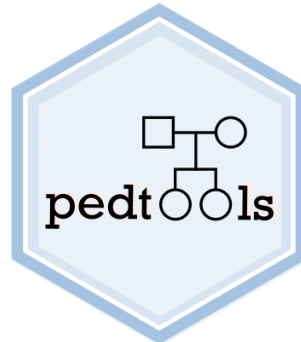
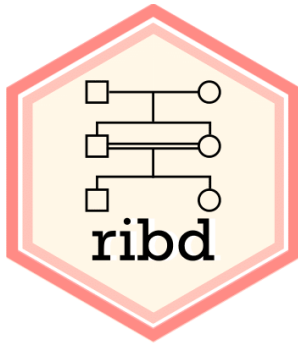
Ingredients for likelihood computations



- Monday 2

Introduction II: Pedigree analysis in R with the ped suite

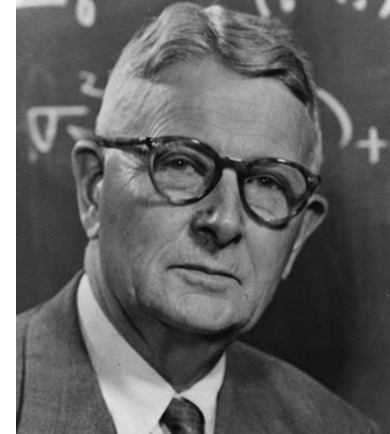
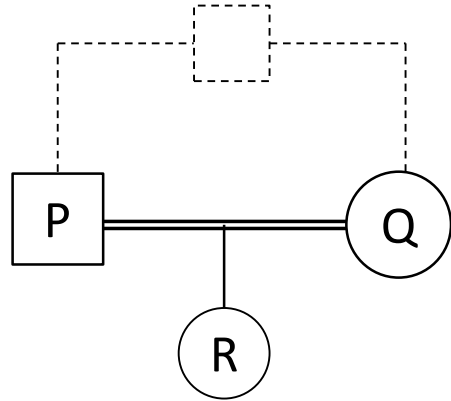
The *ped suite*



- Tuesday 1

IBD and coefficients of relatedness and R

Relatedness yellow belt: Coefficient of kinship/inbreeding



Sewall Wright
(1889 - 1988)

- Wright (1921): The kinship coefficient φ between P and Q

$$\varphi_{P,Q} = P(\text{random allele of P IBD with random allele of Q})$$

$$= P(\text{R is autozygous})$$

$$= f_R$$

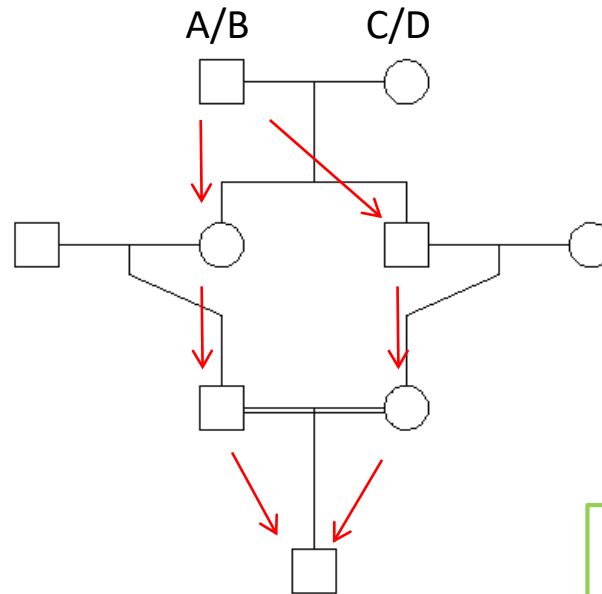
the inbreeding coefficient of R

P and Q related



$$\varphi_{P,Q} > 0$$

Examples



Same for
C and D

$$f = P(\text{A/A autozygous}) \cdot 2 \cdot 2$$

$$= 0.5^6 \cdot 2 \cdot 2$$

$$= \frac{1}{64} \cdot 4 = \frac{1}{16}$$

A or B

The relatedness triangle

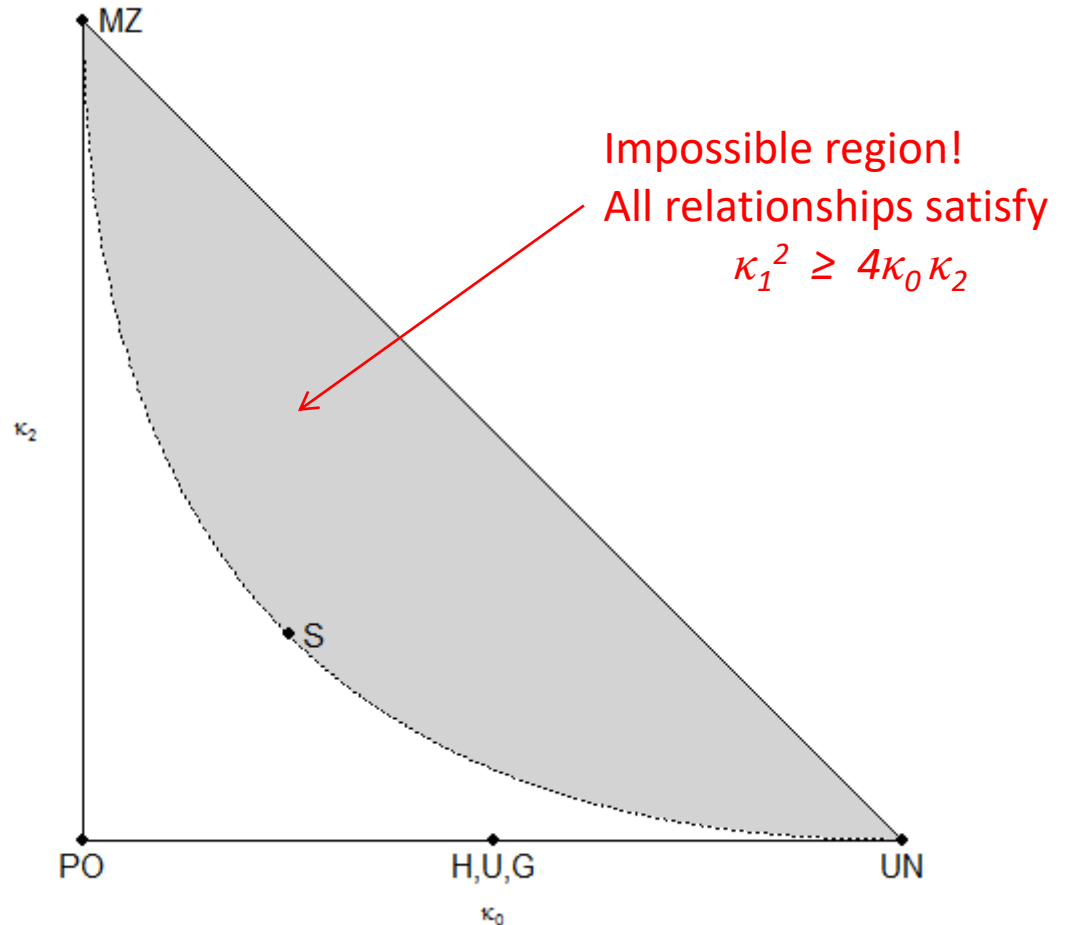
For random autosomal locus

$\kappa_0 = P(0 \text{ alleles IBD})$

$\kappa_1 = P(1 \text{ alleles IBD})$

$\kappa_2 = P(2 \text{ alleles IBD})$

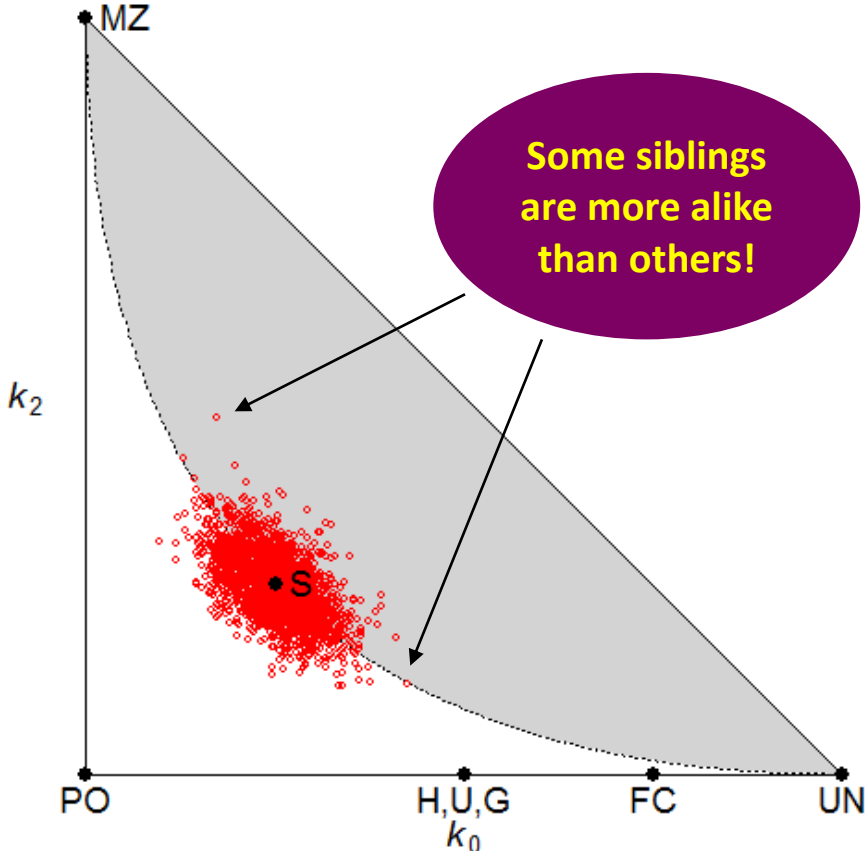
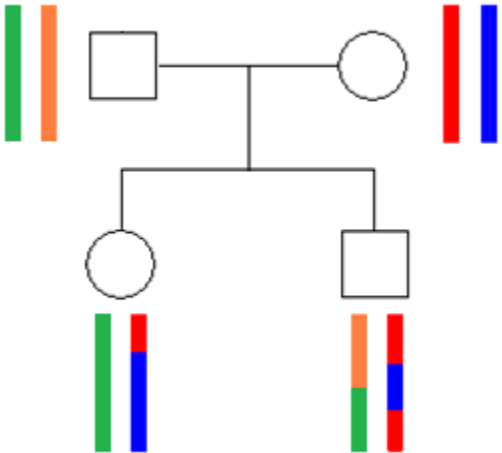
$$\kappa_0 + \kappa_1 + \kappa_2 = 1$$



- Tuesday 2

Why some siblings are more equal than others

Distribution of realised IBD coefficients



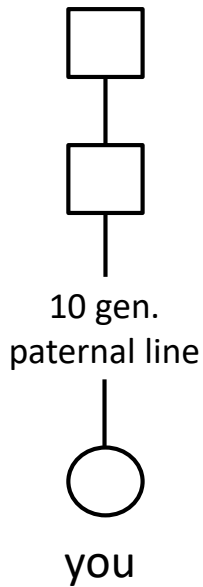
1000 simulations

- Wednesday 1

Recombination and genetic linkage



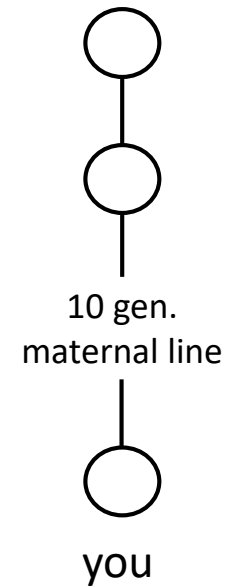
Napoleon Bonaparte (1769 - 1821)



$P(\text{any IBD sharing}) \approx 19\%$



Jane Austen (1775 - 1817)



$P(\text{IBD sharing}) \approx 33\%$

- Wednesday 2

Linkage analysis in medical genetics

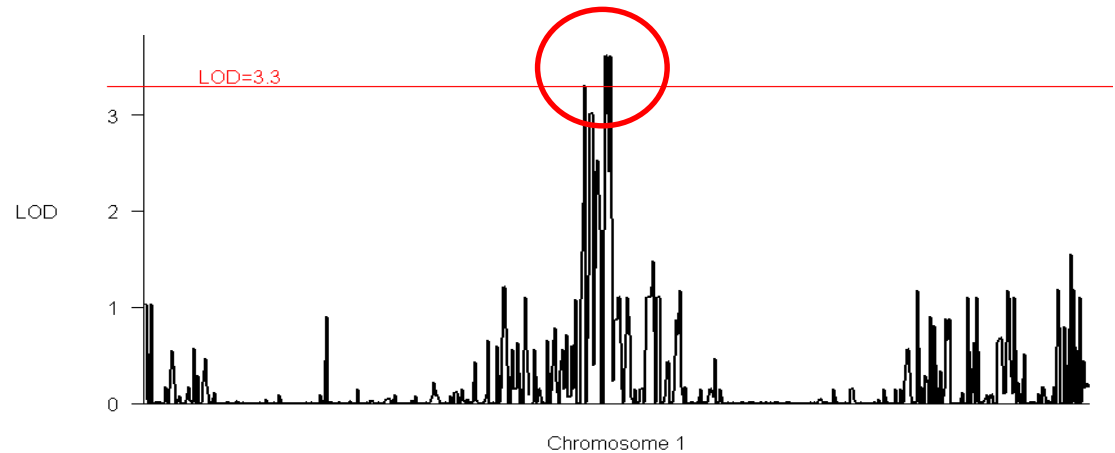
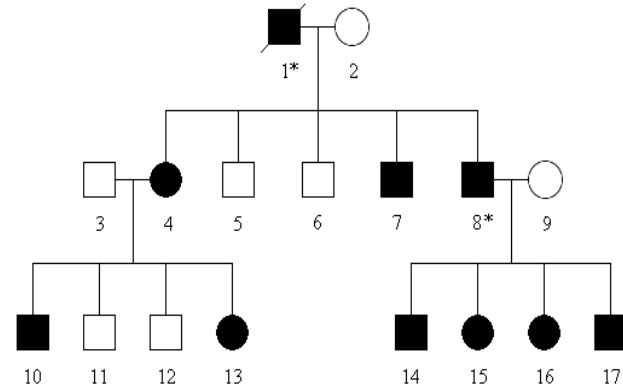
Linkage analysis workflow

1. Collect (large) affected families

2. SNP genotyping

3. Parametric linkage analysis

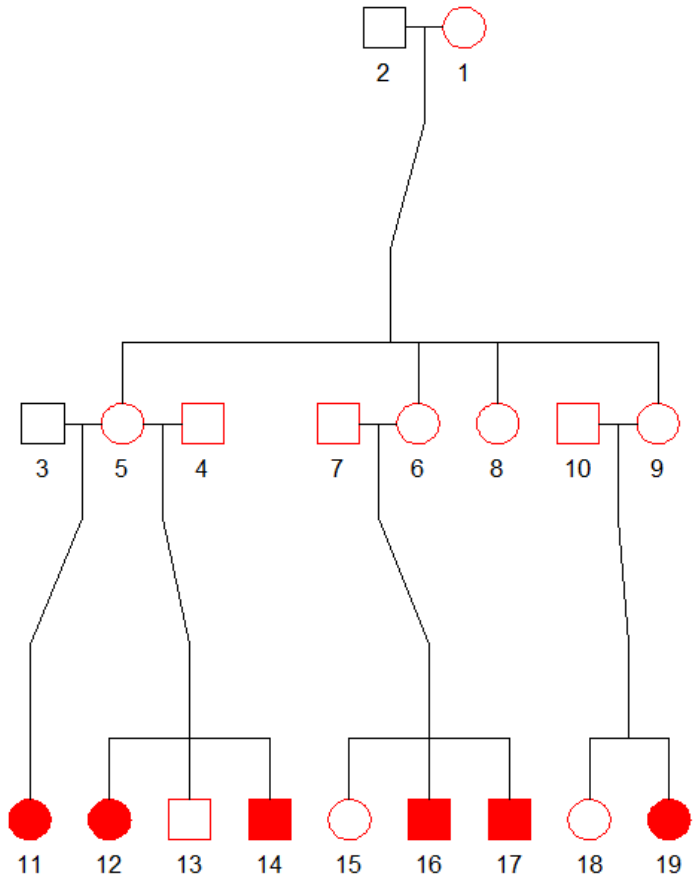
4. Sequence genes in linkage peak → identify causal mutation



- Thursday 1

*Inference of pairwise relatedness
and
Pedigree reconstruction*

Family 22

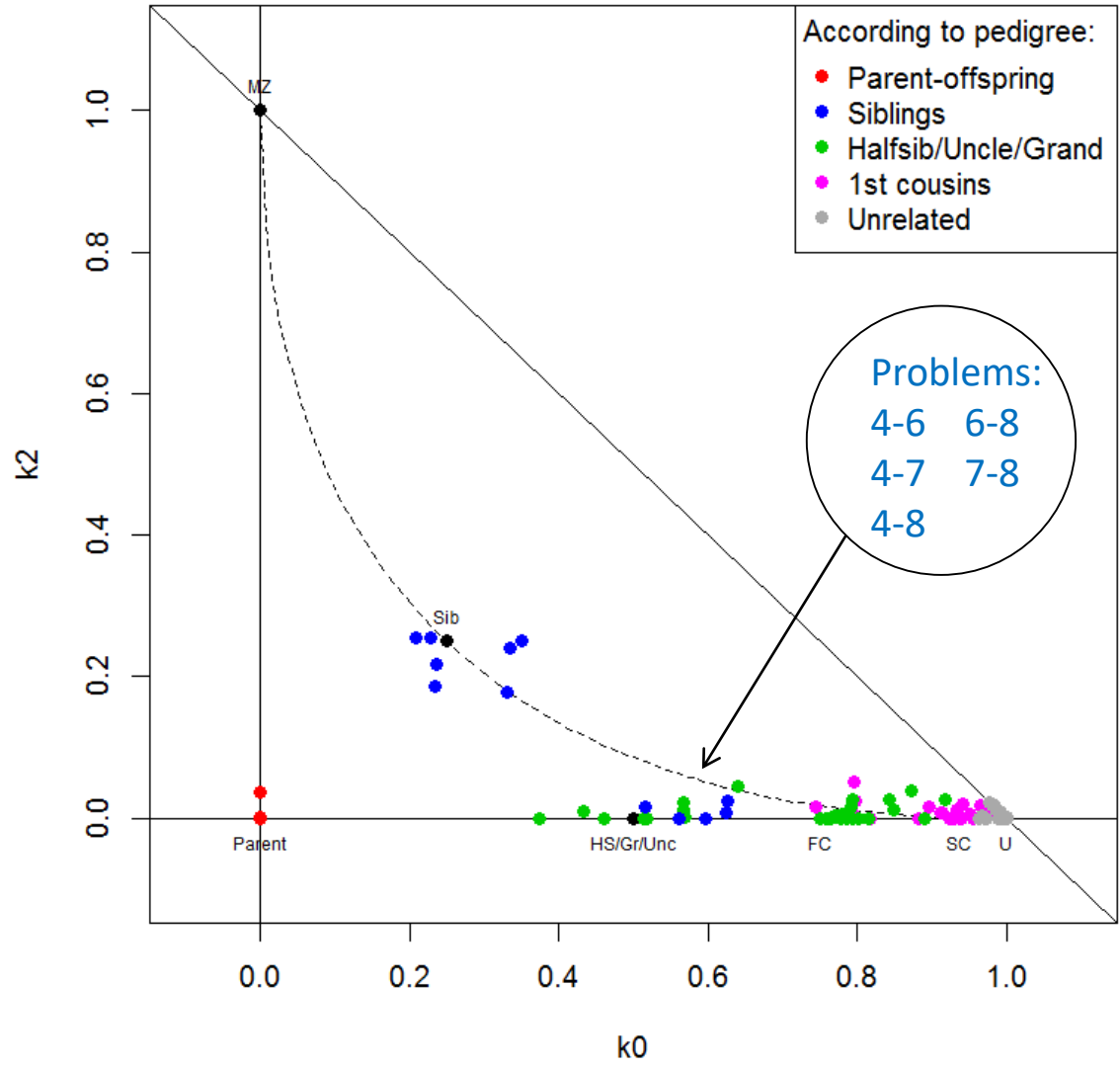
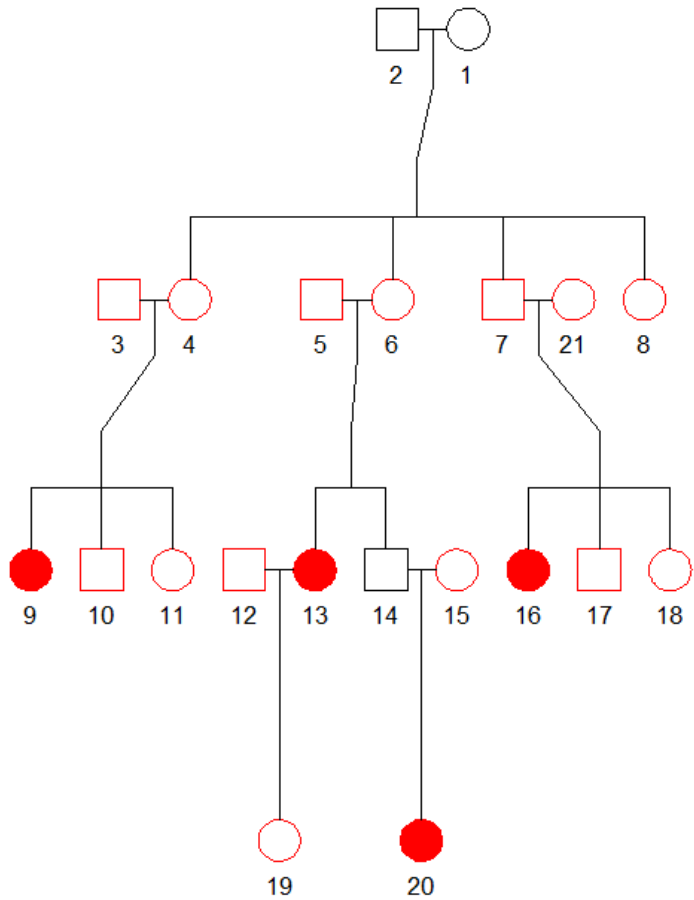


```
> library(forrel)
```

```
> k = IBDEstimate(x)
```

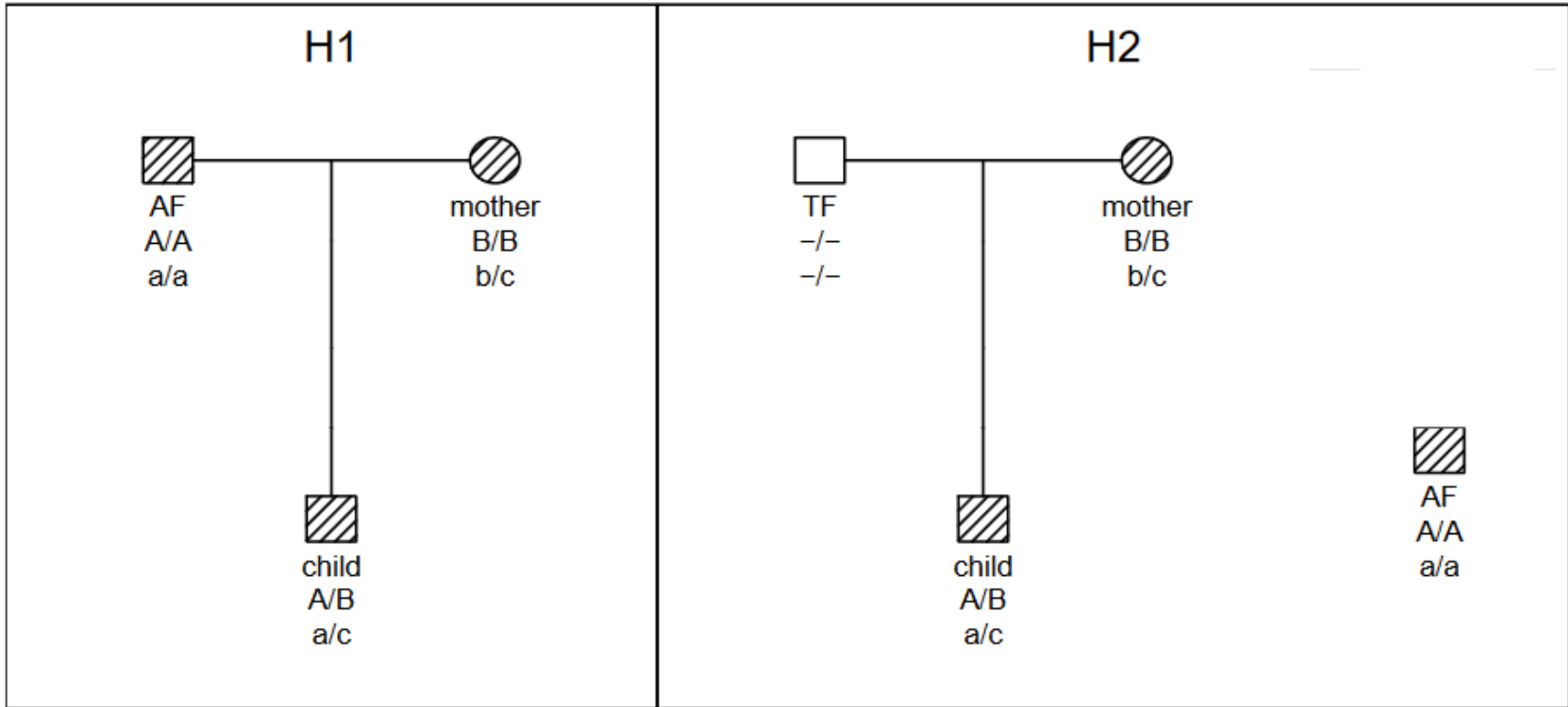
```
> showInTriangle(k)
```

Family 16



- Thursday 2 + Friday 1

Forensics genetics



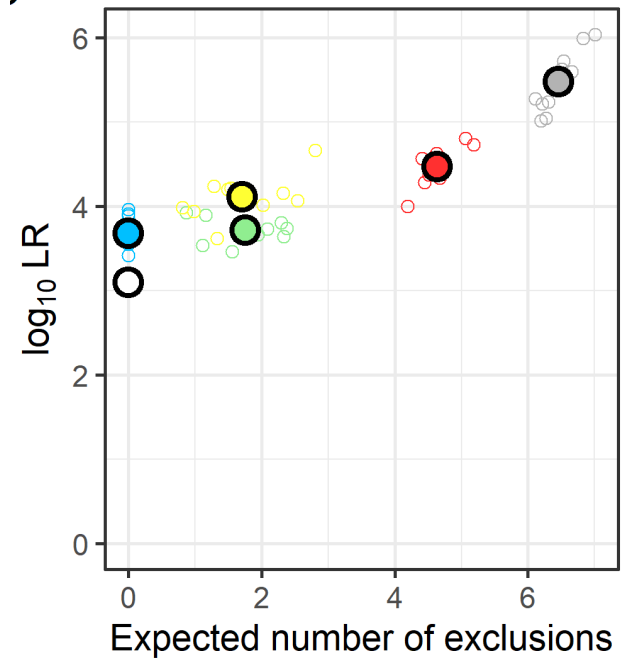
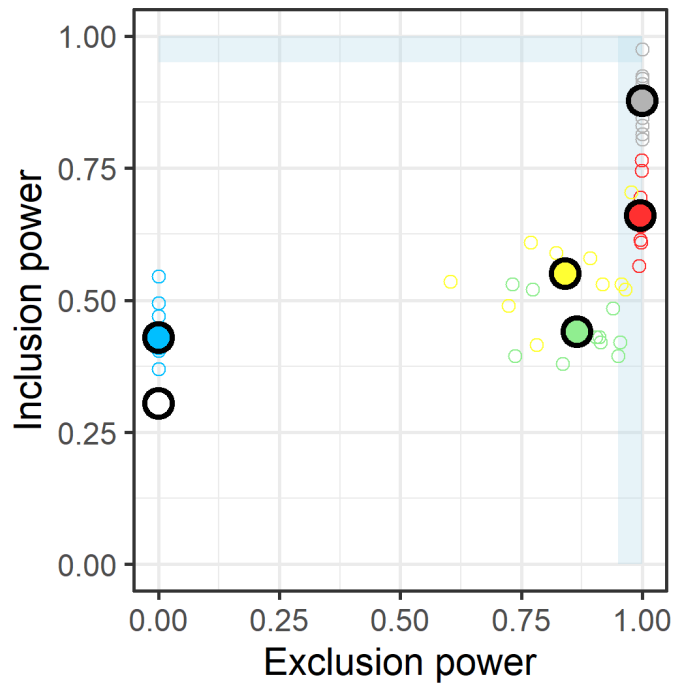
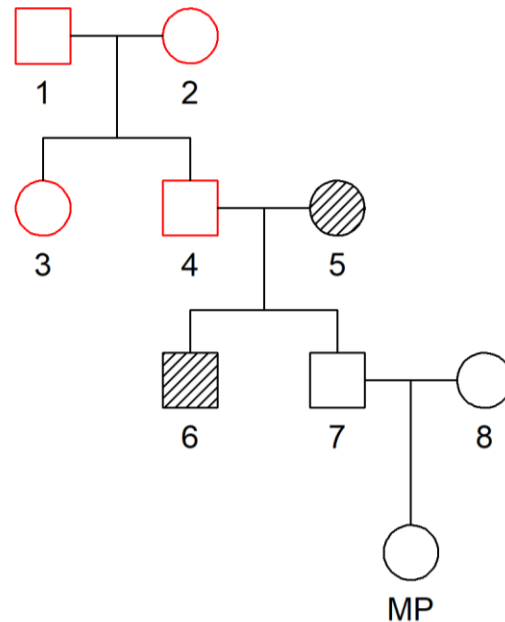
```
res = kinshipLR(list(H1, H2), ref = 2)
```

$$LR = \frac{P(\text{data} | H1)}{P(\text{data} | H2)}$$

- Friday 2

Case study: Argentina

Who should you choose?



The end!