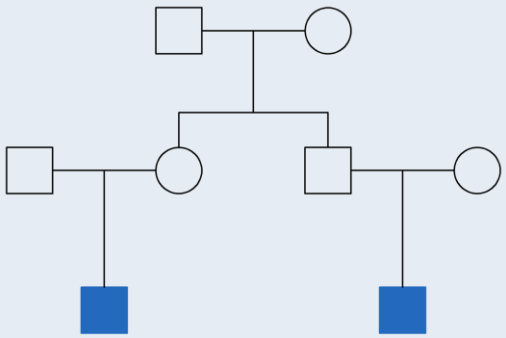


Lecture 3: Relatedness inference and pedigree reconstruction

ISFG Congress 2026 Montreal - Workshop

Pedigree analysis and relatedness inference: Advanced

Magnus Dehli Vigeland

Home page:
https://magnusdv.github.io/pedsuite/articles/web_only/course-isfg2026.html

Schedule

Morning session – Pedigree analysis: Basic

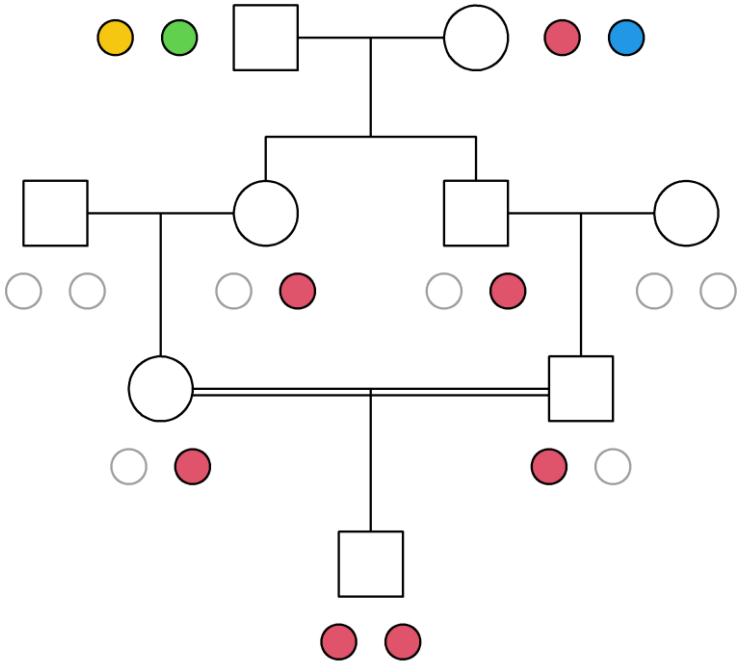
- 08:30–09:30 **Pedigrees and measures of relatedness** (MDV)
- 09:30–10:30 Exercises I
- 10:30–10:45 *Coffee break*
- 10:45–11:30 **Kinship testing** (TE)
- 11:30–12:15 Exercises II
- 12:15–12:30 Summary and discussion

Lunch break 12:30 - 14:00

Afternoon session – Pedigree analysis: Advanced

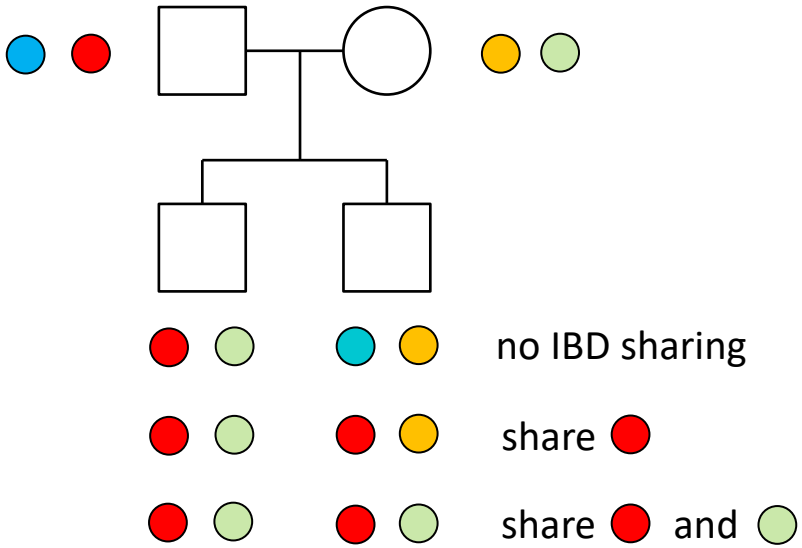
- 14:00–15:00 **Relatedness inference and pedigree reconstruction** (MDV)
- 15:00–15:45 Exercises III
- 15:45–16:00 *Coffee break*
- 16:00–17:00 **Disaster victim identification** (TE)
- 17:00–17:45 Exercises IV
- 17:45–18:00 Summary and discussion

Recap: Identity by descent (IBD)



Recap: IBD coefficients

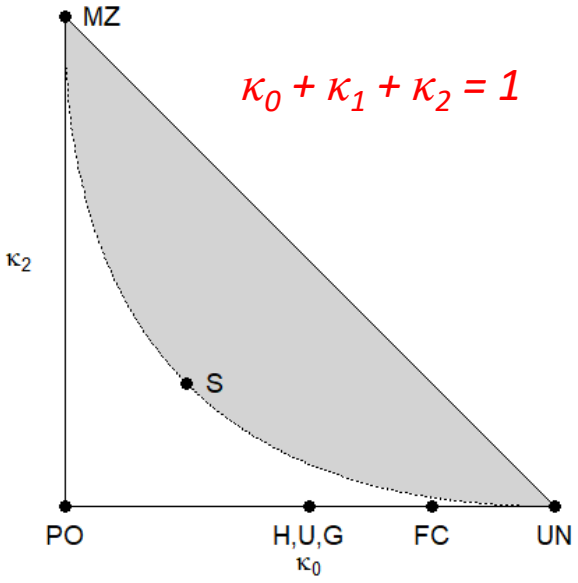
- How many alleles are IBD in each locus?



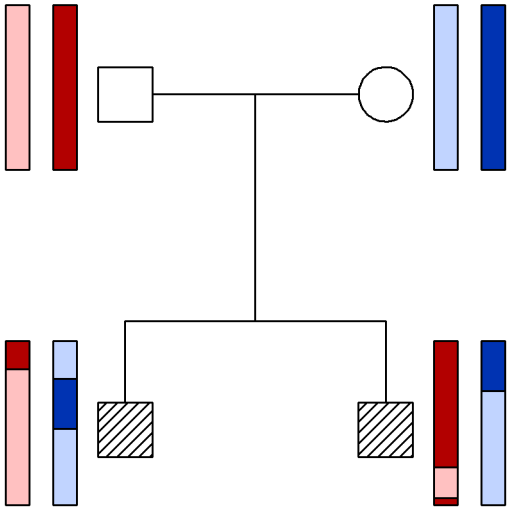
Definition

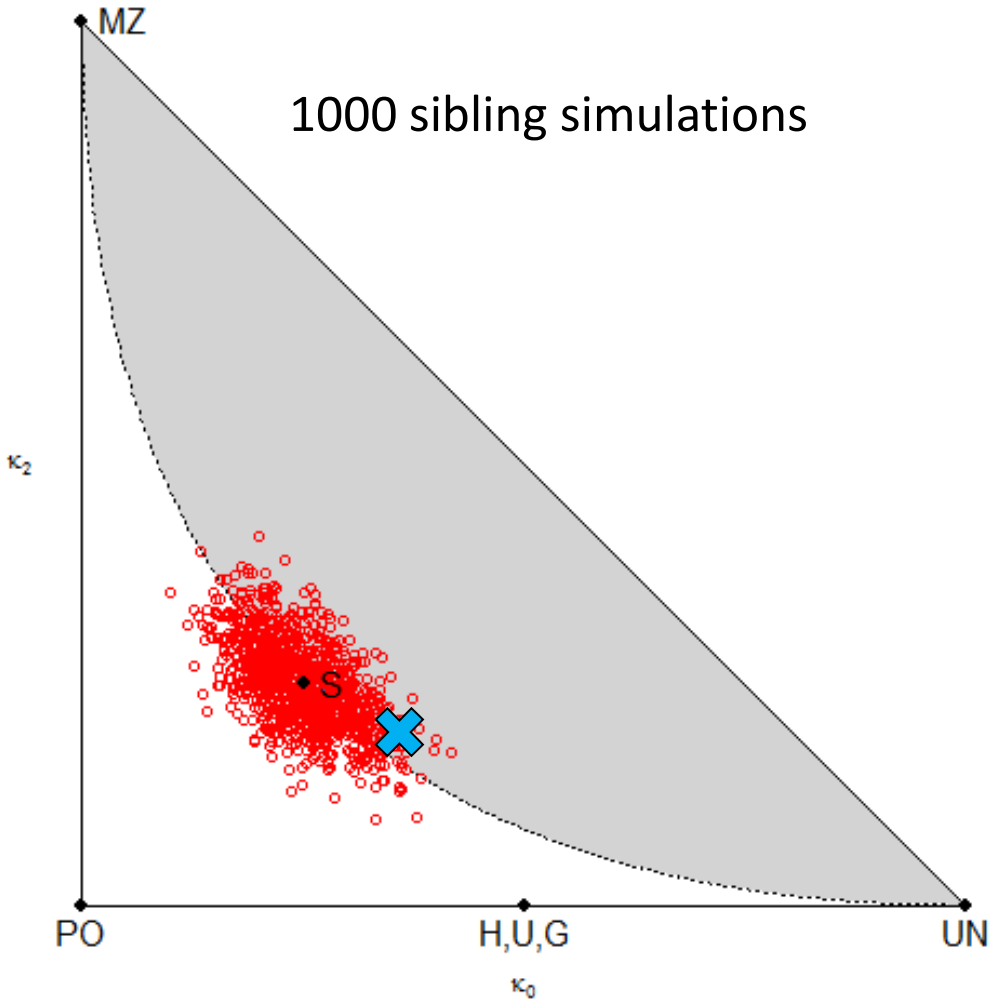
- κ_0 = *P*(0 alleles IBD)
- κ_1 = *P*(1 alleles IBD)
- κ_2 = *P*(2 alleles IBD)

(at random autosomal locus)



Recap: Realised relatedness





Part I: Inference of ***pairwise*** relatedness

Pairwise inference: Main approaches

A. Based on IBD coefficients

- Typically with STR markers
- Maximum-likelihood estim.
- Assumes independence

- Complexity: **Easy**
- Accuracy: **Poor** (except PO/MZ)
- Scope: **Close relationships**

OLD SCHOOL

B. Based on IBD segments

- Requires lots of SNPs
- Two steps:
  1) SNPs → IBD segments
  2) IBD segments → relatedness

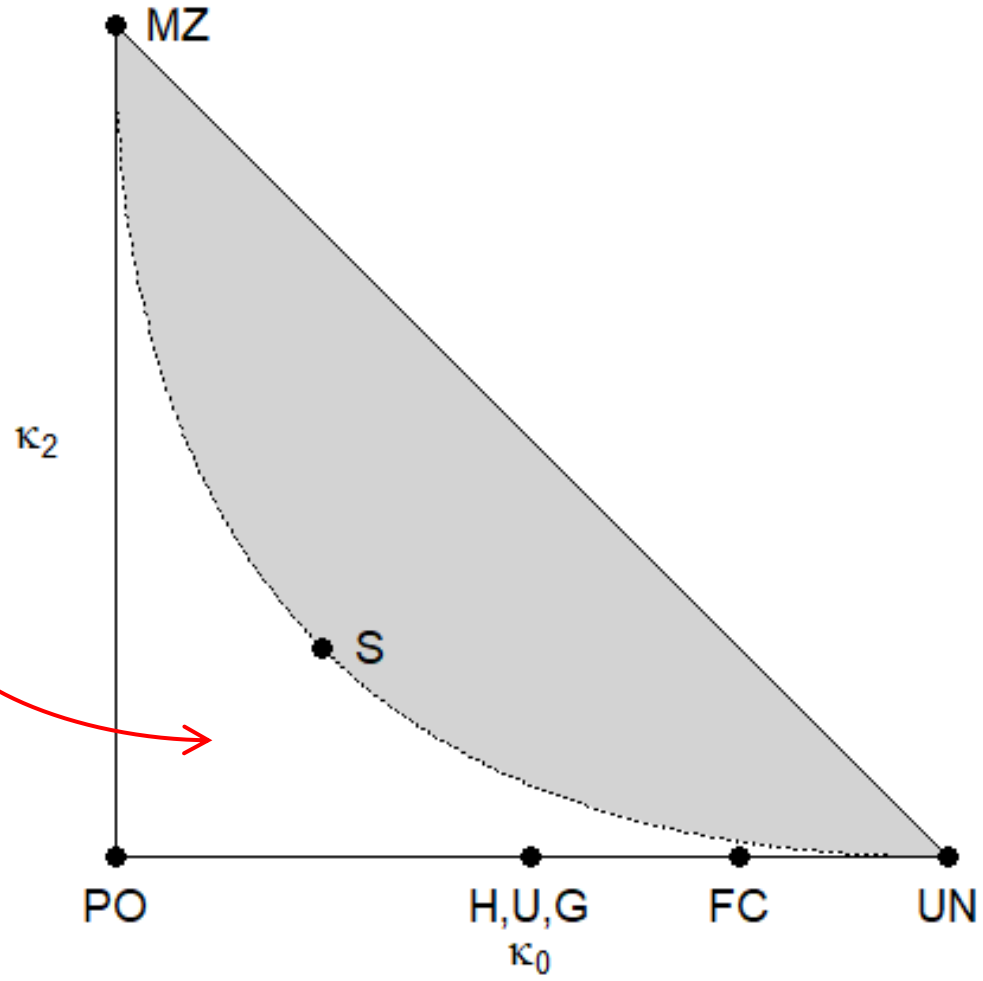
  (Often different software)

- Complexity: **Medium/high**
- Accuracy: **Better**
- Scope: **Close + distant**

MODERN

Approach A

where??



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

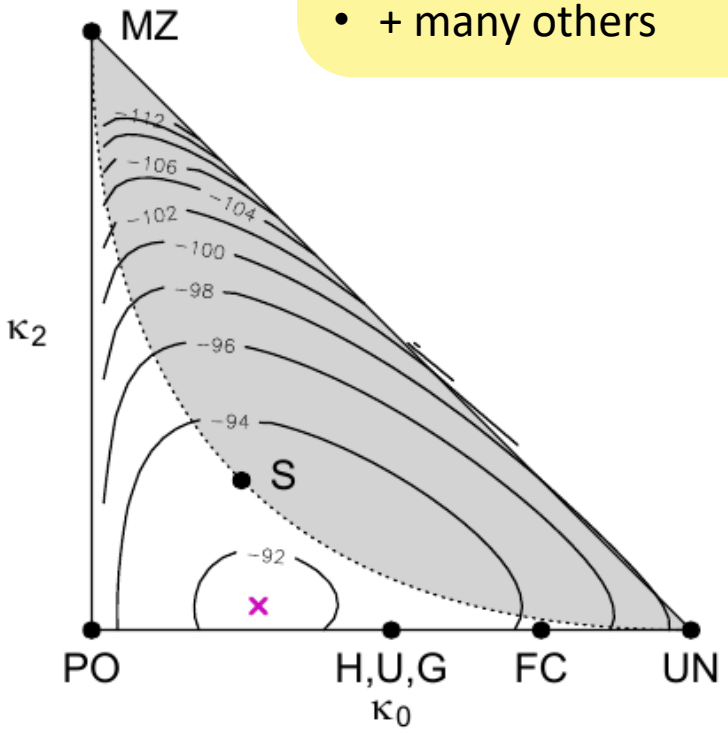
- Thompson (1975)
  - Given: marker genotypes for two individuals
  - The likelihood function

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

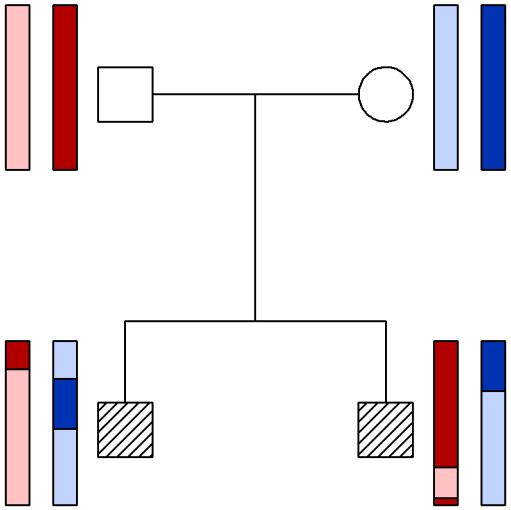
- Find the point κ which maximizes $L(\kappa)$!

- Assumptions:
  - known allele freqs
  - HWE
  - no inbreeding

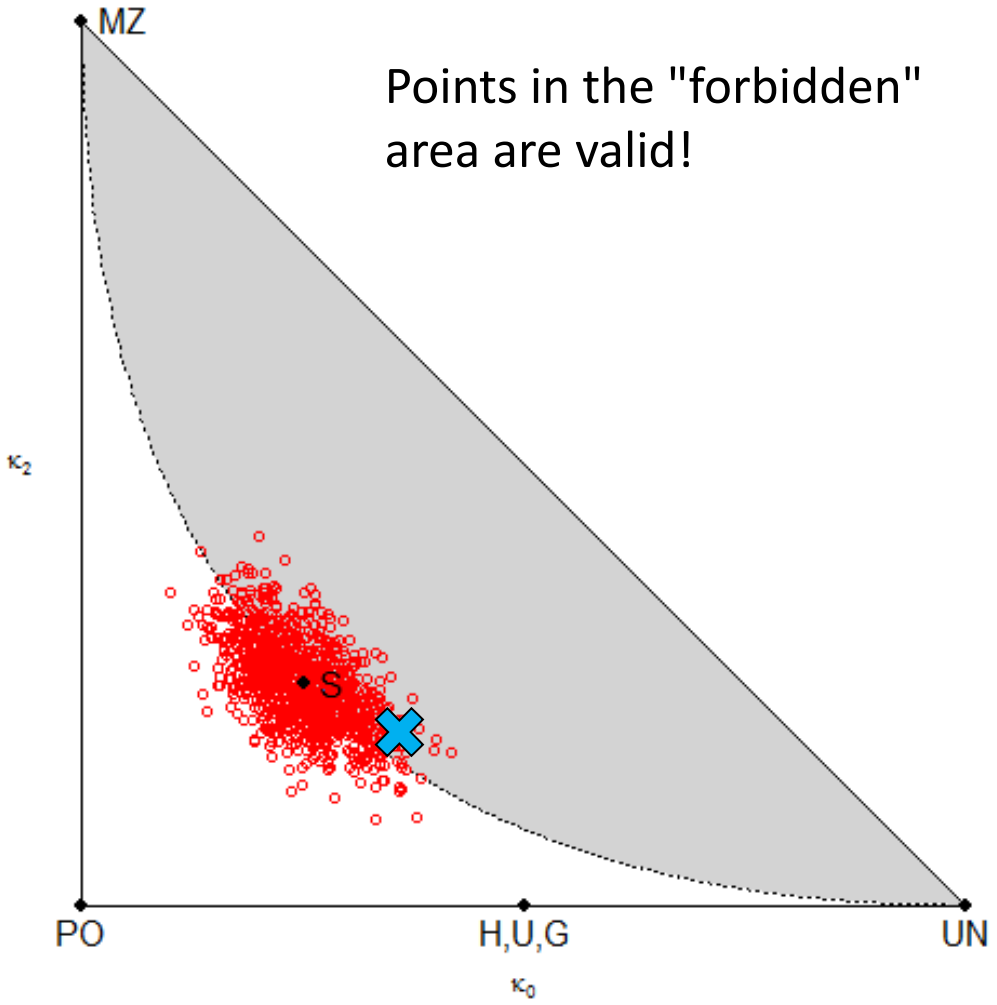
Implementations:
- R/pedsuite
- DIVIANA
- Familias
- + many others



What are we estimating?



Answer: The *realised relatedness*



Pairwise inference in R

- Key functions

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

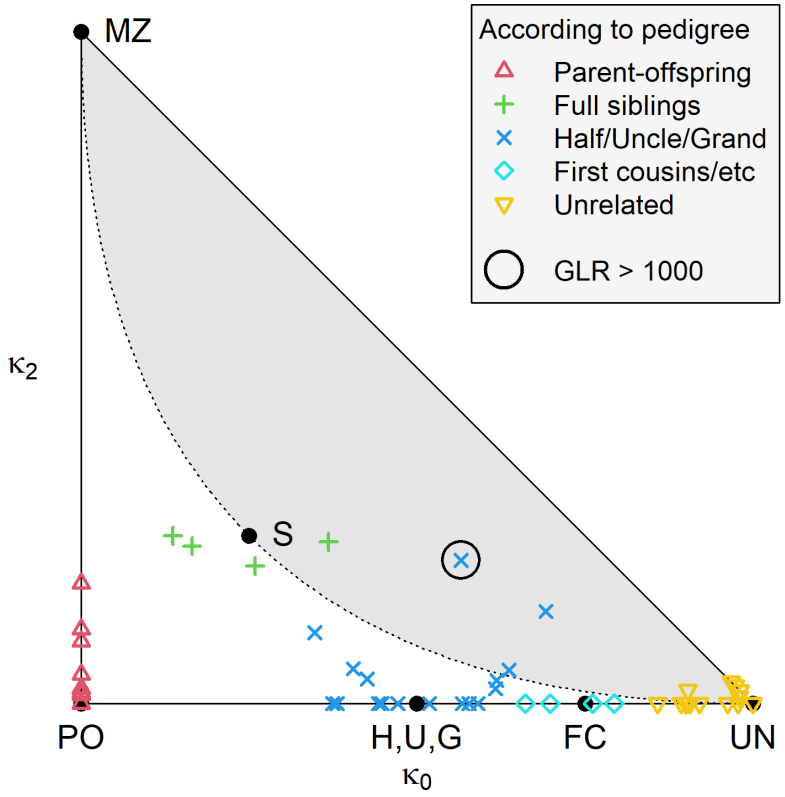
- Simulation

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

(Both of these support conditioning on known genotypes)

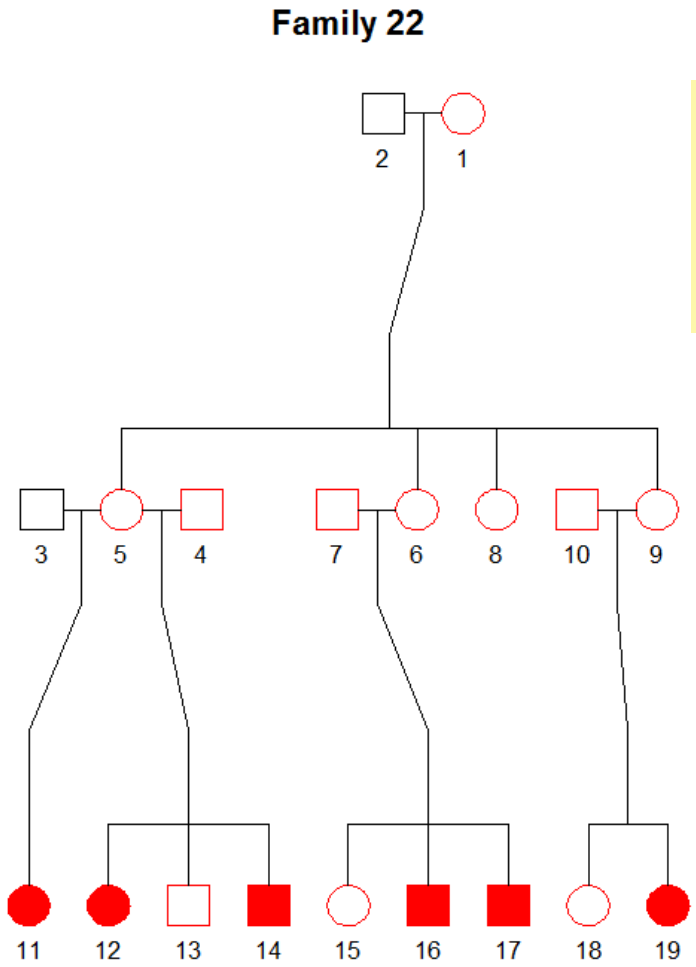
Application: Pedigree error detection

- For each pair of individuals, compute:
  1) pedigree-based kappa
  2) marker-based kappa estimate
- Compare them with an LR (actually: generalised LR*)
- Color-coded plot:
  - Position shows estimate
  - Color/shape shows pedigree claim

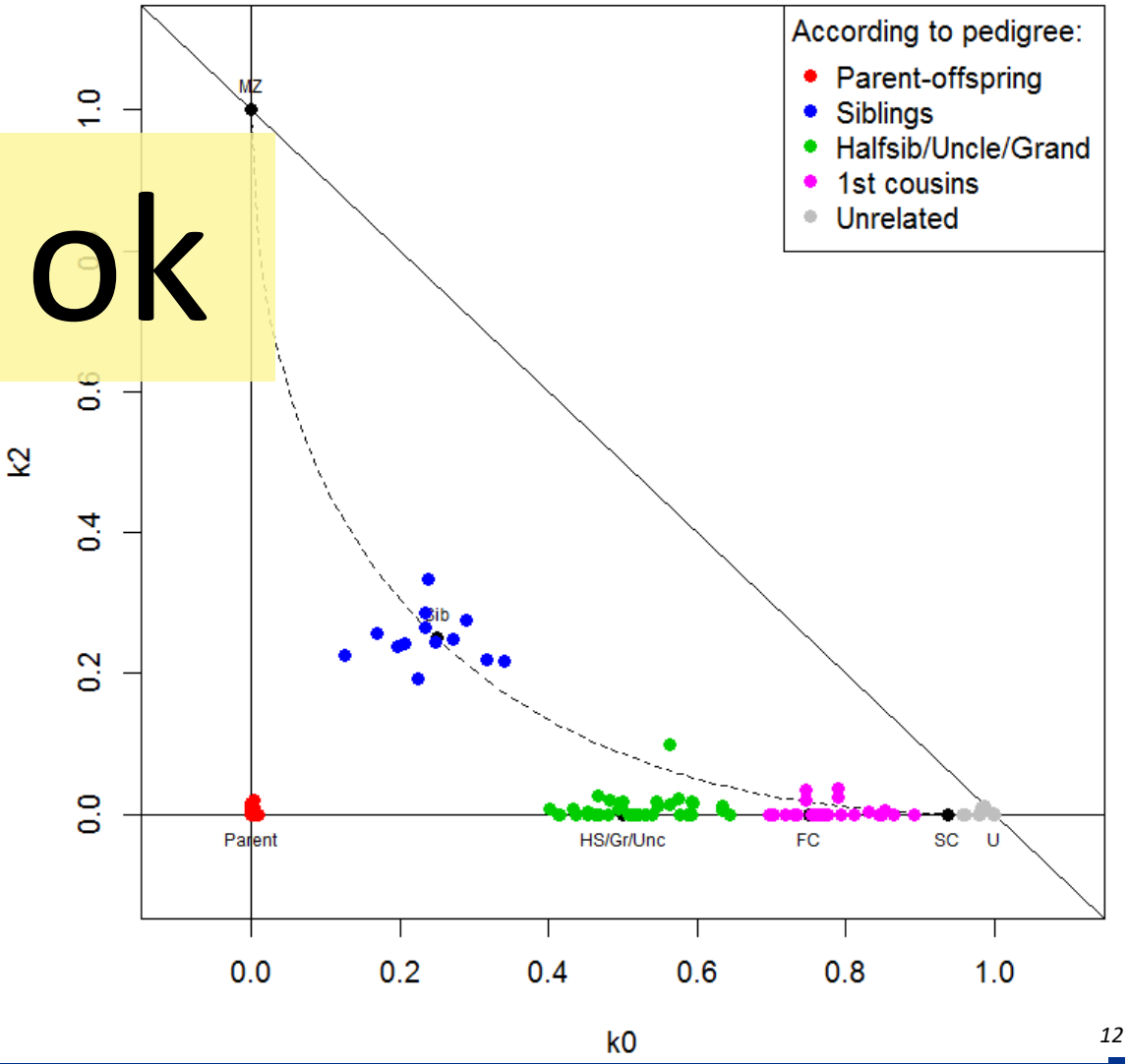


*Egeland & Vigeland (FSI:Genetics, 2025): *Kinship cases with partially specified hypotheses.*

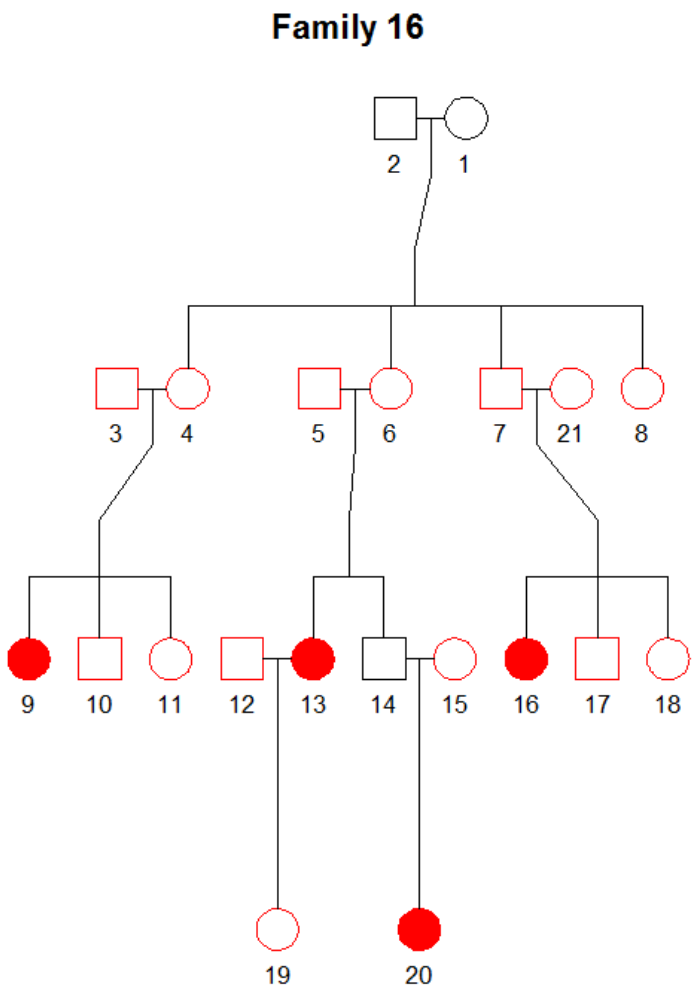
Example 1

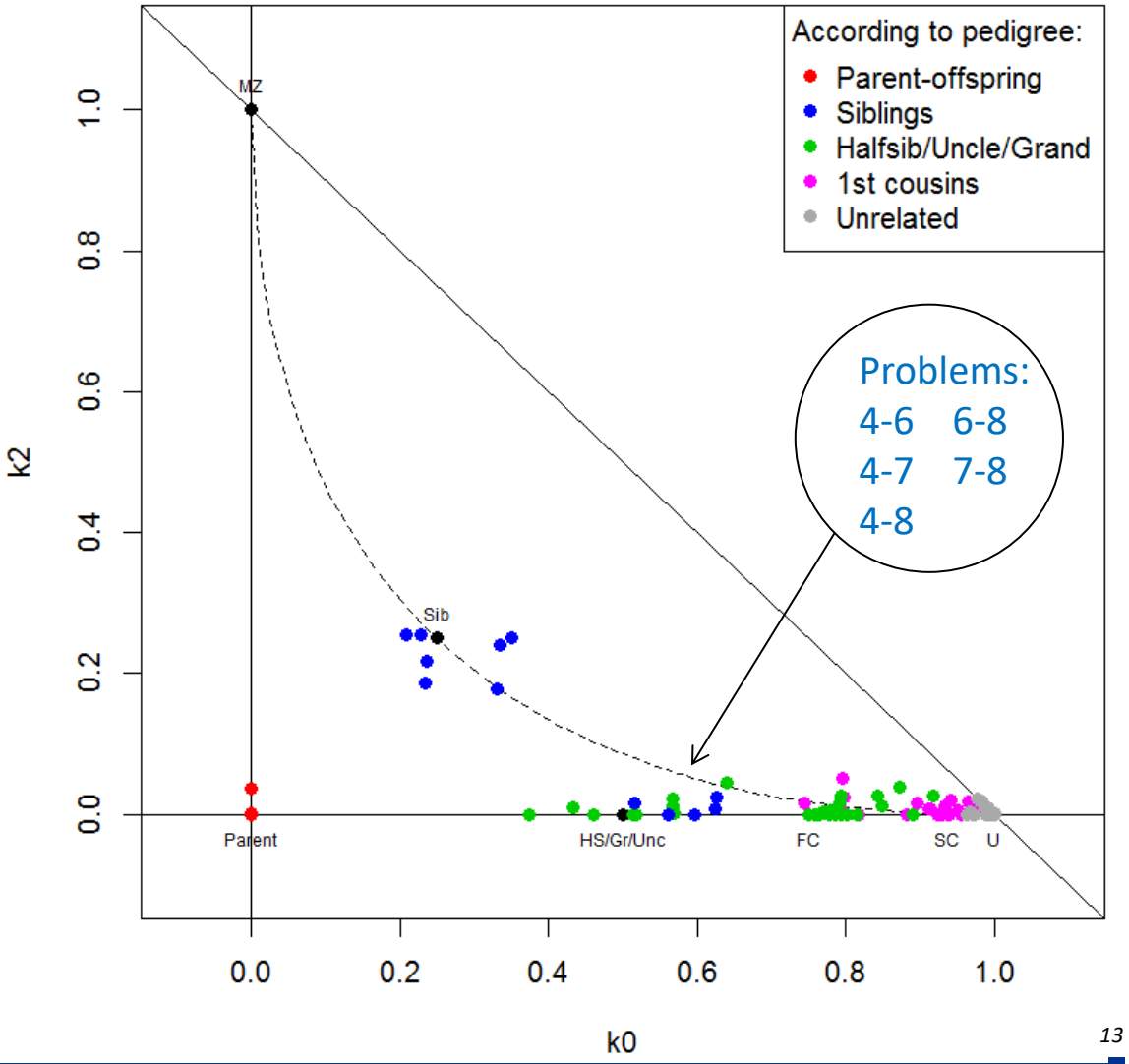

Family 22
2
1
3
5
4
7
6
8
10
9
11
12
13
14
15
16
17
18
19

ok

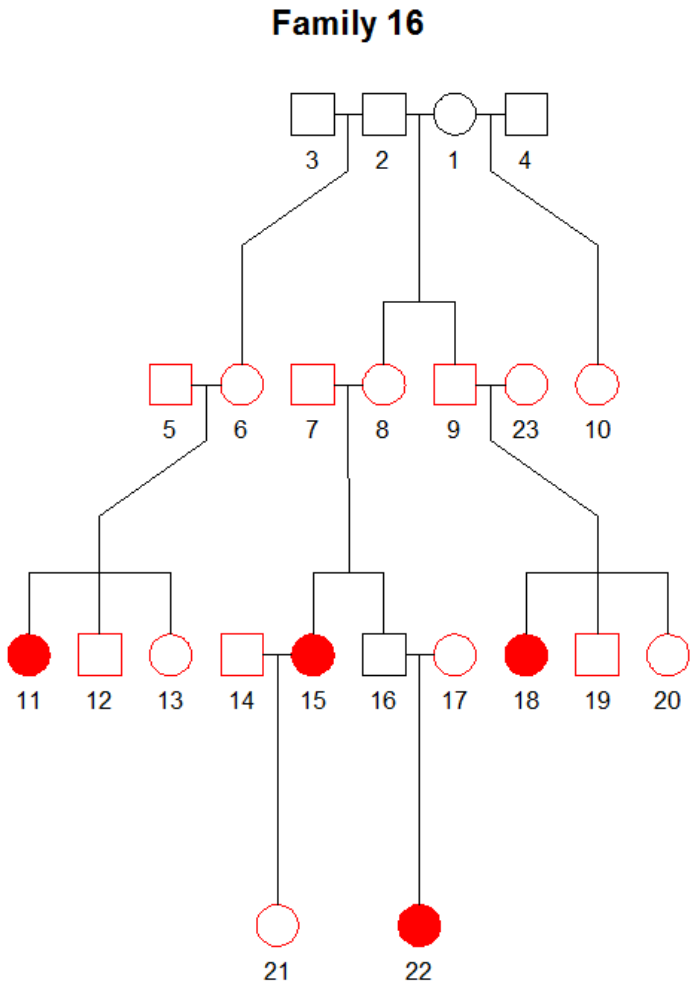

According to pedigree:
Parent-offspring
Siblings
Halfsib/Uncle/Grand
1st cousins
Unrelated
MZ
Sib
Parent
HS/Gr/Unc
FC
SC
U
k2
k0
0.0
0.2
0.4
0.6
0.8
1.0

Example 2

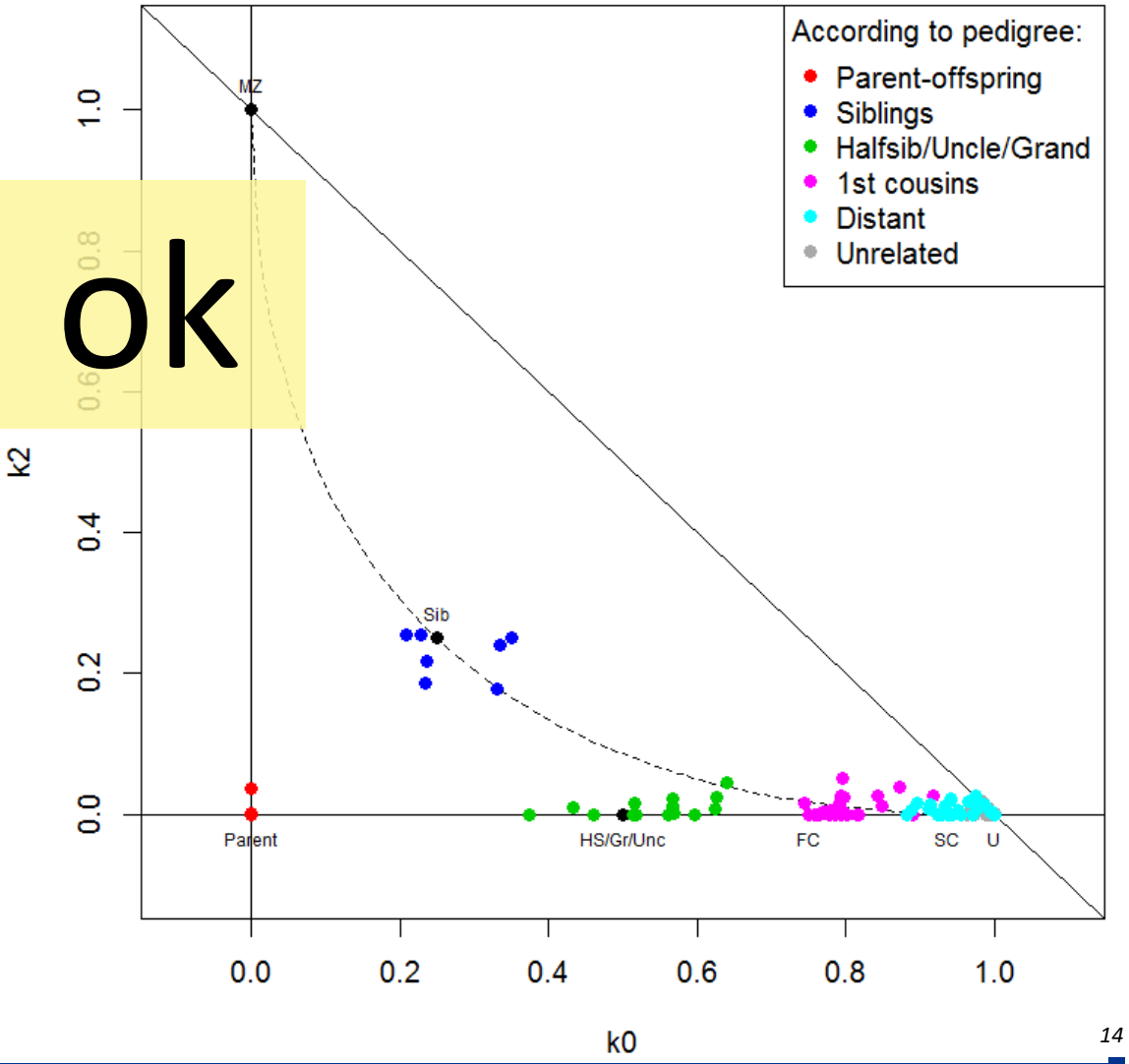

Family 16
2
1
3
4
5
6
7
21
8
9
10
11
12
13
14
15
16
17
18
19
20


According to pedigree:
Parent-offspring
Siblings
Halfsib/Uncle/Grand
1st cousins
Unrelated
Problems:
4-6 6-8
4-7 7-8
4-8
MZ
Sib
Parent
HS/Gr/Unc
FC
SC
U
k2
k0
0.0
0.2
0.4
0.6
0.8
1.0

Example 2 - corrected



ok



Pairwise inference: Main approaches

A. Based on coefficients
- Typically with STR markers
- Maximum-likelihood estim.
- Assumes independence

- Complexity: **Easy**
- Accuracy: **Poor** (except PO/MZ)
- Scope: **Close relationships**

OLD SCHOOL

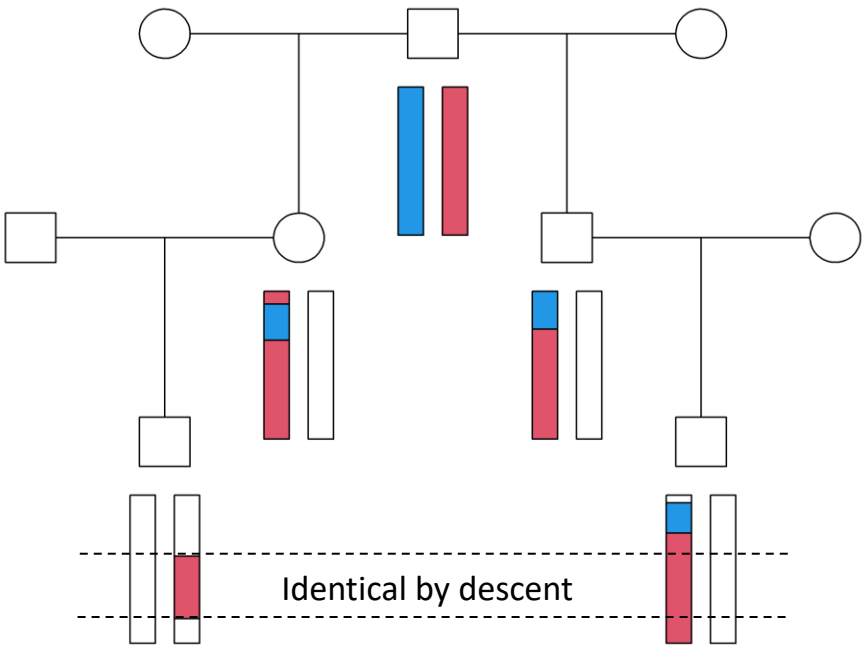
B. Based on IBD segments
- Requires lots of SNPs
- Two steps:
  1) SNPs → IBD segments
  2) IBD segments → relatedness

  (Often different software)

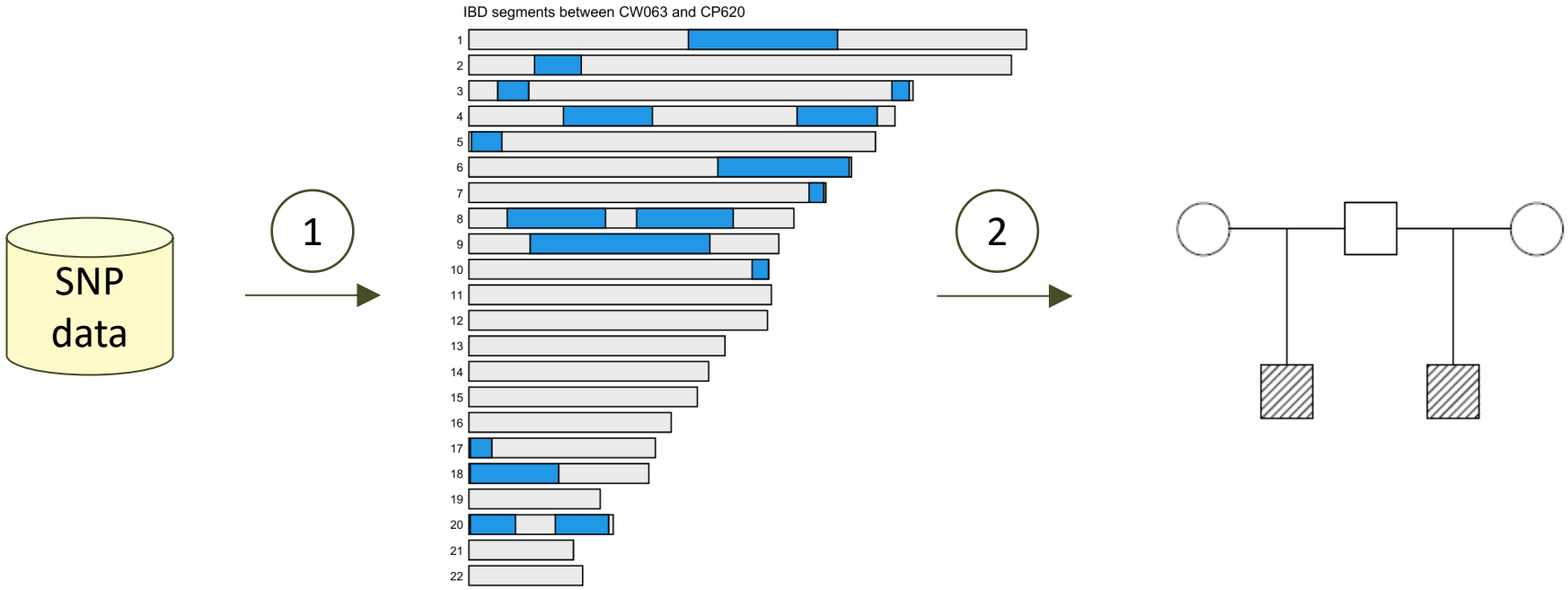
- Complexity: **Medium/high**
- Accuracy: **Better**
- Scope: **Close + distant**

MODERN

Remember: IBD comes in segments



Idea: Use observed IBD segments to infer the relationship


SNP data
1
IBD segments between CW063 and CP620
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
2

1 Detection of IBD segments

- At least 35 published programs!
- But beware: Some of them ...
  - require > N markers (N = huge number)
  - require > M samples
  - require phased data
  - require massive installation+setup
  - don't work
  - don't exist

Many unsuitable for forensic applications

	Software	Reference	Comments
2	PLINK	Purcell et al., 2007	Probabilistic,
3	BEAGLE HMM	Browning and Browning, 2007	Probabilistic,
4	GERMLINE	Gusev et al., 2009	Not probabili
5	Relate	Albrechtsen et al., 2009	Probabilistic,
6	KING	Manichaikul et al., 2010	
7	BEAGLE IBD	Browning and Browning, 2010	Probabilstic,
8	fastIBD	Browning and Browning, 2011	
9	DASH	Gusev et al., 2011	
10	MCMC_IBDfinder	Moltke et al., 2011	
11	IBDLD	Han and Abney, 2011	Probabilistic,
12	ERSA	Huff et al., 2011	
13	IBD_Haplo	Brown et al., 2012	Probabilistic,
14	Refined IBD	Browning and Browning, 2013a	
15	IBDseq	Browning and Browning, 2013b	
16	IBD-Groupon	He, 2013	
17	HapFABIA	Hochreiter, 2013	
18	PREST-plus	Sun and Dimitromanolakis, 2014	Likelihood
19	Parente2	Rodriguez et al., 2015	
20	FISHR	Bjelland et al., 2017	
21	hmmIBD	Schaffner et al., 2018	Hidden Mark
22	DRUID	Ramstetter et al., 2018	
23	TRUFFLE	Dimitromanolakis et al., 2019	
24	RaPID	Naseri et al., 2019	PBWT on pha
25	TRIBES	Twine et al., 2019	
26	FastSMC	Nait Saada et al., 2020	Hash/Extend
27	IBIS	Seidman et al., 2020	Sliding Windo
28	Hap-IBD	Zhou et al., 2020	Error adjuste
29	iLash	Shemirani et al., 2021	Locality sensi
30	Phaseibd/TPBWT	Freyman et al., 2021	TPWBT
31	ancIBD	Ringbauer et al., 2024	
32	COANCESTRY	Wang, 2011	
33	NGSremix		
34	NgsRelate		
35	IBDMAP	Bercovici et al., 2010	HMM

Popular choice w/ many SNPs: **IBIS**

GED match

Source code, docs & examples:
https://github.com/magnusdv/ibdfindr

ibdfindr:

An R package for detecting IBD segments from SNP data

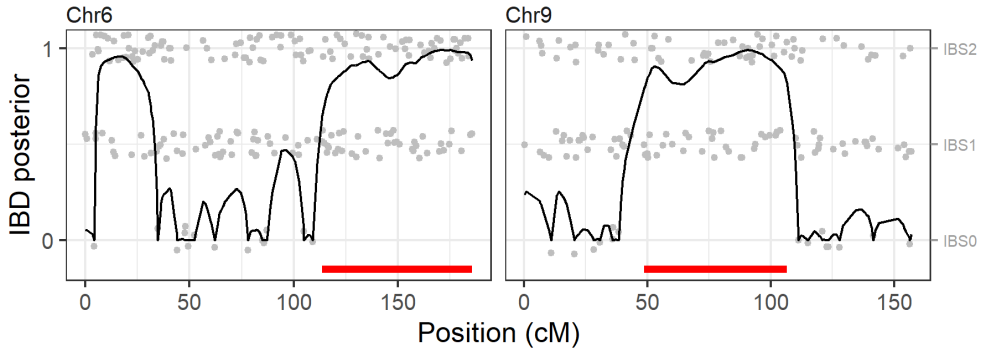
Input data

CHR	MARKER	CM	A1	FREQ1	CW063	CP620
1	rs944..	0.00	G	0.61	AG	AG
1	rs464..	0.33	C	0.64	AC	AC
1	rs1091..	1.24	T	0.63	CC	TT
1	rs669..	2.00	C	0.57	CC	CC
1	rs376..	4.81	T	0.59	GG	GT
1	rs736..	5.86	C	0.67	CT	CT
1	rs667..	6.77	G	0.56	AA	GG

R code

```
library(ibdfindr)
ibd = findIBD(data)
plotIBD(ibd, chrom = c(6,9))
```

- Fits a continuous-time **Hidden Markov Model**
- Predicts IBD segments (Viterbi algorithm)
- Finds posterior IBD probs (forward-backward algorithm)



2 Segments → relationships

Henrik Nordtorp

IBDrel

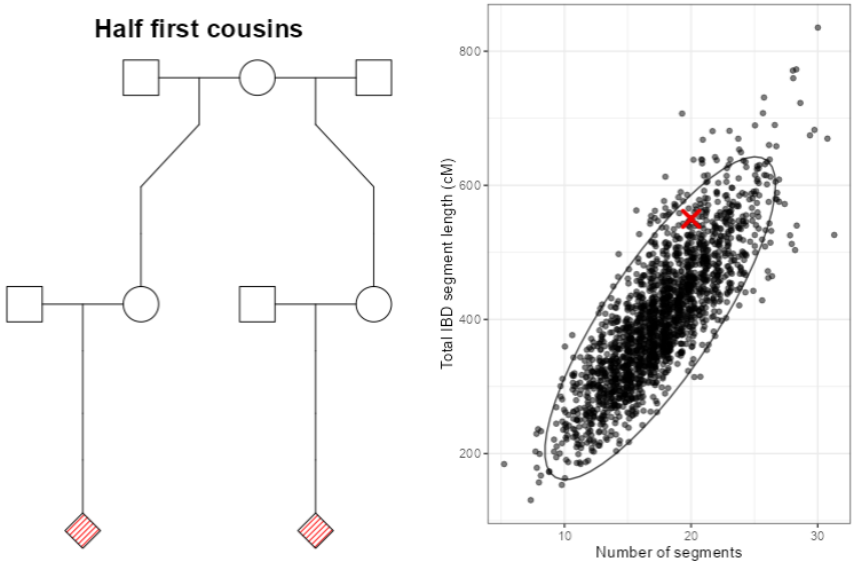
IBD Data

20 segments above threshold 7 cM with 550 cM total length.

Individual segments: 8, 13, 15, 19, 19, 20, 21, 24, 24, 25, 25, 26, 28, 29, 31, 33, 35, 45, 49, 61

Plot

Pedigree 1/2



The selected class contains 2 indistinguishable pedigrees.

Relatedness

Sex | Rel | Kap | Kin | Deg

		Rel	Sex	Posterior	κ	ϕ	Degree
1	✓	G^2 grandparent	m	**0.2**	(7/8,1/8,0)	1/32	4
2	✓	Half 1st cousins	m	**0.17**	(7/8,1/8,0)	1/32	4
3	✓	G^2 avuncular	m	**0.12**	(7/8,1/8,0)	1/32	4
4	✓	G grandparent	m	**0.1**	(3/4,1/4,0)	1/16	3
5	✓	Half Avuncular	mp	**0.07**	(3/4,1/4,0)	1/16	3
6	✓	1st cousins 1R	m	**0.05**	(7/8,1/8,0)	1/32	4
7	✓	1st cousins 1R	p	**0.04**	(7/8,1/8,0)	1/32	4
8	✓	Half Avuncular	m	**0.04**	(3/4,1/4,0)	1/16	3
9	✓	G avuncular	mp	**0.04**	(3/4,1/4,0)	1/16	3
10	✓	Half Avuncular	pm	**0.04**	(3/4,1/4,0)	1/16	3
11	✓	G avuncular	p	**0.03**	(3/4,1/4,0)	1/16	3
12	✓	G^2 avuncular	p	**0.03**	(7/8,1/8,0)	1/32	4
13	✓	G avuncular	pm	**0.02**	(3/4,1/4,0)	1/16	3
14	✓	1st cousins	p	**0.02**	(3/4,1/4,0)	1/16	3
15	✓	G avuncular	m	**0.02**	(3/4,1/4,0)	1/16	3
16	✓	G grandparent	mp	**0.01**	(3/4,1/4,0)	1/16	3

☑ **Normalize** ☐ **Filter**

2 Segments → relationships

Henrik Nordtorp

IBDrel

IBD Data

20 segments above threshold 7 cM with 550 cM total length.

Individual segments: 8, 13, 15, 19, 19, 20, 21, 24, 24, 25, 25, 26, 28, 29, 31, 33, 35, 45, 49, 61

Relatedness

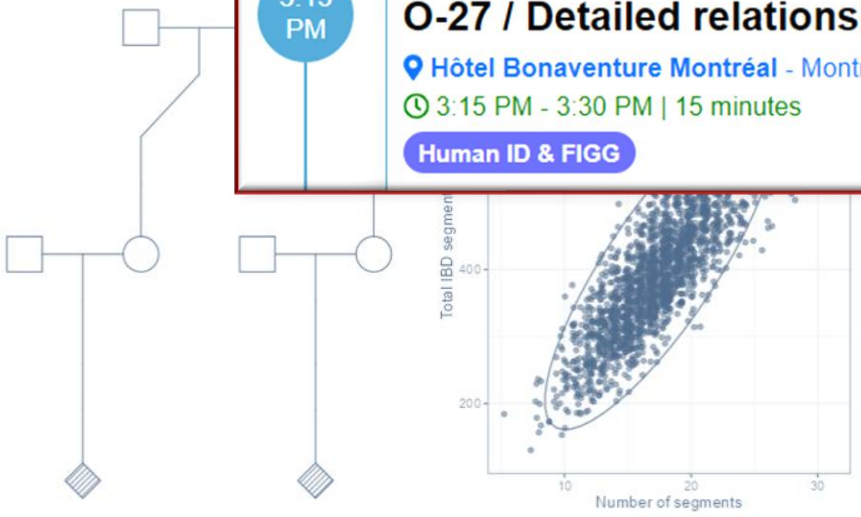
Sex Rel Kap Kin Deg

		Rel	Sex	Posterior	κ	ϕ	Degree
1	☑	G^2 grandparent	m	0.2	(7/8,1/8,0)	1/32	4
11	☑	G avuncular	p	0.03	(3/4,1/4,0)	1/16	3
12	☑	G^2 avuncular	p	0.03	(7/8,1/8,0)	1/32	4
13	☑	G avuncular	pm	0.02	(3/4,1/4,0)	1/16	3
14	☑	1st cousins	p	0.02	(3/4,1/4,0)	1/16	3
15	☑	G avuncular	m	0.02	(3/4,1/4,0)	1/16	3

Normalize Filter

Plot Pedigree 1/2

Half first co



The selected class contains 2 indistinguishable pedigrees.

Thursday August 20th, 2026

Time Zone: (GMT-04:00) Eastern Time (US & Canada)

3:15 PM

O-27 / Detailed relationship predictions for FIGG with ibdrel

Hôtel Bonaventure Montréal - Montréal Ballroom

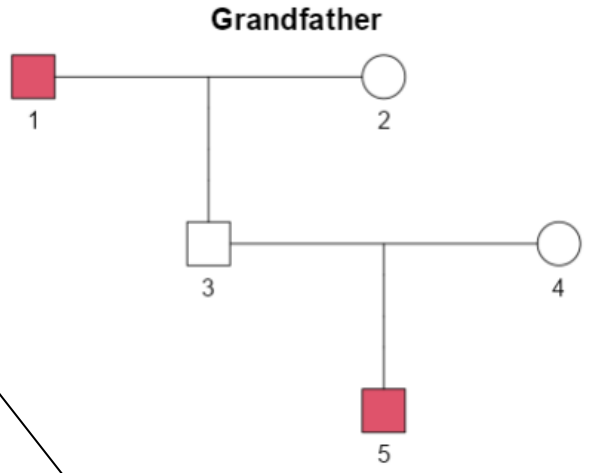
3:15 PM - 3:30 PM | 15 minutes

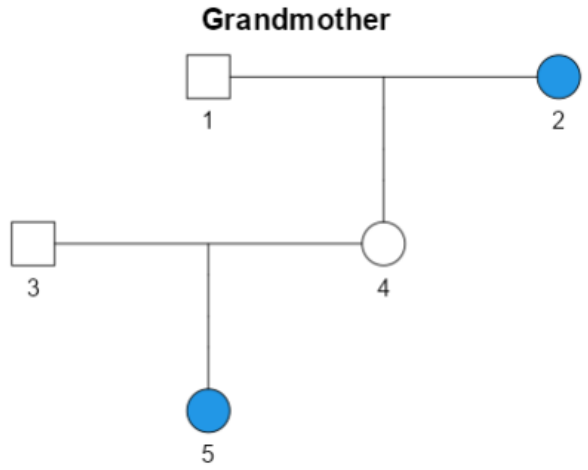
Human ID & FIGG

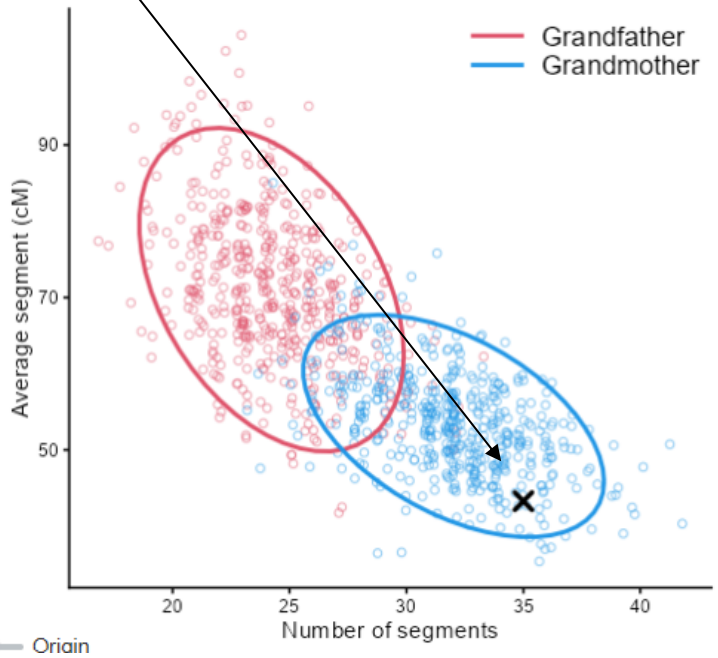
2 Alternative: ibdsim2 (For comparing specific hypotheses)

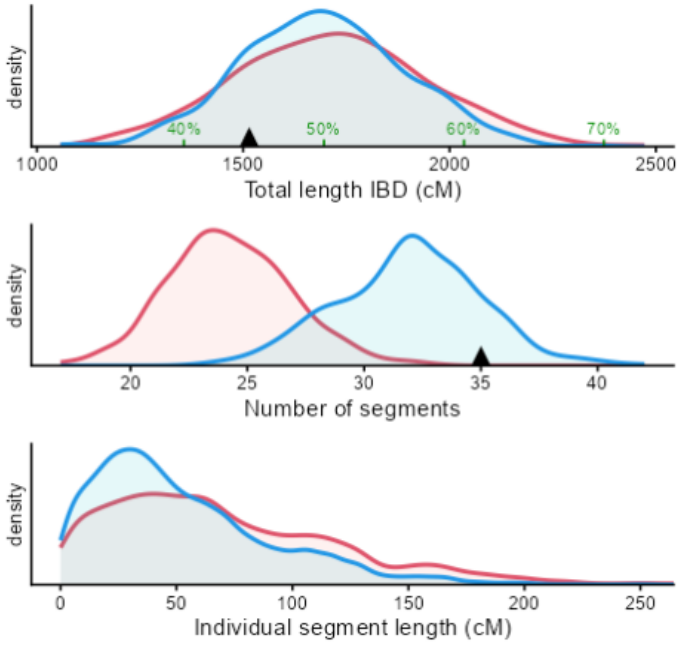
Example:
35 segments
Total 1500 cM

Conclusion:
Grandmother

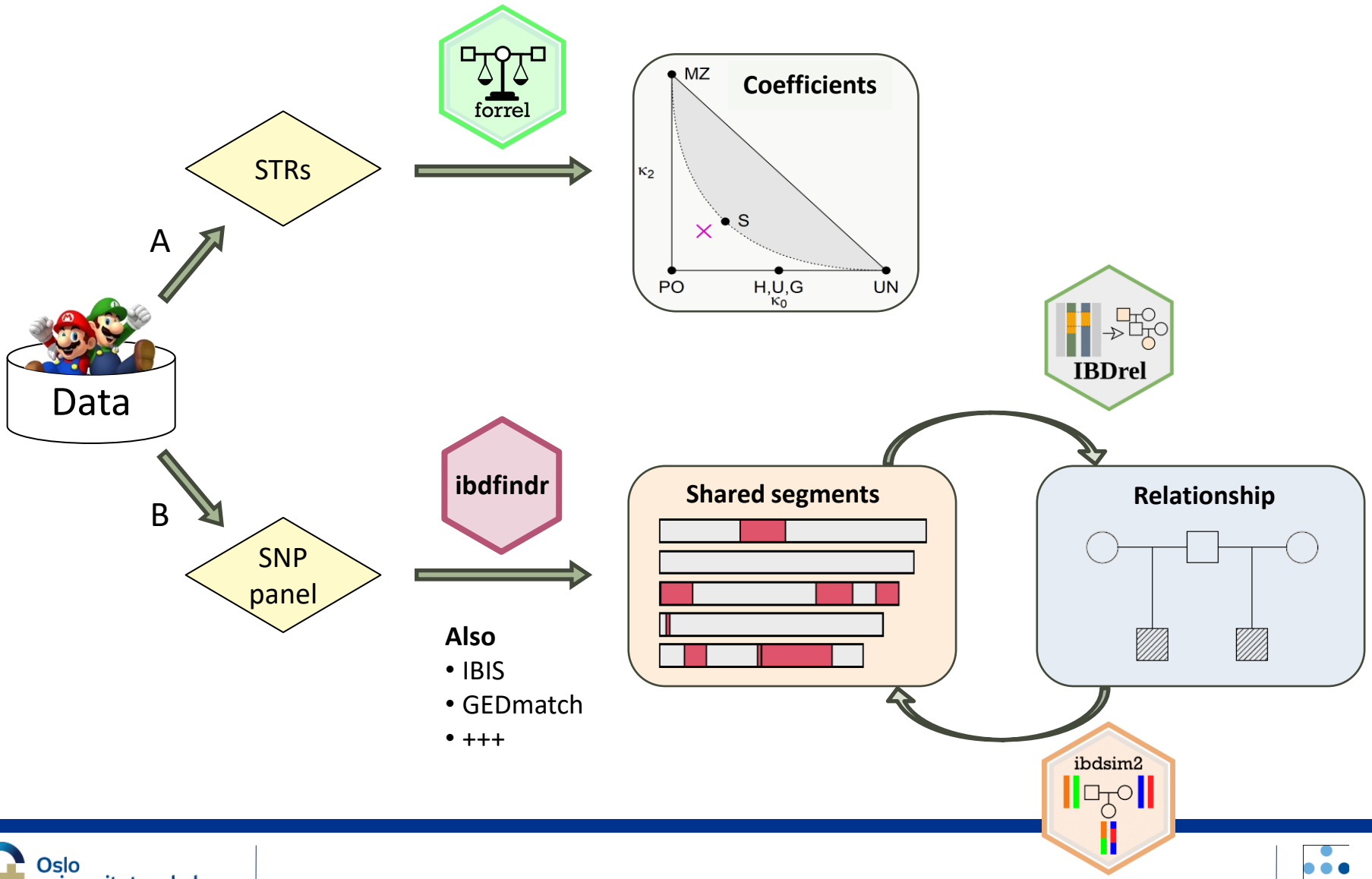






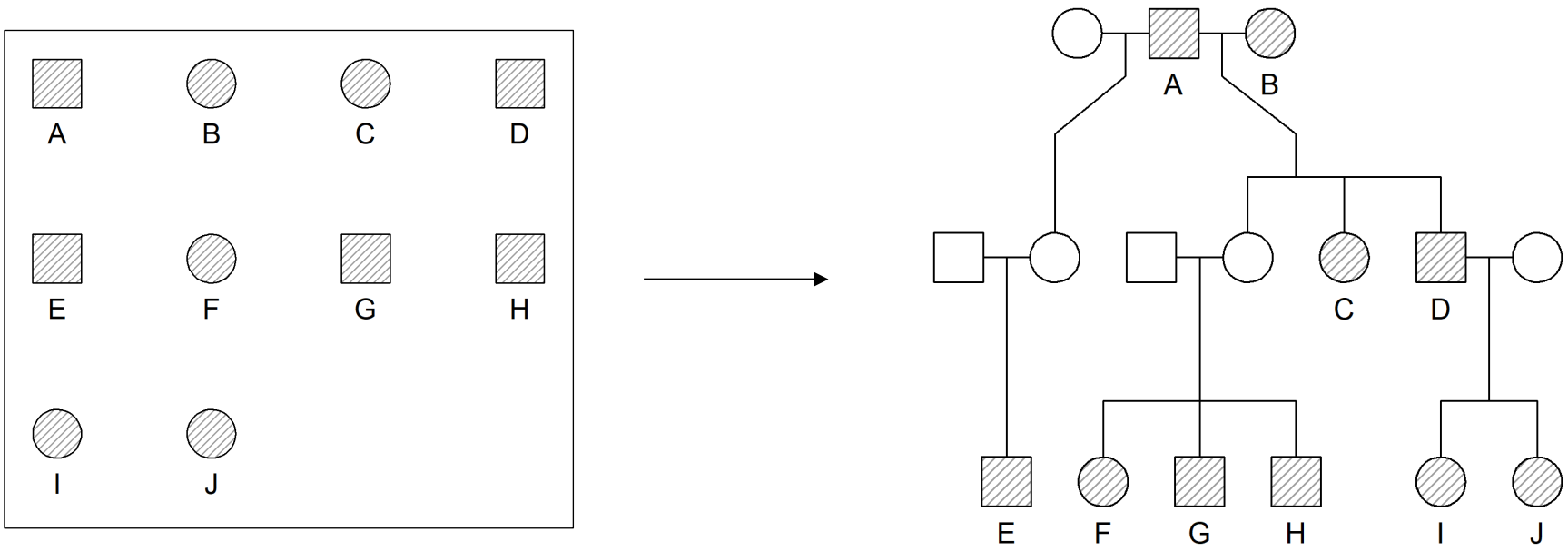


Summary: Tools for pairwise kinship inference



Part II: Pedigree reconstruction

Pedigree reconstruction: Ultimate goal



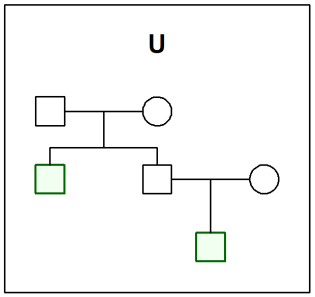
Generally impossible - even in theory!

Data

A/A	A/C
G/A	A/A
C/C	C/C
C/G	G/G
...	...

?

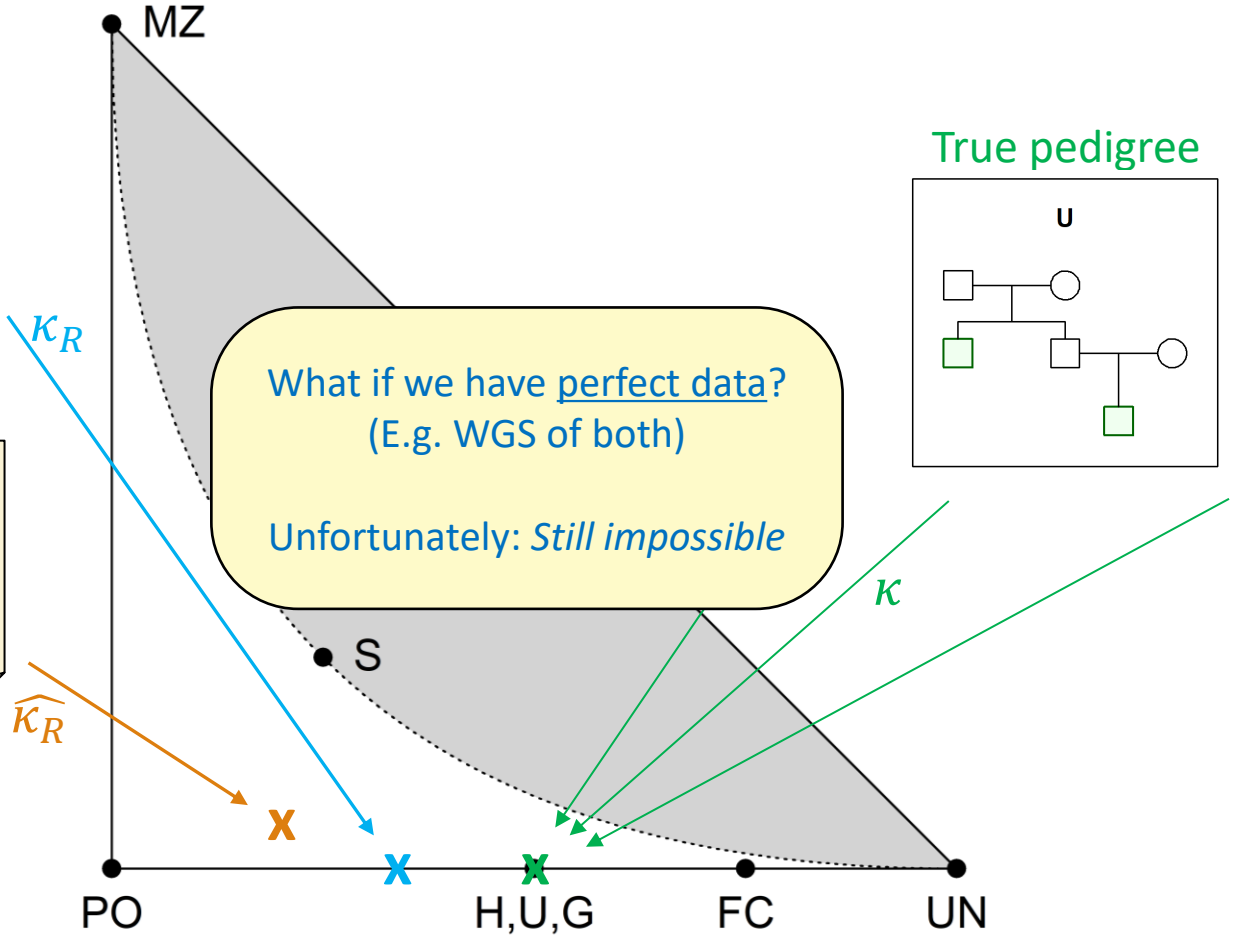
True pedigree



Generally impossible - even in theory!

Data

A/A	A/C
G/A	A/A
C/C	C/C
C/G	G/G
...	...

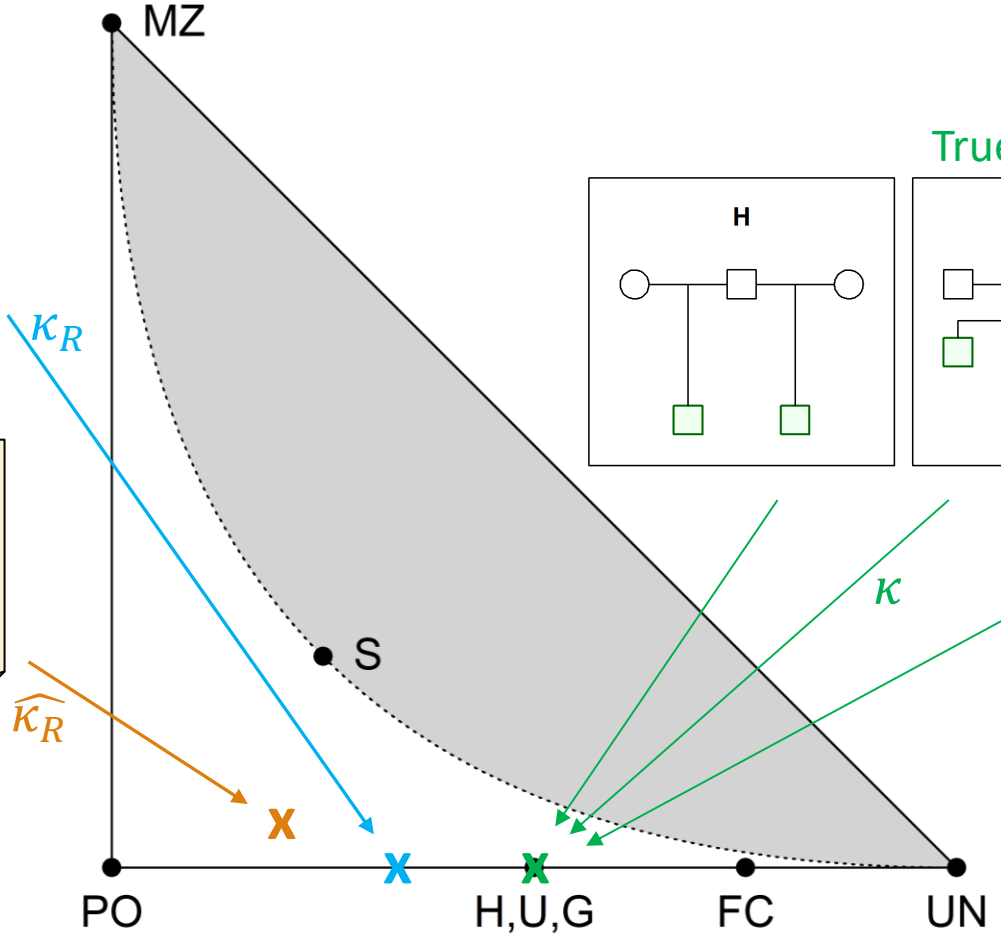


True pedigree

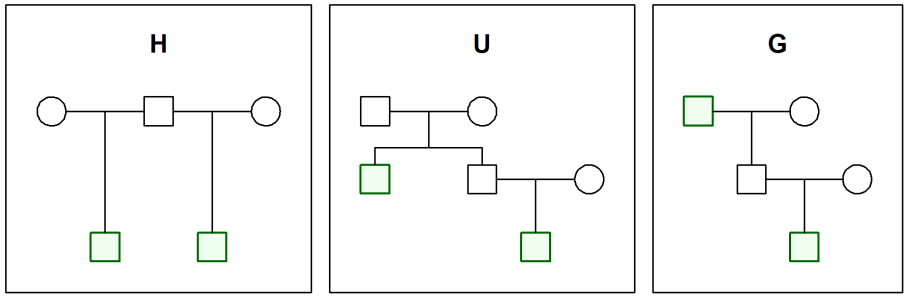
Generally impossible - even in theory!

Data

A/A	A/C
G/A	A/A
C/C	C/C
C/G	G/G
...	...

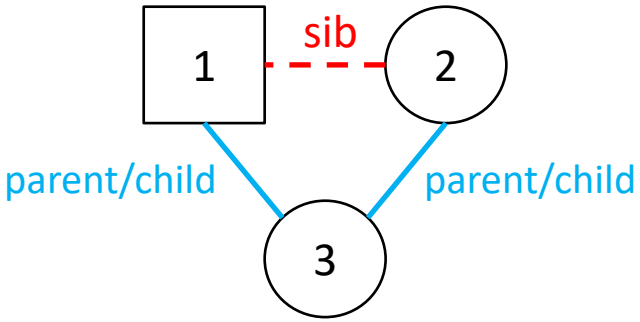


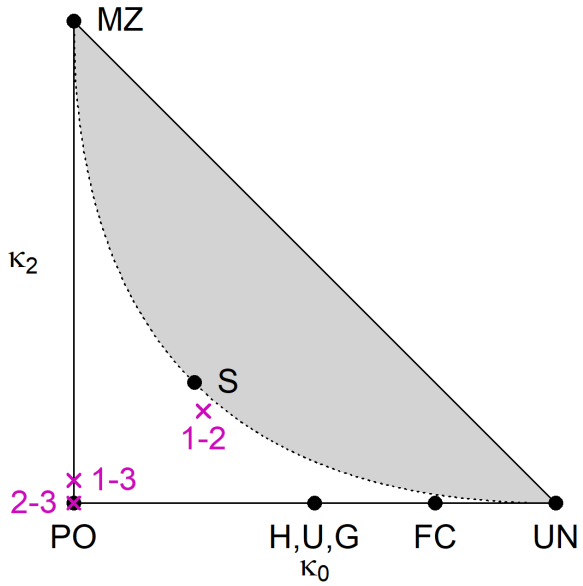
True pedigree

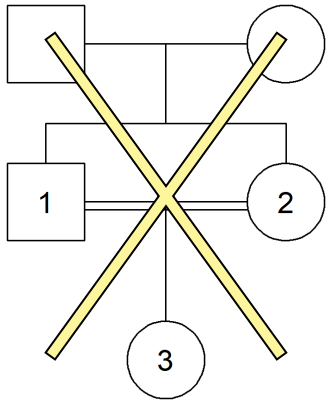


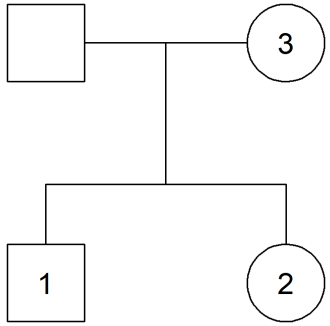
What *can* we trust?
- Parent-offspring
- Full sibs if good data

Pedigree reconstruction: The "puzzle" approach









Step 1: Sex

Step 2: Pairwise estimates

- Connect parent-child
- Exploit siblings

Step 3: Solve the puzzle!

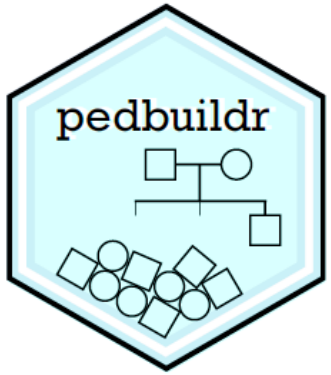
Better approach: Maximum likelihood

Idea:

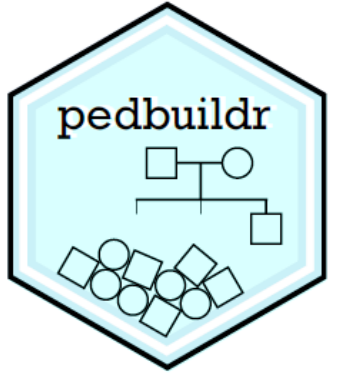
- Generate a list of "all possible" pedigrees connecting the individuals
- Compute the likelihood of each pedigree
- Sort and output the best pedigrees

Key functions:

```
> buildPeds()     # generate pedigrees

> reconstruct()   # main function!
> plot()          # plot top hits
```

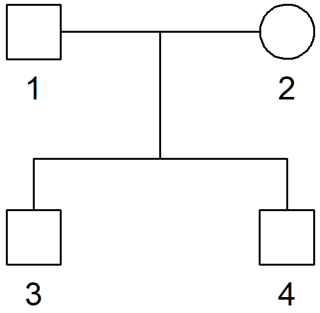


pedbuildr: Example



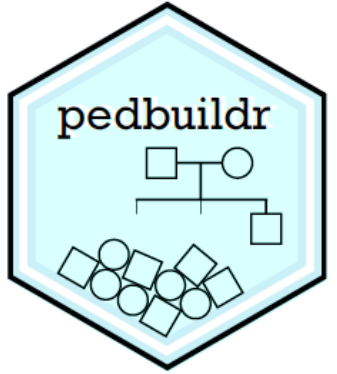
Simulate 23 STRs in family quartet

```
>   library(pedsuite)

>   # Frequency database
>   db = norSTR::norwayDB[1:23]

>   # Simulate
>   x = nuclearPed(2) |>
      profileSim(markers = db, seed = 1)

>   # Inspect the result
>   x
  id fid mid sex D1S1656 TPOX D2S1360
   1   *   *   1 12/15.3 8/11   22/23
   2   *   *   2   13/15  8/8   26/28
   3   1   2   1 15/15.3 8/11   23/28
   4   1   2   1   12/15 8/11   23/28
 Only 3 (out of 23) markers are shown.
```



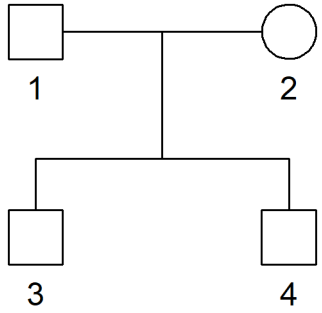
Next: Reconstruct the pedigree from this data!

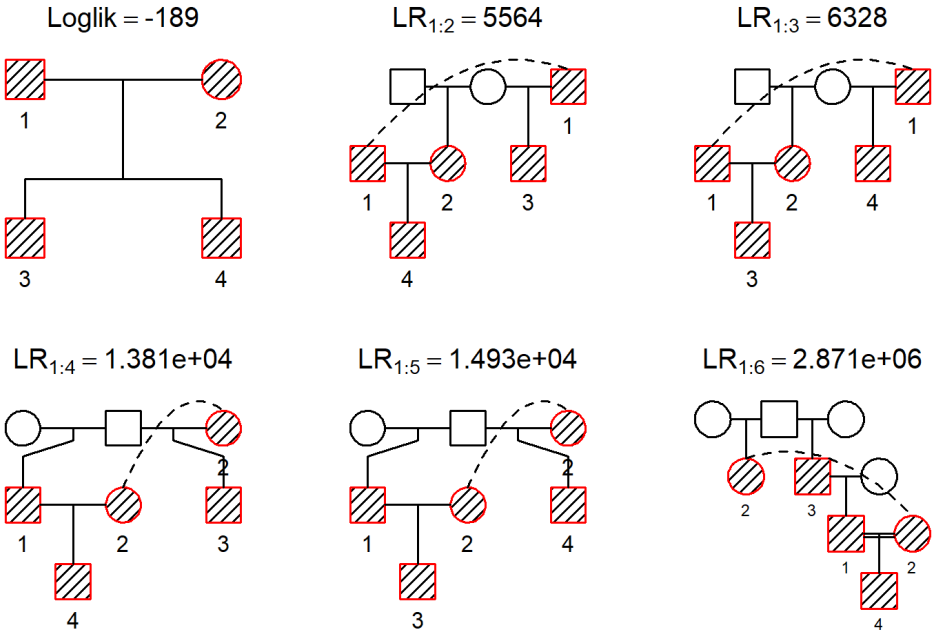
pedbuildr: Example



Reconstruct the pedigree

```
>   library(pedbuildr)

>   r = reconstruct(x)
Pedigree parameters:
  ID labels: 1, 2, 3, 4
  Sex: 1, 2, 1, 1
  Extra: parents
  Age info: -
  Connected only: TRUE
  Max inbreeding: 0.0625
  ...

Building pedigree list:
  ...
Computing the likelihood of 983 pedigrees

>   plot(r, top = 6)
```

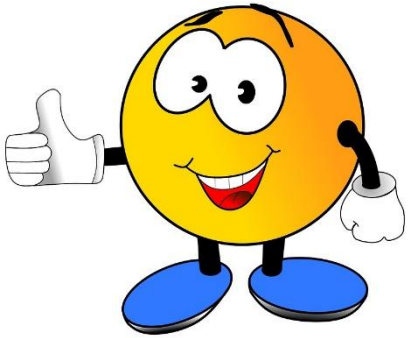


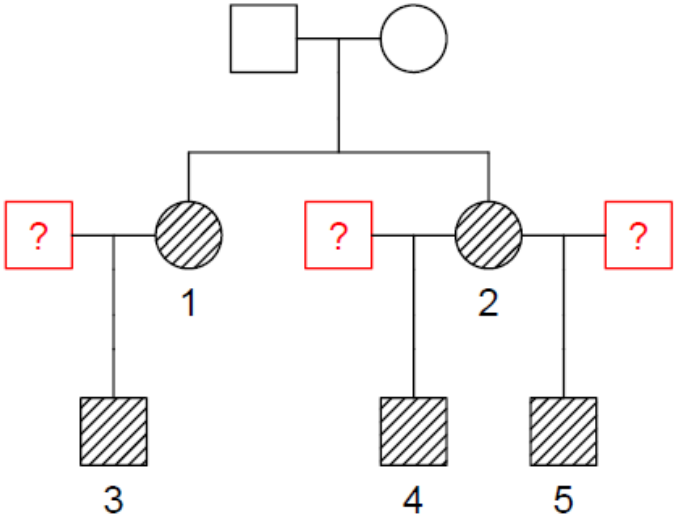


Optional parameters for restricting the search space

- **`extra`**: The max number of connecting individuals
  - default: **`extra = "parents"`** (suitable for small datasets)
- **`maxInbreeding:`** Default: 0.0625 (= first cousins)
- **`age`**: A vector of (relative) ages OR age inequalities, e.g. "Al > Bob"
- **`inferPO`**: If TRUE, an initial stage of pairwise IBD estimation is done
- **`knownPO`**: Known parent–offspring pairs
- **`allKnown`**: Is **`knownPO`** the *complete* list of POs?
- **`notPO`**: Pairs known not to be parent–offspring
- **`noChildren`**: Individuals known to have no children
- **`linearInb`**: Max incestuous generation gap (default: 0)
- **`connected`**: Set to FALSE to allow disconnected pedigrees
- **`sexSymmetry`**: Remove 'symmetric' versions. Default: TRUE

Your turn: Exercises!





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