

Lecture 6:

Inference of pairwise relatedness

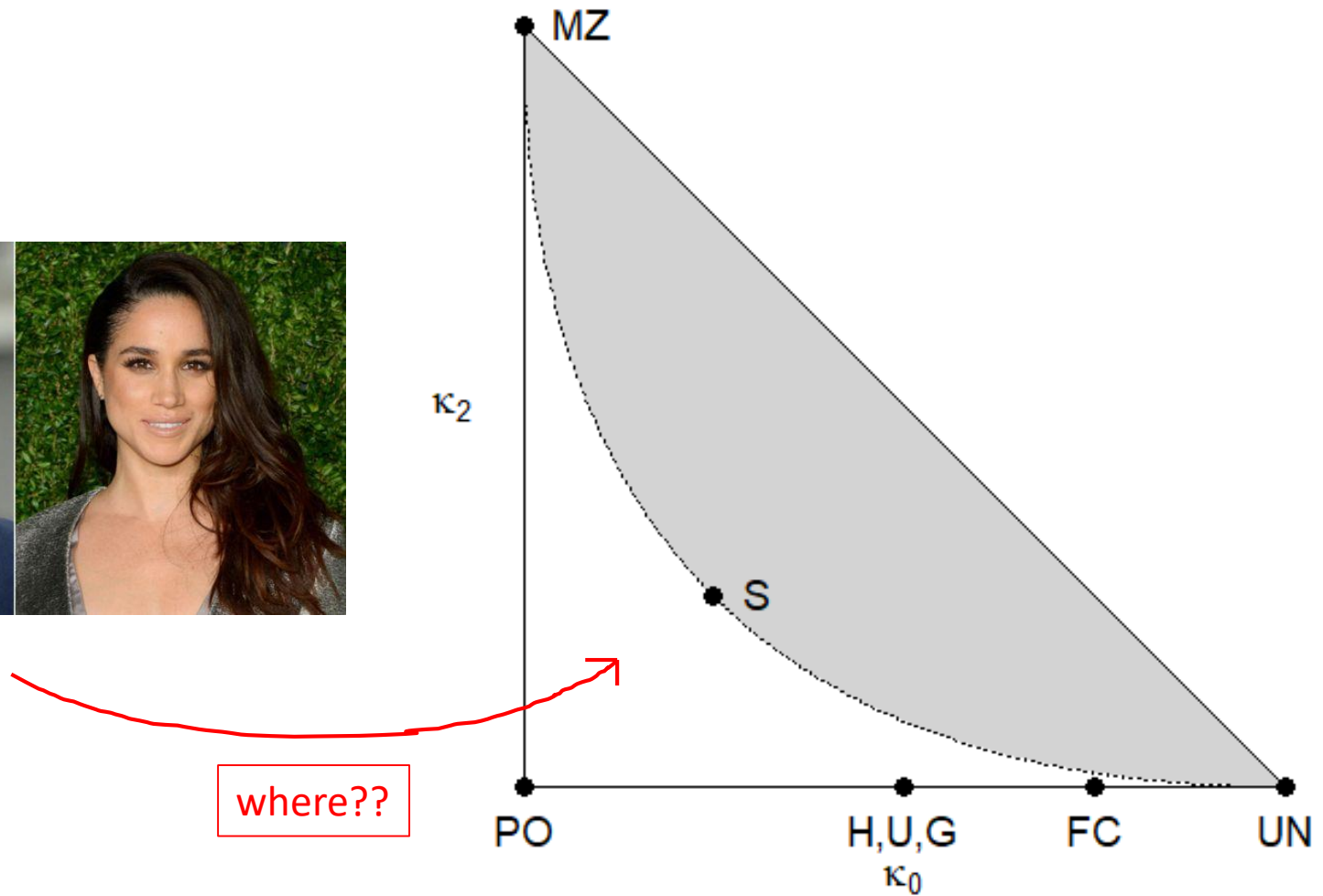
Pedigree reconstruction

Statistical methods in genetic relatedness and pedigree analysis

NORBIS course, 13th – 17th of June 2022, Oslo

Magnus Dehli Vigeland

Part 1: Pairwise inference



Two main approaches to relatedness inference

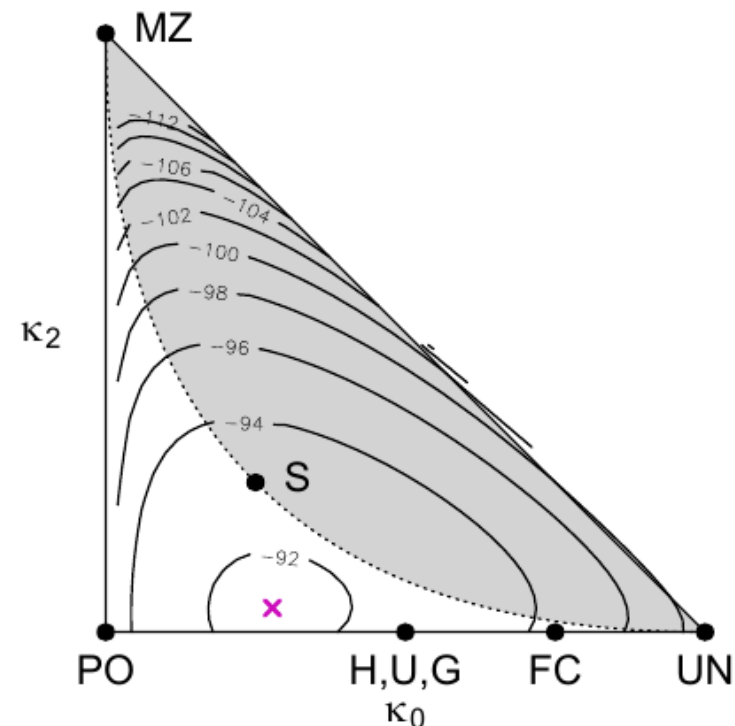
- 1) Maximum likelihood estimation from marker data
- 2) Classification based on IBD segments
(sequence -> detect IBD -> classify)

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



$G_1 = a/b$ $G_2 = a/a$

- 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*

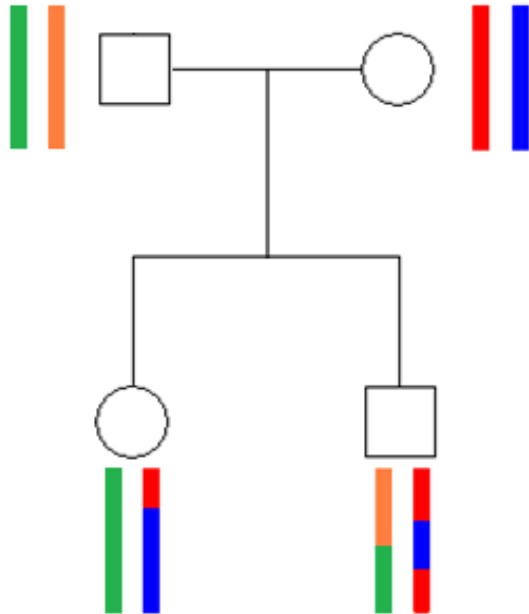
$$\begin{aligned} L(\kappa) &= P(G_1, G_2 \mid \kappa) \\ &= P(G_1, G_2 \mid UN) \kappa_0 + P(G_1, G_2 \mid PO) \kappa_1 + P(G_1, G_2 \mid MZ) \kappa_2 \end{aligned}$$

UN = unrelated
PO = parent/offspr
MZ = monozygotic

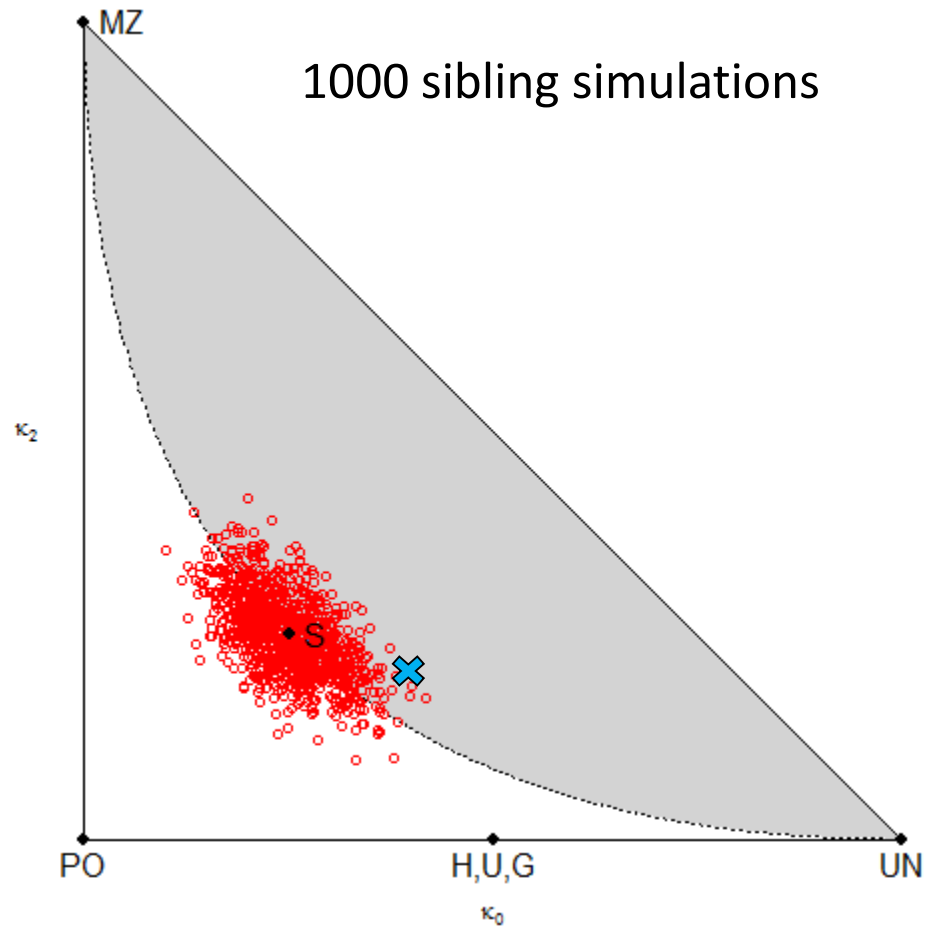
- With several independent markers:

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

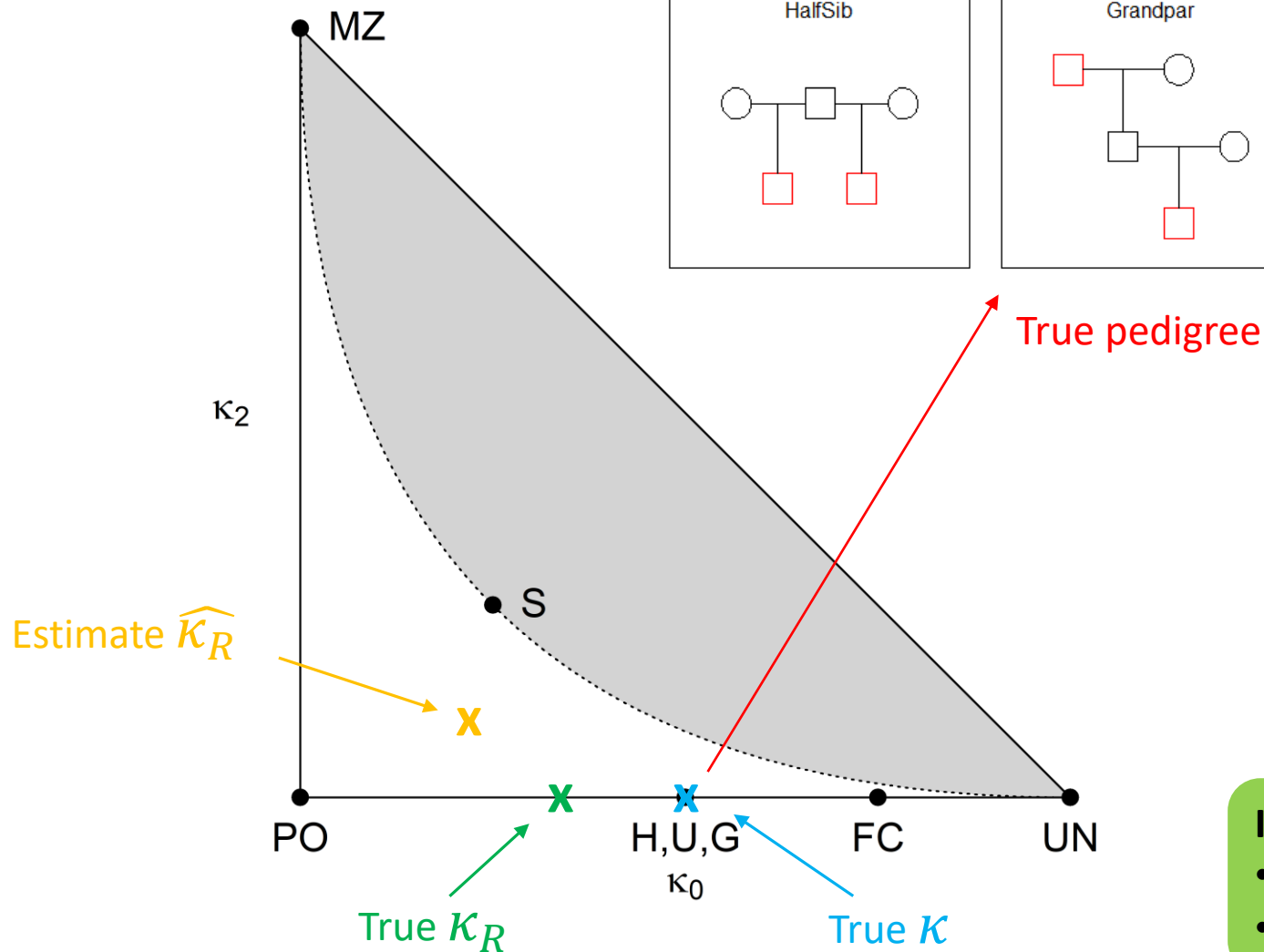
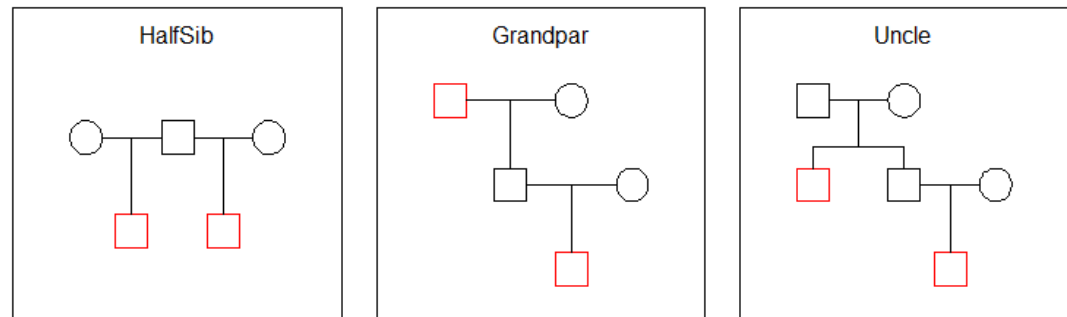
What are we estimating?



Answer: The *realised* coefficients!



Can we recover the pedigree?



In practice

- PO usually ok
- Sibs if good data

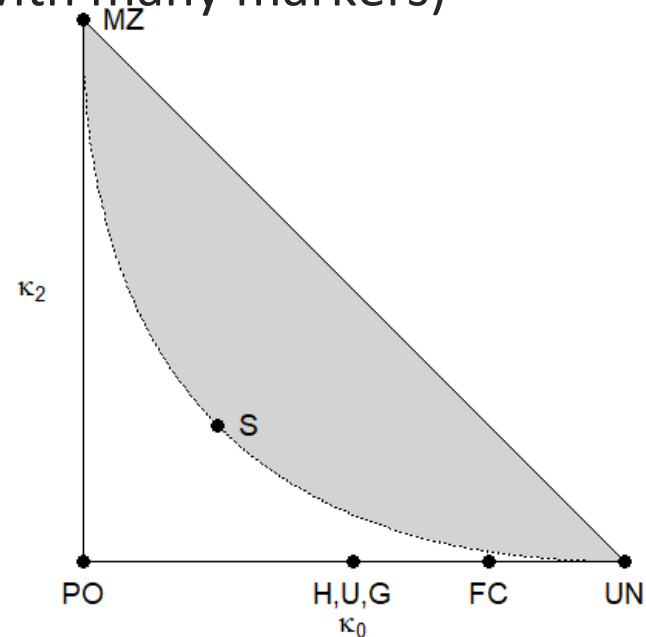
Implementations

- **R**

- pedsuite (package: **forrel**)
- GENESIS (Bioconductor)
- GWASTools (optimized for association studies)
- CrypticIBDcheck (as above, slow with many markers)

- Other

- KING
- PLINK
- Beagle
- +++



Pairwise inference with forrel



- Key functions

```
> ibdEstimate()           # estimate kappa
> showInTriangle()        # visualize!
> ibdBootstrap()          # bootstrap confidence
> checkPairwise()         # detect pedigree errors
```

- Simulation

```
> markerSim()             # iid markers
> profileSim()            # complete profiles
```

(Both of these support conditioning on known genotypes)

Pairwise inference with forrel: Example



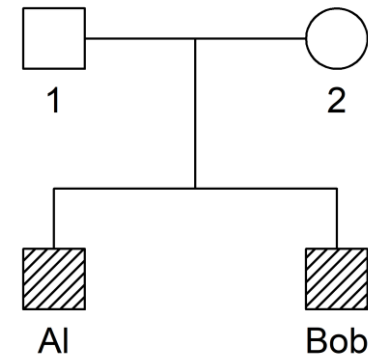
Simulate 100 SNPs for a pair of siblings

```
> ids = c("Al", "Bob")
> x = nuclearPed(children = ids)
> x = markerSim(x, N = 100, ids = ids,
               alleles = 1:2, seed = 1234)
> x
```

id	fid	mid	sex	<1>	<2>	<3>	<4>	<5>
1	*	*	1	-/-	-/-	-/-	-/-	-/-
2	*	*	2	-/-	-/-	-/-	-/-	-/-
Al	1	2	1	1/1	1/2	1/1	1/2	2/2
Bob	1	2	1	1/1	1/2	1/1	1/2	2/2

Only 5 (out of 100) markers are shown.

```
> dat = list(subset(x, "Al"),
             subset(x, "Bob"))
```

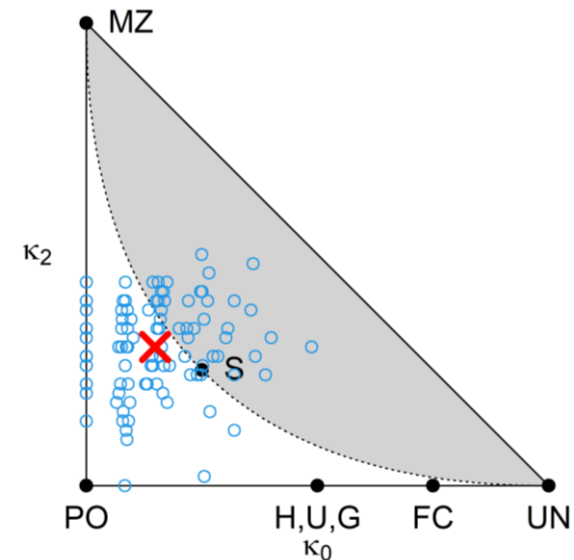
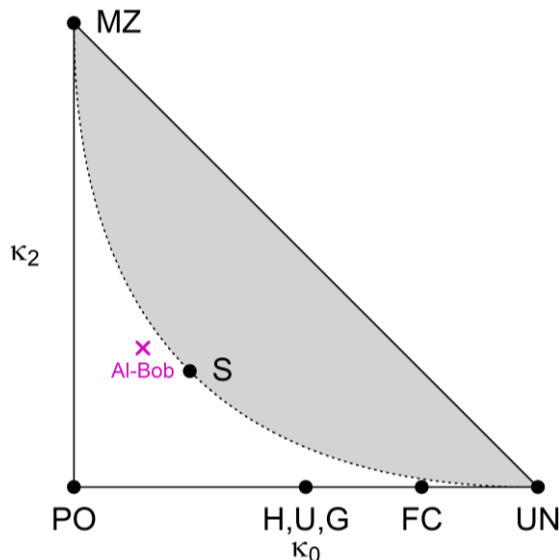
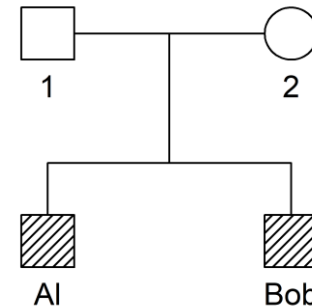


Pairwise inference with forrel: Example



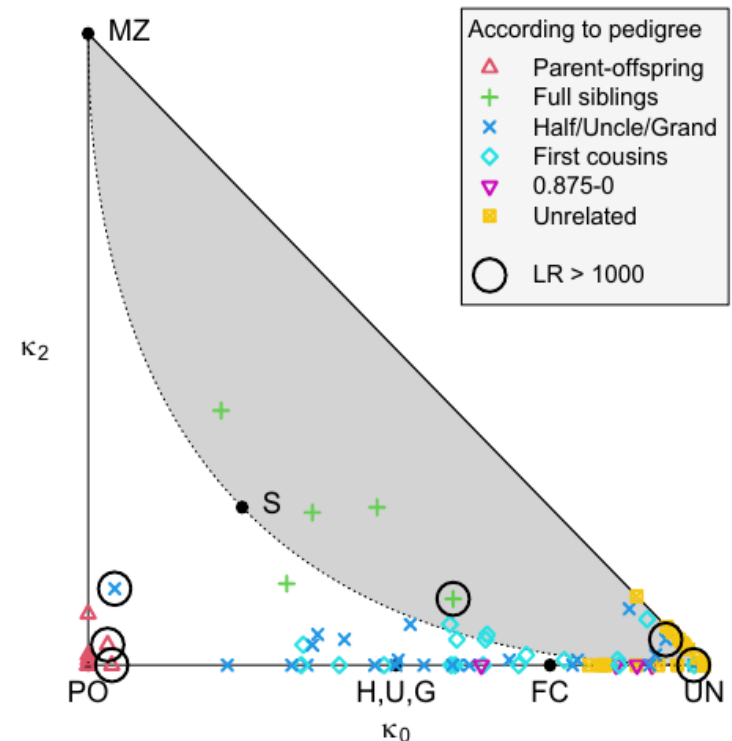
Estimate kappa from the data

```
> k = ibdEstimate(dat)
> k
  id1 id2   N    k0    k1    k2
1  Al Bob 100 0.1486 0.55139 0.30002
> showInTriangle(k, labels = T)
> bs = ibdBootstrap(dat, ids, N = 100,
  param = "kappa")
```



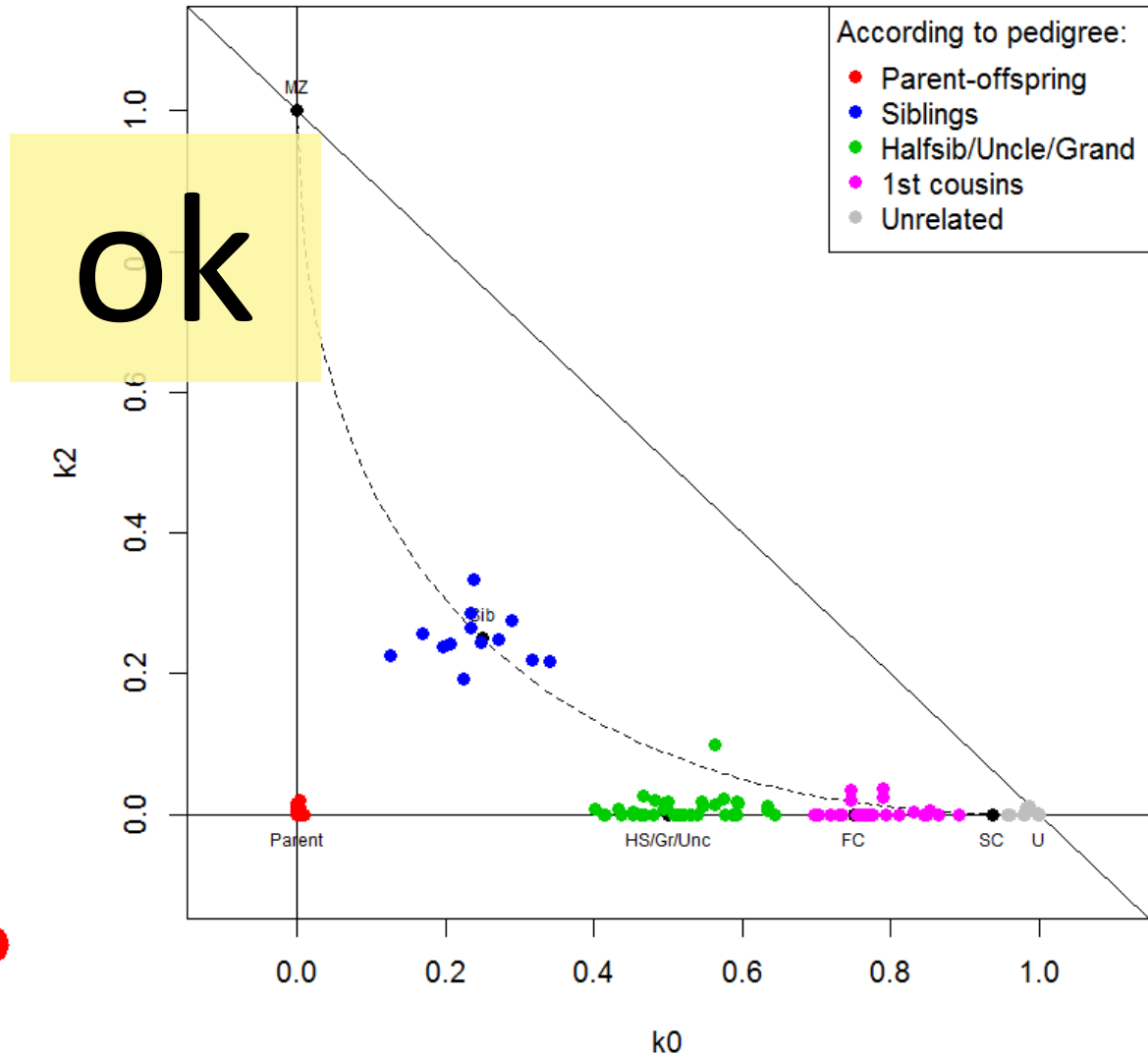
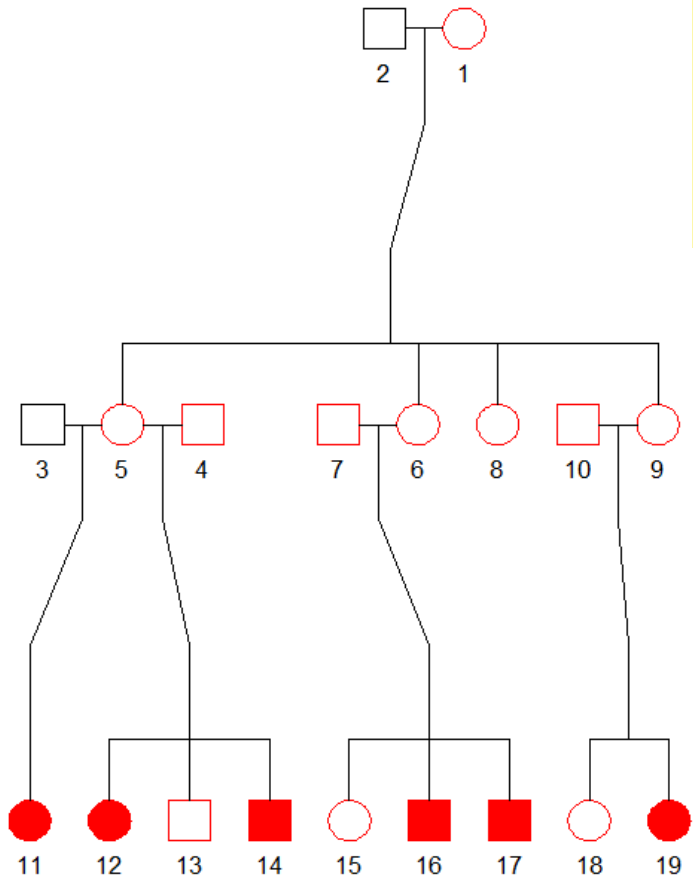
Application: Detecting pedigree errors

- Let \mathbf{x} be a pedigree object with markers
- Then `checkPairwise(\mathbf{x})` computes:
 - pedigree-based kappa for all pairs: `kappaIBD(\mathbf{x})`
 - marker-based kappa estimates for all pairs: `ibdEstimate(\mathbf{x})`
 - LR comparing the two
 - Color-coded plot according to relationship claimed by pedigree



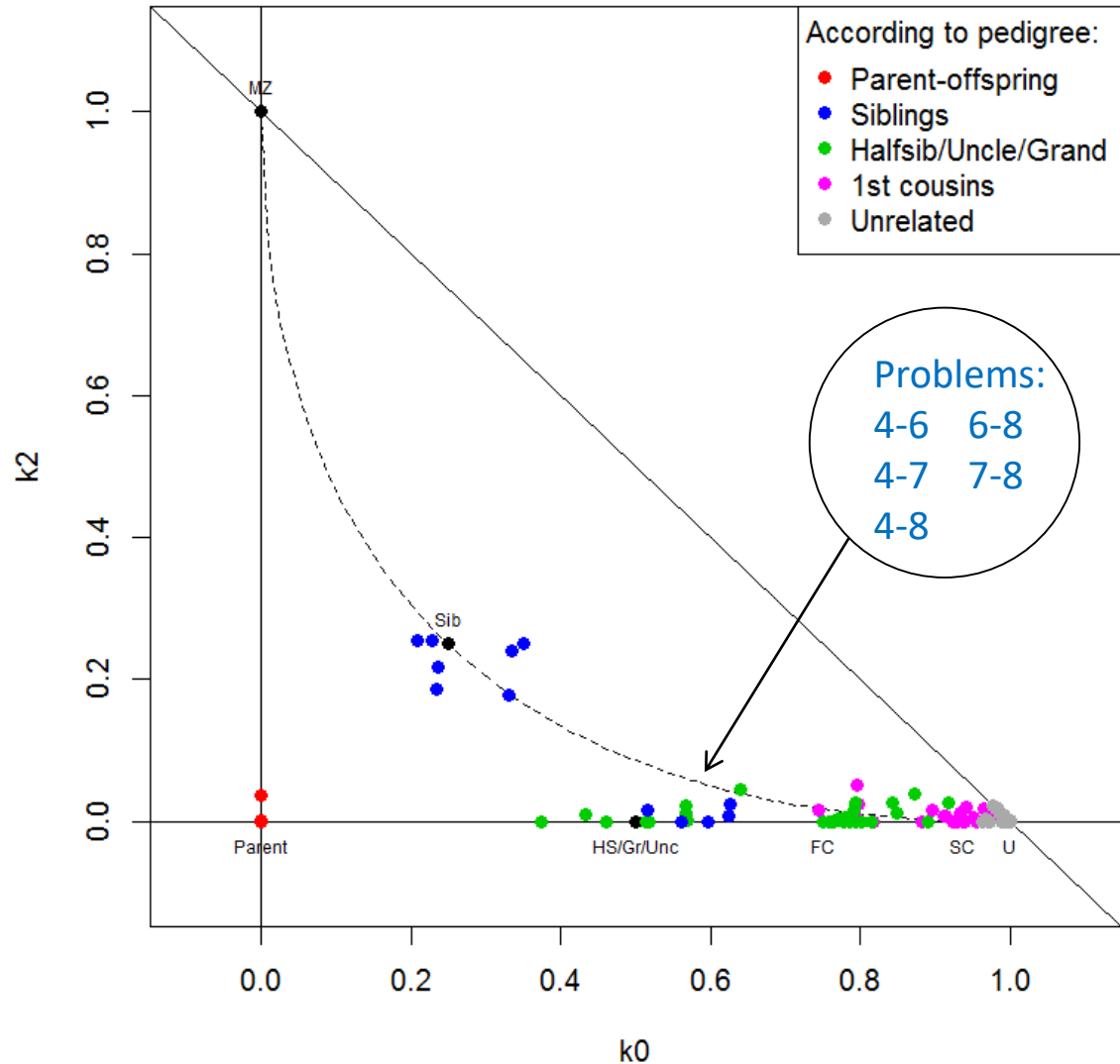
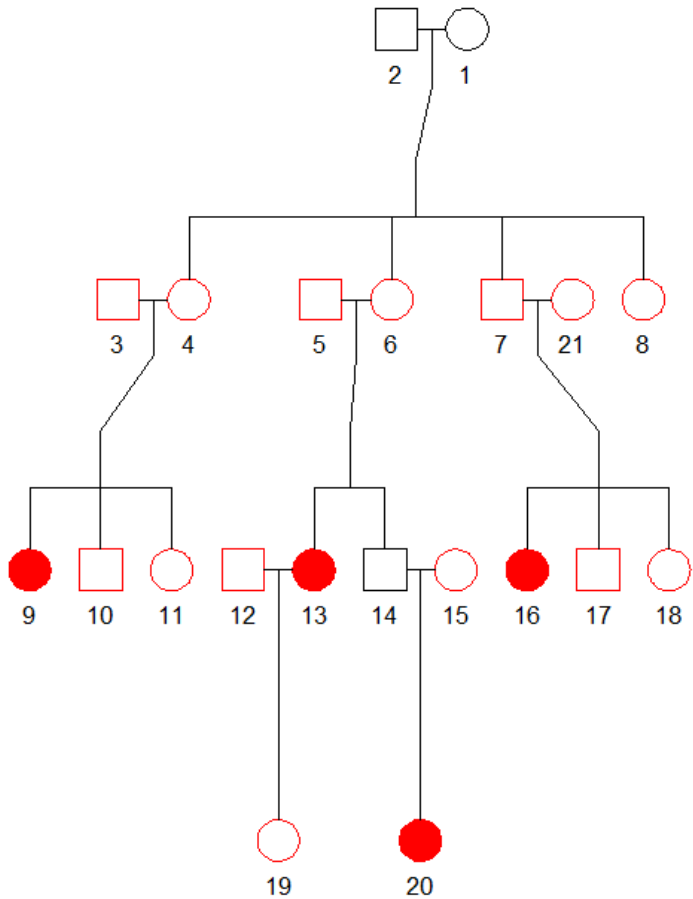
checkPairwise(): Example 1

Family 22



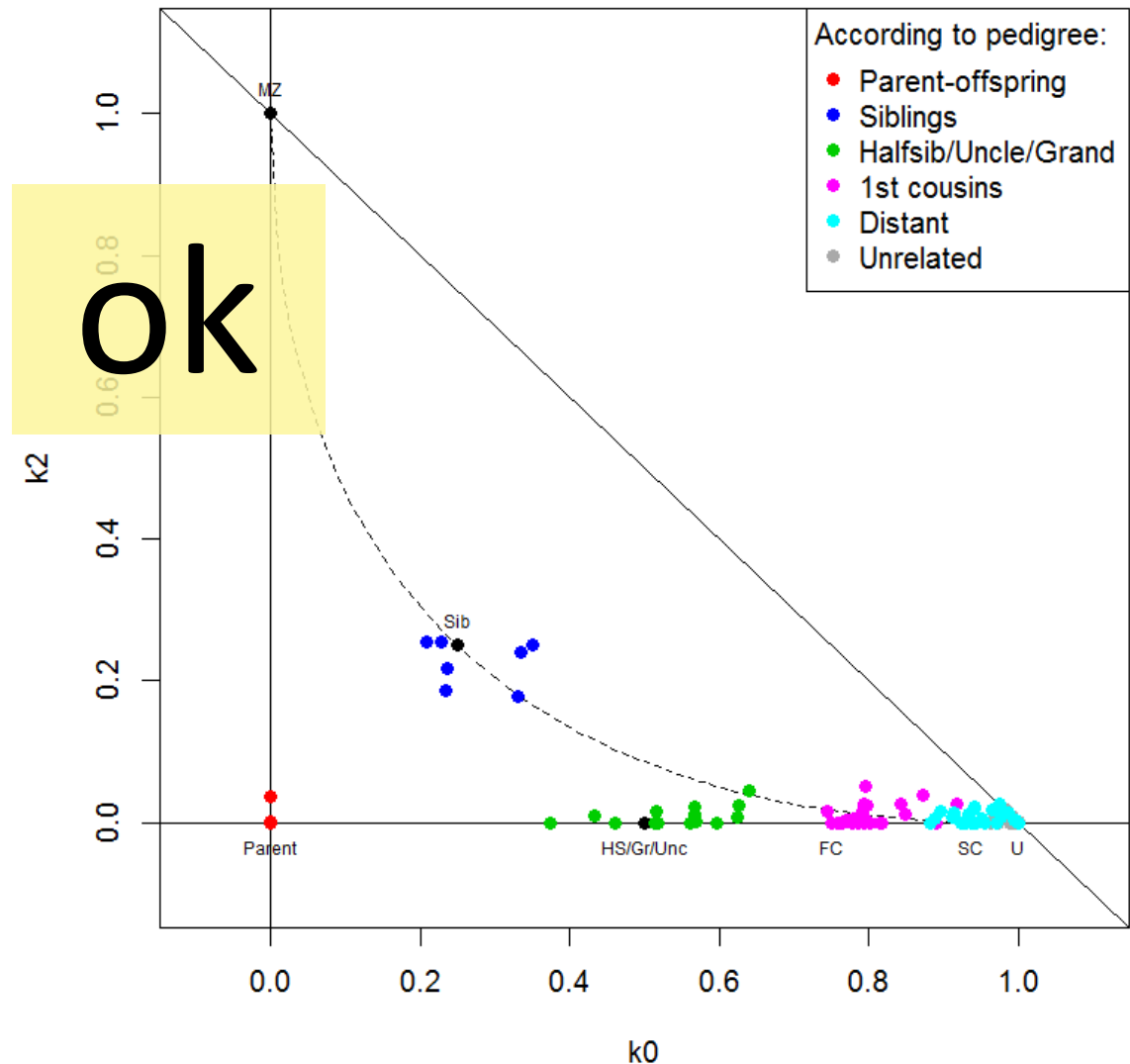
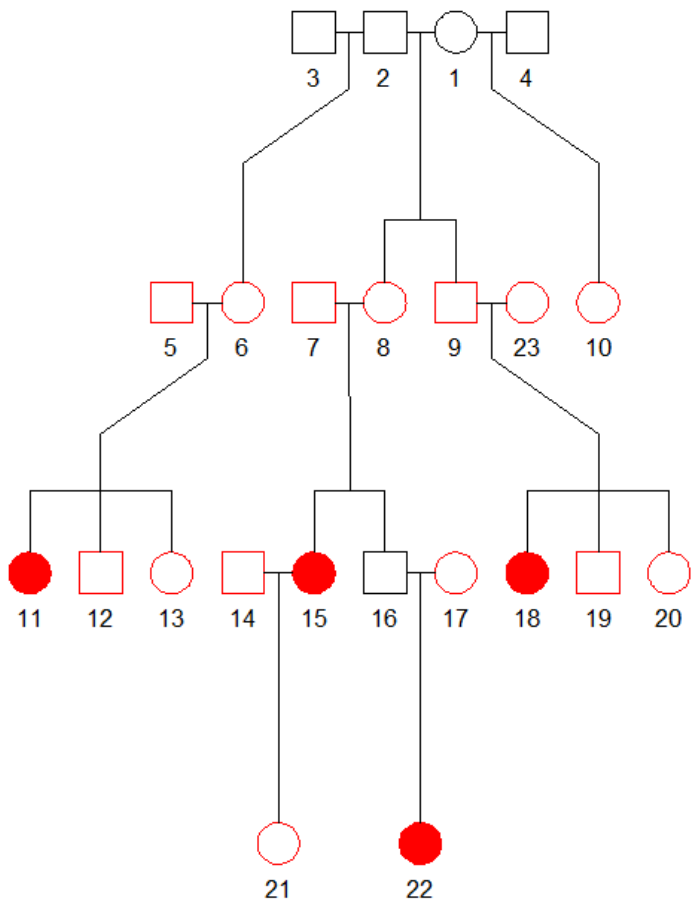
checkPairwise(): Example 2

Family 16



checkPairwise(): Example 2 - corrected

Family 16



Relatedness inference vs. allele frequencies

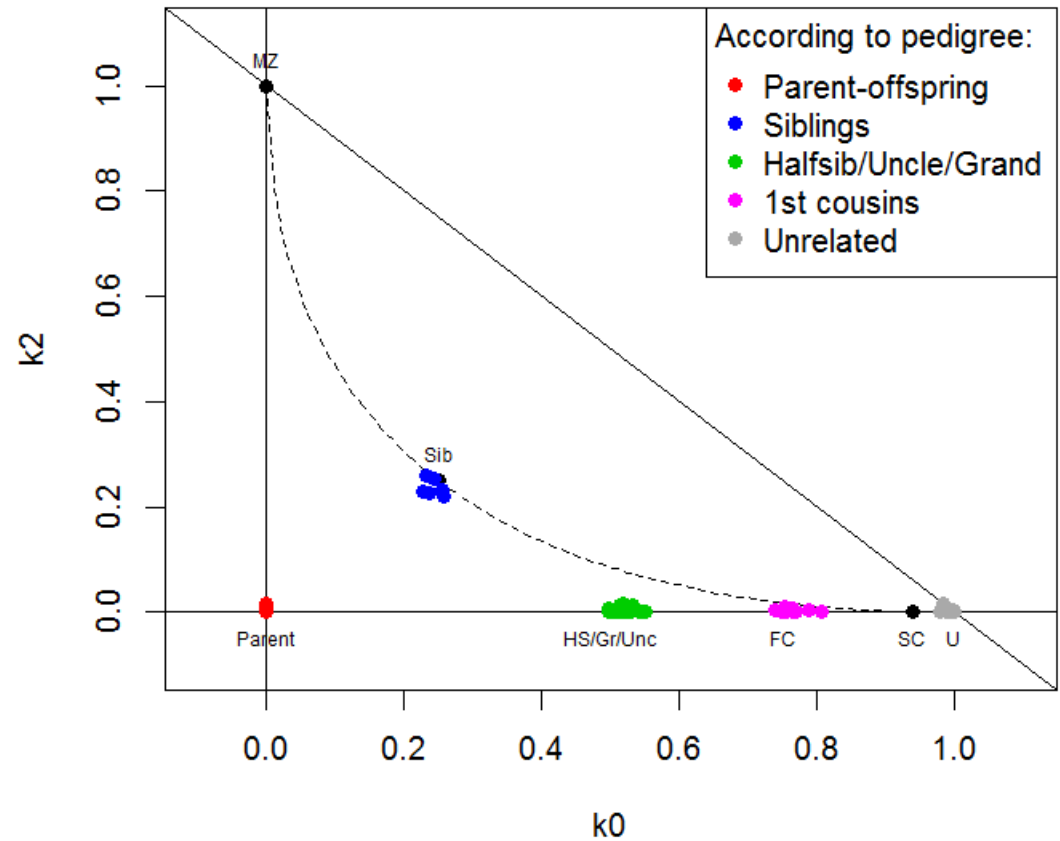
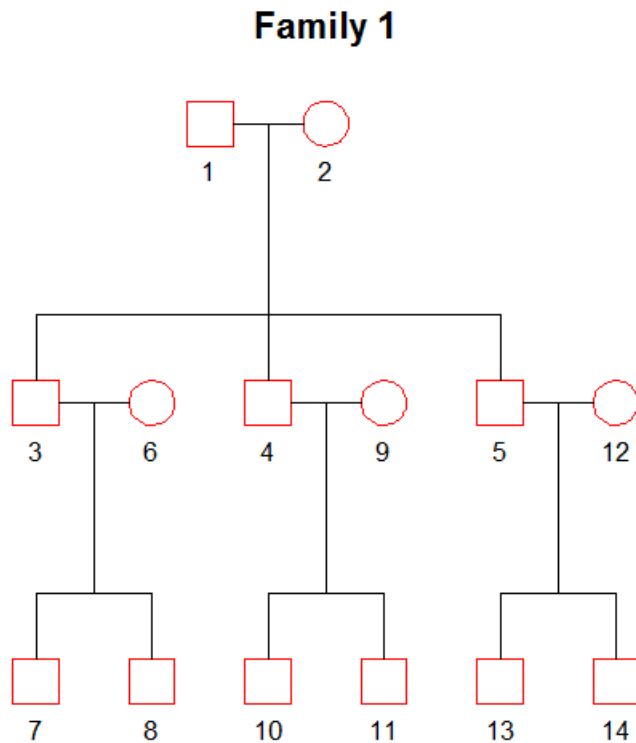
- A little simulation experiment!

Simulation example

SNPs: 10 000

True frequency distr: Unif(0,1)

Frequencies used: Correct

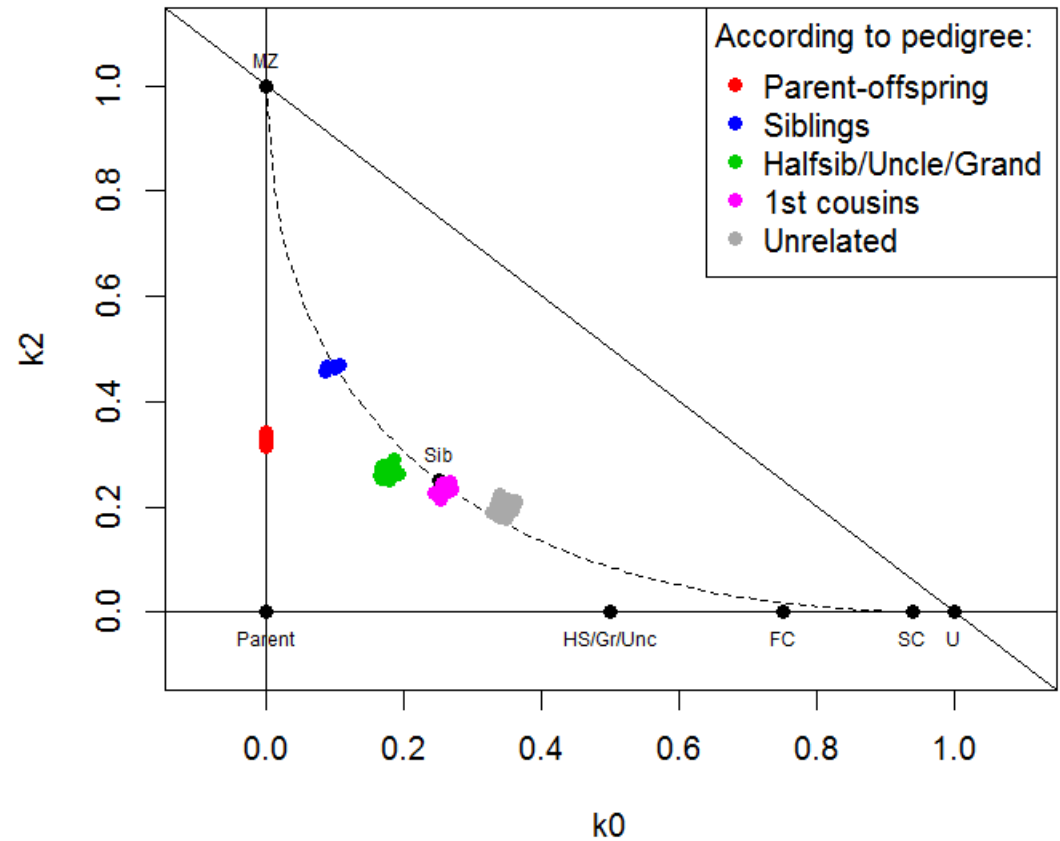
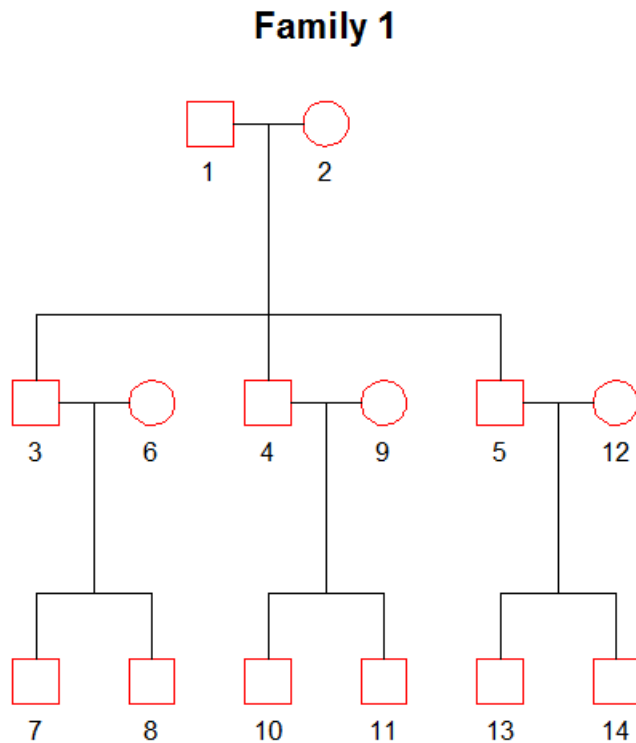


Simulation example

SNPs: 10 000

True frequency distr: Unif(0,1)

Frequencies used: All = 0.5

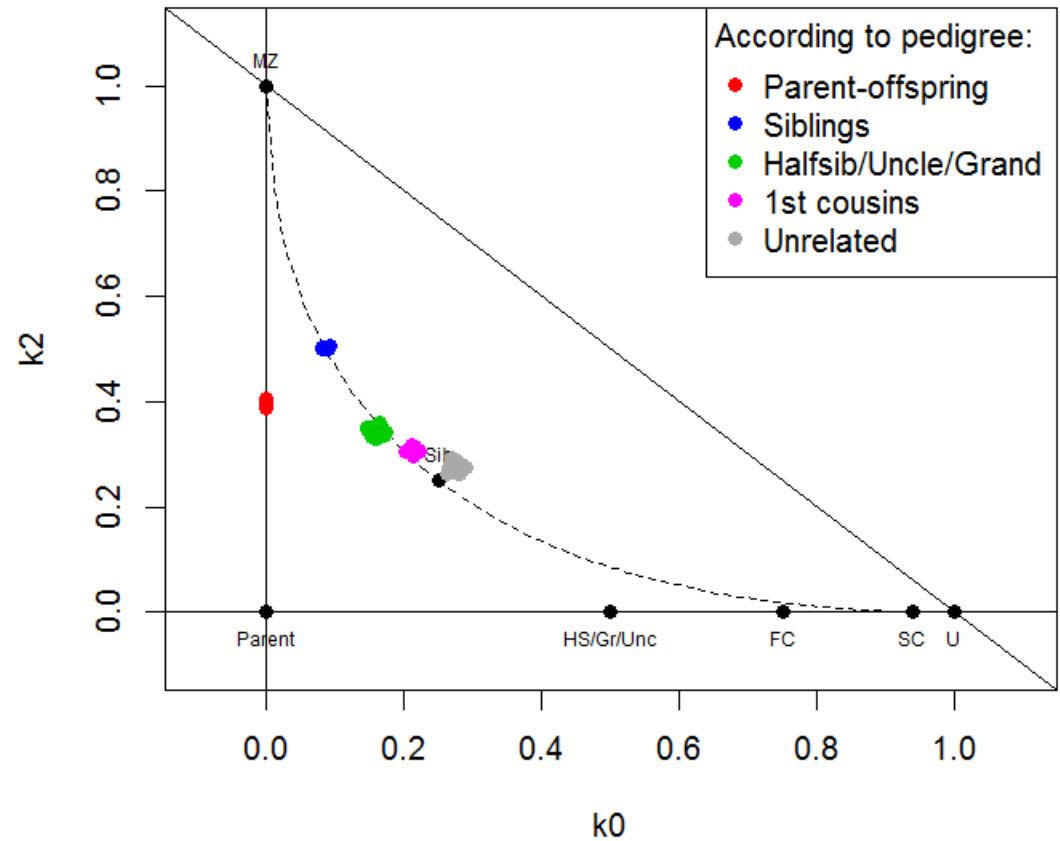
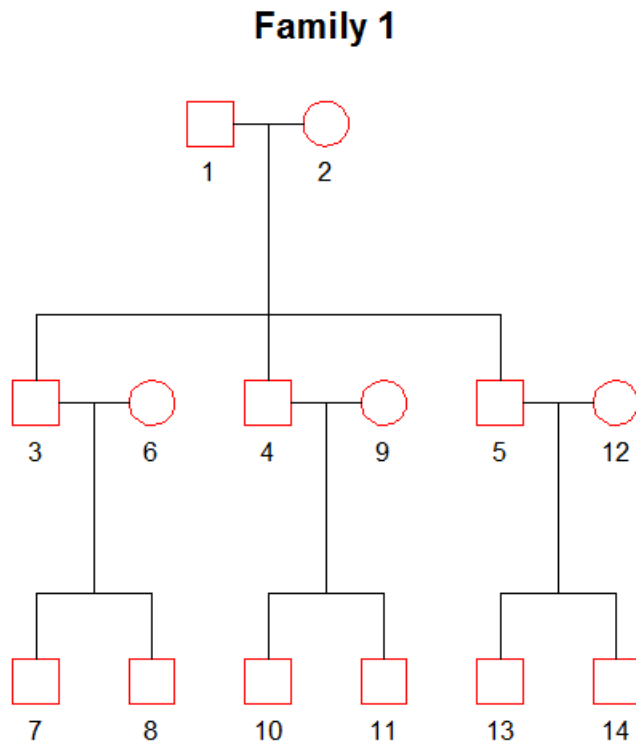


Simulation example

SNPs: 10 000

True frequency distr: Unif(0,1)

Frequencies used: Unif(0,1)

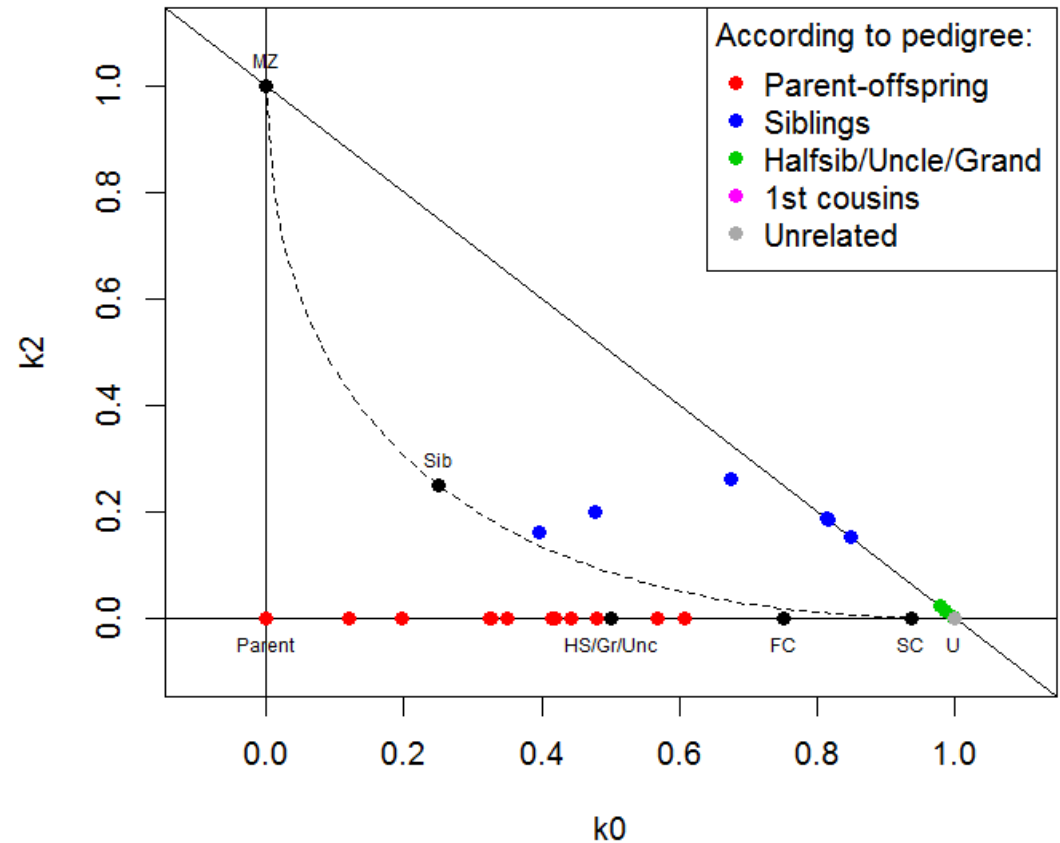
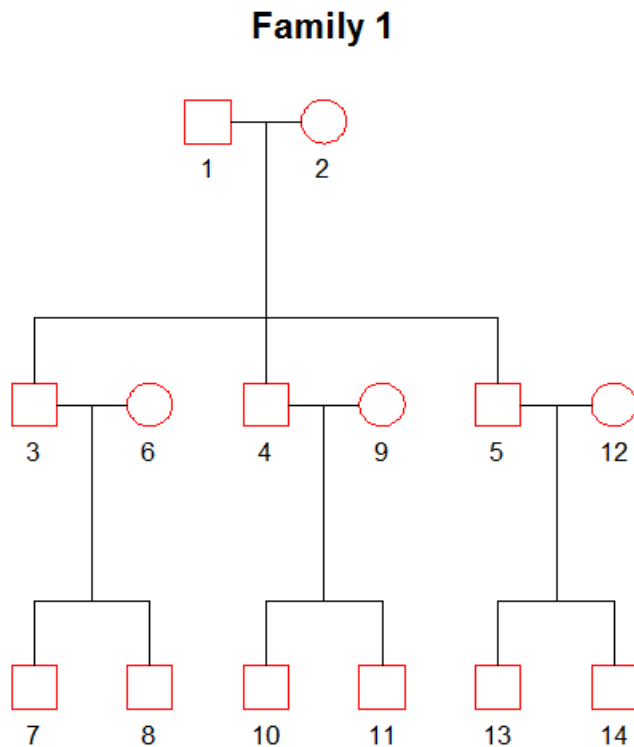


Simulation example

SNPs: 10 000

True frequency distr: Unif(0,1)

Frequencies used: Family estimate



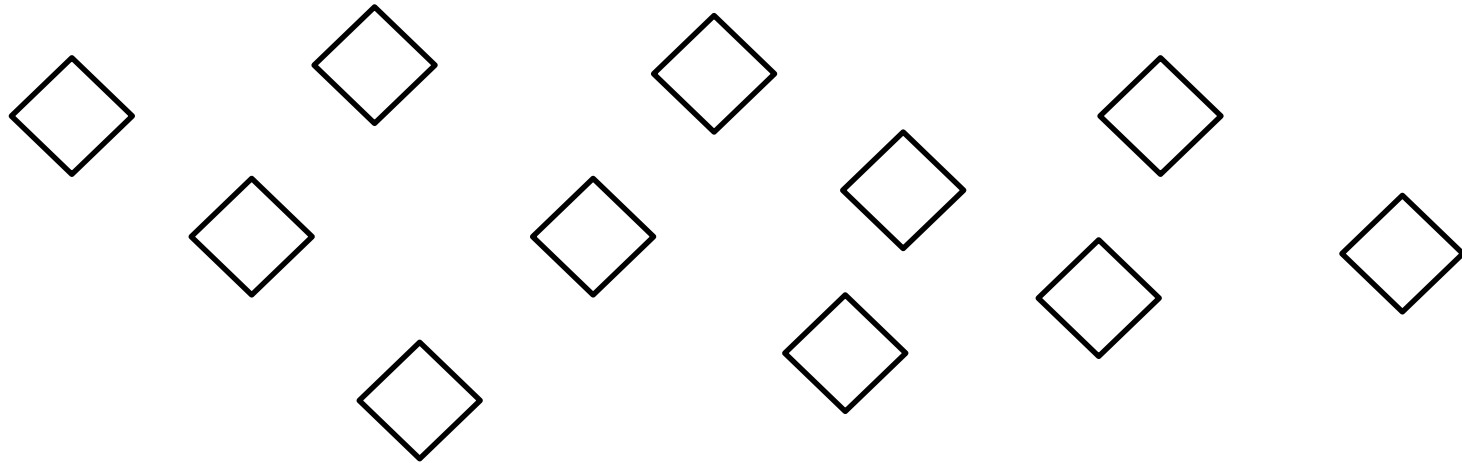
Conclusion from these simulations:

Pairwise inference is highly sensitive to wrong allele frequencies



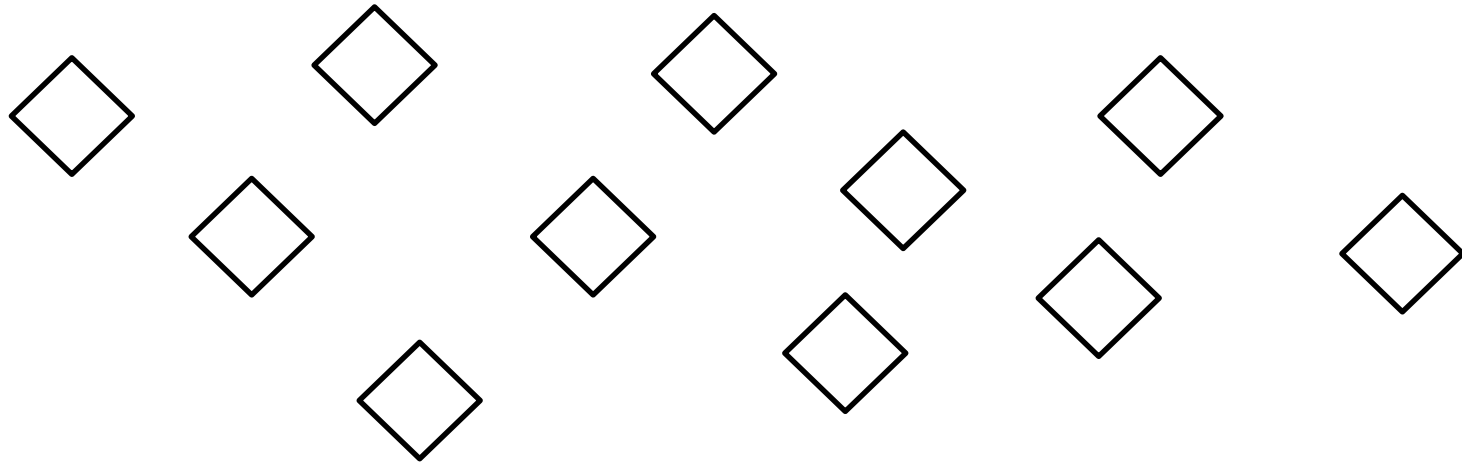
Part 4: Pedigree reconstruction

Pedigree reconstruction



Goal: Reconstruct the complete pedigree from genotype data

Pedigree reconstruction



Naive approach

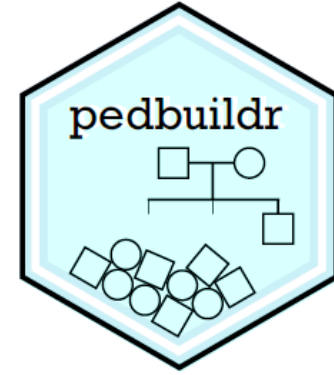
Step 1: Genders

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

Key functions:

```
> buildPeds()      # generate pedigrees

> reconstruct()    # main function!

> plot()           # plot top hits
```

pedbuildr: Example

Same dataset as before:

Simulate 100 SNPs for a pair of siblings

```
> ids = c("Al", "Bob")
> x = nuclearPed(children = ids)

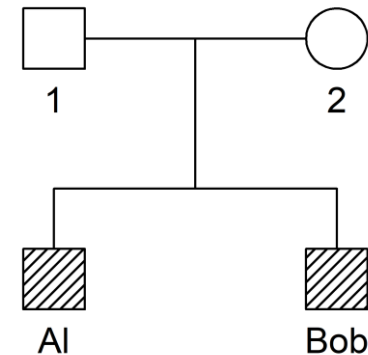
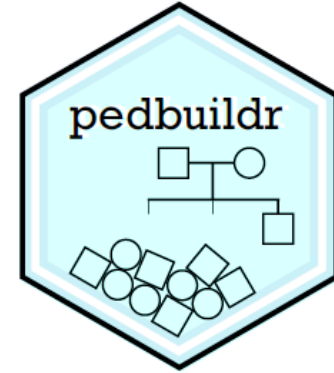
> x = markerSim(x, N = 100, ids = ids,
               alleles = 1:2, seed = 1234)

> x
```

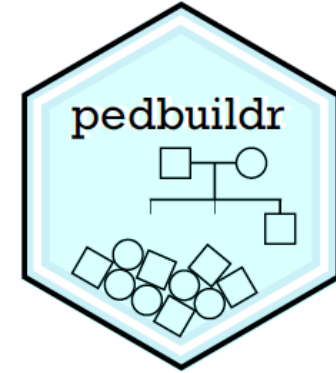
id	fid	mid	sex	<1>	<2>	<3>	<4>	<5>
1	*	*	1	-/-	-/-	-/-	-/-	-/-
2	*	*	2	-/-	-/-	-/-	-/-	-/-
Al	1	2	1	1/1	1/2	1/1	1/2	2/2
Bob	1	2	1	1/1	1/2	1/1	1/2	2/2

Only 5 (out of 100) markers are shown.

```
> dat = list(subset(x, "Al"),
              subset(x, "Bob"))
```



pedbuildr: Example



Reconstruct the most likely

```
> library(pedbuildr)
```

```
> r = reconstruct(dat)
```

Pedigree parameters:

ID labels: Al, Bob

Sex: 1, 1

Extra: parents

Age info: -

Known PO: -

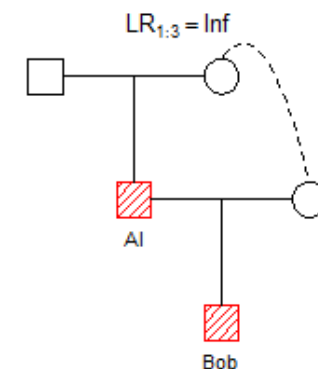
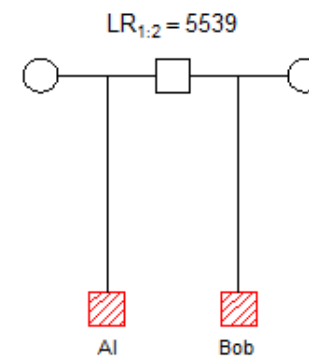
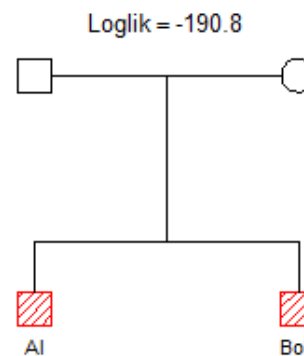
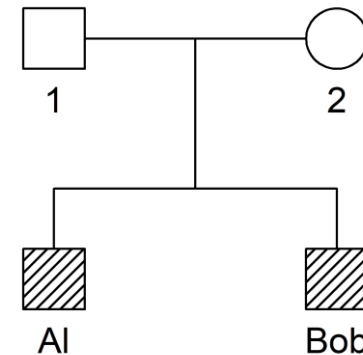
...

Building pedigree list:

...

Computing the likelihood of 6 pedigrees.

```
> plot(r, top = 3)
```

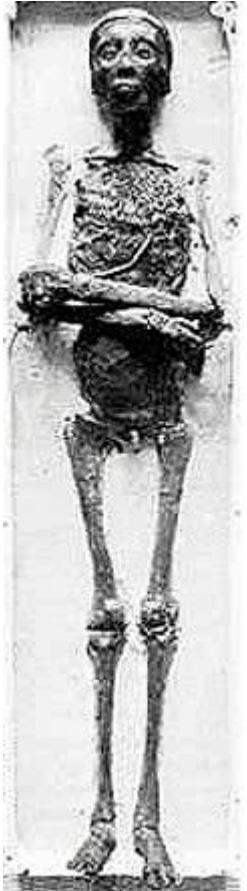


Parameters for restricting the search space

- **connected**: If TRUE (default), only connected pedigrees are considered
- **extra**: The number of extra individuals allowed to connect the original individuals
- **noChildren**: Individuals known to have no children
- **inferPO**: If TRUE, an initial stage of pairwise IBD estimation is done
- **knownPO**: Known parent–offspring pairs
- **age**: A numerical age vector, or a character vector describing age inequalities
- **notPO**: Pairs known not to be parent–offspring
- **allKnown**: If TRUE, then knownPO is the complete list of parent–offspring pairs
- **linearInb**: Set to FALSE to disallow inbreeding between linear descendants
- **linearInb**: Background inbreeding level

Reconstructing the pedigree of
Tuthankamon







Akhenaten



Tutankhamon



Nefertiti

Original Contribution

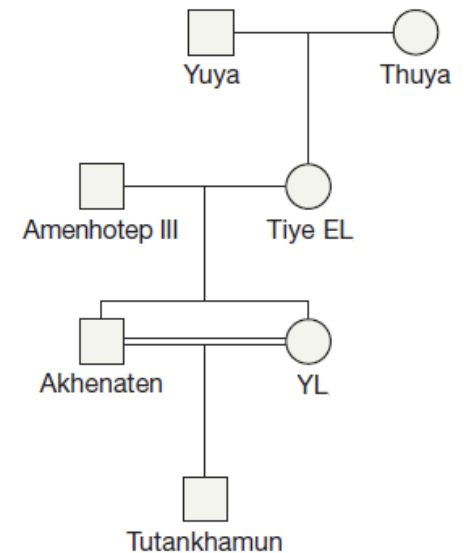
February 17, 2010

JAMA[®]

Ancestry and Pathology in King Tutankhamun's Family

Zahi Hawass, PhD; Yehia Z. Gad, MD; Somaia Ismail, PhD; Rabab Khairat, MSc; Dina Fathalla, MSc; Naglaa Hasan, MSc; Amal Ahmed, BPharm; Hisham Elleithy, MA; Markus Ball, MSc; Fawzi Gaballah, PhD; Sally Wasef, MSc; Mohamed Fateen, MD; Hany Amer, PhD; Paul Gostner, MD; Ashraf Selim, MD; Albert Zink, PhD; Carsten M. Pusch, PhD

	D13S317	D7S820	D2S1338	D21S11	D16S539	D18S51	CSF1PO	FGA
Thuya	9 12	10 13	19 26	26 35	11 13	8 19	7 12	24 26
Yuya	11 13	6 15	22 27	29 34	6 10	12 22	9 12	20 25
Tiye EL	11 12	10 15	22 26	26 29	6 11	19 22	9 12	20 26
Amenhotep III	10 16	6 15	16 27	25 34	8 13	16 22	6 9	23 31
Akhenaten	10 12	15 15	16 26	29 34	11 13	16 19	9 12	20 23
YL	10 12	6 10	16 26	25 29	8 11	16 19	6 12	20 23
Tutankhamun	10 12	10 15	16 26	29 34	8 13	19 19	6 12	23 23





CURSE OF THE PHARAOH'S DNA

Some researchers claim to have analysed DNA from Egyptian mummies. Others say that's impossible. Could new sequencing methods bridge the divide?

BY JO MARCHANT

Cameras roll as ancient-DNA experts Carsten Pusch and Albert Zink scrutinize a row of coloured peaks on their computer screen. There is a dramatic pause. "My god!" whispers Pusch, the words muffled by his surgical mask. Then the two hug and shake hands, accompanied by the laughter and applause of their Egyptian colleagues. They have every right to be pleased with themselves. After months of painstaking work, they have finally completed their analysis of 3,300-year-old DNA from the mummy of King Tutankhamun.

Featured in the Discovery Channel documentary *King Tut Unwrapped* last year and published in the *Journal of the American Medical Association (JAMA)*¹, their analysis — of Tutankhamun and ten of his relatives — was the latest in a string of studies reporting the analysis of DNA from ancient Egyptian mummies. Apparently revealing the mummies' family relationships as well as their afflictions, such as tuberculosis and malaria, the work seems to be providing unprecedented insight into the lives and health of ancient Egyptians and is ushering in a new era of 'molecular Egyptology'. Except that half of the researchers in the field challenge every word of it.

Enter the world of ancient Egyptian DNA and you are asked to choose between two alternate realities: one in which DNA analysis is routine, and the other in which it is impossible. "The ancient-DNA field is split absolutely in half," says Tom Gilbert, who heads two research groups at the Center for GeoGenetics in Copenhagen, one of the world's foremost ancient-DNA labs.

Mummies found in King Tutankhamun's tomb are at the centre of a dispute over DNA analysis.



ENiM

Égypte Nilotique et Méditerranéenne

L'ADN de la famille royale amarnienne et les sources égyptiennes De la complémentarité des méthodes et des résultats

Marc Gabolde

Équipe Égypte Nilotique et Méditerranéenne

UMR 5140 (CNRS - Université Paul-Valéry - Montpellier III)

Gabolde's alternative

Thoutmosis IV

D13S317	10
ou	
D7S820	6
ou	15
D2S1338	16
D21S11	25
D16S539	8
ou	13
D18S51	12
ou	16
CSF1PO	6
ou	23
FGA	31

Parent putatif 1 de KV 35 YL*

D13S317	10
D7S820	10
D2S1338	16
D21S11	25
D16S539	8
D18S51	16
CSF1PO	6
FGA	23

Amenhotep III (KV 35)

D13S317	10
D7S820	6
D2S1338	16
D21S11	25
D16S539	8
D18S51	16
CSF1PO	6
FGA	23

KV 35 Young Lady Nefertiti (?)

D13S317	10
D7S820	6
D2S1338	16
D21S11	25
D16S539	8
D18S51	16
CSF1PO	6
FGA	23

Moutemouiya KV21A

D13S317	10
ou	16
D7S820	6
ou	15
D2S1338	27
D21S11	34
D16S539	8
ou	13
D18S51	12
ou	22
CSF1PO	9
FGA	23

KV 35 Elder Lady (Tiyi)

D13S317	11
D7S820	10
D2S1338	22
D21S11	26
D16S539	6
D18S51	19
CSF1PO	9
FGA	20

Toutânkhamon KV 62

D13S317	10
D7S820	10
D2S1338	16
D21S11	29
D16S539	8
D18S51	19
CSF1PO	6
FGA	23

Allèles restitués et assemblés des père et mère de Youya et Moutemouiya

D13S317	11	13
D7S820	6	15
D2S1338	22	27
D21S11	34	29
D16S539	6	10
D18S51	12	22
CSF1PO	9	12
FGA	23	25

Yuya (KV 46[A])

D13S317	11
D7S820	6
D2S1338	22
D21S11	29
D16S539	6
D18S51	12
CSF1PO	12
FGA	20

Parent putatif 2 de KV 35 YL*

D13S317	11
D7S820	6
D2S1338	27
D21S11	26
D16S539	6
D18S51	12
CSF1PO	12
FGA	20

Amenhotep IV KV 55

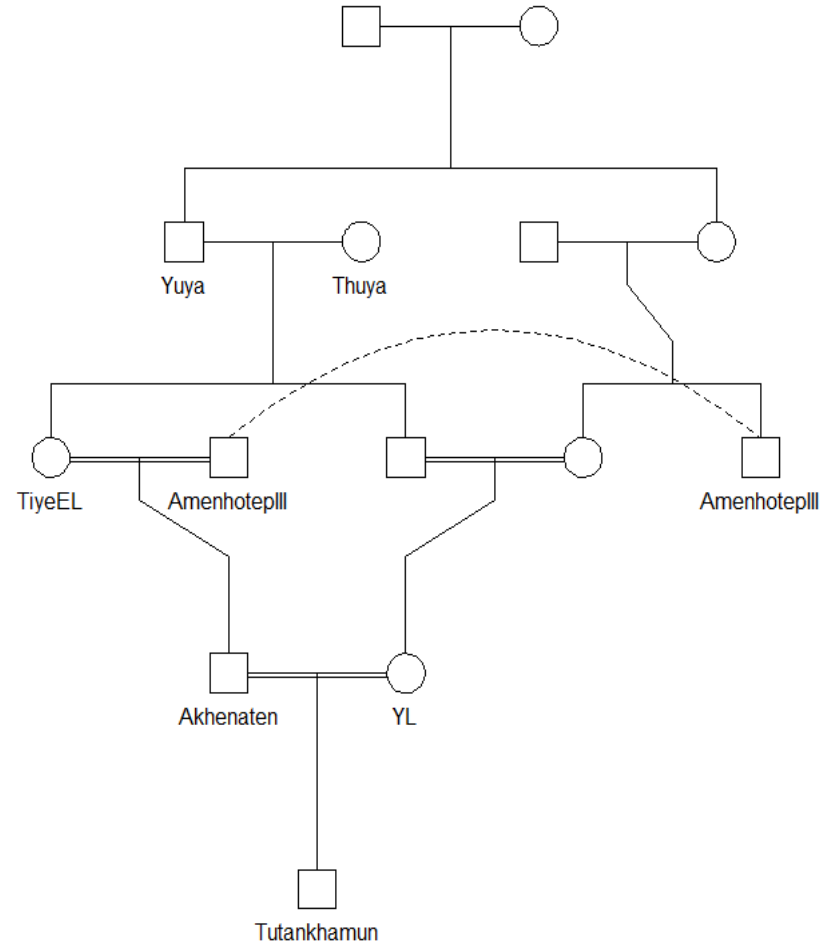
D13S317	10
D7S820	15
D2S1338	16
D21S11	29
D16S539	11
D18S51	16
CSF1PO	9
FGA	20

Thouya (KV 46)

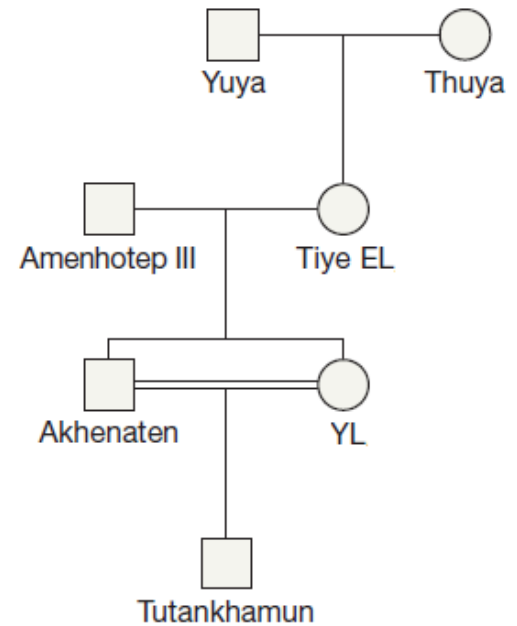
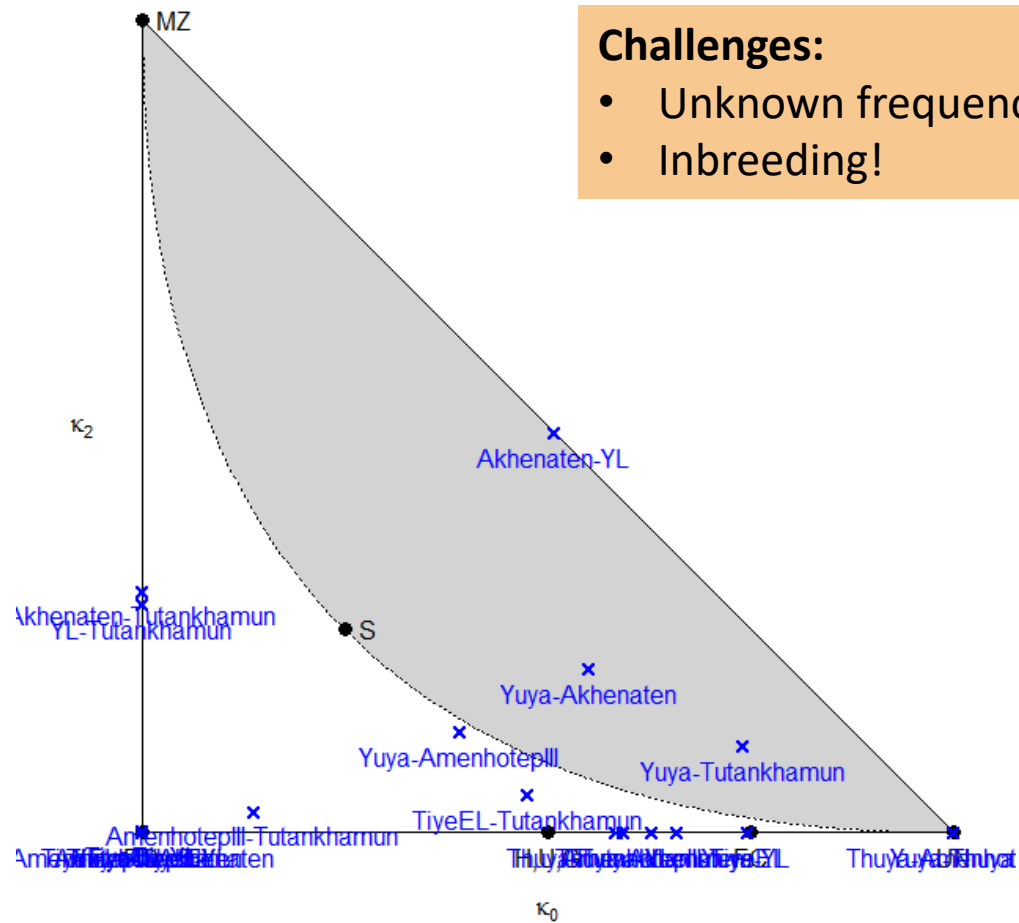
D13S317	9
D7S820	10
D2S1338	19
D21S11	26
D16S539	11
D18S51	8
CSF1PO	7
FGA	24

D13S317	11
D7S820	6
D2S1338	27
D21S11	29
D16S539	6
D18S51	12
CSF1PO	12
FGA	20

Mariages ==
Filiations —



Pairwise estimates



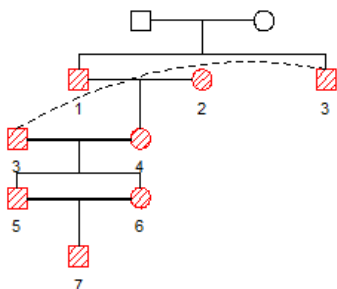
Pedigree reconstruction with pedbuilder

Infer PO pairwise

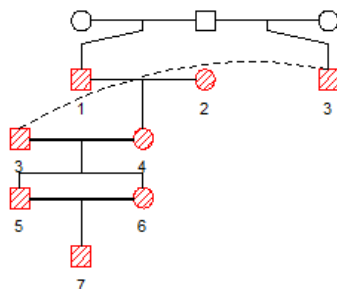
Mutation rate = 2%

Founder inbreeding = 5%

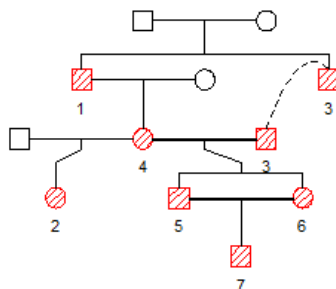
Loglik = -200.83



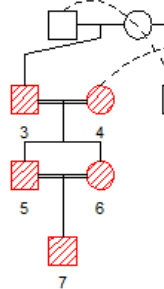
Loglik = -204.21



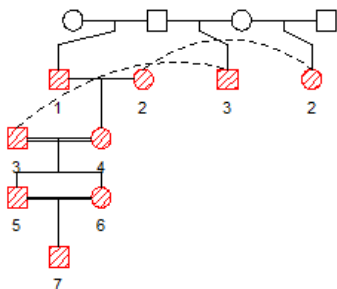
Loglik = -205.76



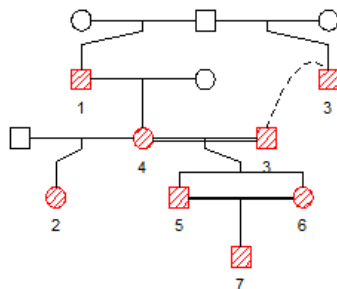
Loglik =



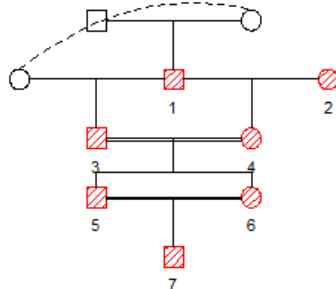
Loglik = -208.55



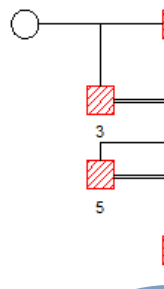
Loglik = -209.14



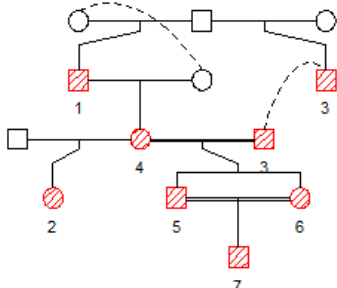
Loglik = -209.88



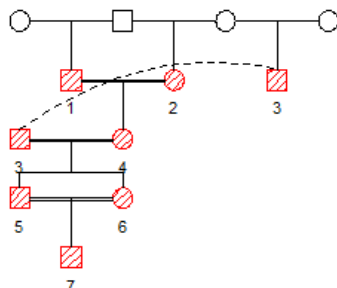
Loglik =



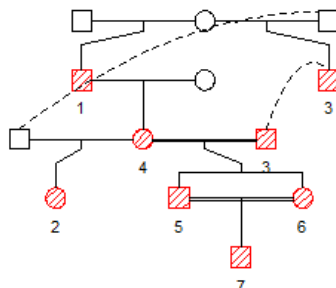
Loglik = -213.27



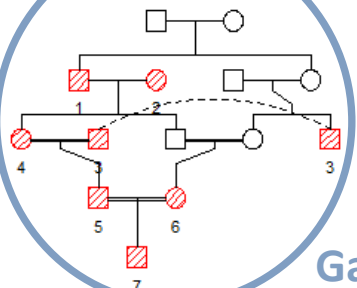
Loglik = -213.37



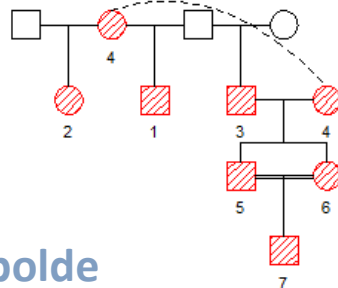
Loglik = -213.47



Loglik = -213.73

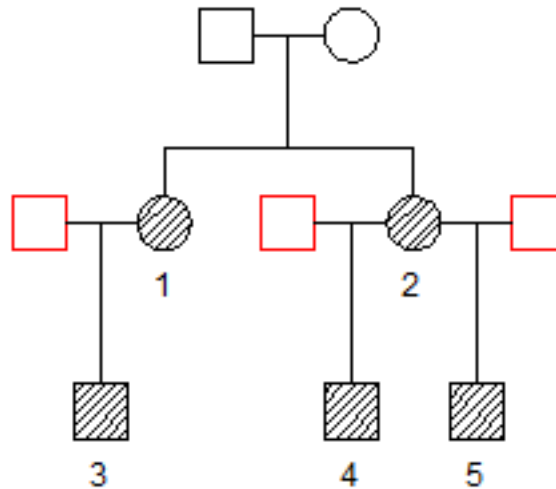
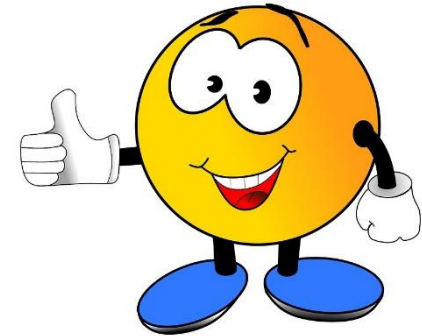


Loglik = -214.18



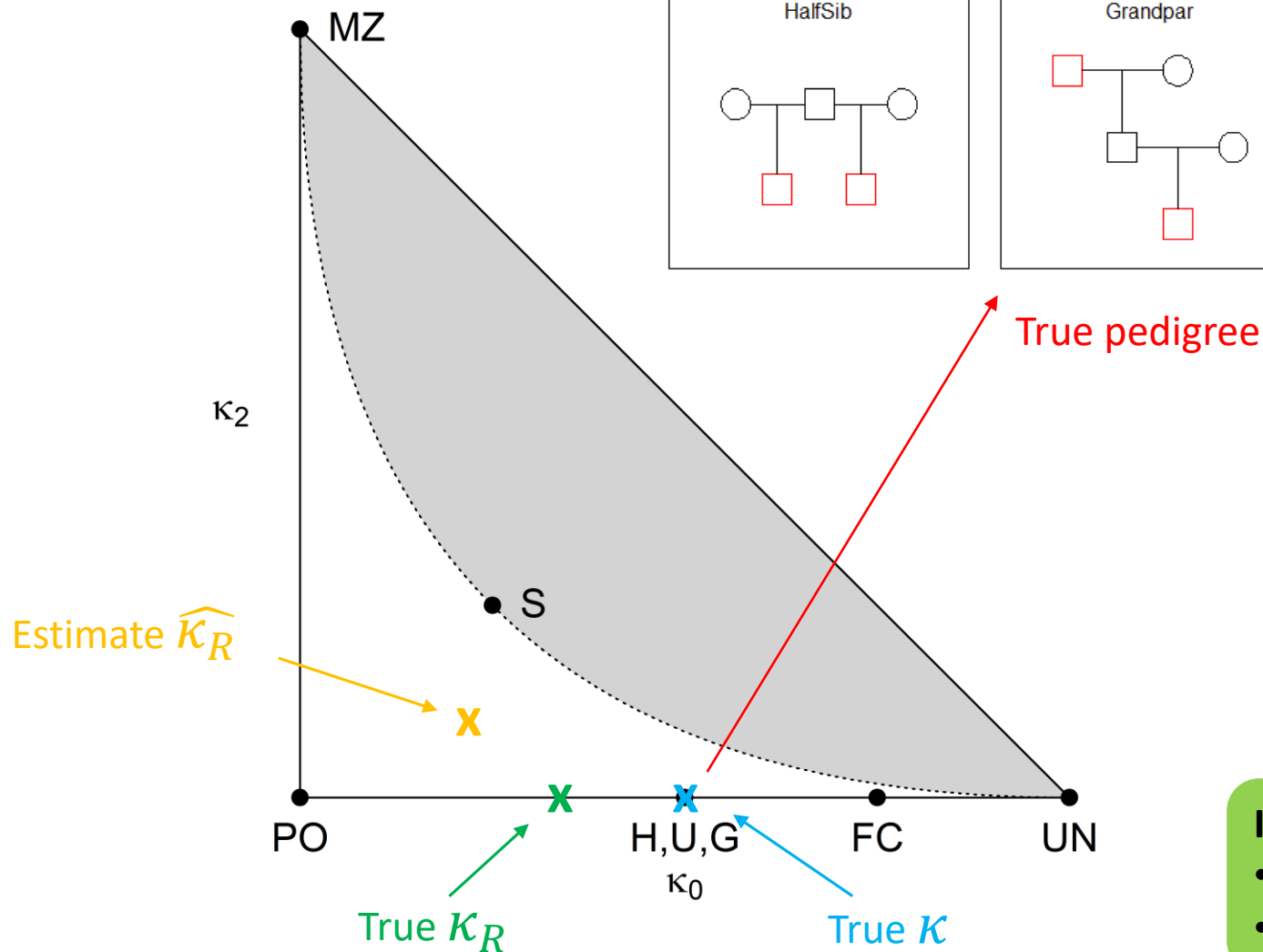
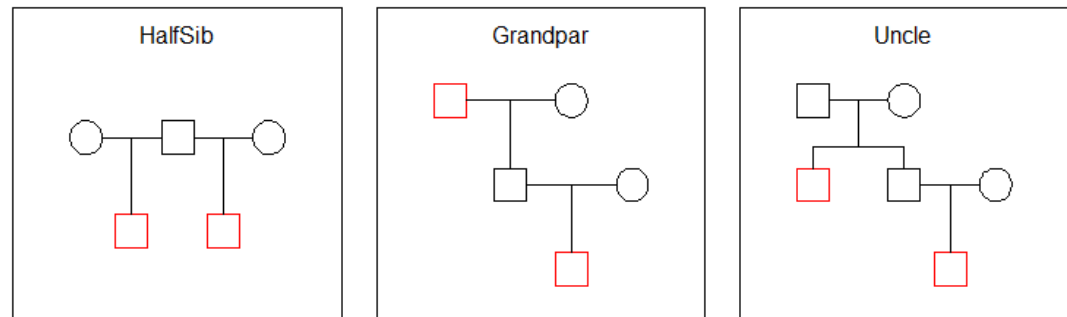
Gabolde

Your turn: Exercises!



Q: Do any of the children have the same father?

Can we recover the pedigree?



In practice

- PO usually ok
- Sibs if good data