

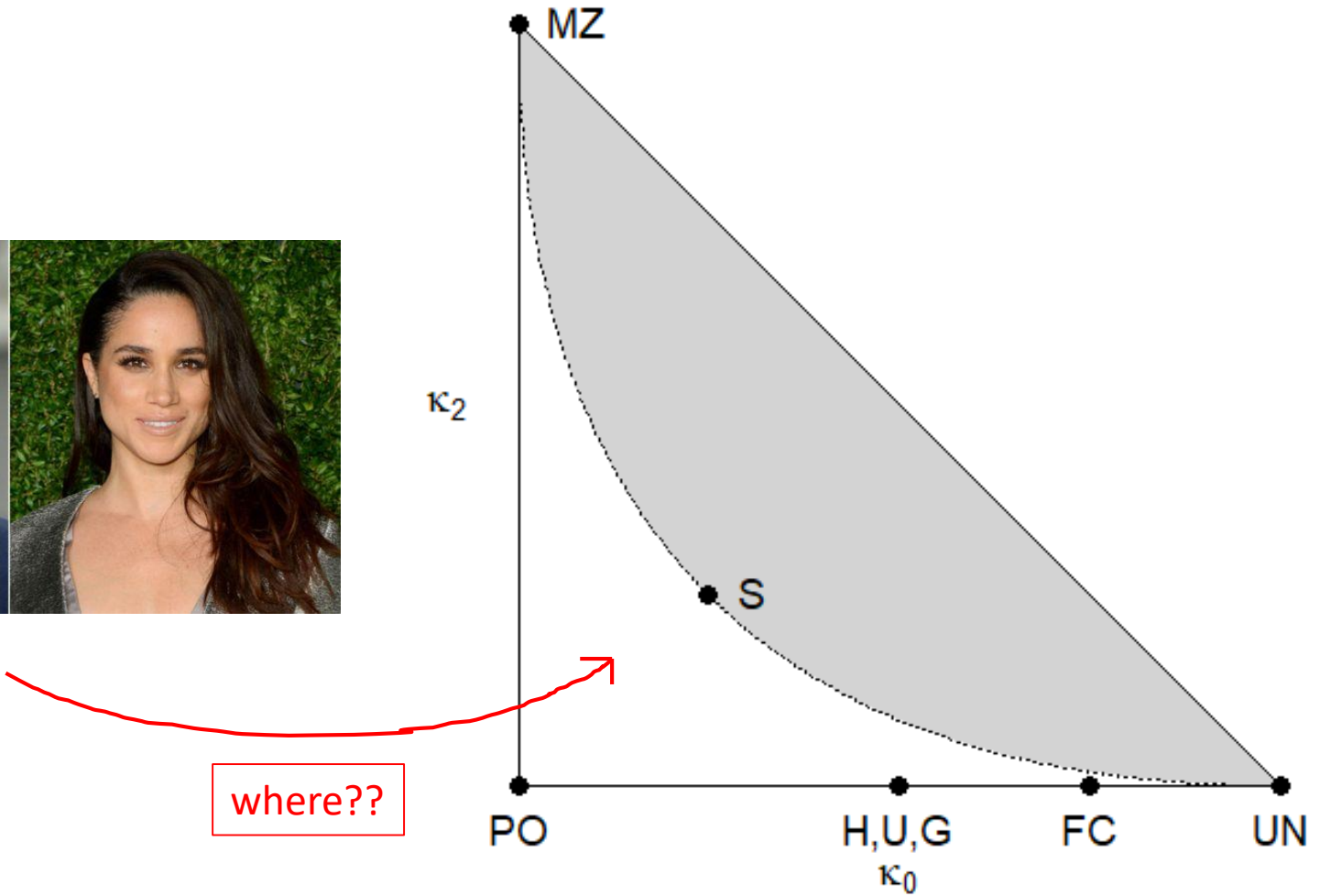


Lecture 7: Estimation of pairwise relatedness and Pedigree reconstruction

Magnus Dehli Vigeland

Statistical methods in genetic relatedness and pedigree analysis

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

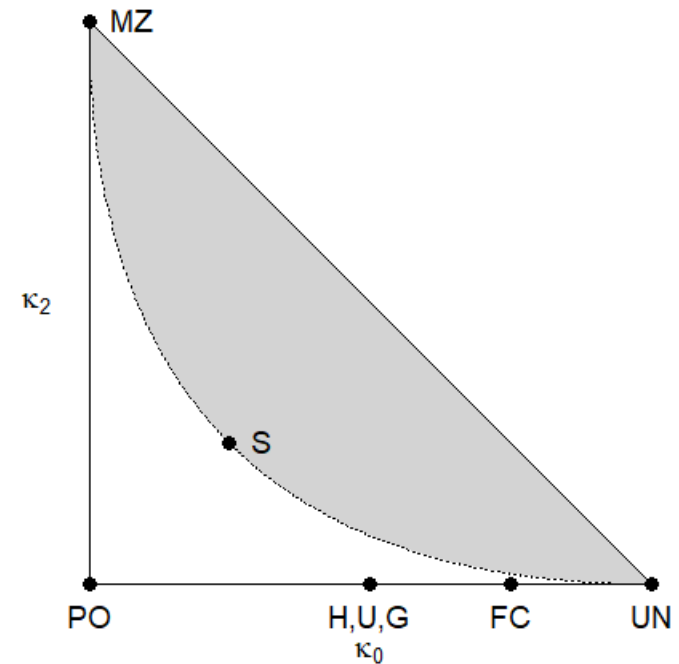


Maximum likelihood estimation of $\kappa = (\kappa_0, \kappa_1, \kappa_2)$

- Thompson (1975)
 - Given: marker genotypes for two individuals
 - The likelihood function

$$L(\kappa) = P(\text{genotypes} \mid \kappa)$$

- Find the point k which maximizes L !
 - Called the maximum likelihood estimate (MLE)
- Assumptions:
 - known allele freqs
 - HWE
 - no inbreeding



The likelihood function

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

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

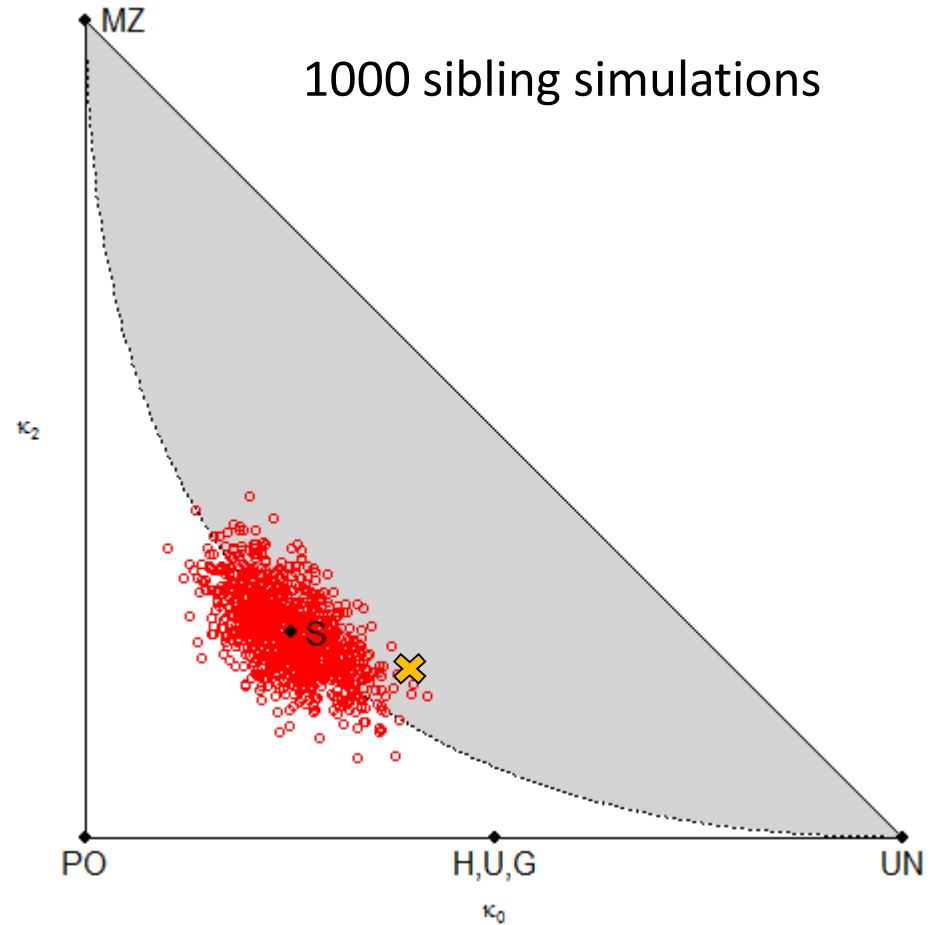
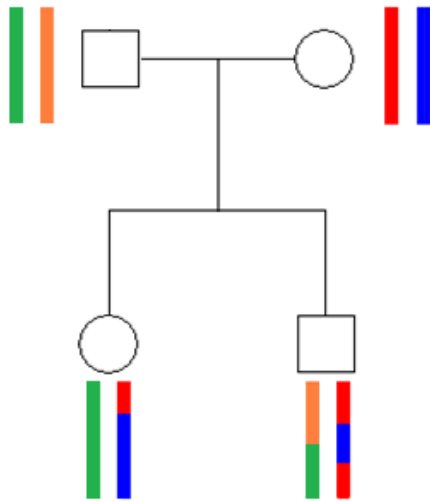
UN = unrelated
PO = parent/offspr
MZ = monozygotic

- With several independent markers:

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

What are we estimating?

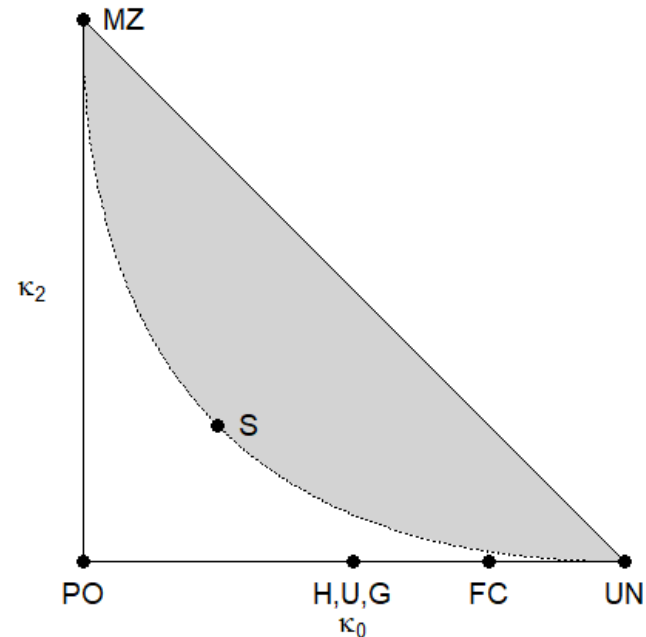
Answer: The *realised* coefficients!



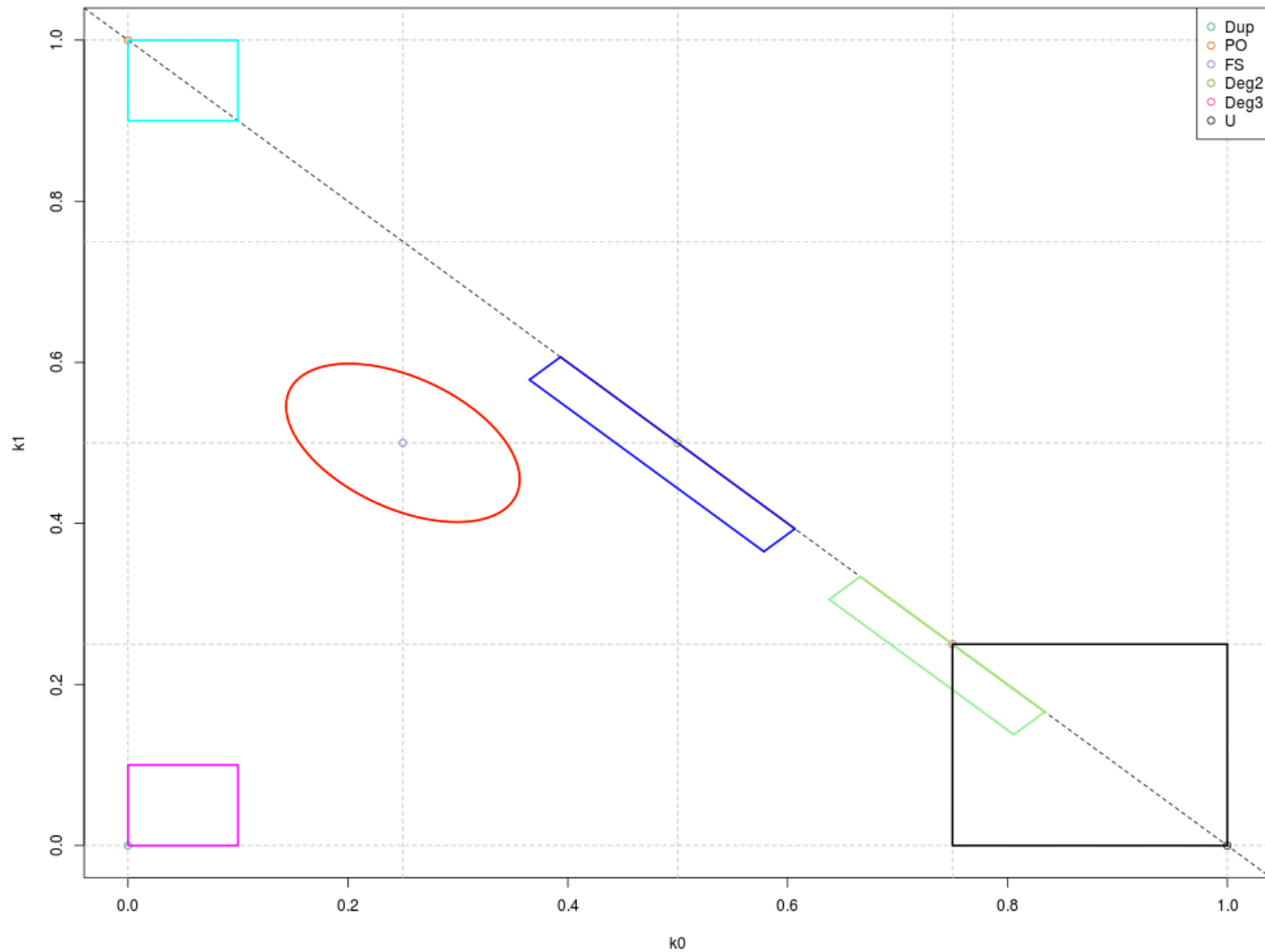
Implementations

- R
 - pedtools + forrel (imports package **maxLik** for the ML-estimation)
 - **SNPrelate, GWASTools** (optimized for association studies)
 - **CrypticIBDcheck** (as above, slow with many markers)
 - +++

- Other
 - PLINK
 - KING
 - Beagle
 - +++

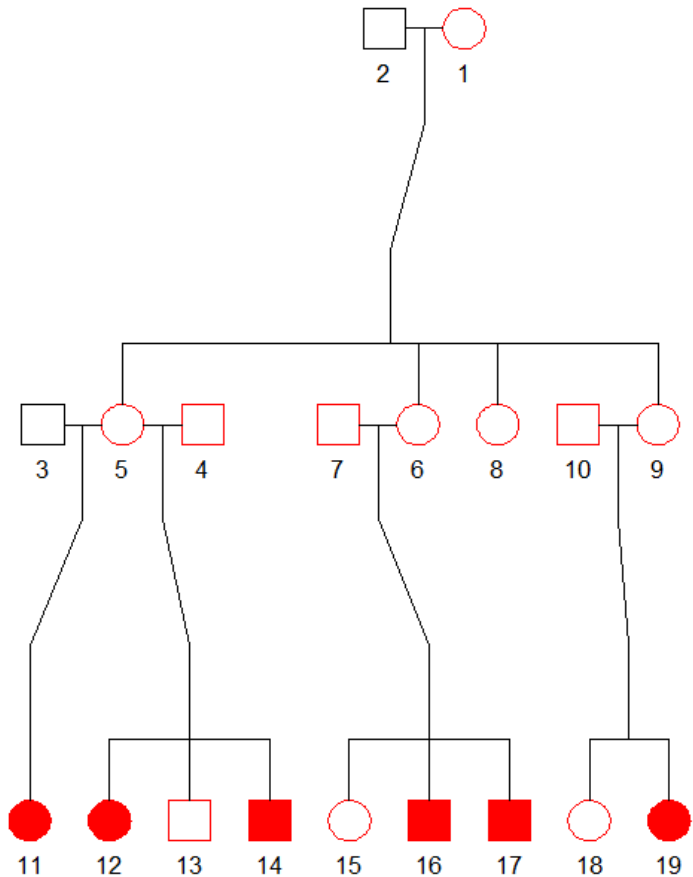


Example: From the manual of GWAStools



Let's look at some examples!

Family 22



```
> library(forrel)
```

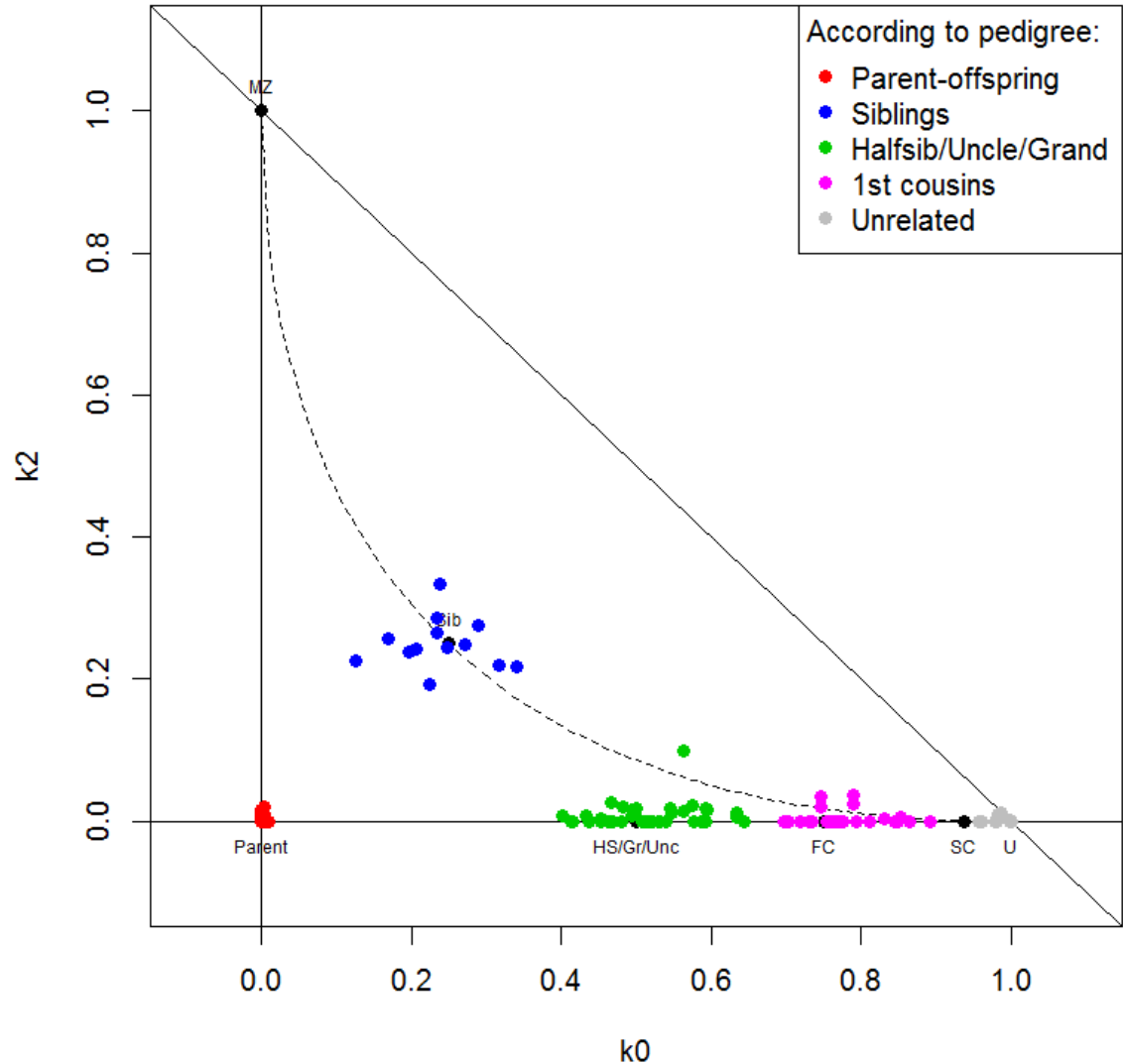
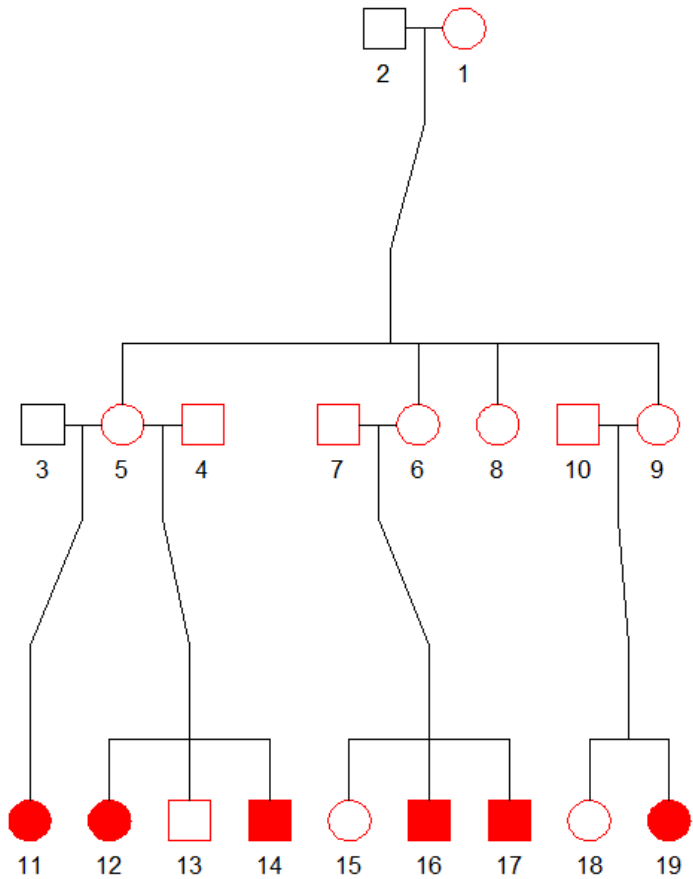
```
> x = ... (pedigree with markers)
```

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

```
> showInTriangle(k)
```

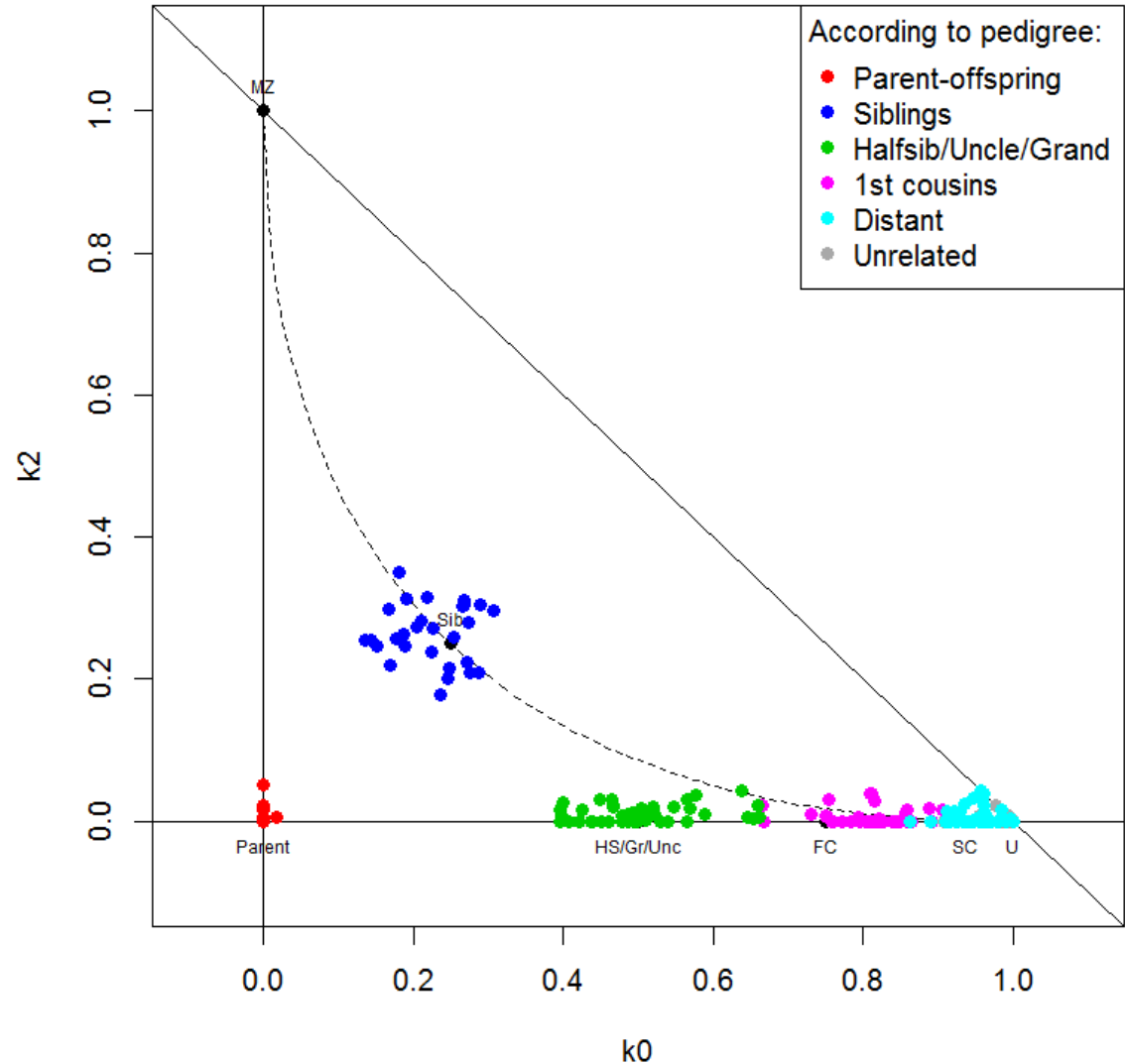
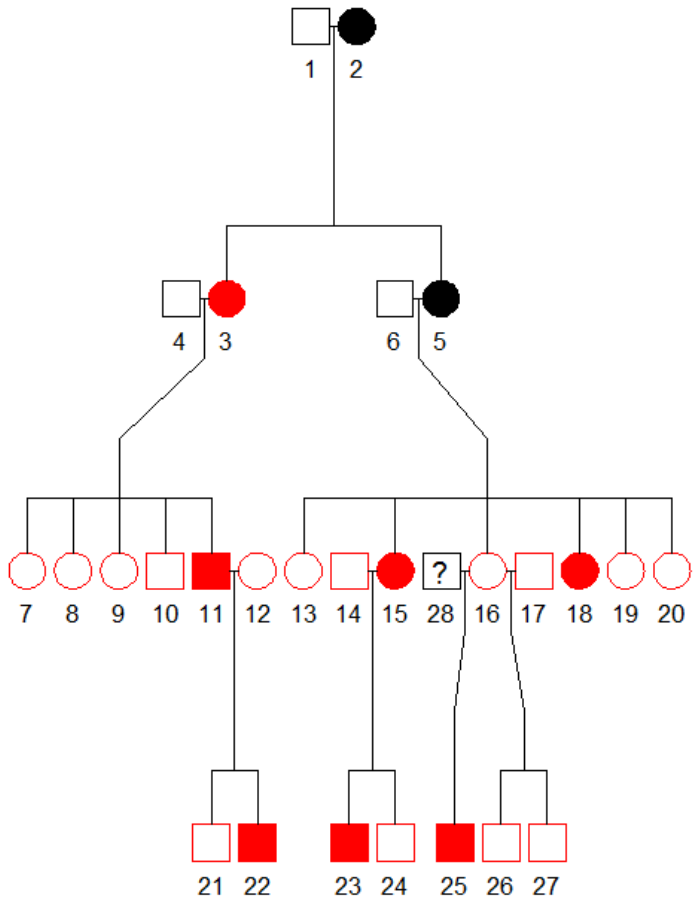
Pedigree 22 - OK

Family 22



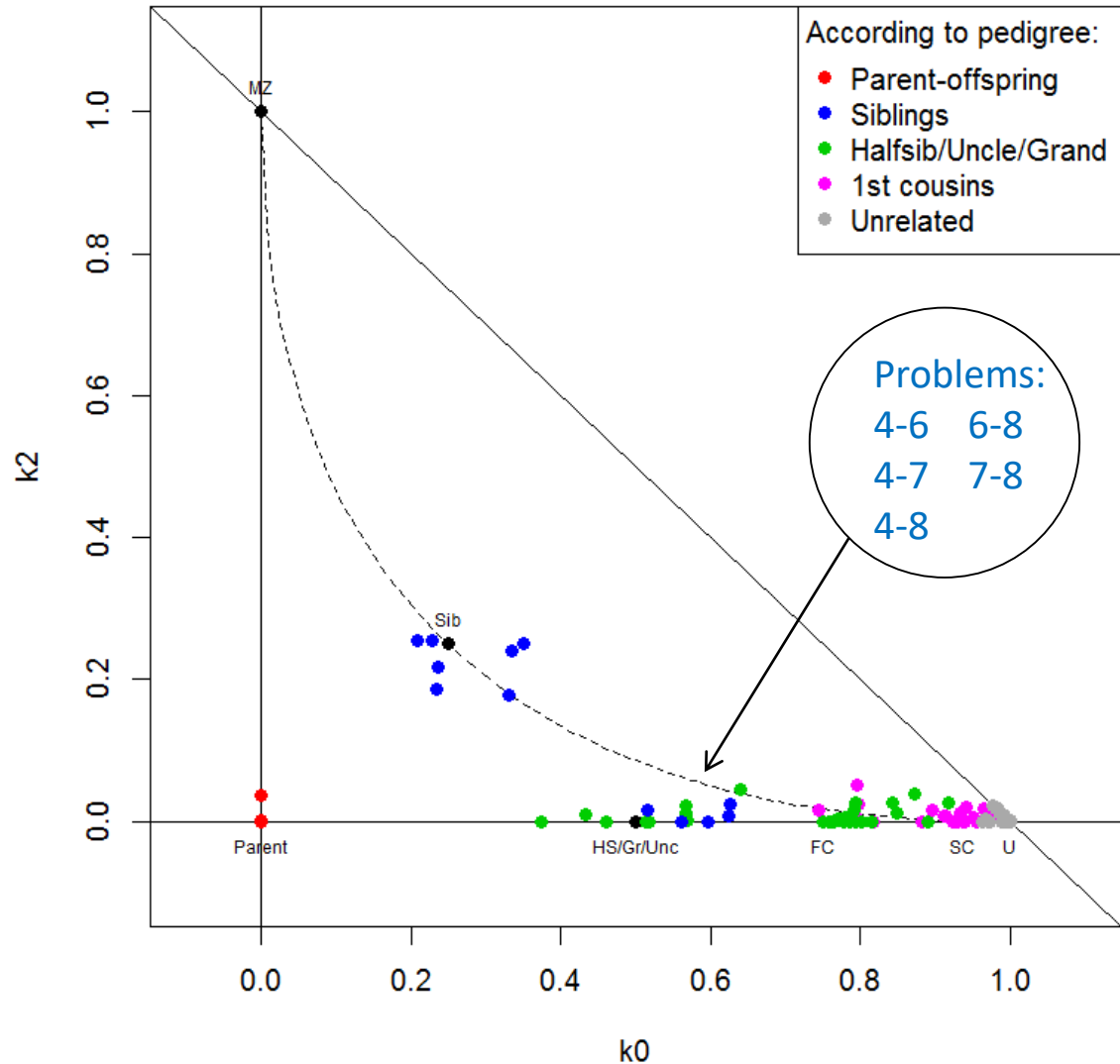
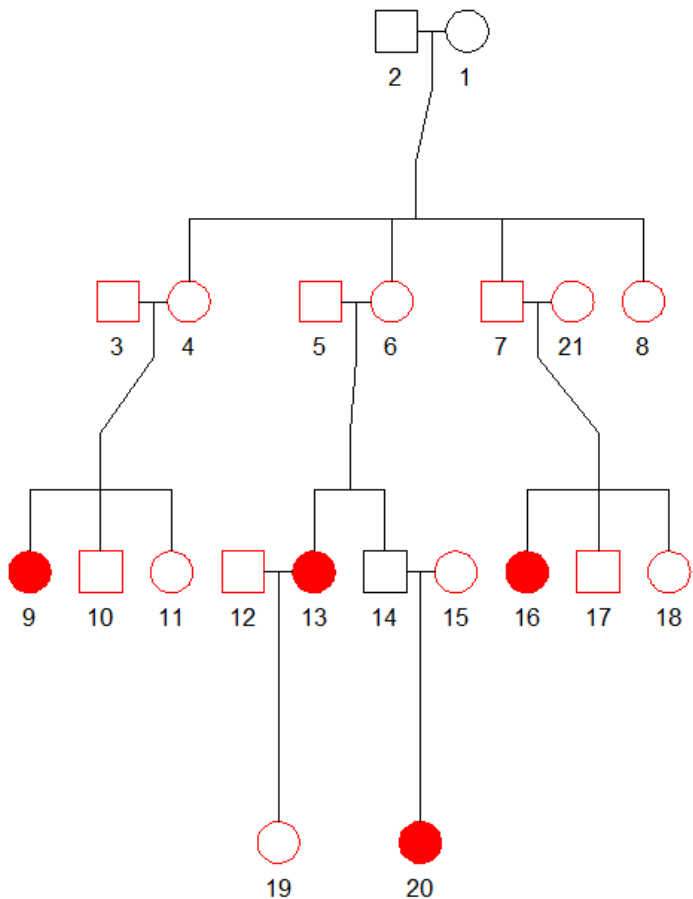
Pedigree 32 - OK

Family 32



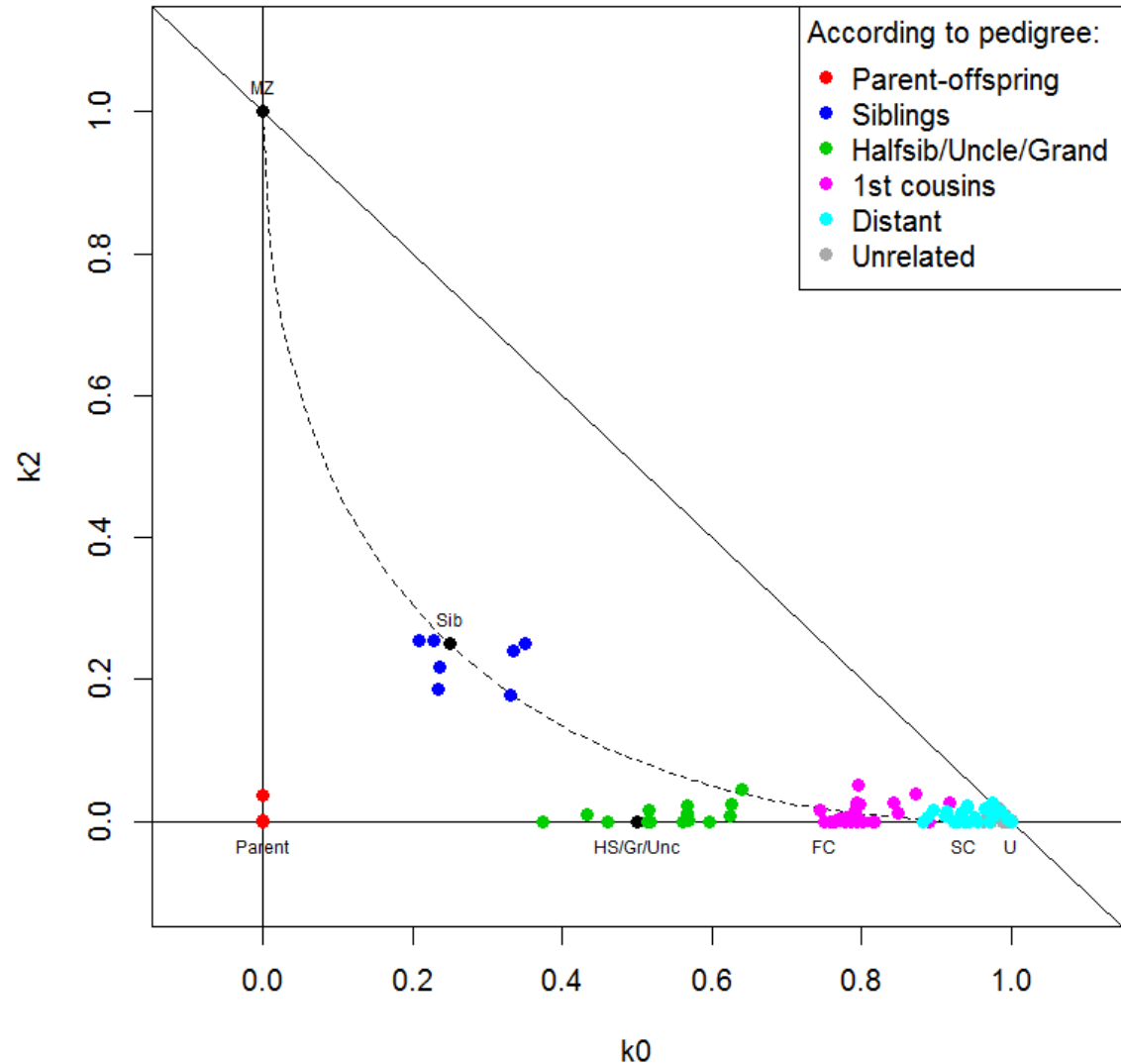
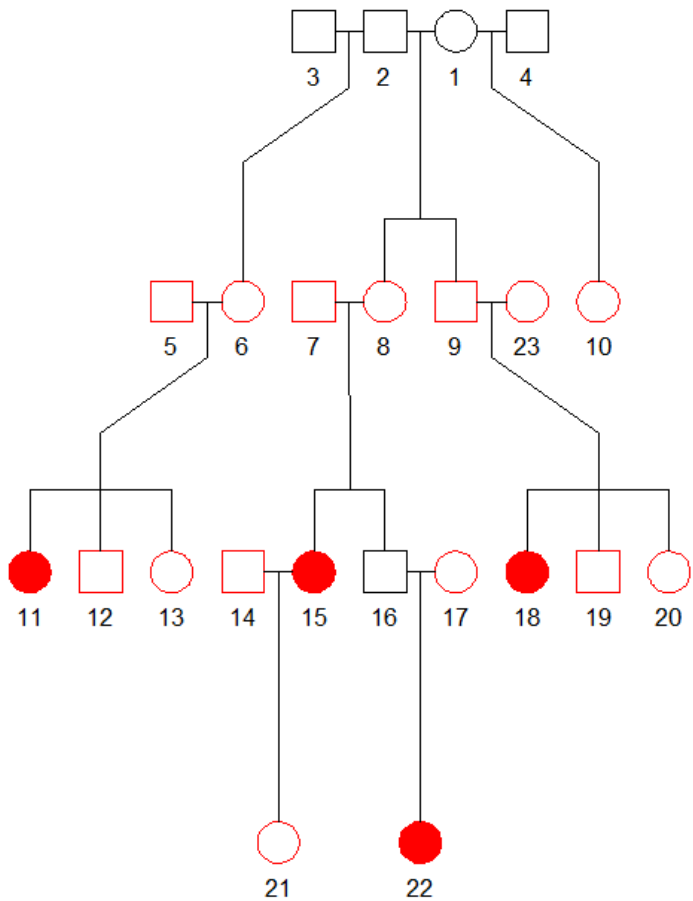
Pedigree 16...something's wrong

Family 16



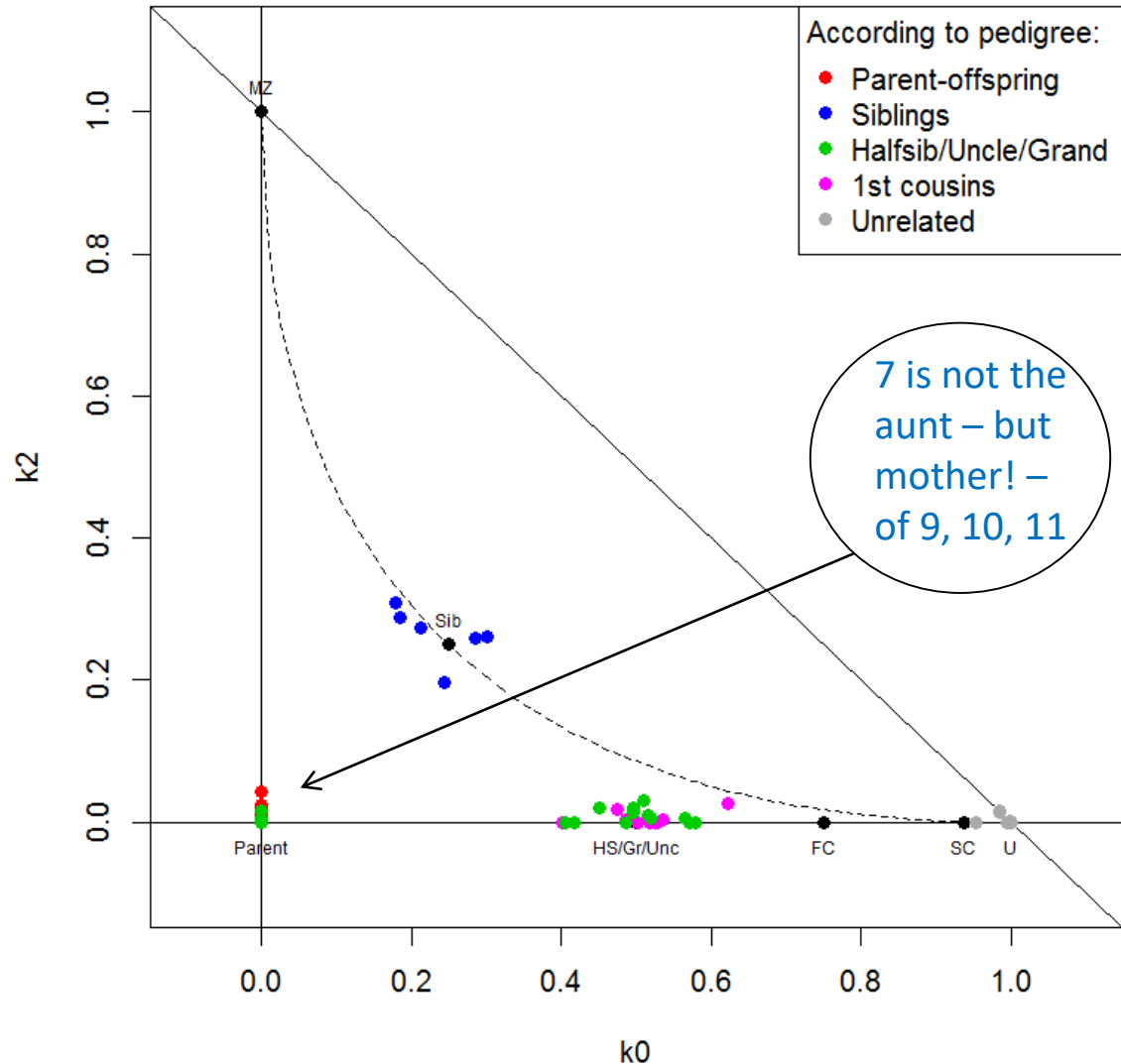
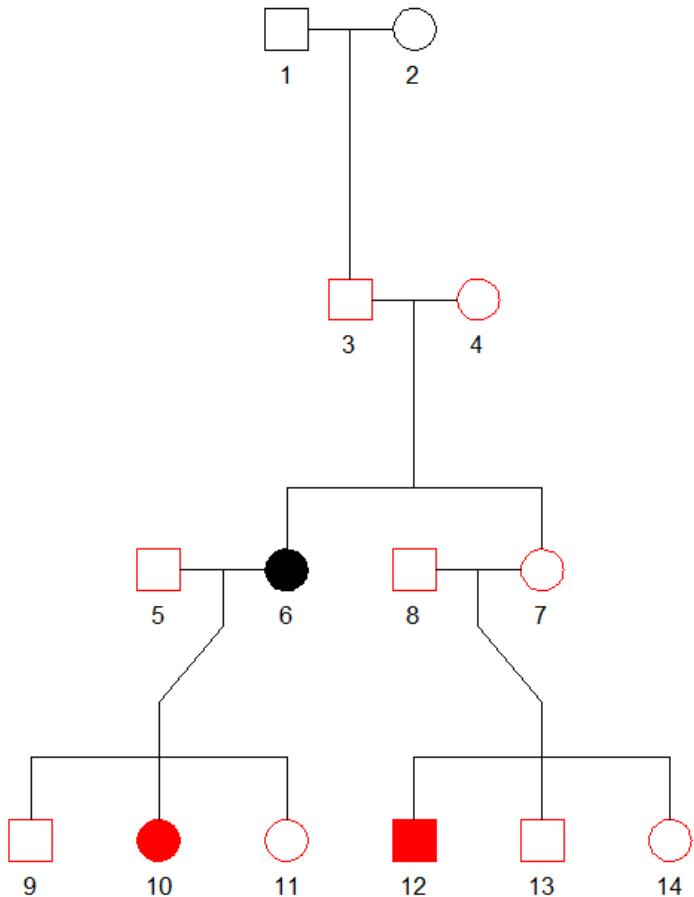
Pedigree 16 - corrected

Family 16



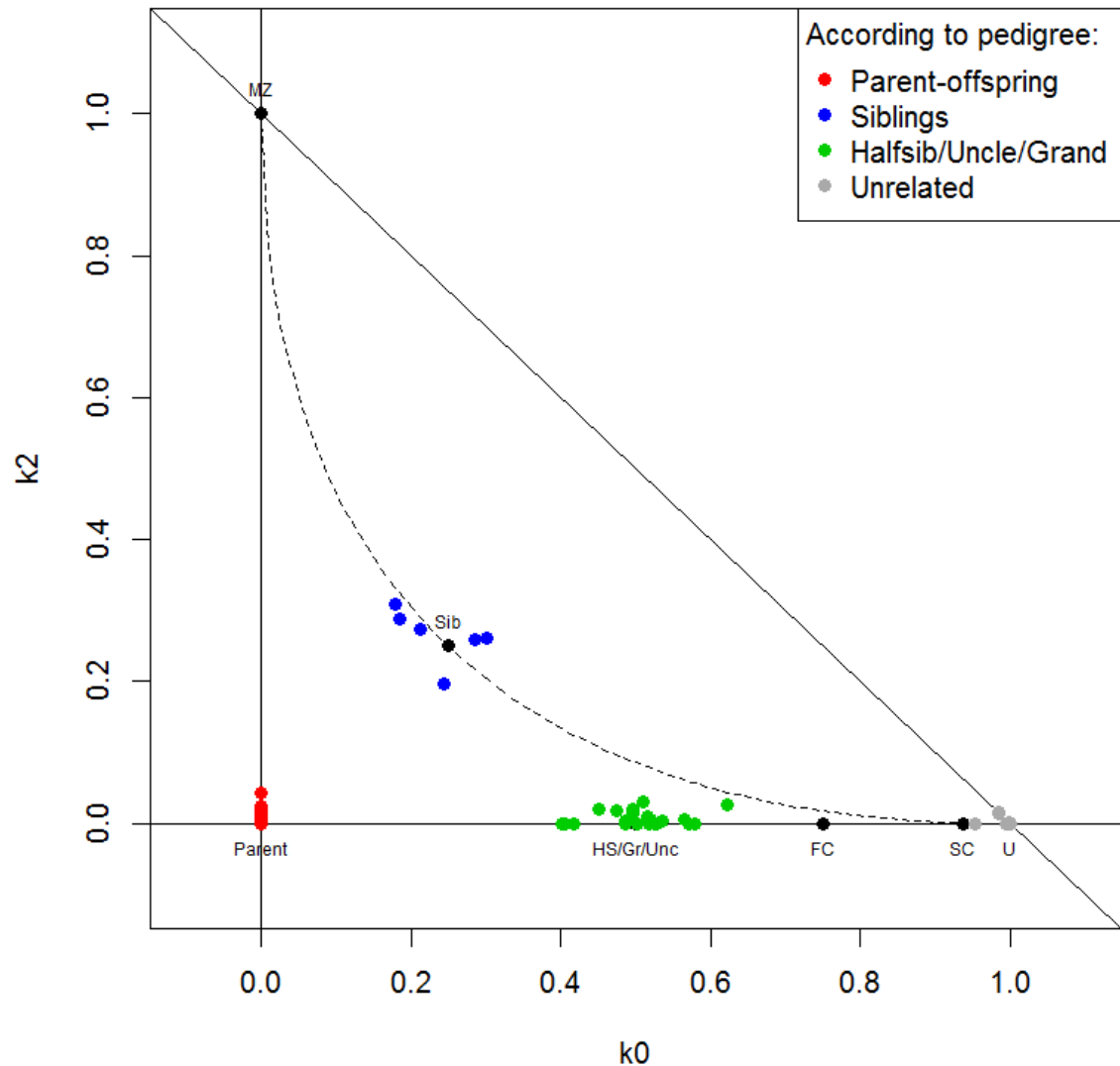
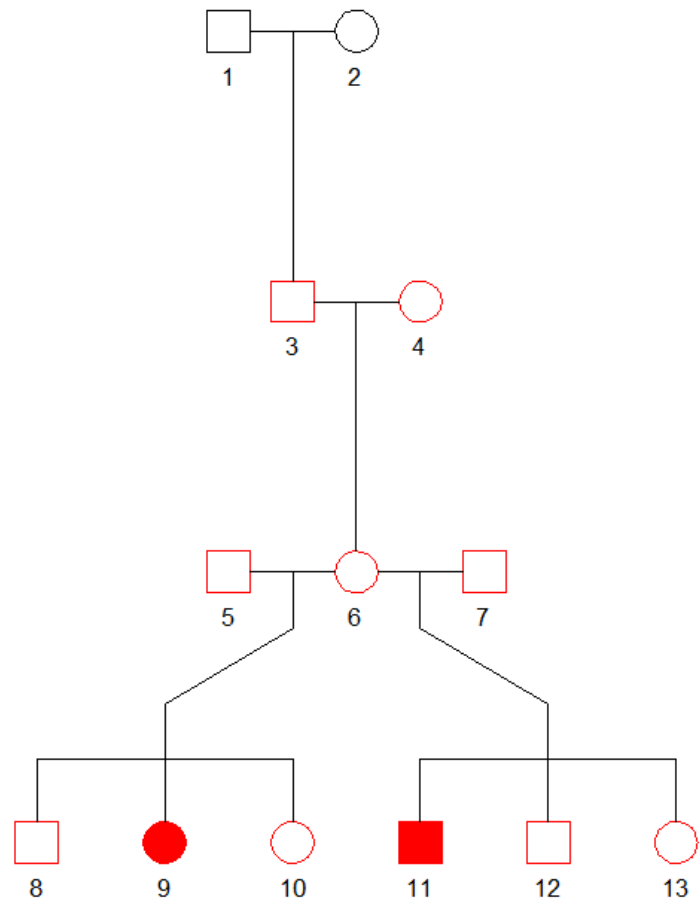
Pedigree 41...another error!?

Family 41

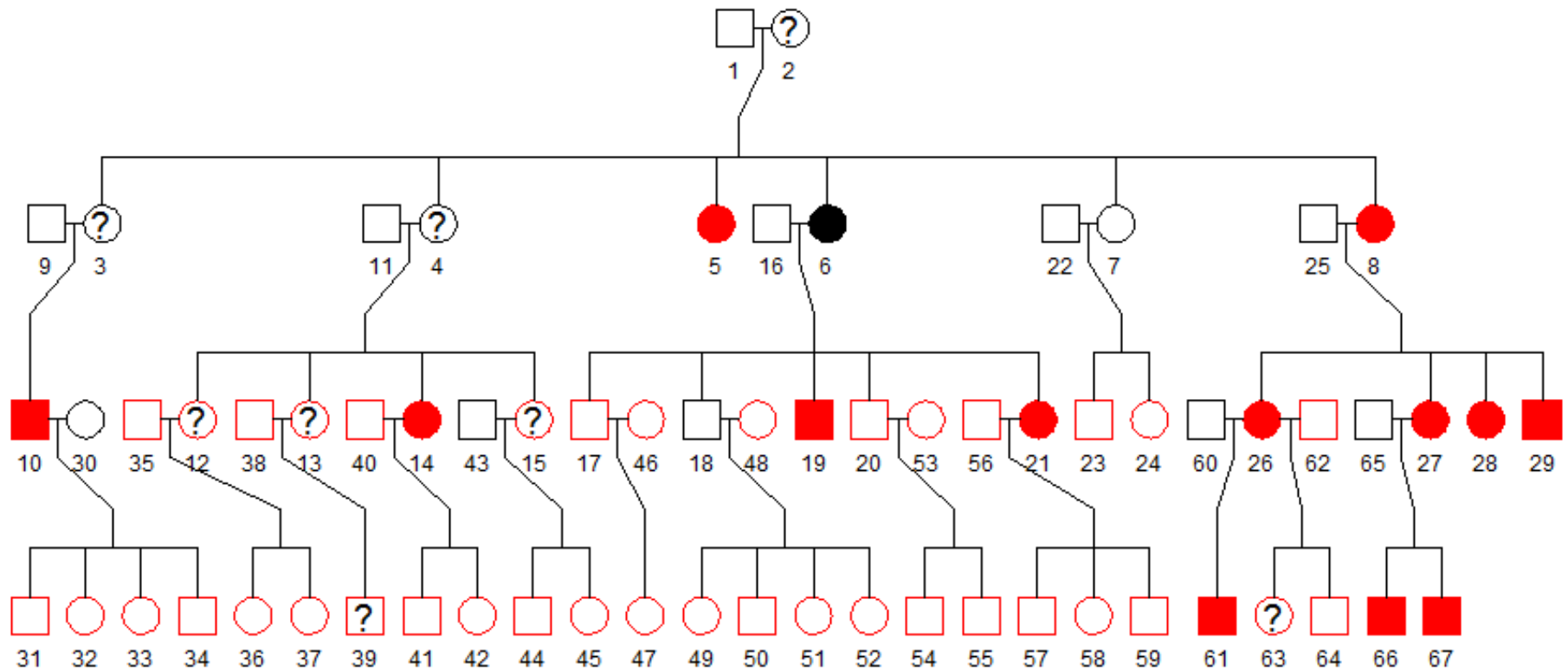


Example: Pedigree 41 - corrected

Family 41

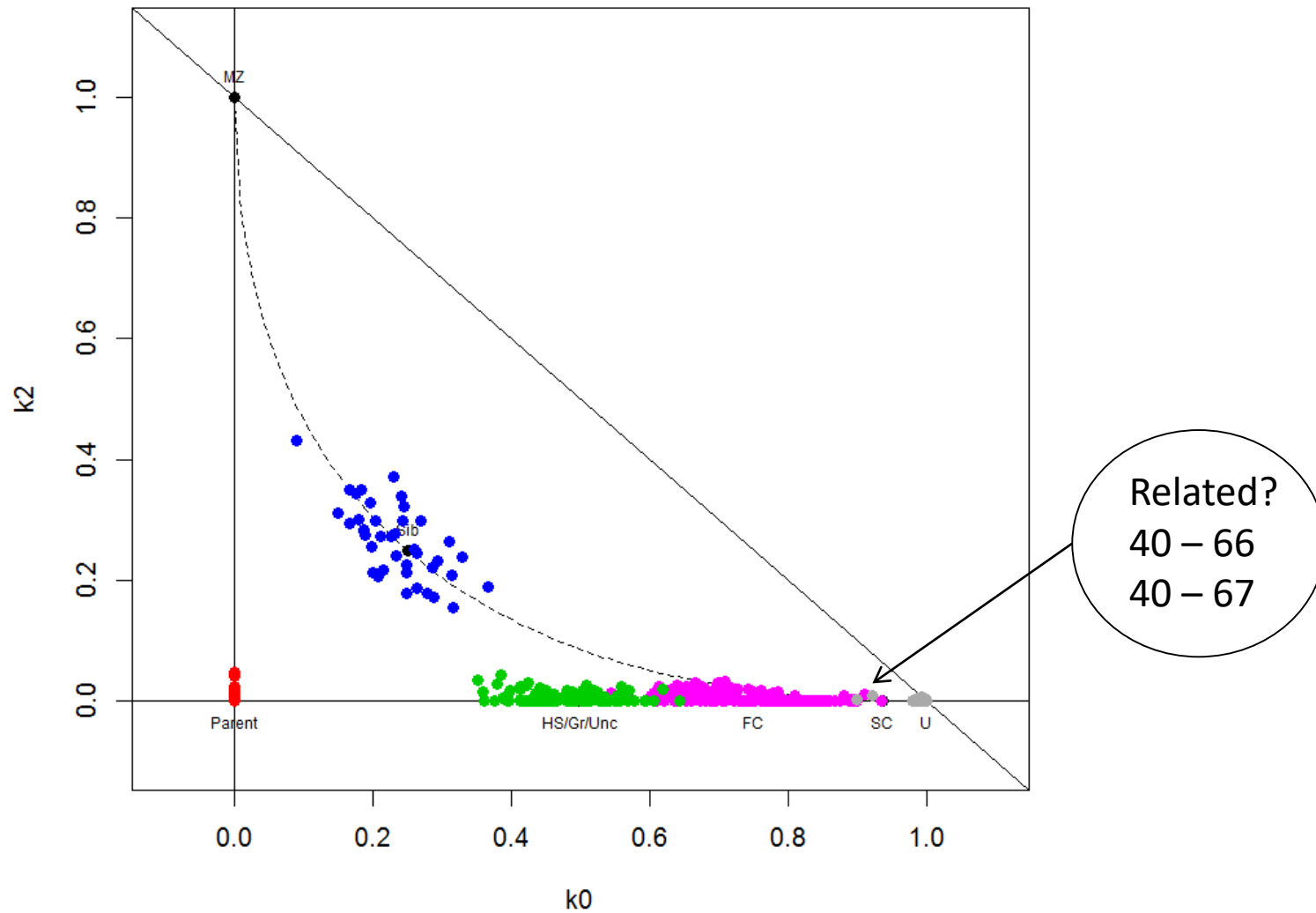


A large pedigree



Genotyped for linkage analysis: Affymetrix 50k array

Results



Potential problems

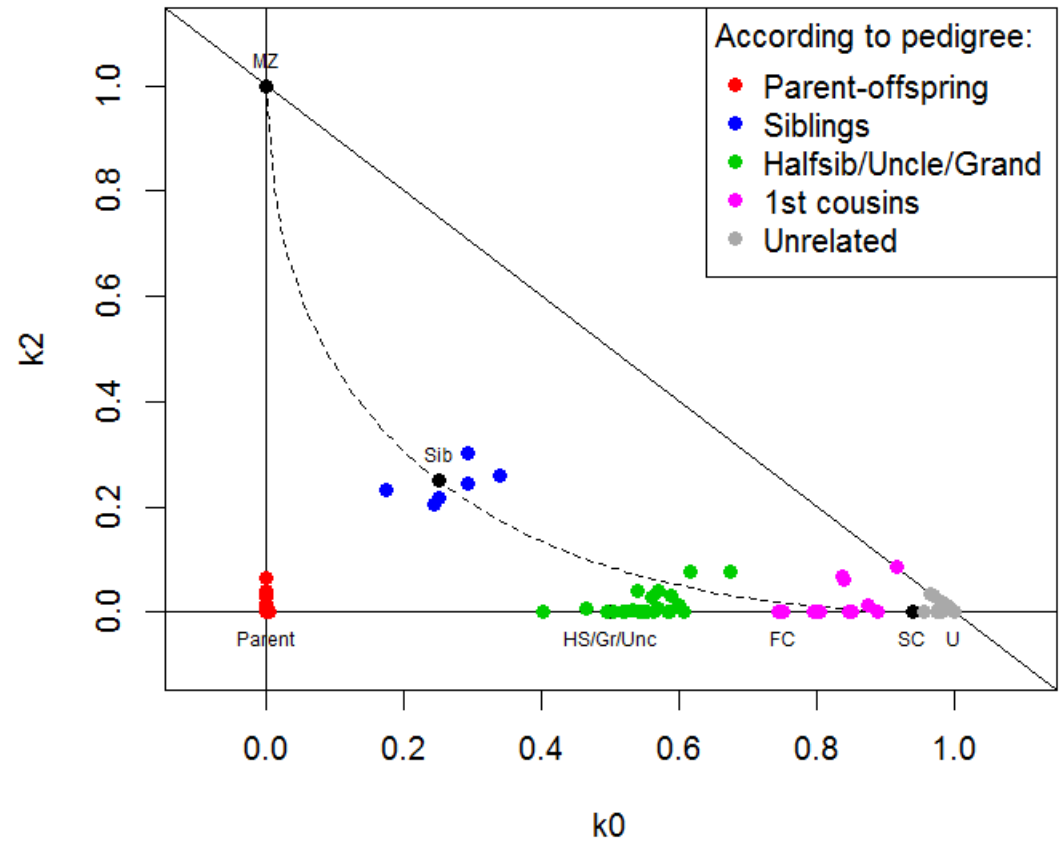
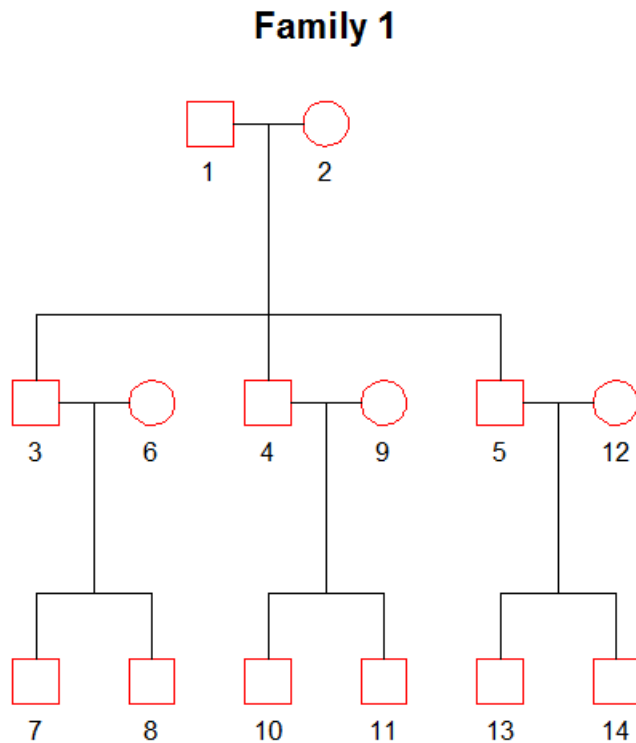
- The maximum likelihood calculation is sensitive to
 - number of markers
 - correct allele frequencies
- Use simulation to investigating the impact of these

Simulation example

SNPs: 1 000

True frequency distr: Unif(0,1)

Frequencies used: Correct

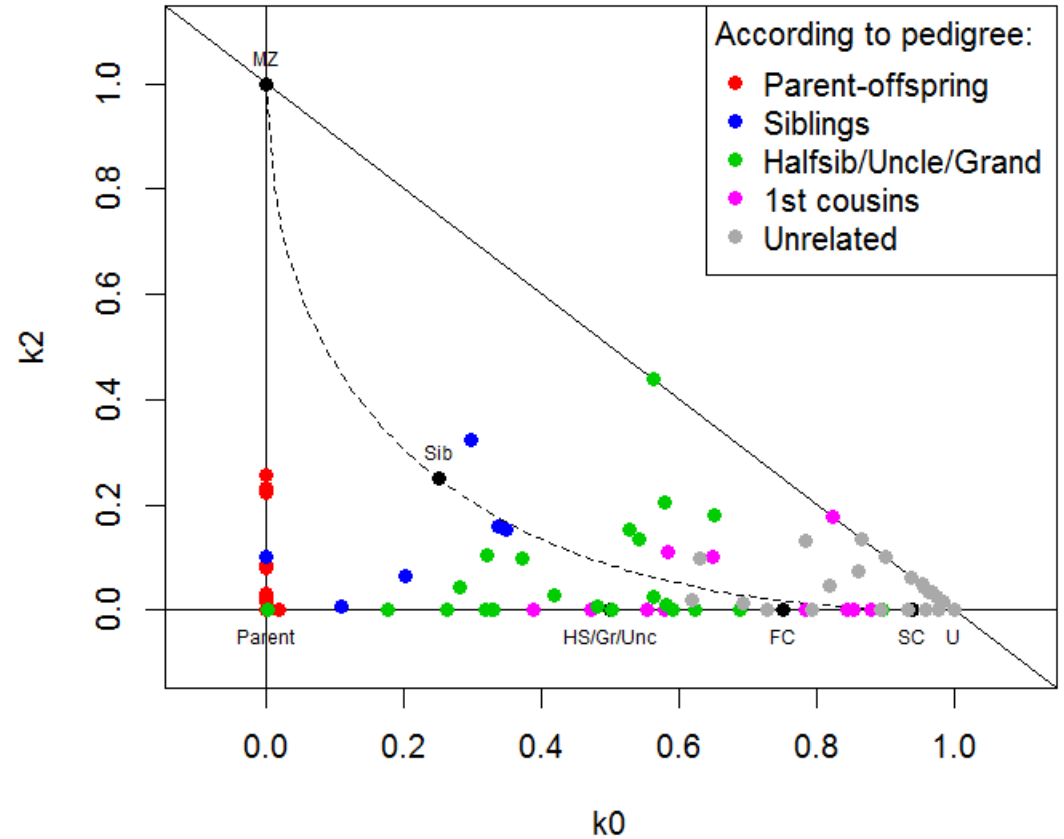
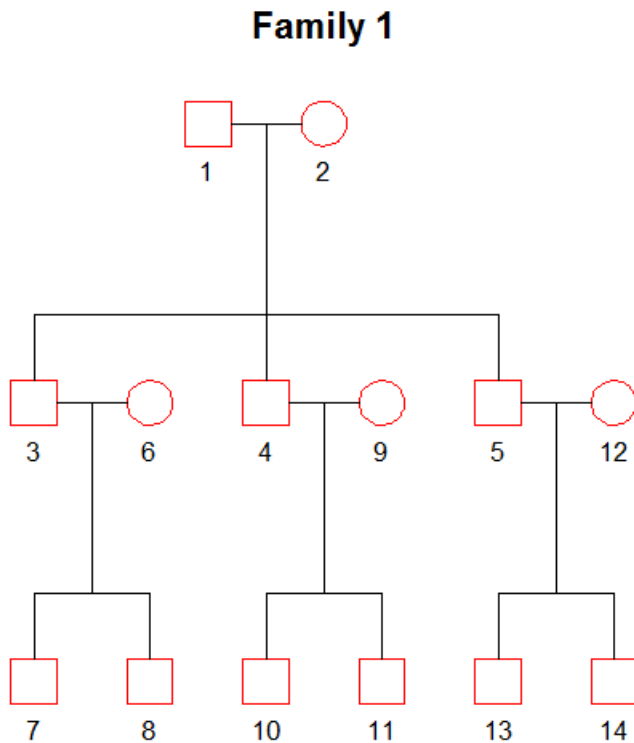


Simulation example

SNPs: 100

True frequency distr: Unif(0,1)

Frequencies used: Correct

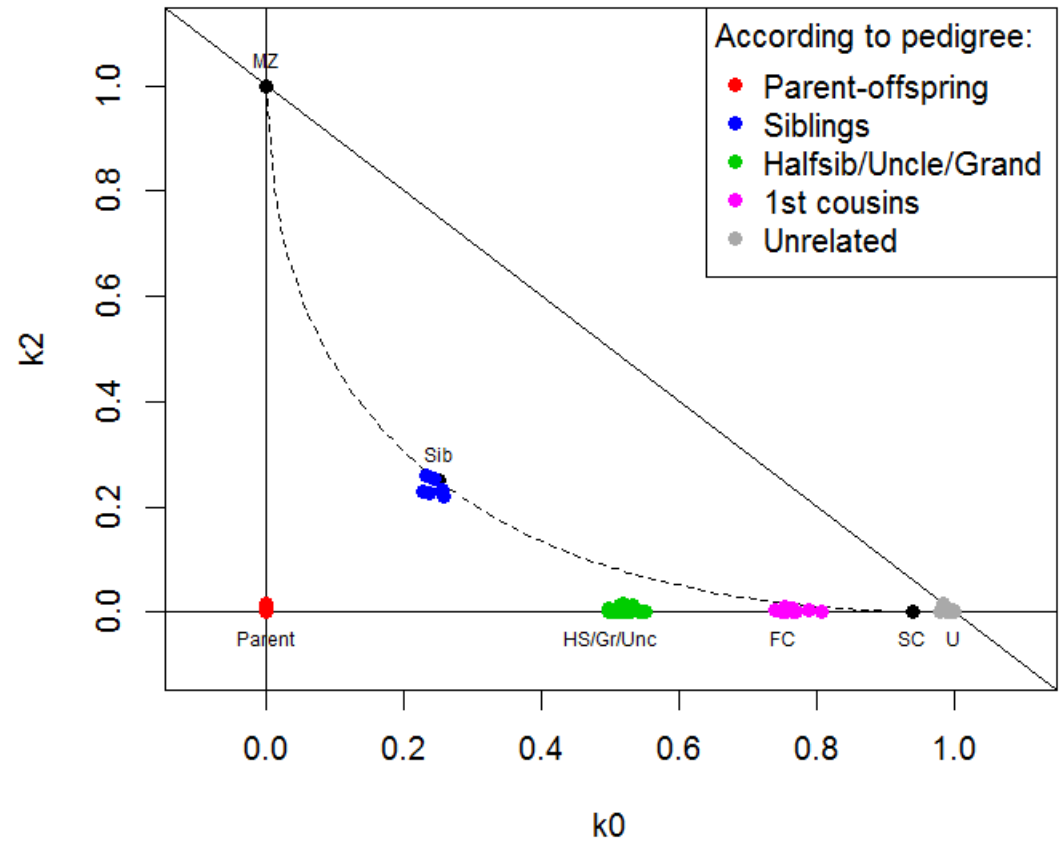
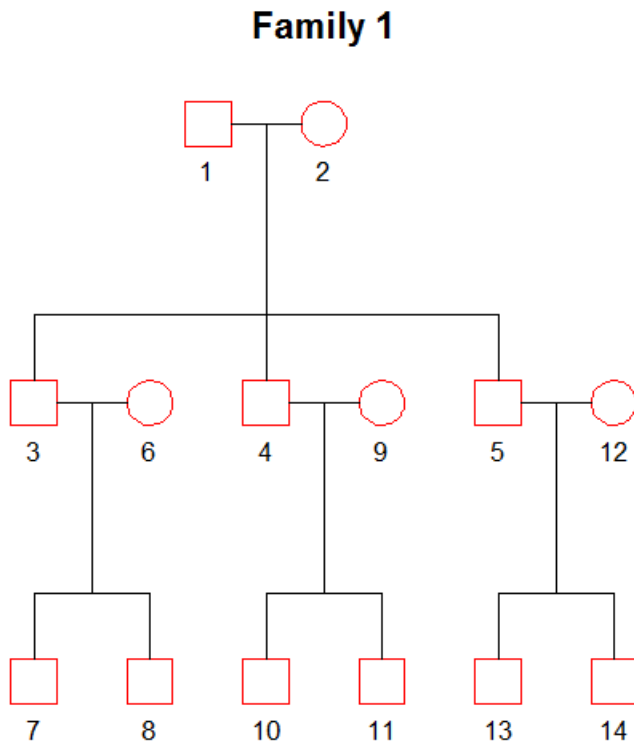


Simulation example

SNPs: 10 000

True frequency distr: Unif(0,1)

Frequencies used: Correct

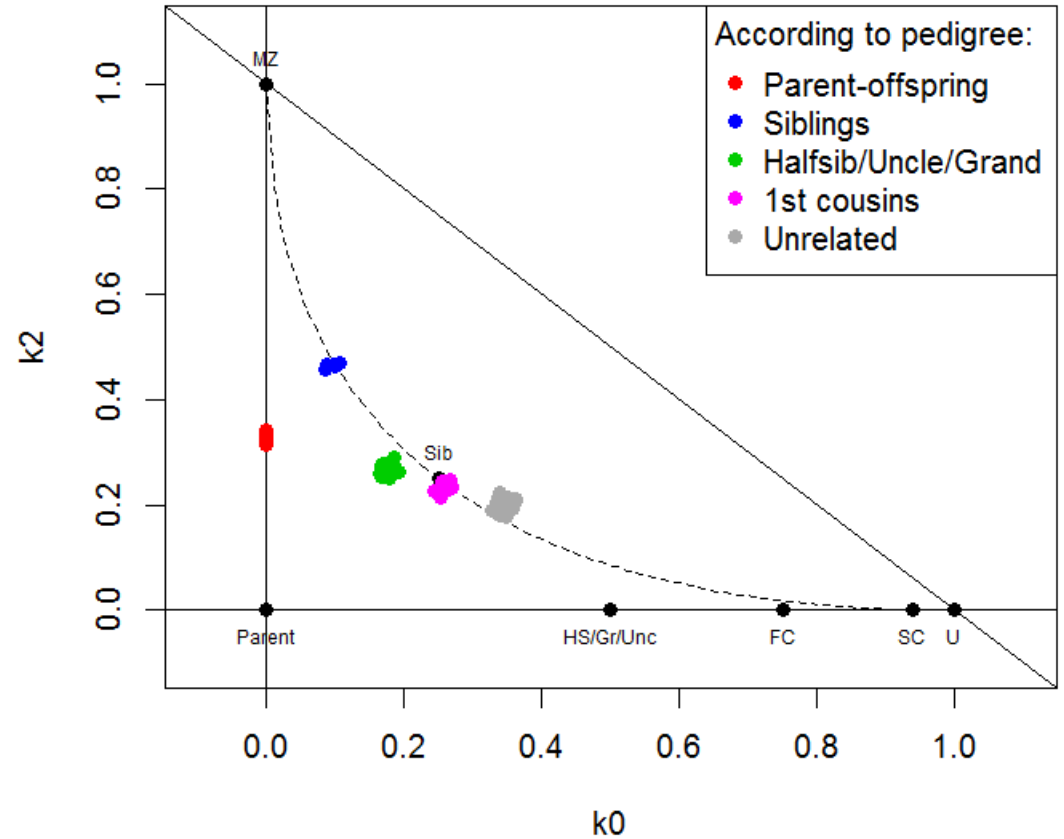
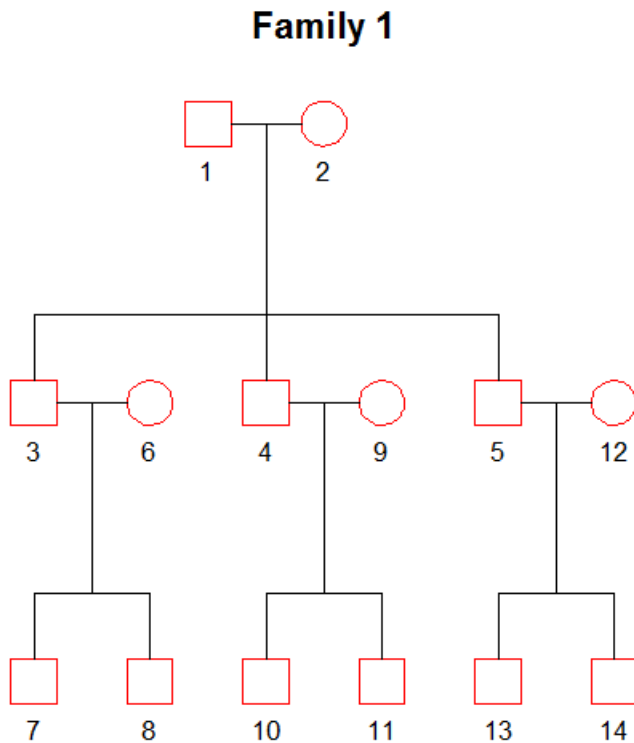


Simulation example

SNPs: 10 000

True frequency distr: Unif(0,1)

Frequencies used: All = 0.5

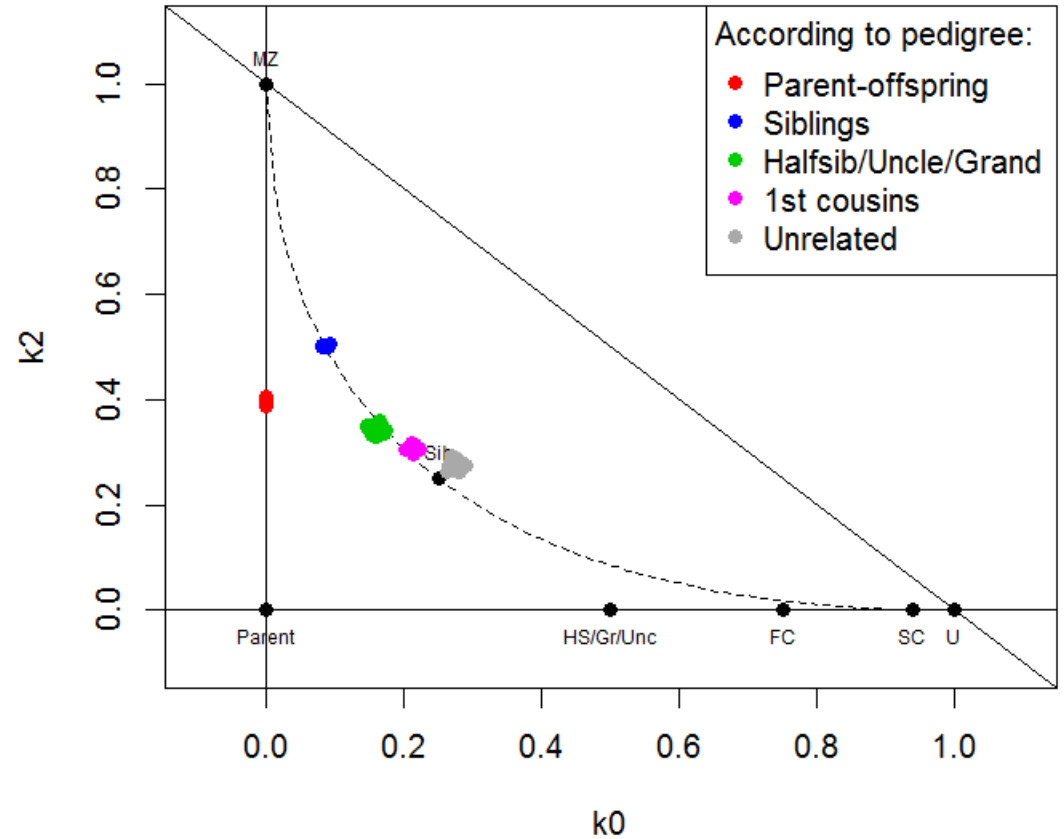
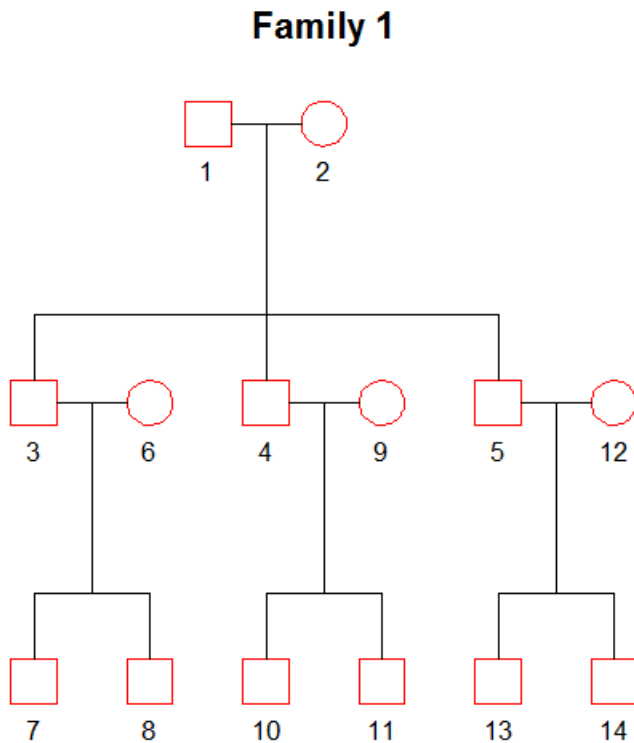


Simulation example

SNPs: 10 000

True frequency distr: Unif(0,1)

Frequencies used: Unif(0,1)

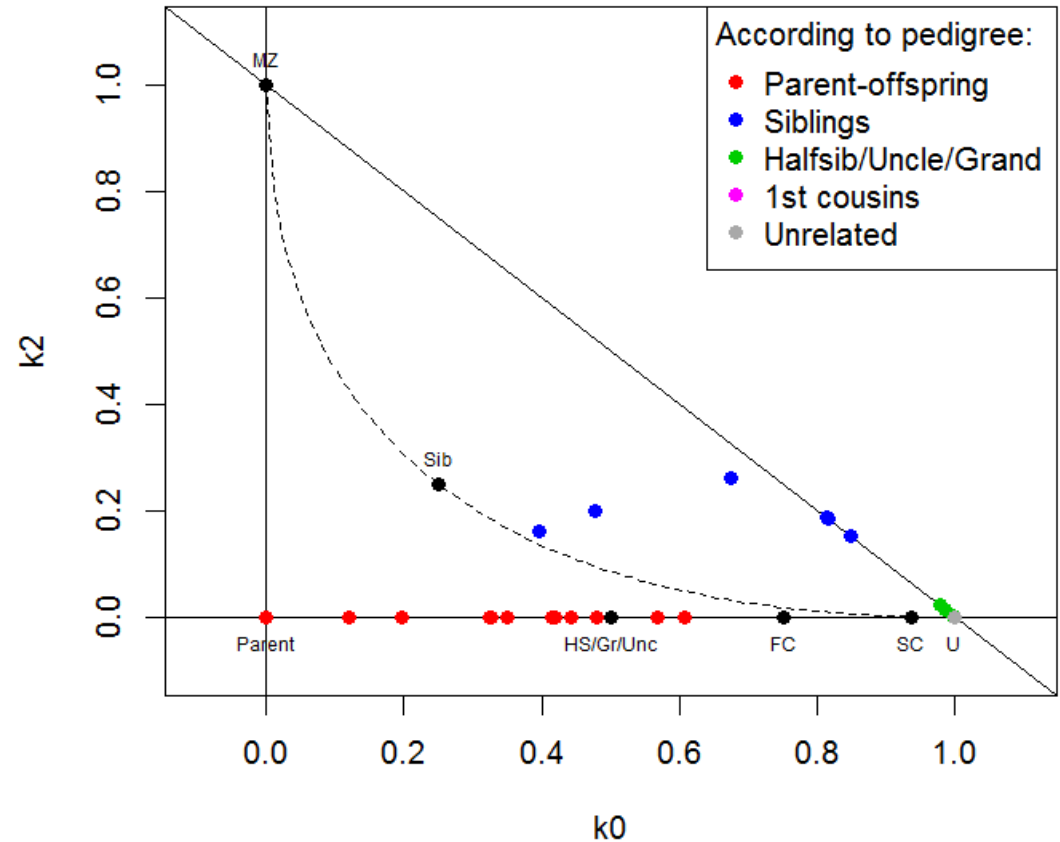
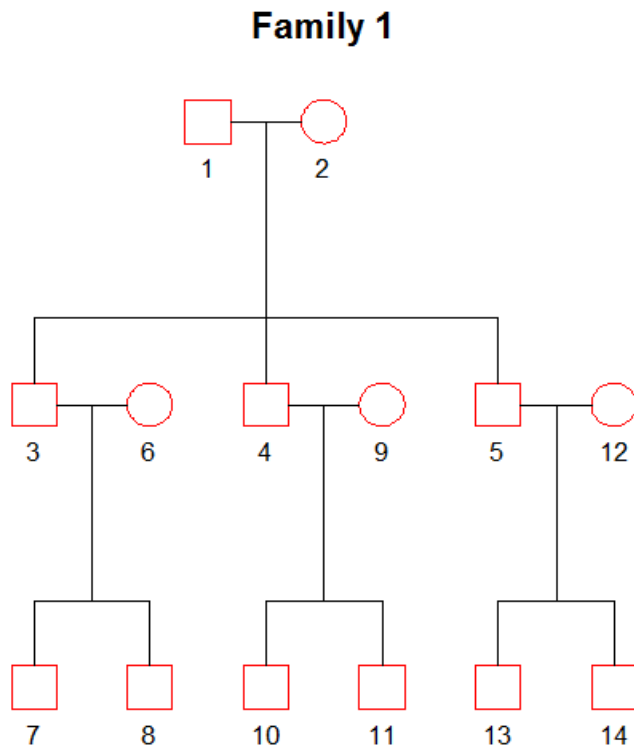


Simulation example

SNPs: 10 000

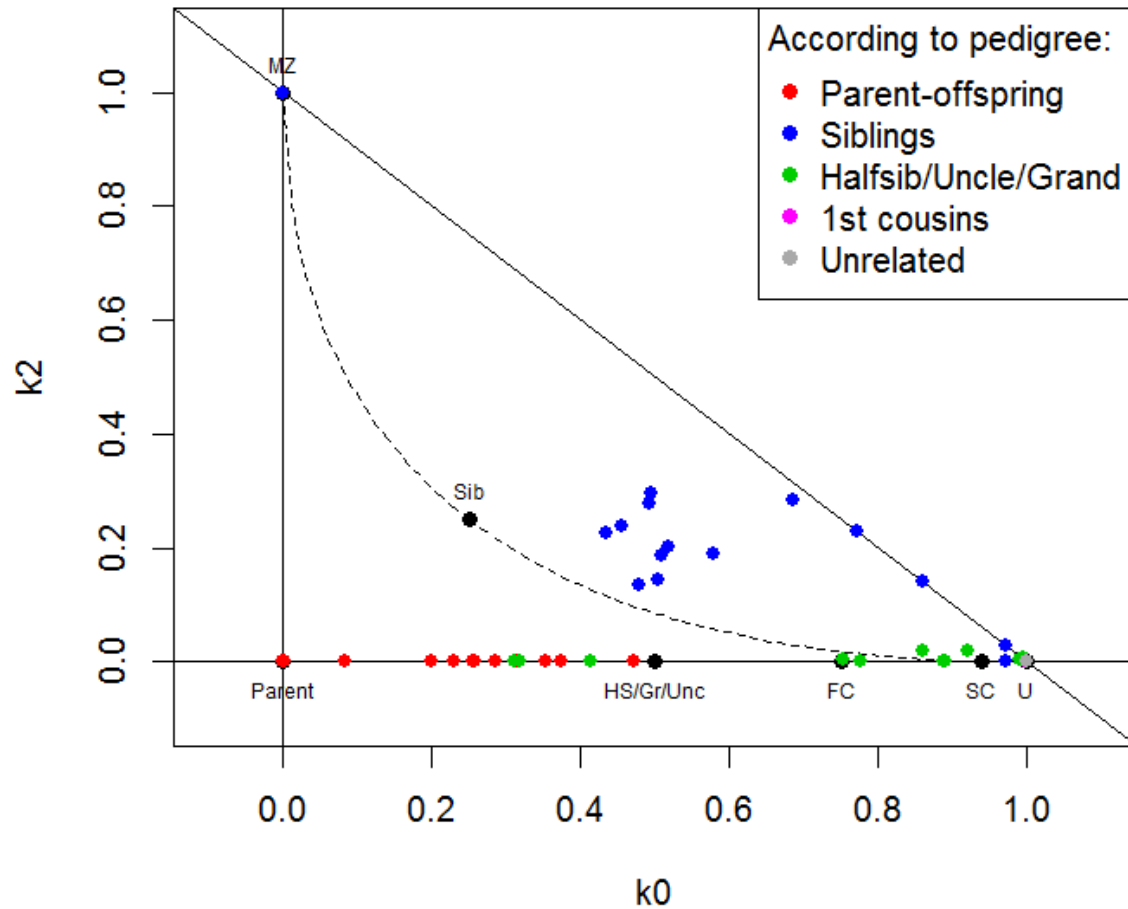
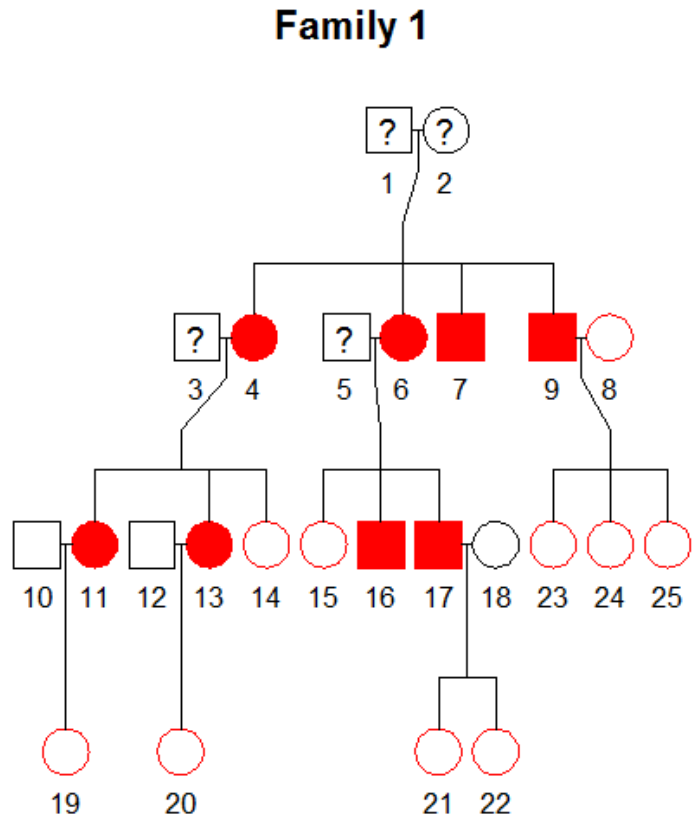
True frequency distr: Unif(0,1)

Frequencies used: Family estimate



- These simulations are good to have in mind when encountering weird results

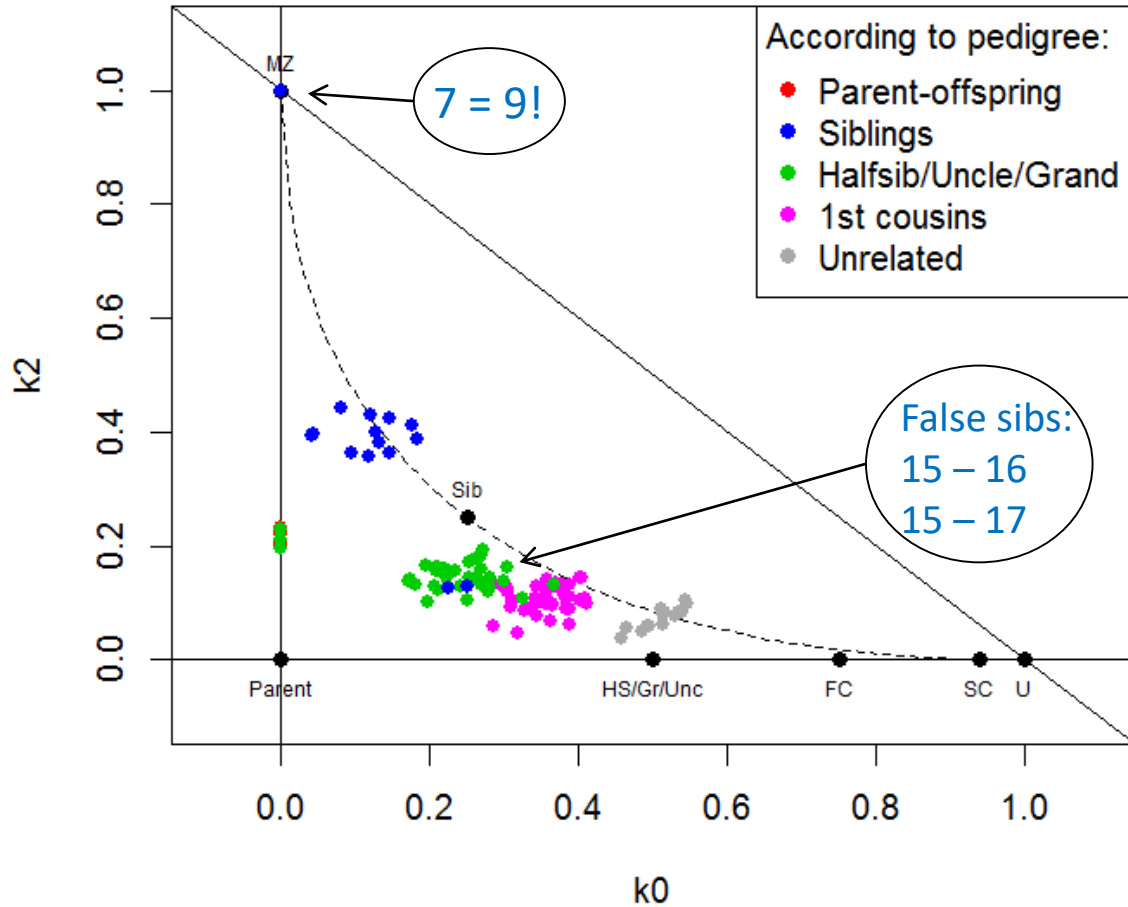
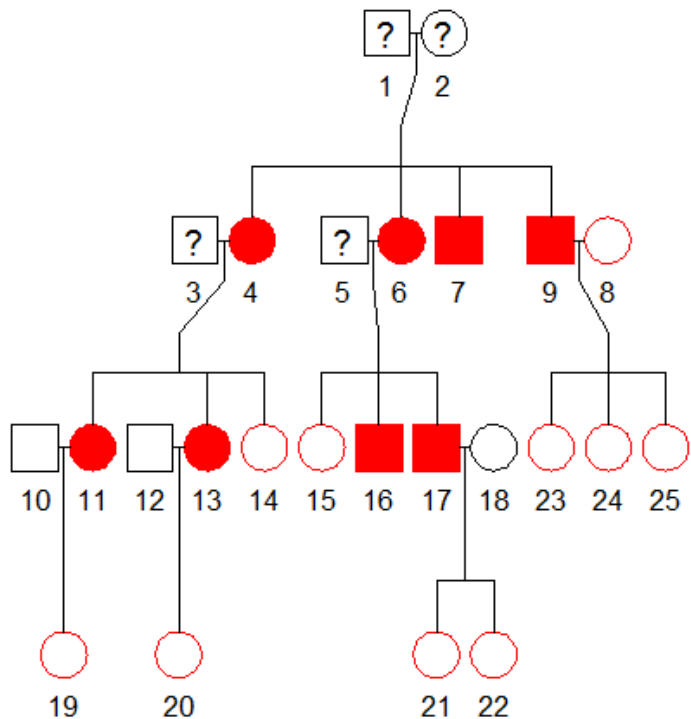
Example: Pedigree L2875 (hypercholesterolemia)



Genotyped with 2000 SNPs

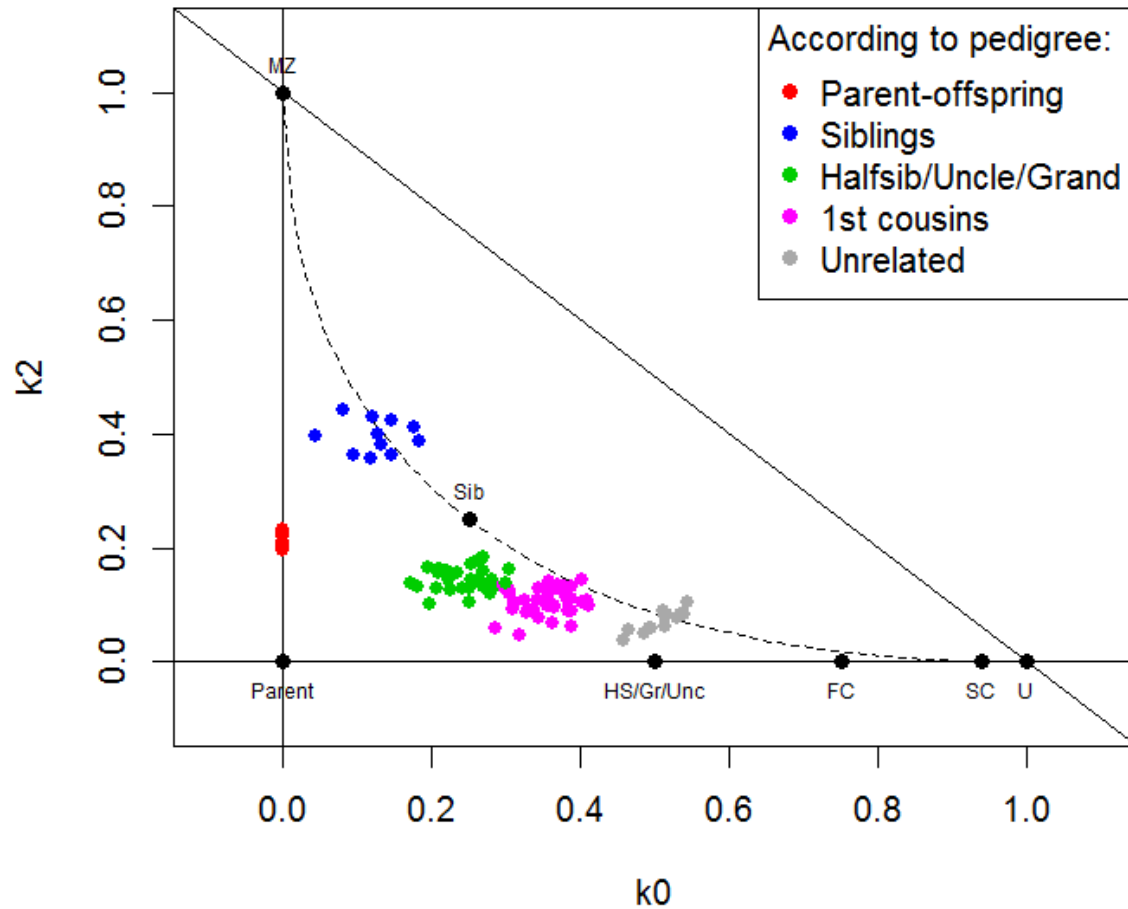
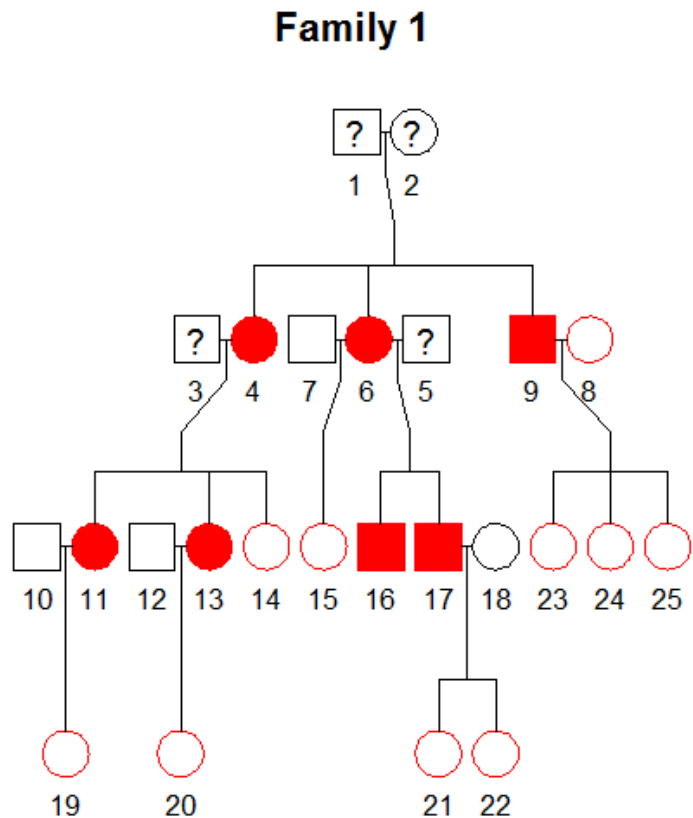
Example: Pedigree L2875 (hypercholesterolemia)

Family 1



Assuming equiprequent SNPs

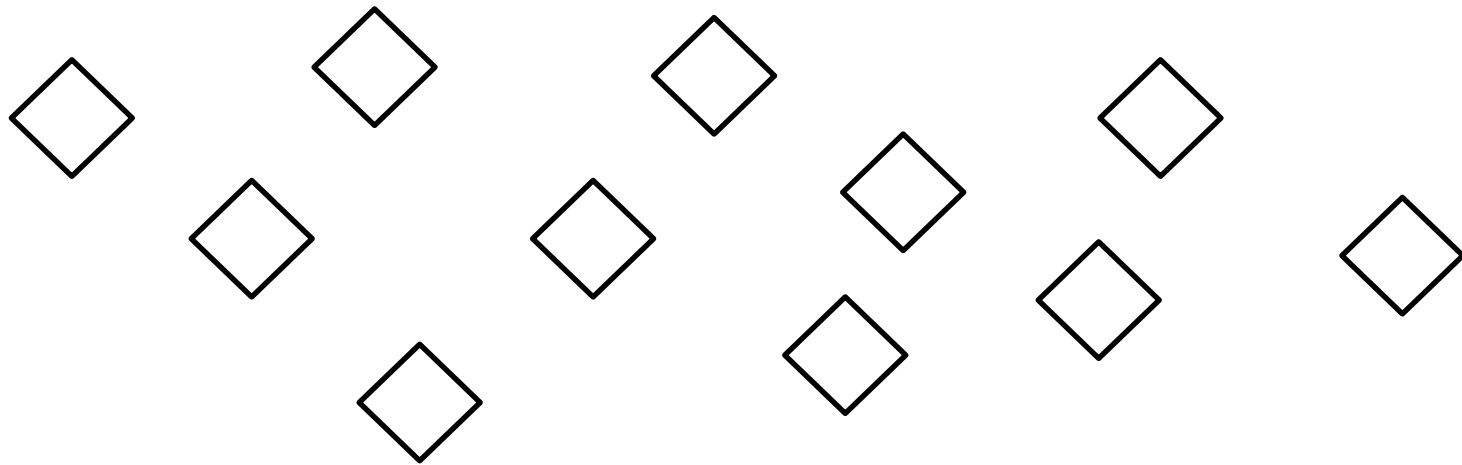
Example: Pedigree L2875 - corrected



Pedigree reconstruction

Goal:

Reconstruct the complete pedigree from genotype data



- Step 1: Genders
- Step 2: Estimate pairwise relationships
 - Connect parent-child
 - Exploit siblings
- Step 3: Solve the puzzle

Alternative method: R/pedbuildr

Idea:

- Generate a list of *all possible* pedigrees connecting the individuals
- Compute the likelihood of each pedigree
- Sort and output the best pedigrees