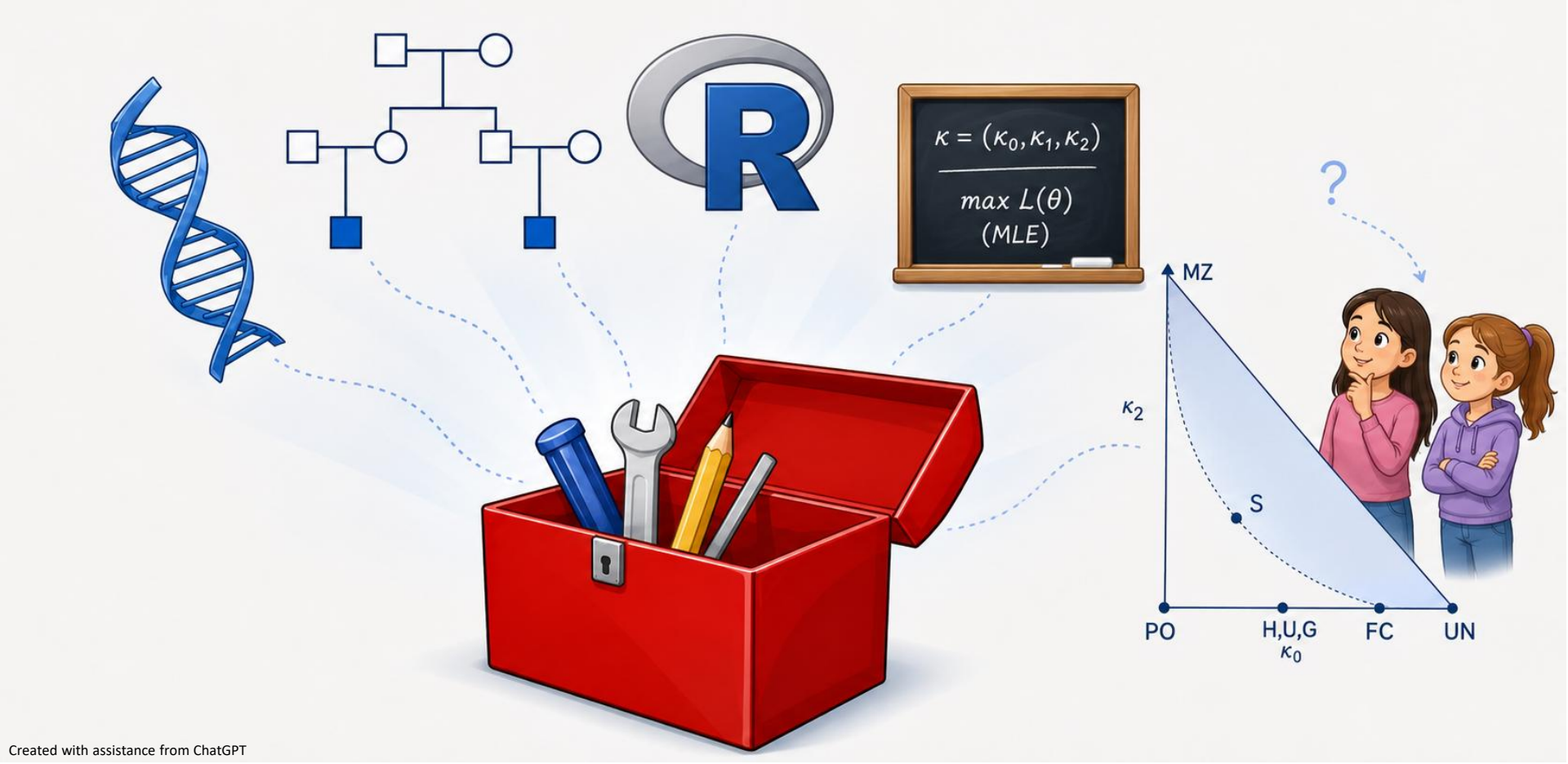


Created with assistance from ChatGPT

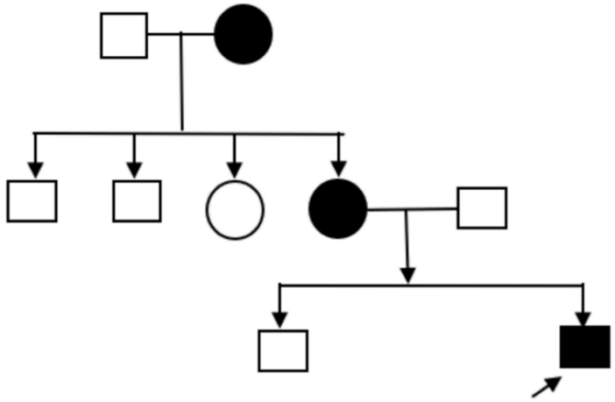
Lecture 2: Statistical tools for kinship and DVI

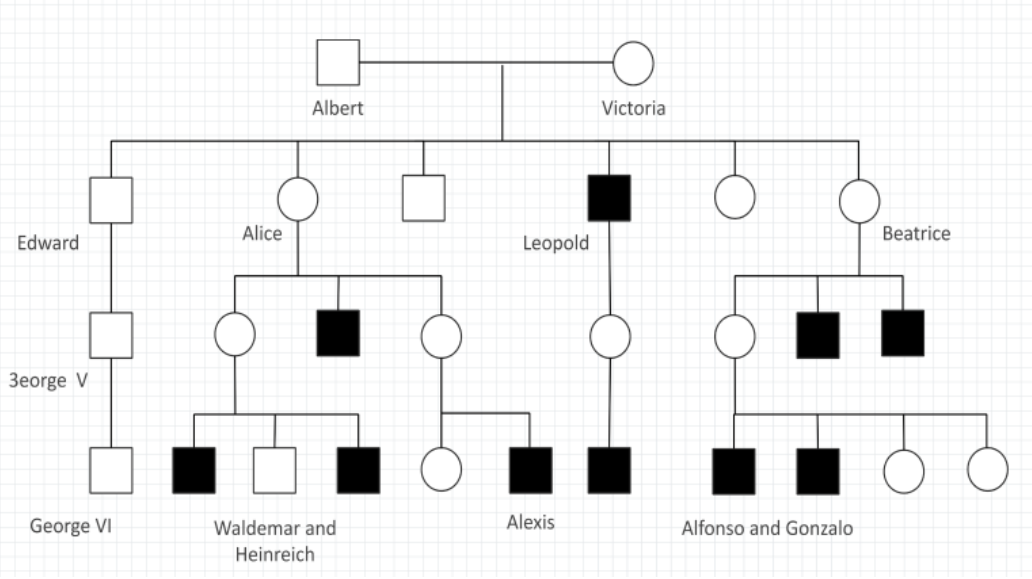
Magnus Dehli Vigeland

GHEP-ISFG 2026, Mexico

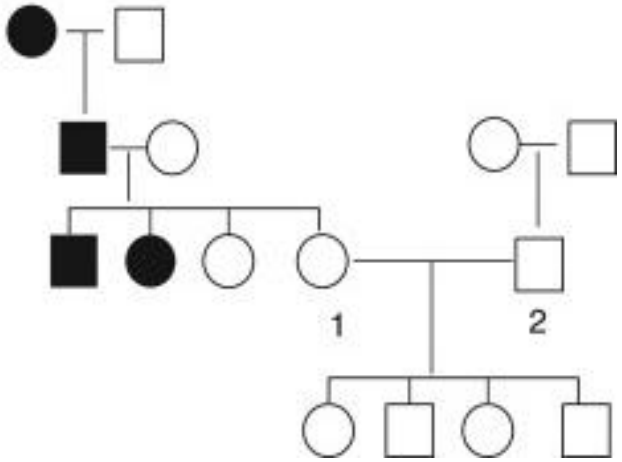
Outline

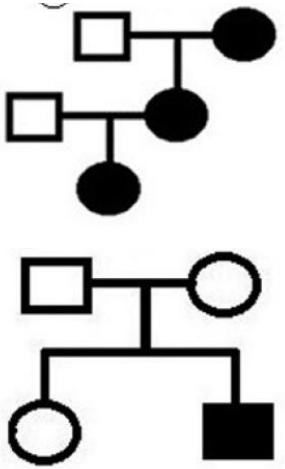
- Tool 1: *Pedigrees*
- Tool 2: *Measures of relatedness*
- Tool 3: *The relatedness triangle*
- Tool 4: *Estimating relatedness*

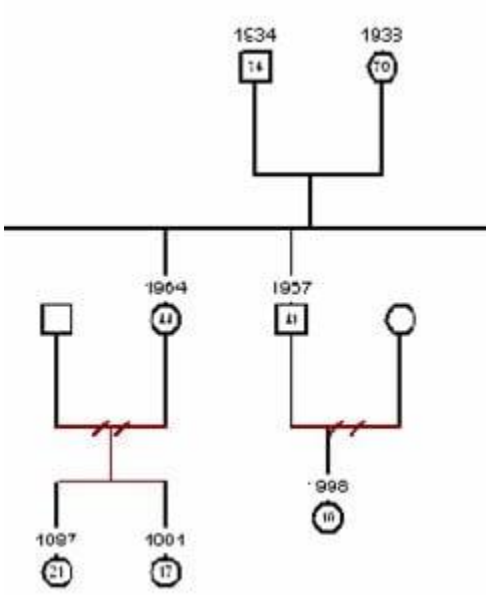




Tool 1: Pedigrees







What software exists to create pedigrees?

It depends:

- medical genetics
- forensic genetics
- animal pedigrees
- amateur genealogy

In this course:

- QuickPed
- DIVIANA

https://magnusdv.shinyapps.io/quickped/

QuickPed: An Interactive Pedigree Creator

New app design!
Discover the new features
Or stay with the old version: QuickPed3

Purpose: QuickPed lets you rapidly create attractive pedigree plots, save them as images or text files, and analyse the relationships within them.

Instructions: Choose a suitable start pedigree and modify it by clicking on individuals and using appropriate buttons. For example, to add a male child, select the parent(s) and press the icon. Check out the online user manual for various tips and tricks, including an introduction to relatedness coefficients.

Citation: If you use QuickPed in a publication, please cite this paper: Vigeland MD (2022). QuickPed: an online tool for drawing pedigrees and analysing relatedness. *BMC Bioinformatics*, **23**. DOI:10.1186/s12859-022-04759-y.

Quick start 1

Built-in pedigree: Trio

or

Load a ped file

or

Random pedigree

Reset all

Modify 2

Add

Sex

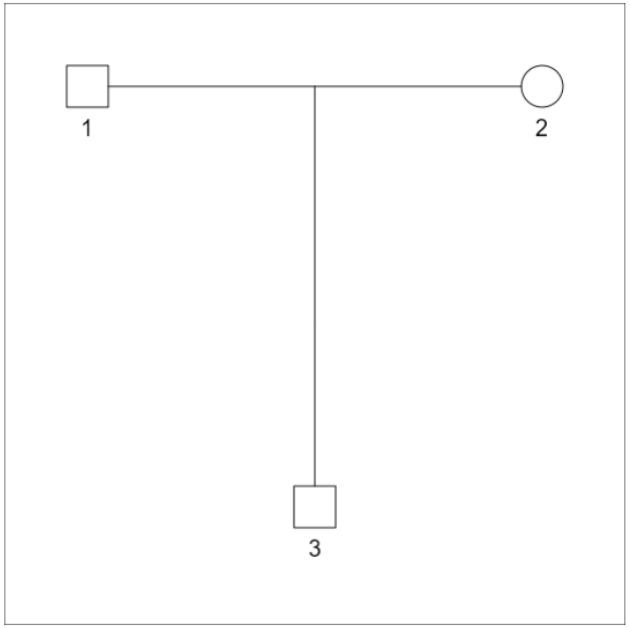
Style

Fill

Twins: MZ / DZ

Remove

Undo



Double-click on an individual to add text

Labels 3

1, 2, 3, .. | I-1, I-2, ..

Show all | Hide all

1
2
3

Update

Plot settings

Width: 430
Height: 430
Cex: 1,4
Symbols: 1
Margins: 3

Other options (beta)
- Straight legs
- Arrows

R code

PNG | PDF

Relationships

4

Ped file 5

Include
- Headers
- Family ID
- Affection status

Save ped file

Oslo universitetssykehus

Tool 2: Measures of relatedness

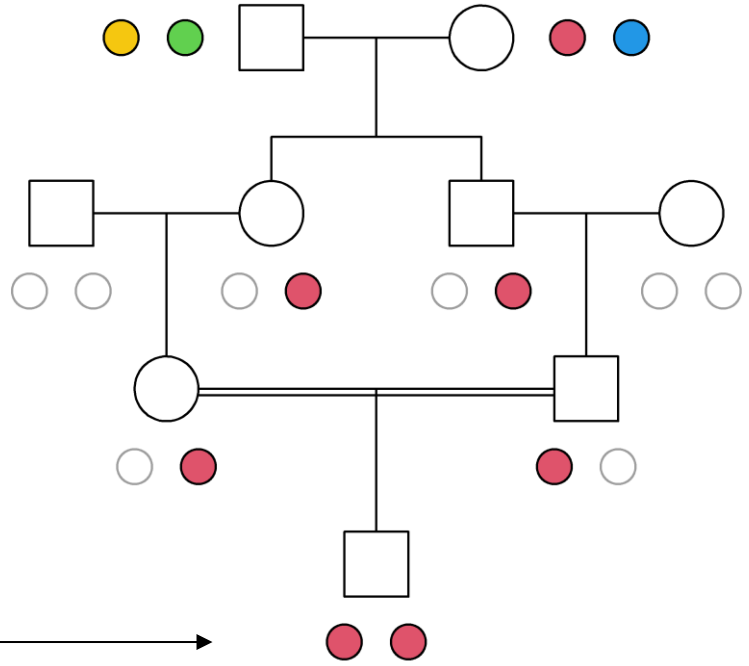
Possible answers

- being connected by family
- having a common ancestor...
  (not too far back)
- sharing DNA ...
  (more than unrelated people)

To make this precise, we need some terminology!

IBD and autozygosity

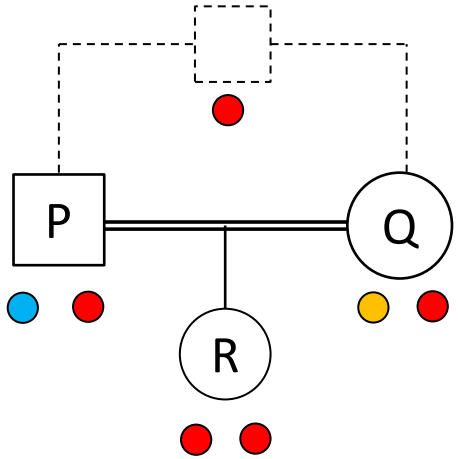
- identity by descent (IBD)
  - when alleles have a common origin **in the given pedigree**



- autozygous
  - homozygous; alleles are IBD

Inbreeding coefficient
$f = Pr(autozygosity)$

Coefficient of kinship/inbreeding



Sewall Wright
(1889 - 1988)

- Wright (1921): The kinship coefficient φ

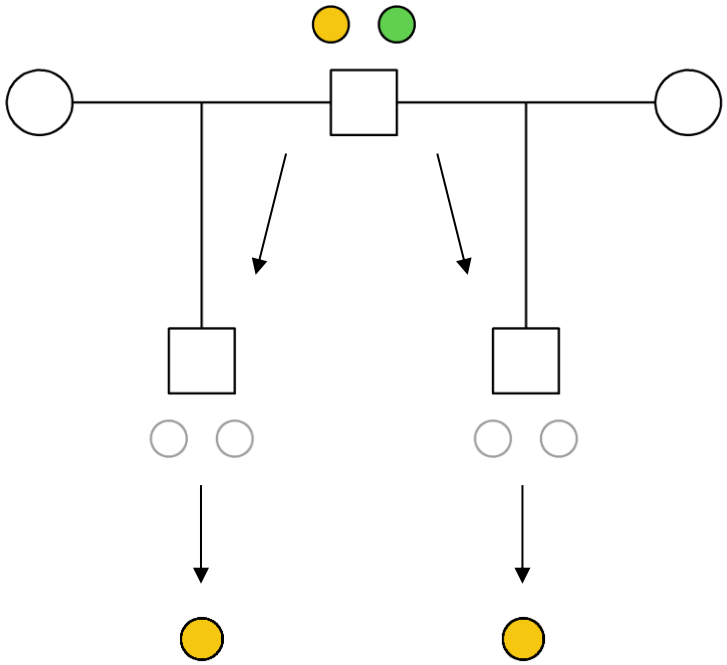
  $\varphi_{P,Q}$ = *Pr*(P and Q emit IBD alleles)

  = *Pr*(R is autozygous)

  = f_R

Kinship of parents = inbreeding of child

Example 1: Kinship coefficient of half siblings

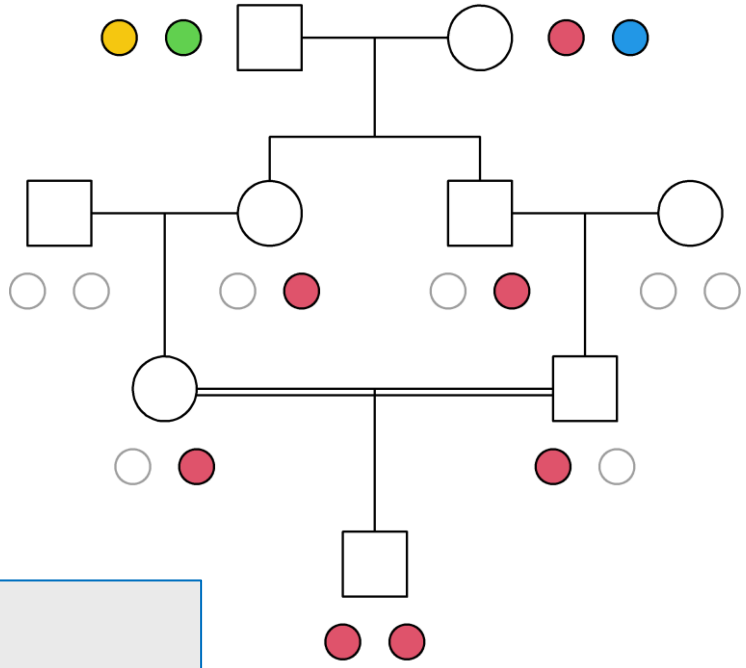


Kinship coefficient

$$\varphi = P(\text{● from both}) \cdot 2$$
$$= 0.5^4 \cdot 2$$
$$= 1/8$$

green

Example 2: Inbreeding coefficient



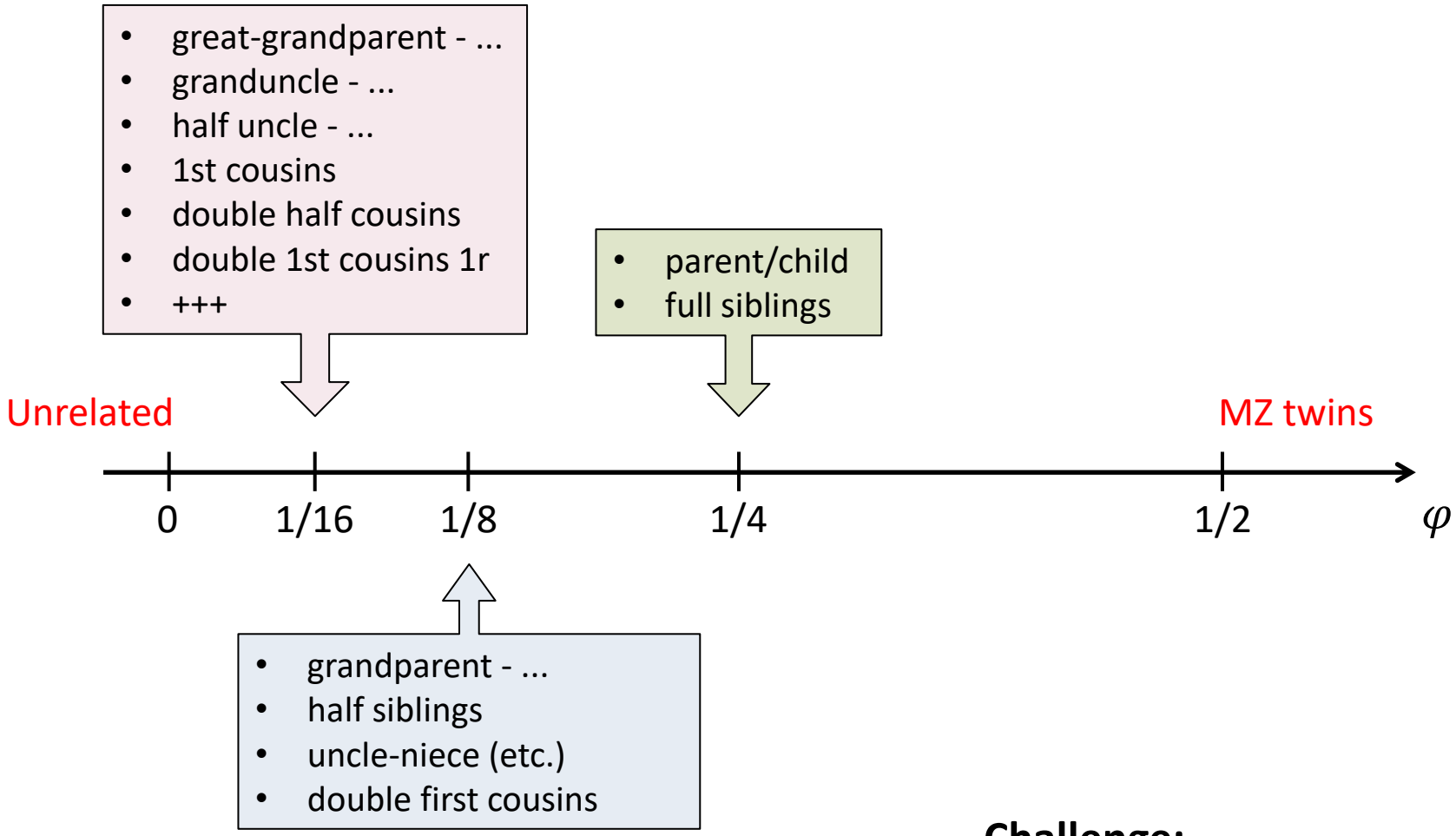
By hand

$$f = P(\text{●} \text{ autozygous}) \cdot 4$$
$$= 0.5^6 \cdot 4$$
$$= 1/16$$

other colors

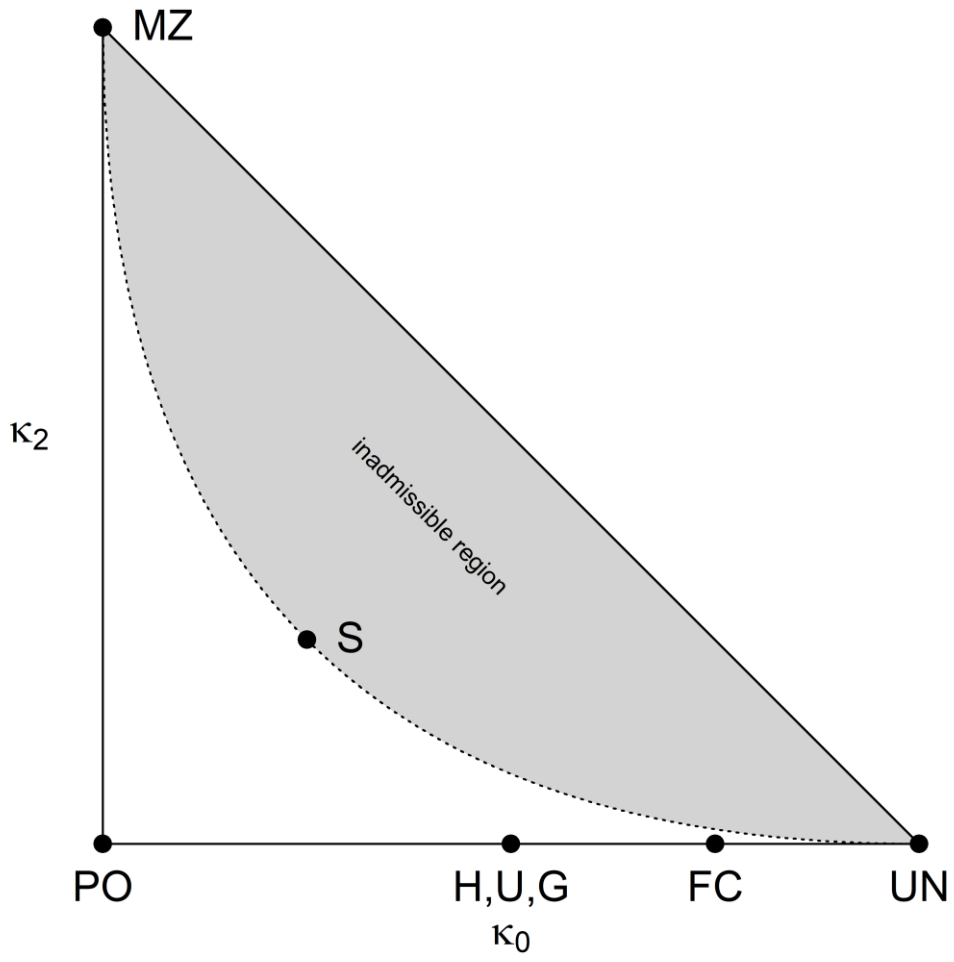
Wright's path formula:

$$\varphi_{P,Q} = \sum_A \sum_v \left(\frac{1}{2}\right)^{|v|+1} (1 + f_A)$$



Challenge:
Many relationships have the same kinship!

Tool 3: The relatedness triangle

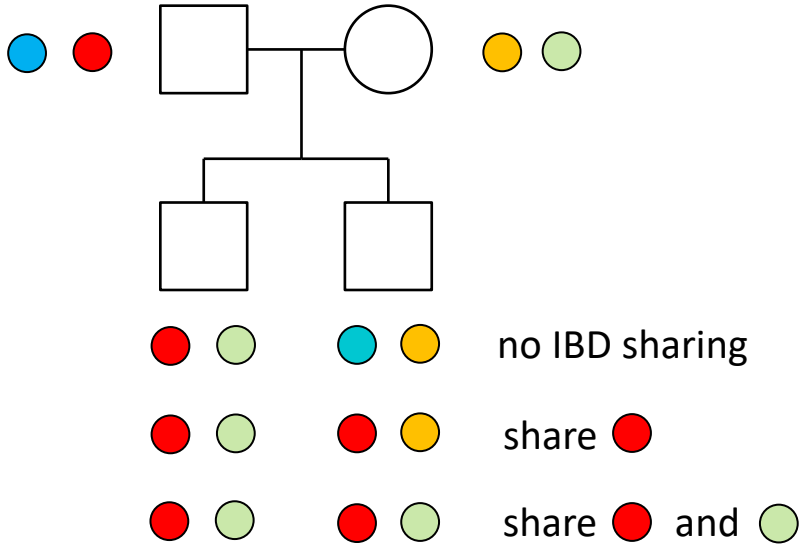


Charles Cotterman
(1914-1989)

Elisabeth Thompson
(1949 -)

IBD coefficients

- Summary so far:
  - Related individuals share alleles by descent (IBD)
  - Kinship coefficient meassures the overall amount of IBD sharing
- Natural generalisation:
  - How *many* alleles are IBD in each locus?

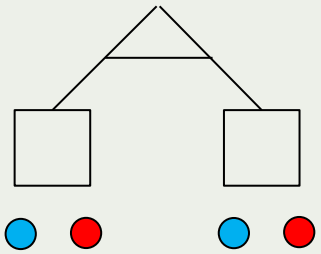


Definition

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

(at random autosomal locus)

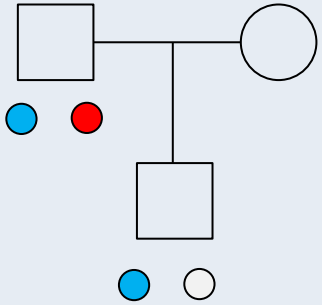
Three trivial relationships



Monozygotic twins (MZ)

$(\kappa_0, \kappa_1, \kappa_2) = (0,0,1)$

Always share 2



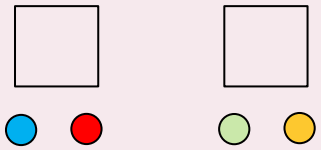
Parent-offspring (PO)

$(\kappa_0, \kappa_1, \kappa_2) = (0,1,0)$

Always share 1

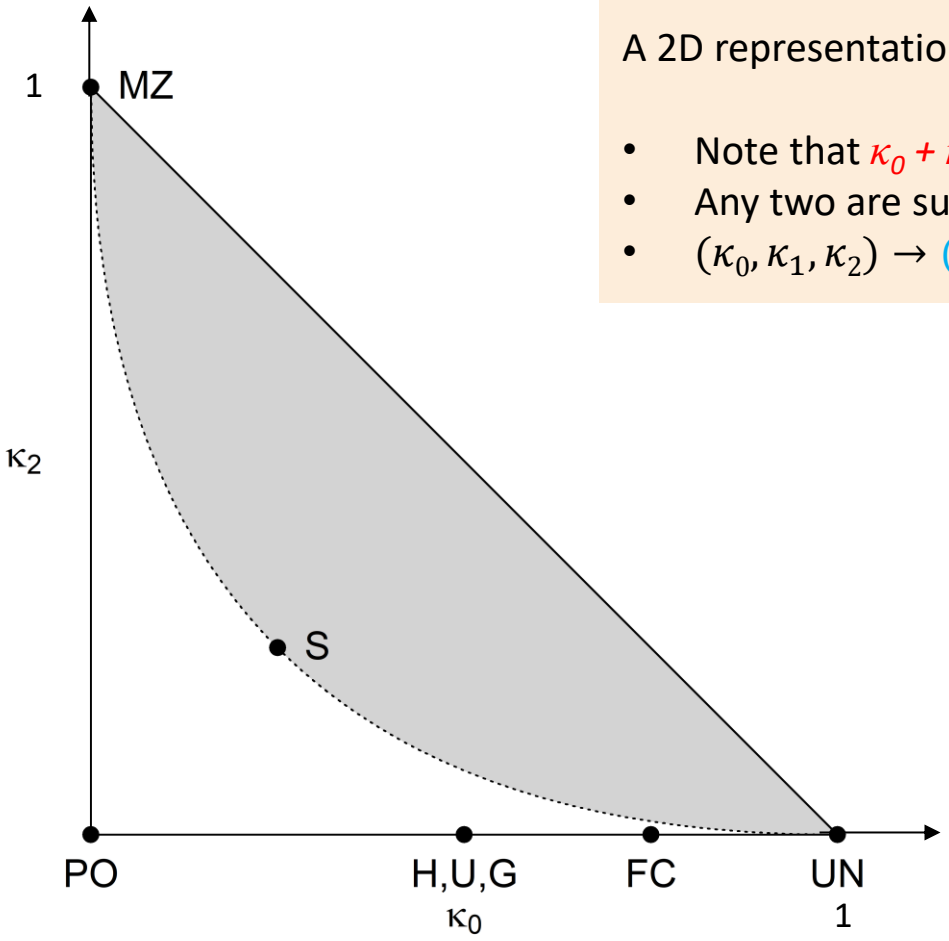
Always share 0

Unrelated (UN)



$(\kappa_0, \kappa_1, \kappa_2) = (1,0,0)$

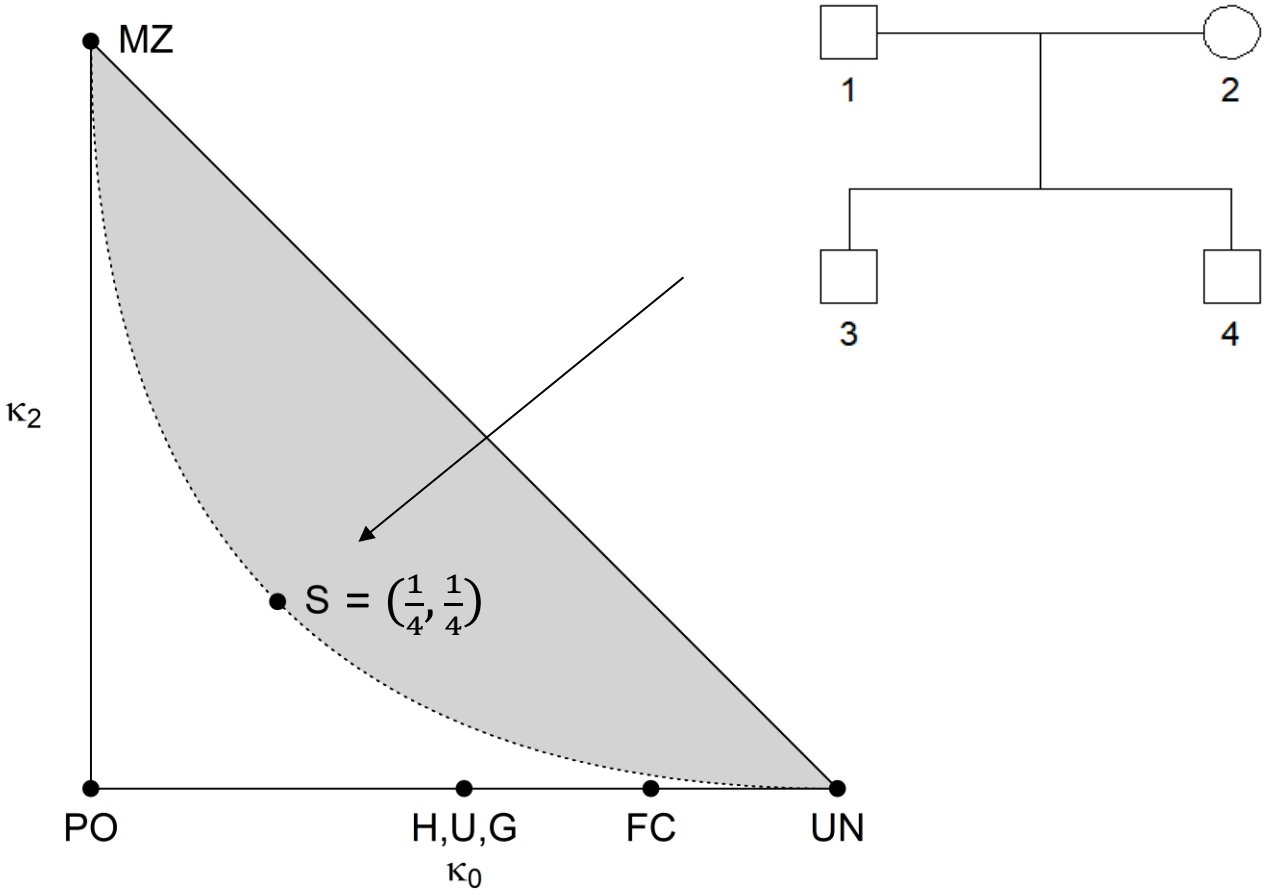
The relatedness triangle



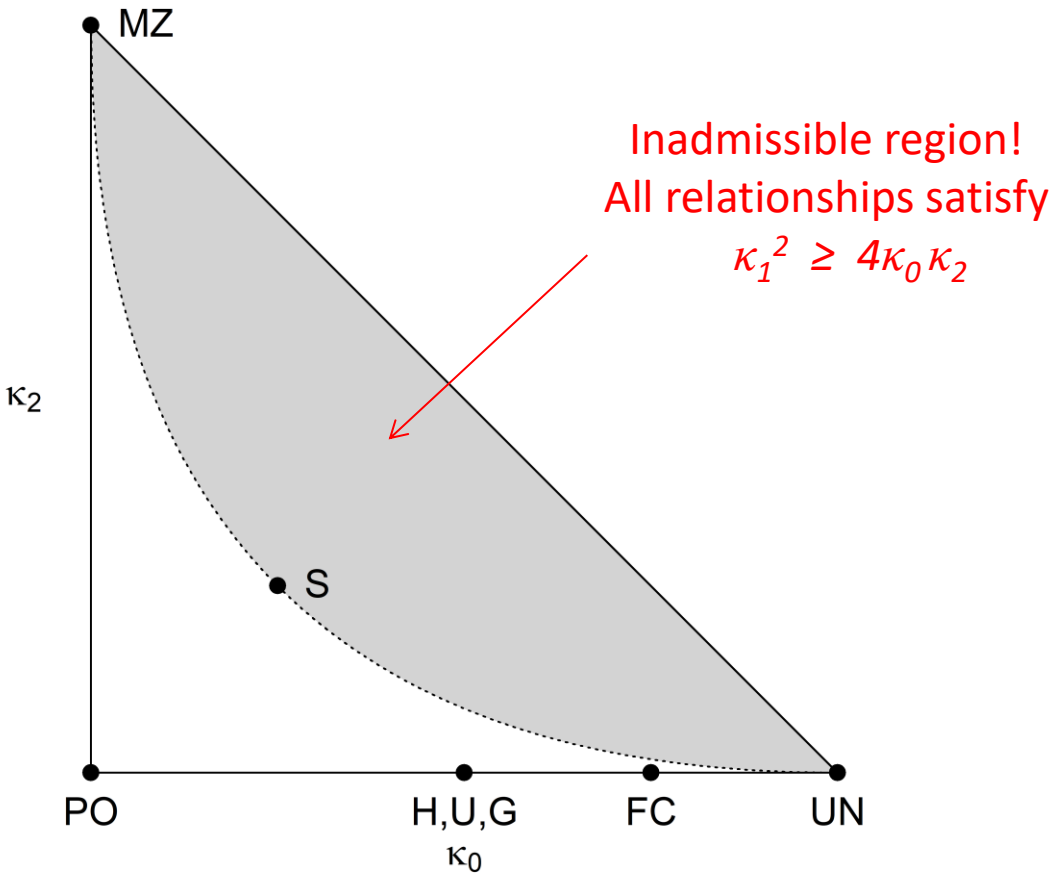
A 2D representation of $(\kappa_0, \kappa_1, \kappa_2)$

- Note that $\kappa_0 + \kappa_1 + \kappa_2 = 1$
- Any two are sufficient
- $(\kappa_0, \kappa_1, \kappa_2) \rightarrow (\kappa_0, \kappa_2)$

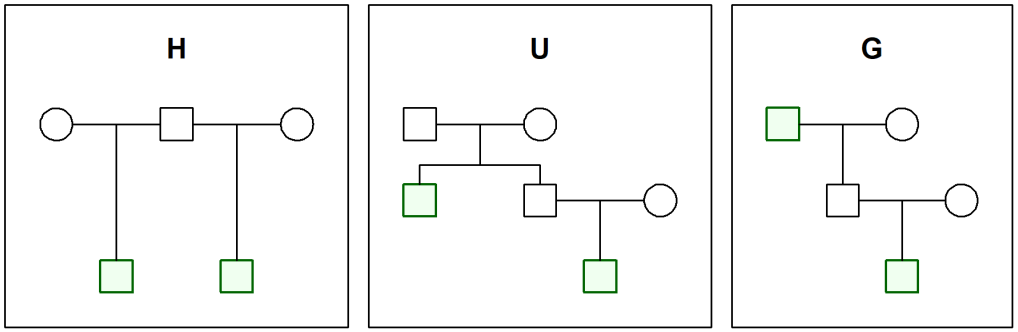
Full siblings: $(\kappa_0, \kappa_1, \kappa_2) = (0.25, 0.5, 0.25)$



Quirks of the relatedness triangle: 1

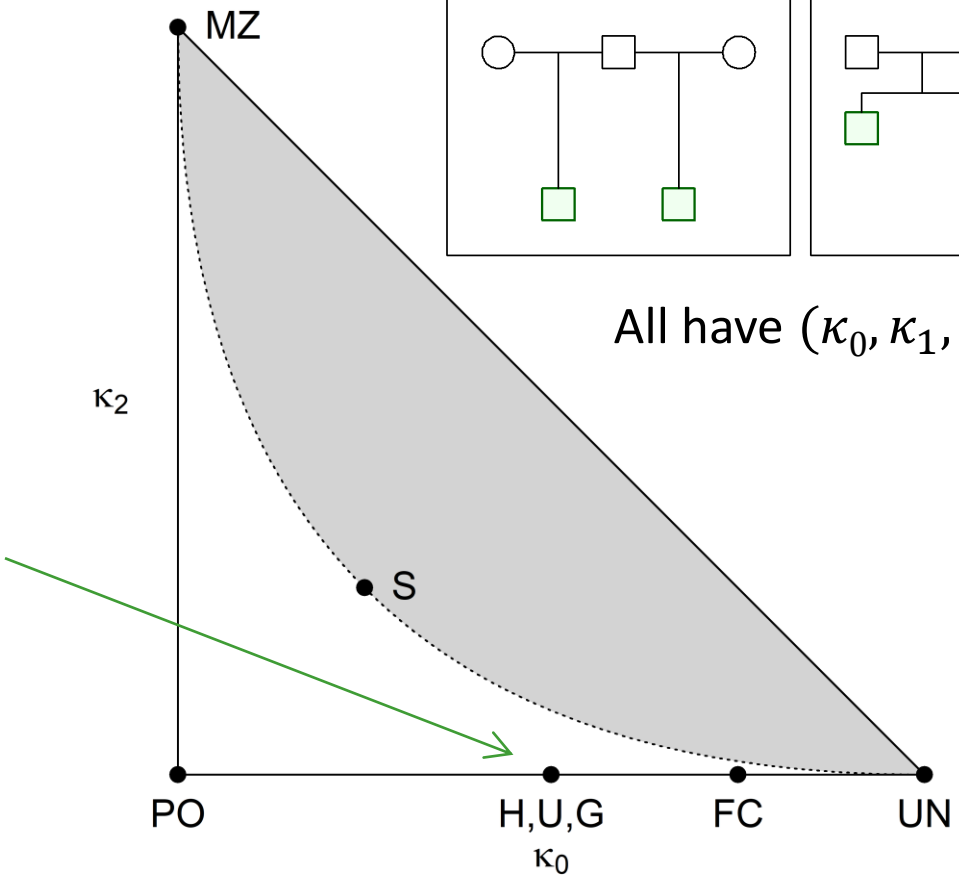


Quirks of the relatedness triangle: 2

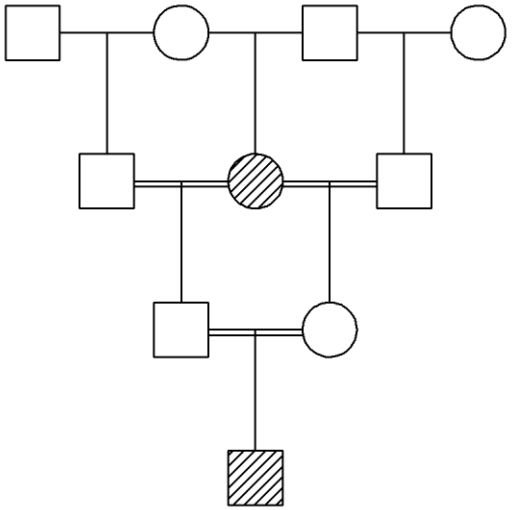


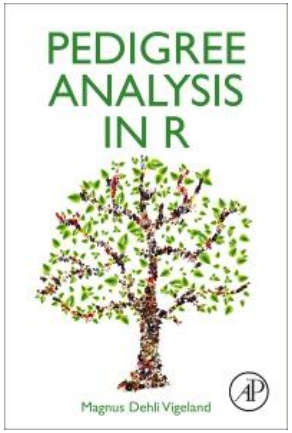
All have $(\kappa_0, \kappa_1, \kappa_2) = (\frac{1}{2}, \frac{1}{2}, 0)$

Some relationships coincide!



A word of caution

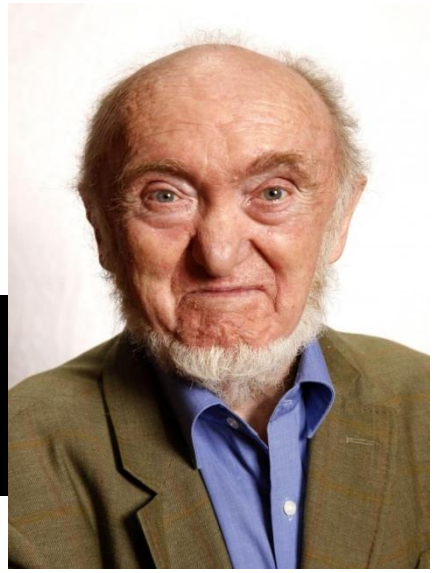




κ is only defined for non-inbred individuals.
For the whole story, we need 9 coefficients!

Jacquard's identity coefficients

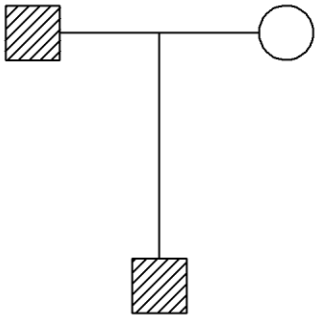
Not suitable for a basic course ...



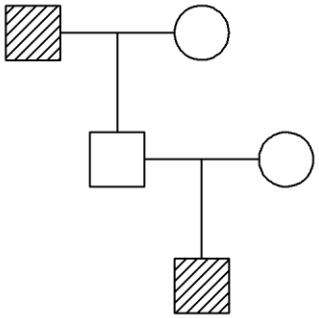
Albert Jacquard
(1925 - 2013)

Another word of caution: *Degree of relatedness*

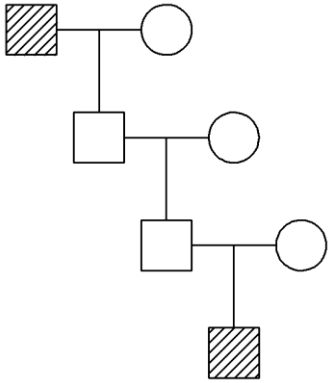
deg = 1



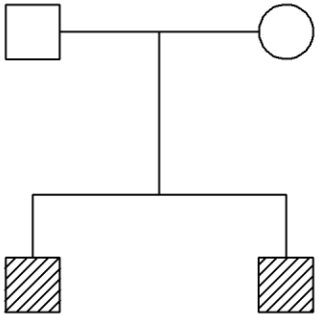
deg = 2



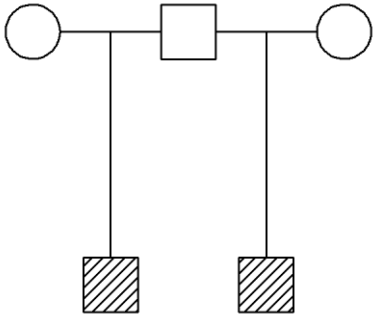
deg = 3



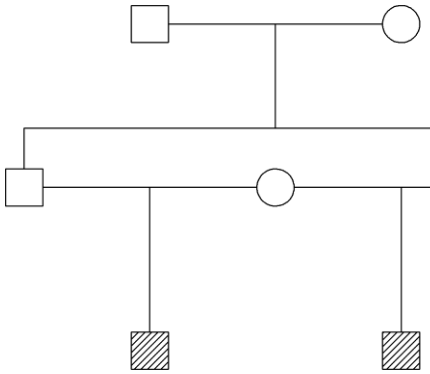
deg = ?



deg = ?



deg = ?



Another word of caution: *Degree of relatedness*

Degree of relationship			Relationship
Genetics	Roman law	Canon law	
0	0	0	self
0	2	1	
1	1	1	[10]
1	3	2	's child
1	2		
2	2		hter / grandson
2	3		
2	2		
2	4		
3	4	2	
3	3	3	father / great-granddaughter /
3	4	3	granduncle / grandniece / grandnephew
3	3	2	half-aunt / half-uncle / half-niece / half-nephew
3	5	3	double-first cousin once removed
3	6	3	quadruple-second cousin

Source: Wikipedia *Coefficient of relationship*

Tool 4: Estimating relatedness

Pairwise inference: Main approaches

More about this tomorrow!

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

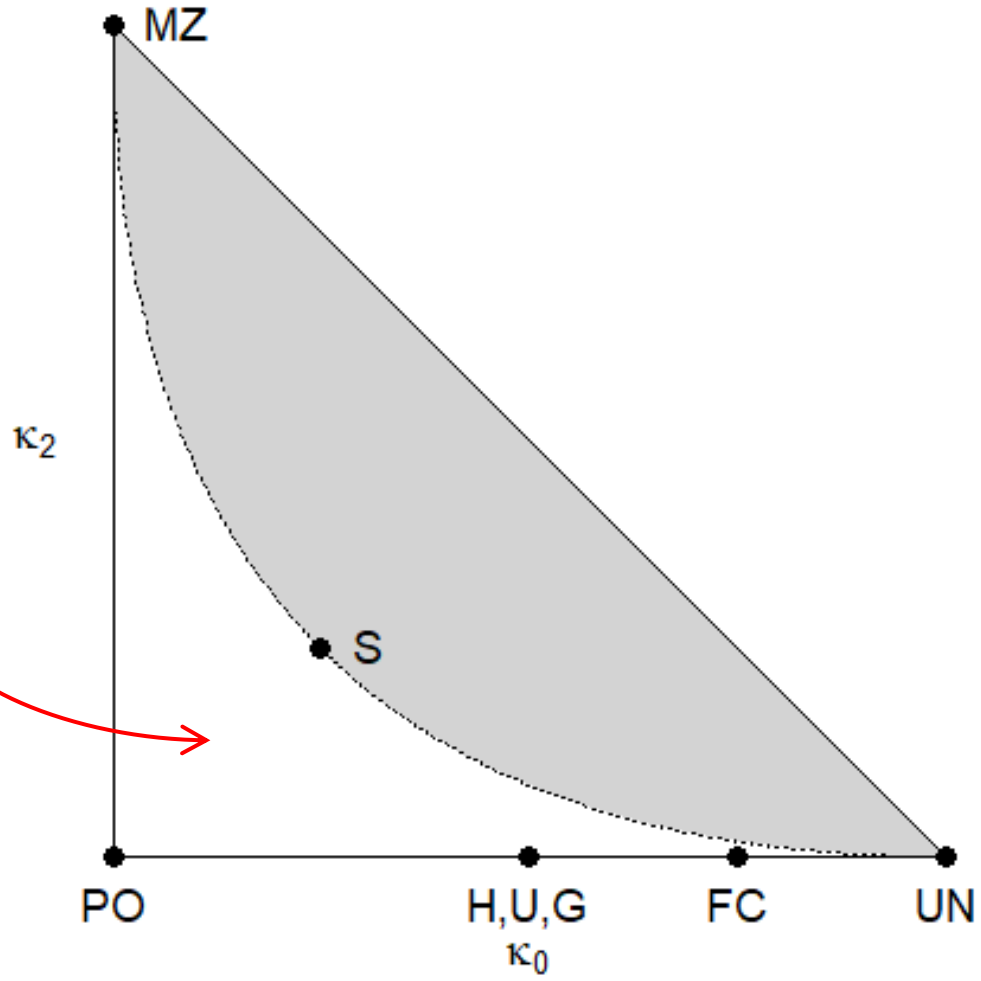
- Requires lots of SNPs
- Two steps:
  1) SNPs → IBD segments
  2) IBD segments → relatedness

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

MODERN

Approach A

where??



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

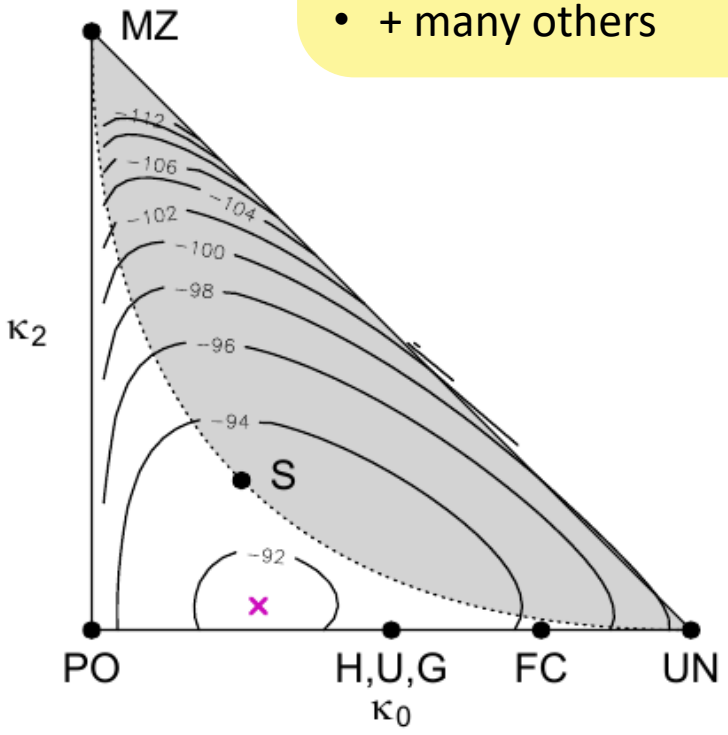
- Thompson (1975)
  - Given: marker genotypes for two individuals
  - The likelihood function

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

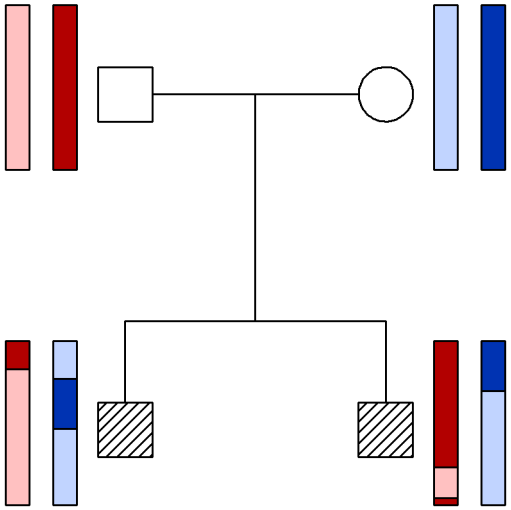
- Find the point κ which maximizes $L(\kappa)$!

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

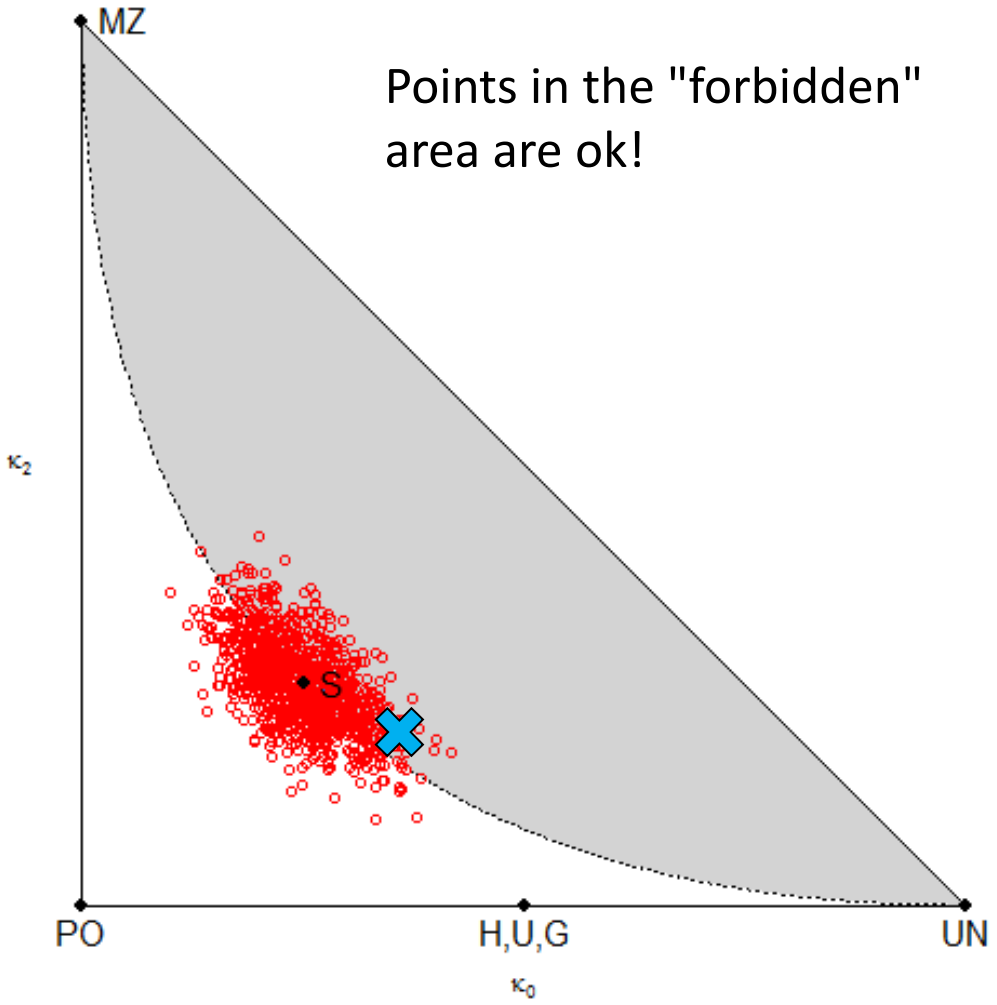
Implementations:
- R/pedsuite
- DIVIANA
- Familias
- + many others



What are we estimating?

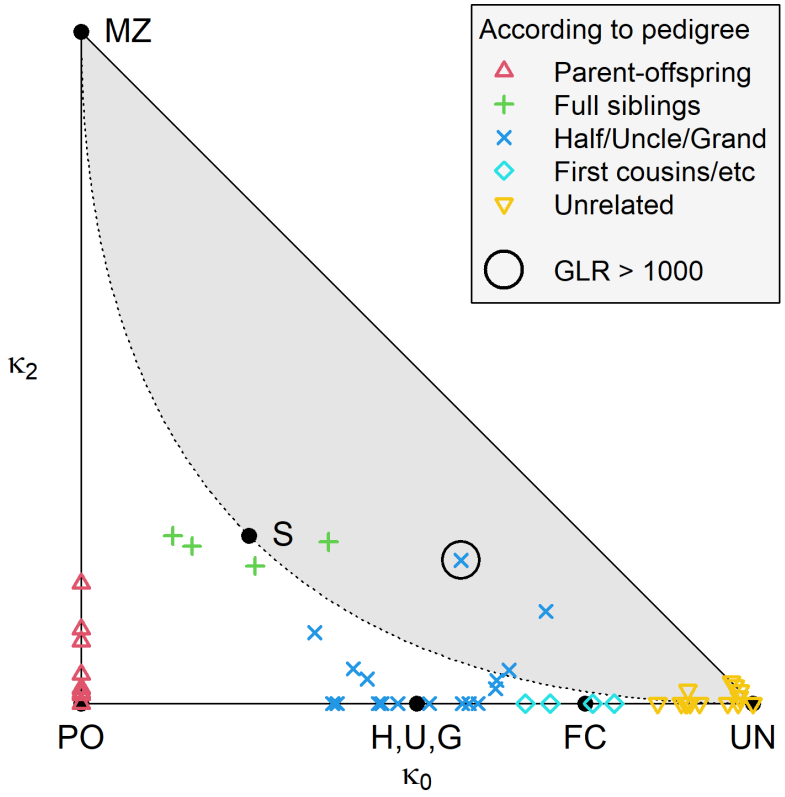


Answer: The *realised relatedness*



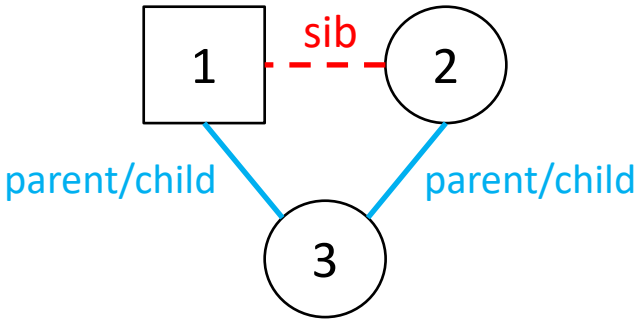
Application 1: Pedigree error detection

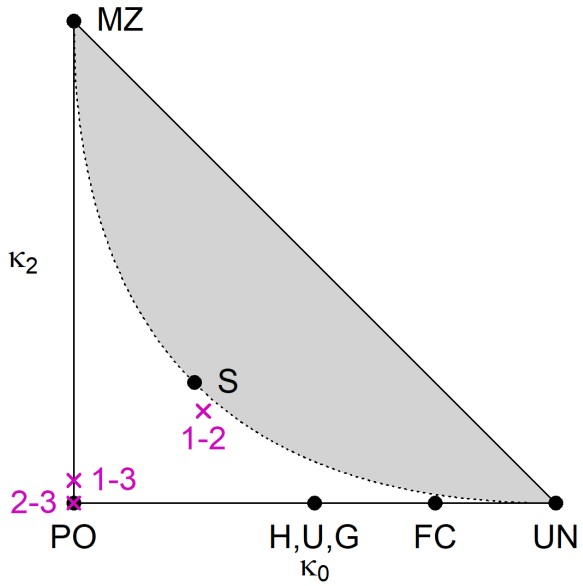
- For each pair of individuals, compute:
  1) pedigree-based kappa
  2) marker-based kappa estimate
- Compare them with an LR (actually: generalised LR*)
- Color-coded plot:
  - Position shows estimate
  - Color/shape shows pedigree claim

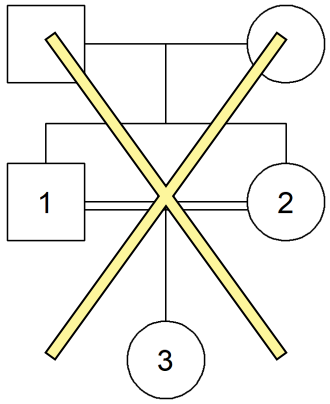


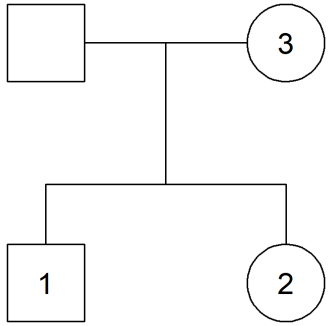
*Egeland & Vigeland (FSI:Genetics, 2025): *Kinship cases with partially specified hypotheses.*

Application 2: Pedigree reconstruction









Step 1: Genders

Step 2: Use pairwise estimates

- Connect parent-child
- Exploit siblings

Step 3: Solve the puzzle!

Your turn: Exercises!