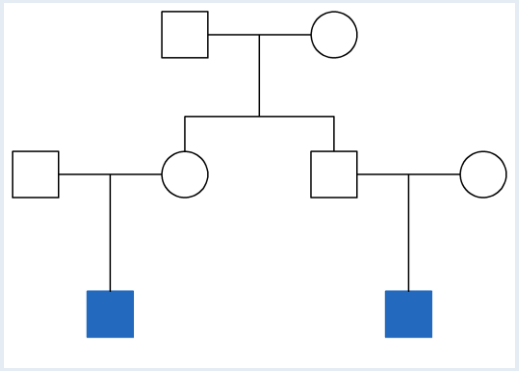
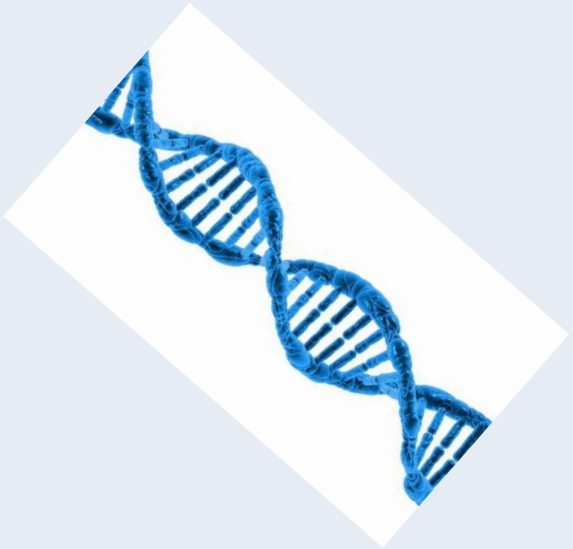




Lecture 1: Introductions

Statistical methods in genetic relatedness and pedigree analysis

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

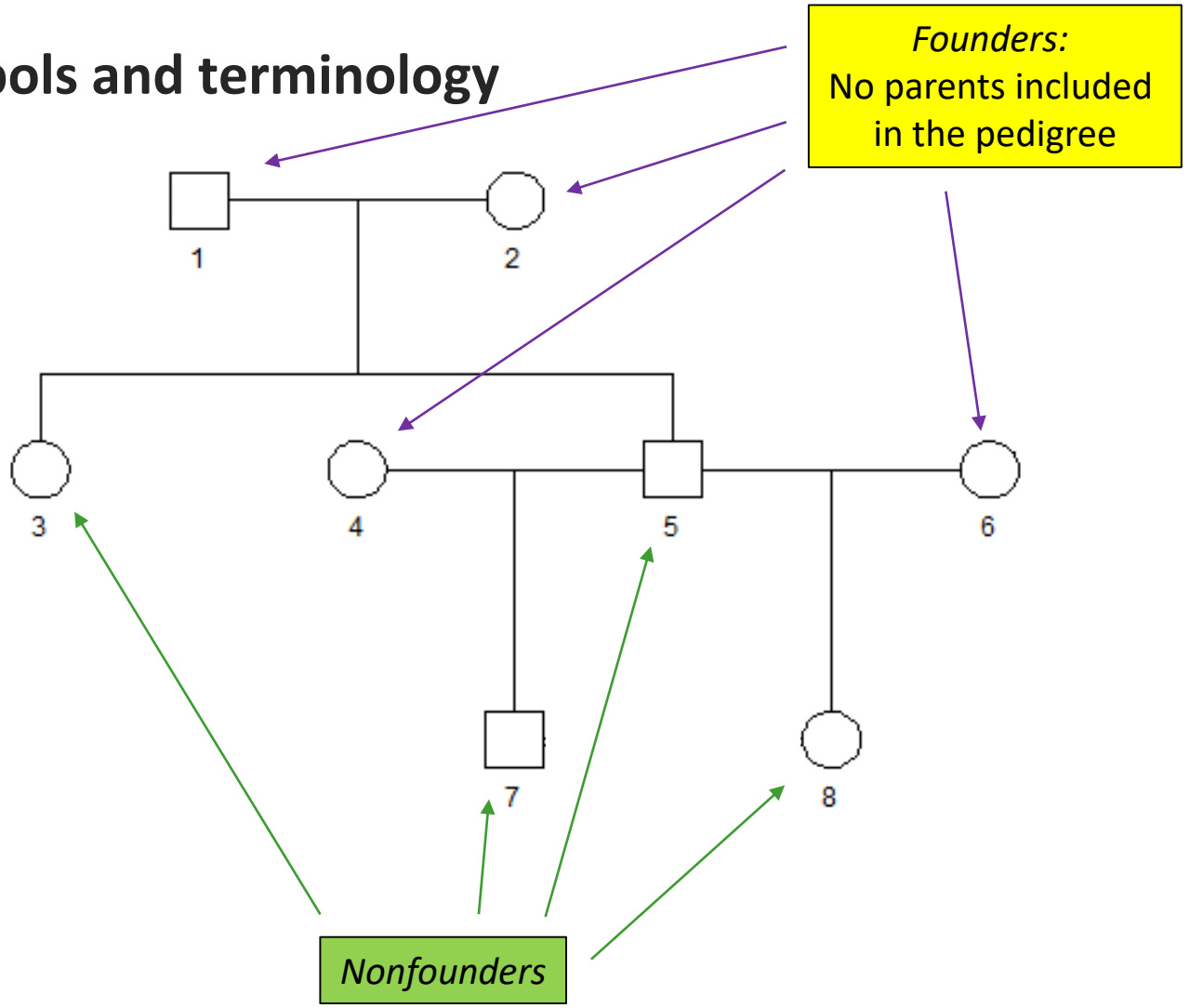
Magnus Dehli Vigeland

Outline

- Part I: *Pedigrees*
  - Pedigree symbols and terminology
  - Some common relationships
  - QuickPed
- Part II: *Genetics*
  - Terminology (Locus, allele, genotype, ...)
  - Mendelian inheritance
- Part III: *Probabilities (on pedigrees)*
  - Motivation: Real-life problems
  - Ingredients:
    - Hardy-Weinberg equilibrium
    - Mendelian transition probabilities
  - Computing pedigree likelihoods

Part I: Pedigrees

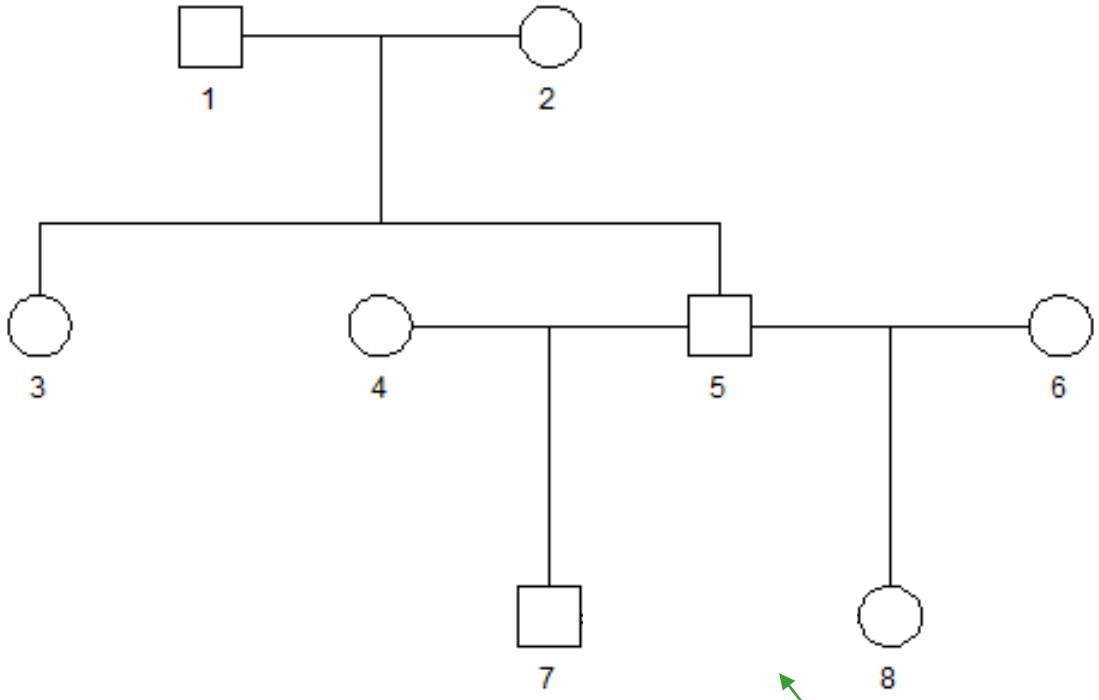
Pedigrees: Symbols and terminology



= male

= female

Pedigrees: Symbols and terminology

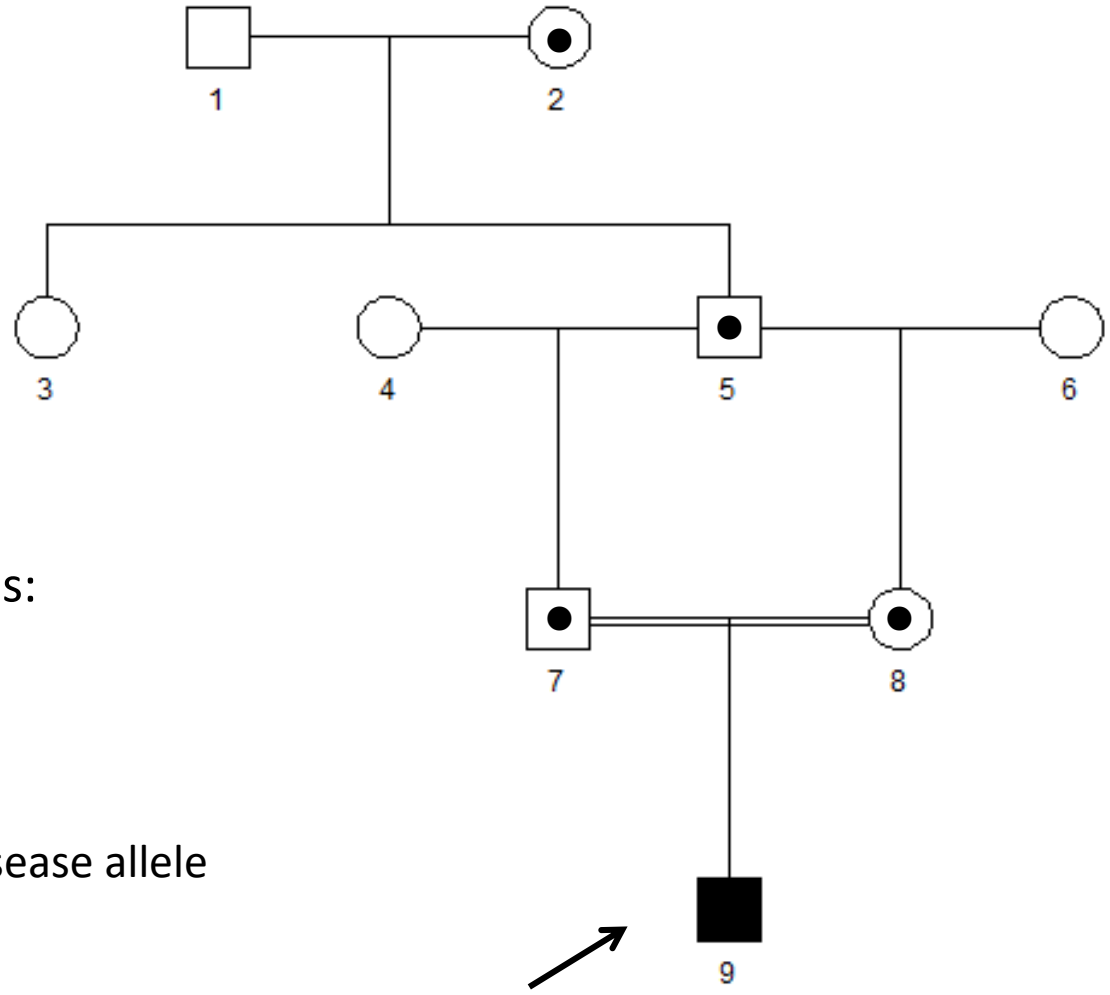


= male

= female

Consanguineous marriage

Pedigrees: Symbols and terminology



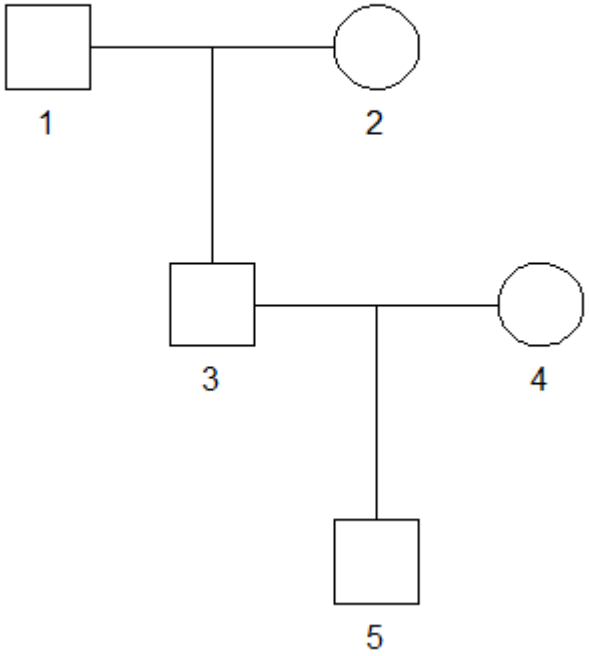
Medical pedigrees:

= affected

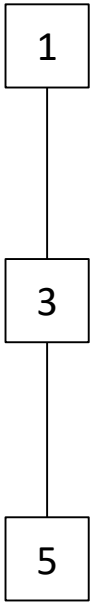
= unaffected

= carrier of disease allele

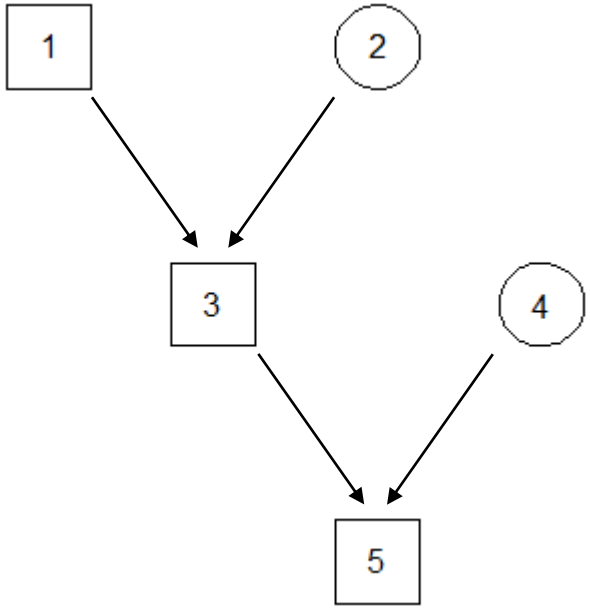
Alternative ways of drawing pedigrees



Standard



Simplified

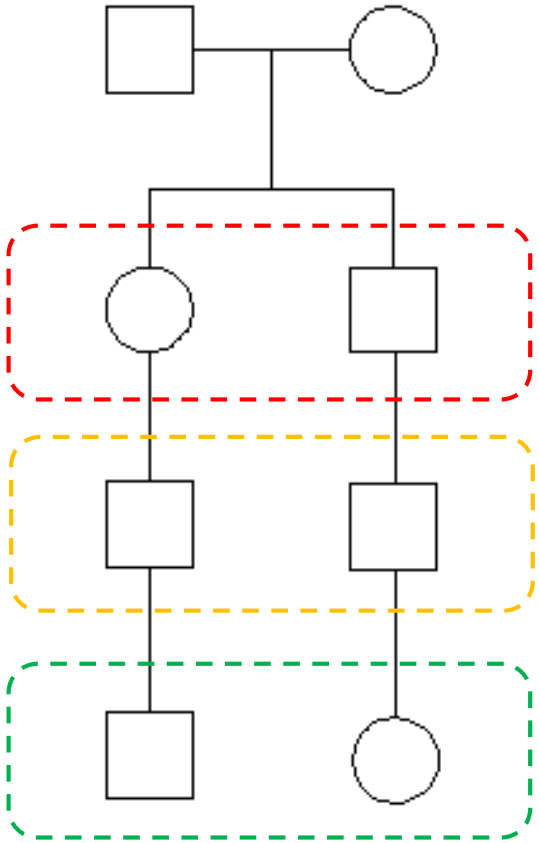


Directed acyclic graph

Some common relationships

(and some less common...)

Cousin relationships

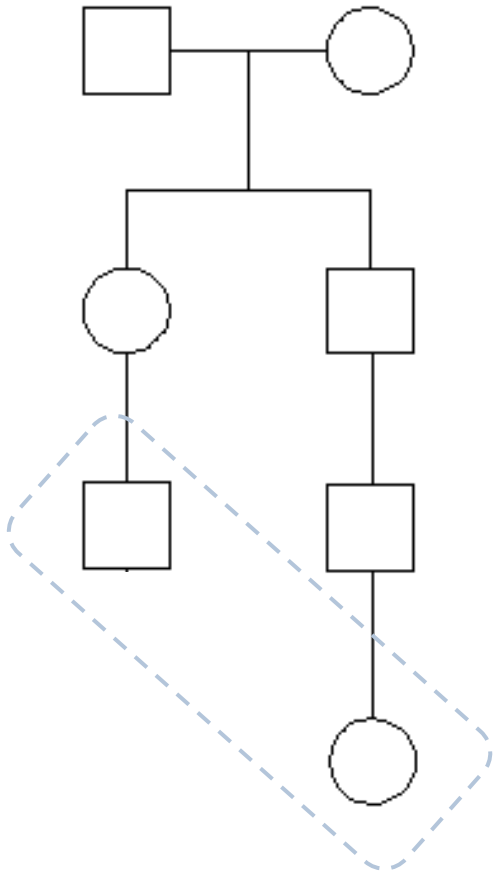


Full siblings

First cousins

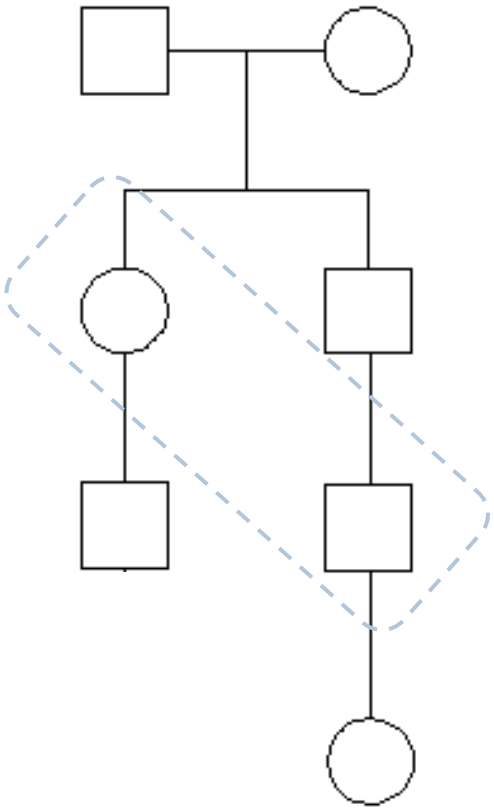
Second cousins

Cousin relationships



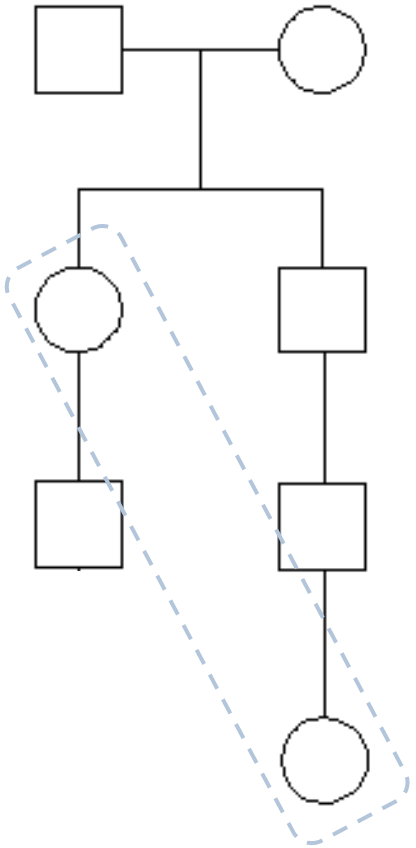
First cousins
once removed

Cousin relationships



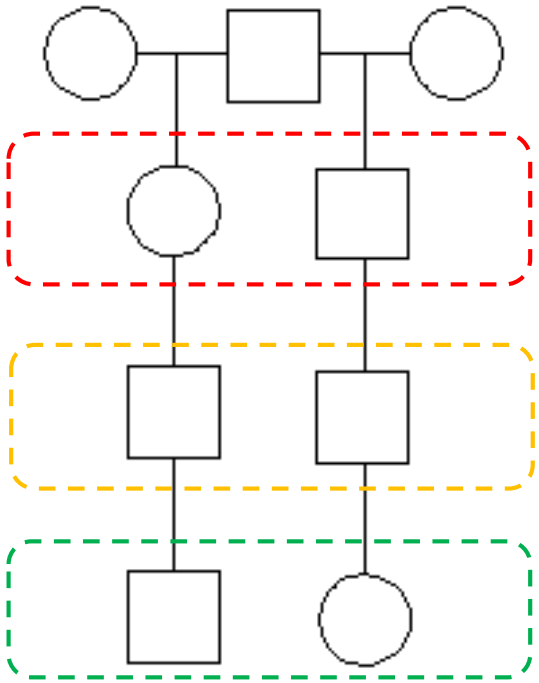
Aunt - nephew

Cousin relationships



Grandaunt

Half cousin relationships

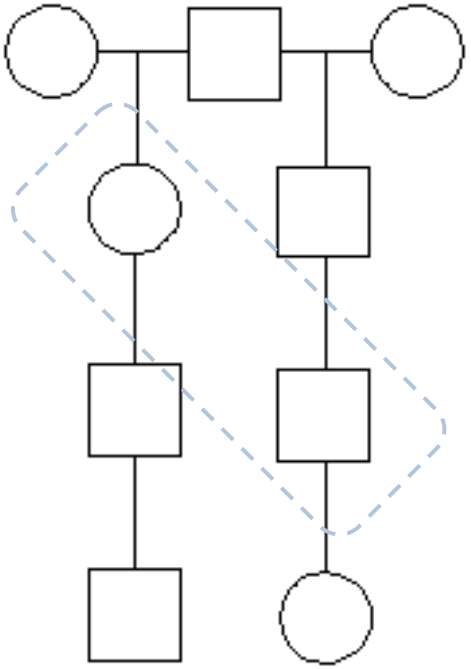


Half siblings (paternal)

Half first cousins

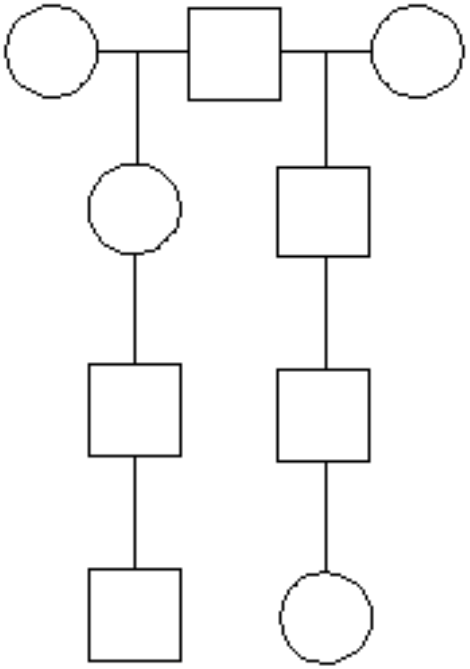
Half second cousins

Half cousin relationships

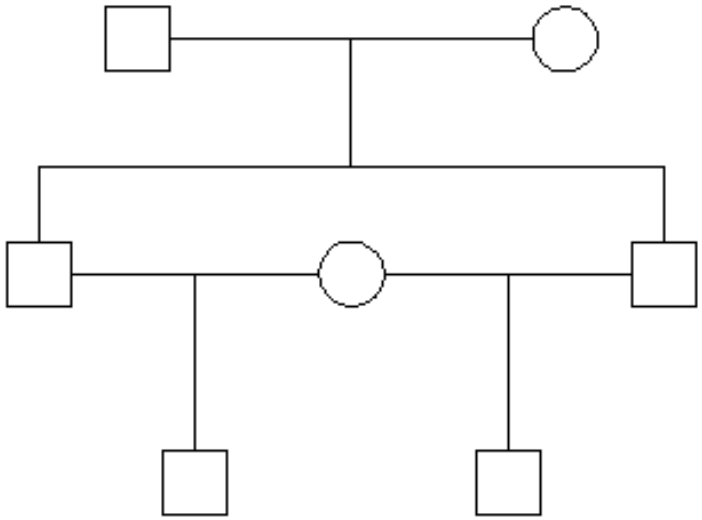


Half aunt /
half nephew

Half cousin relationships

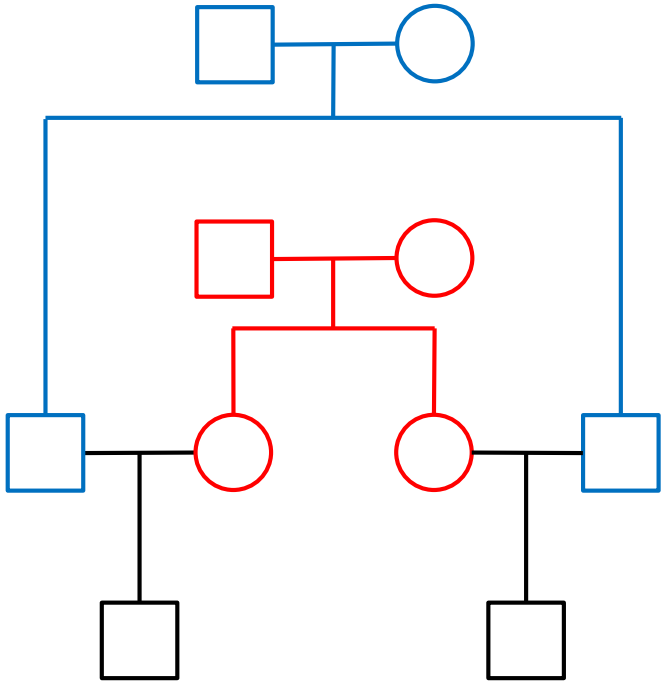


More complicated relationships



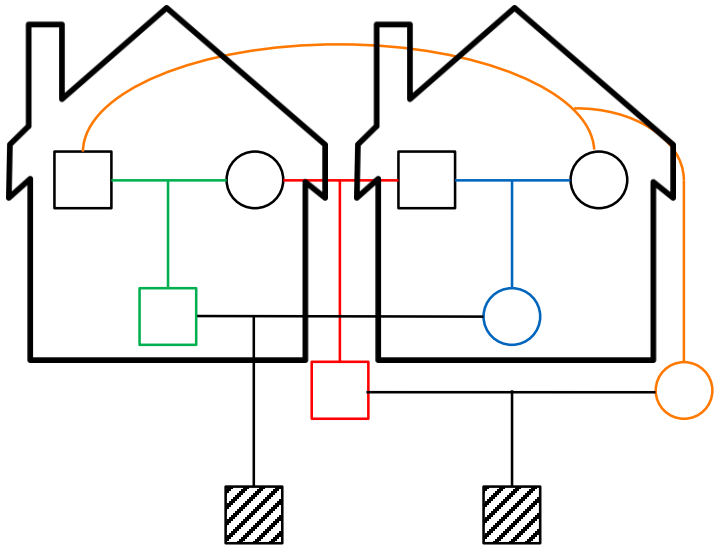
3/4 siblings

What about this?



Double first cousins

The connoisseur's favourite



Quadruple half first cousins!

What software exists to create pedigrees?

It depends!

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

In this course:

- R
- QuickPed

QuickPed: An Interactive Pedigree Creator

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

Purpose: This tool provides a quick way to create pedigree plots and files, and for analysing the relatedness between pedigree members.

Instructions: Choose a suitable start pedigree and modify it by selecting members (by clicking on them in the plot) and using appropriate buttons. For example, to create a new child, select the parents and press *Son* or *Daughter*. (If just one parent is selected, a new spouse is also created.)

More information: Check out the user manual for various tips and tricks, including an introduction to relatedness coefficients. There is also a paper about QuickPed published in BMC Bioinformatics.

Quick start

Built-in pedigree

Trio

or

Load a ped file

Reset all

Modify

Add

Son | Daughter | Sibling | Parents

Remove

Individuals | Selection

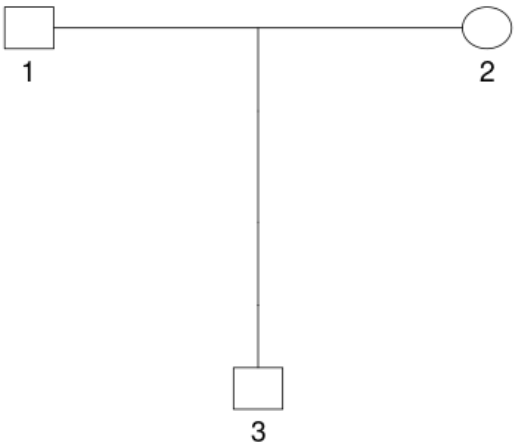
Switch

Sex | Affected | Carrier | Deceased

Twins

MZ | DZ

Undo



Labels

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

Show all | Hide all

1

2

3

Update

Plot settings

Width

Height

Expansion

Symbol size

Margins

PNG | PDF

Relationships

Ped file

Include

- [x] Headers
- [ ] Family ID
- [ ] Affection status

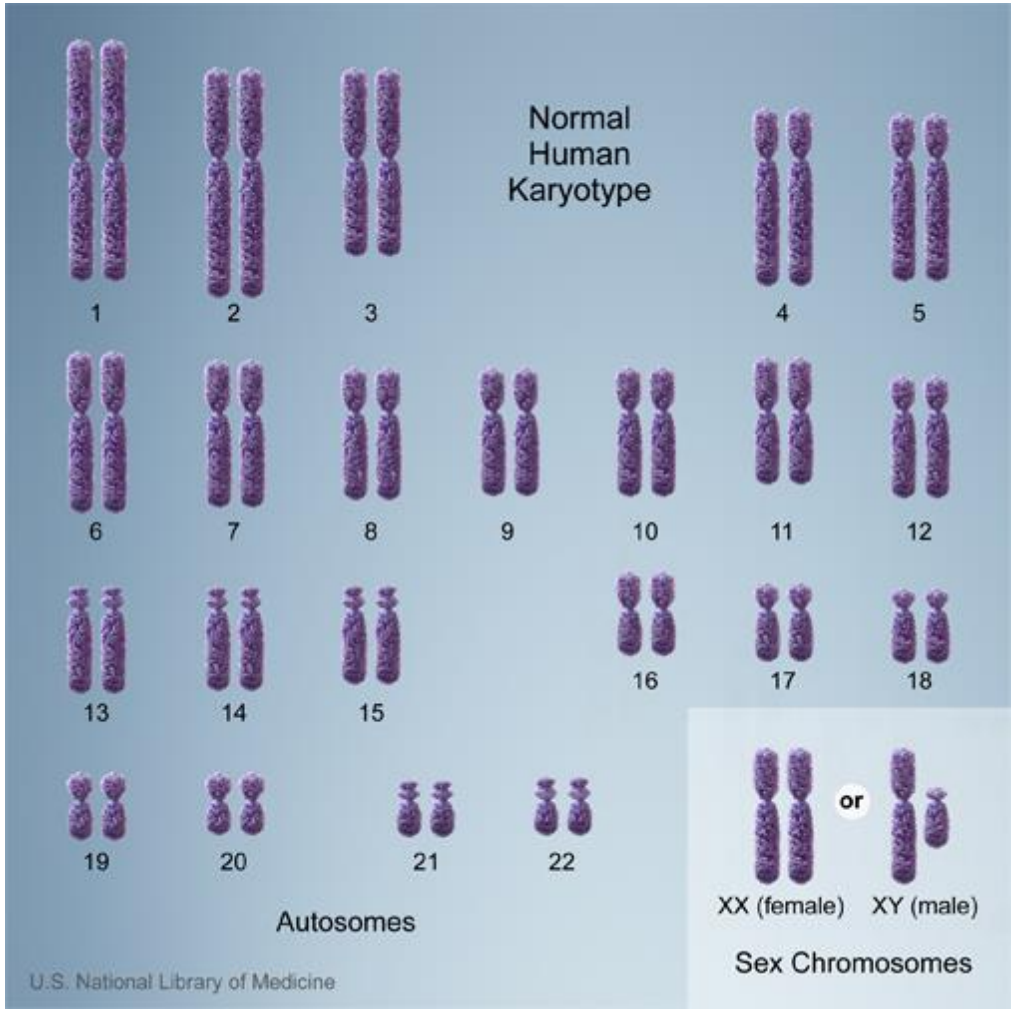
Save ped file

Part II: Genetics

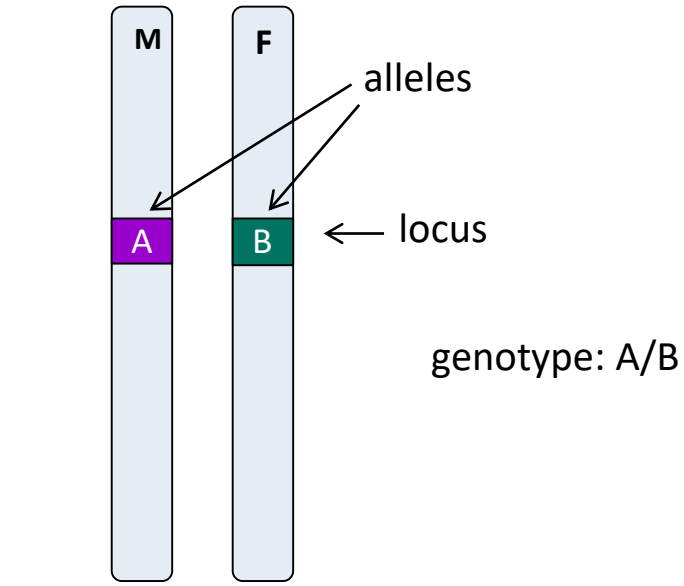
Human Genetics

Some important terms

- Locus
- Allele
- Genotype
- Genetic markers
  - SNPs
  - microsatellites



Locus, allele, genotype



- **LOCUS** = a specific place in the genome
- **ALLELE** = any of the alternative forms of a locus
- **GENOTYPE** = the set (usually: pair) of alleles carried at a given locus

Genetic markers

- Small parts of the genome which ...
  - have known position
  - vary in the population
  - are easy to genotype

- SNPs (single nucleotide polymorphisms)
  - two alleles
  - usual requirement: MAF > 1% = minor allele frequency
  - very common in the genome (millions!)
  - used in medical genetics +++

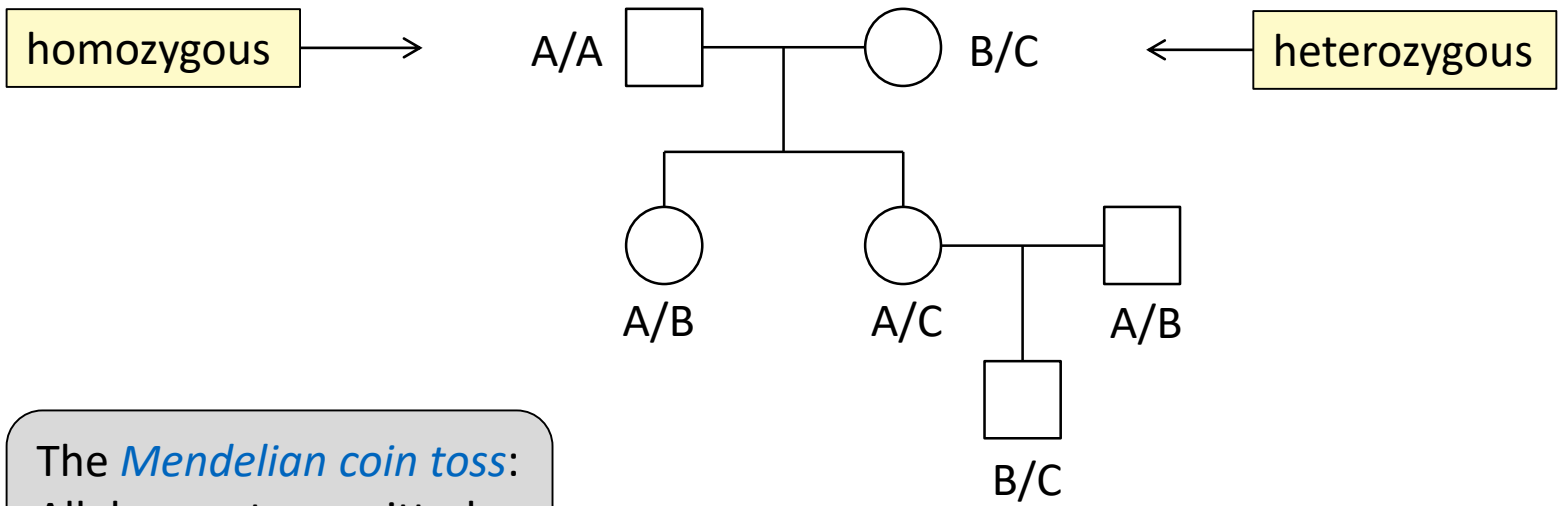
...CCGTTATATGGGC...
...CCGTTAGATGGGC...
...CCGTTATATGGGC...
...CCGTTATATGGGC...
...CCGTTAGATGGGC...

- STRs (short tandem repeats) = microsatellites
  - consecutive repeats of 2-5 bases
  - multiallelic: 5 - 70 alleles
  - allele names: # repeats
  - used in forensics

...ACG TTAG TTAG TTAG TTAG AAC..
...ACG TTAG TTAG AAC..
...ACG TTAG TTAG TTAG TTAG TTAG AAC..

Mendelian inheritance: Autosomal (chromosomes 1-22)

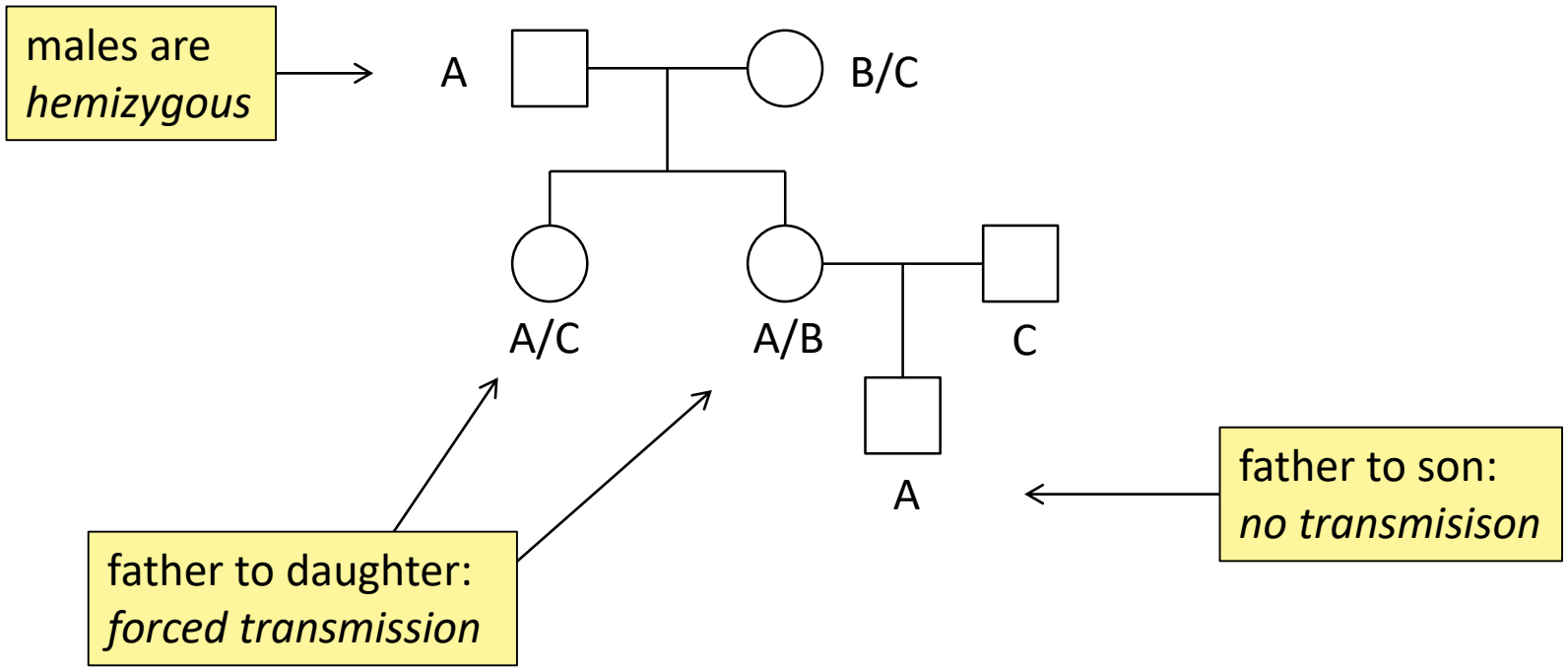
Example: autosomal marker with 3 alleles: A, B, C



The *Mendelian coin toss*: Alleles are transmitted with **50% chance**.

Mendelian inheritance: X-linked

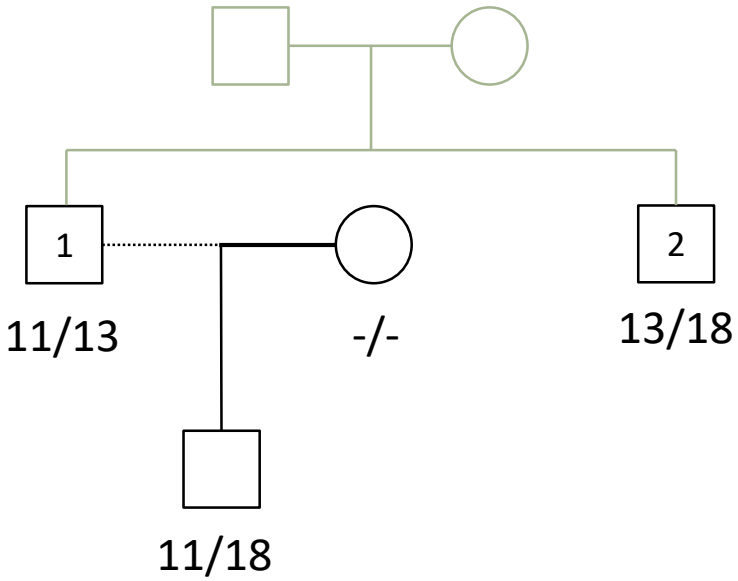
Example: X-linked marker with 3 alleles: A, B, C



Part III: Probabilities on pedigrees

Questions related to pedigrees with genotypes

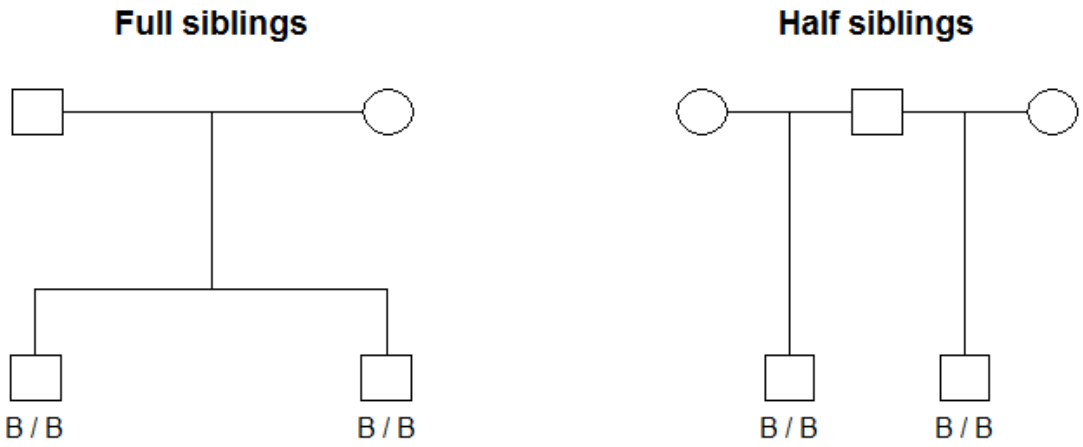
Questions related to pedigrees with genotypes



Suppose:
- 11 is common
- 18 is rare

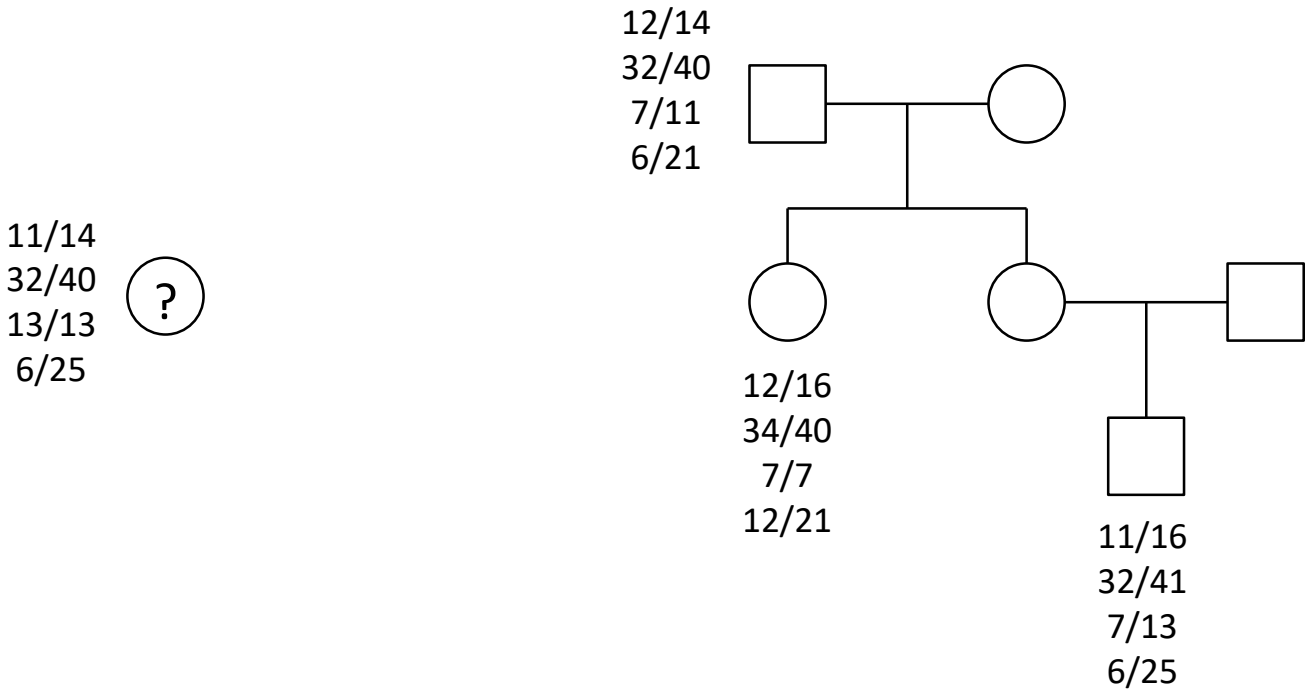
Who is the true father?

Questions related to pedigrees with genotypes



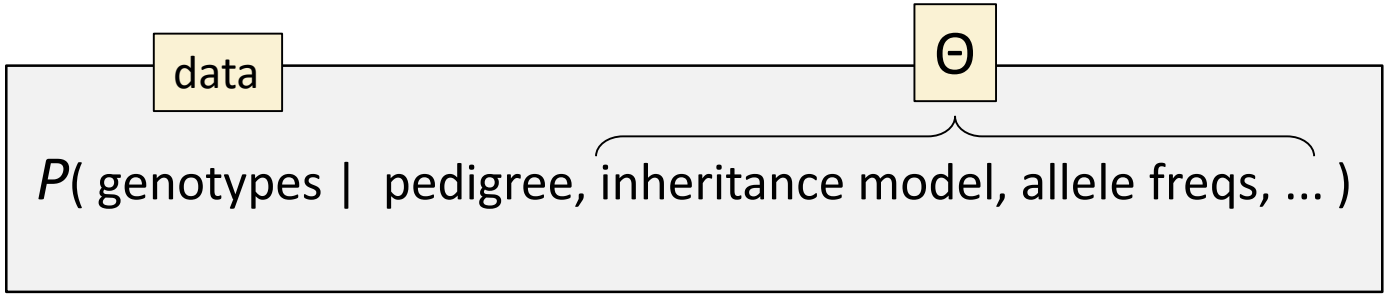
Brothers or half brothers?

Questions related to pedigrees with genotypes



Is this woman related to the family?

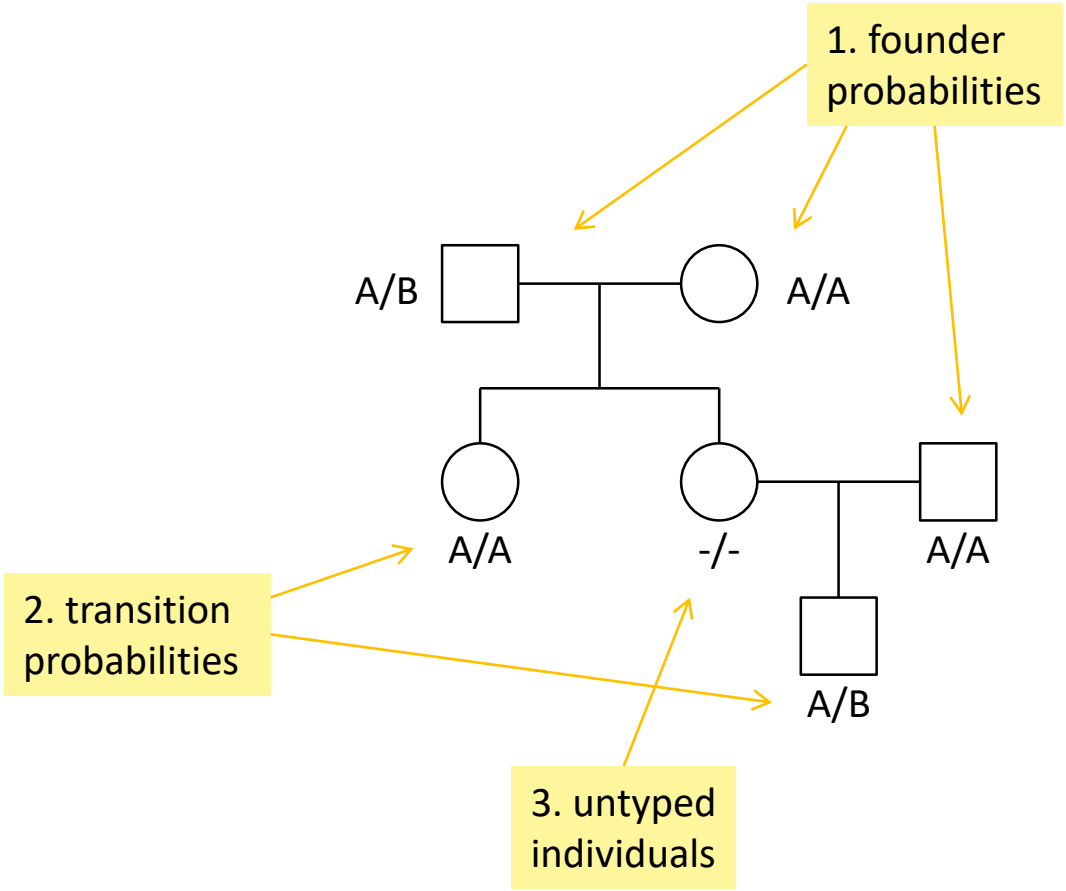
- Many applications involve probabilities of the following form



- Often referred to as a *pedigree likelihood*:

$$L(\text{ pedigree} \mid \text{data}) = P(\text{ data} \mid \text{pedigree}, \Theta)$$

Ingredients for likelihood computations



Ingredient 1: Founder probabilities

- Suppose the allele frequencies are:

$$P(A) = p$$
$$P(B) = q$$

- What are the frequencies of the genotypes *AA*, *AB*, *BB*?
- Under certain assumptions, the alleles can be treated as independent:

$$P(AA) = P(A) * P(A) = p^2$$

$$P(BB) = P(B) * P(B) = q^2$$

$$P(AB) = P(AB\ or\ BA) = pq + qp = 2pq$$

↑

two possible orderings!

The Hardy-Weinberg principle

Assumptions:

- infinite population
- random mating
- no selection
- no mutations
- no migration

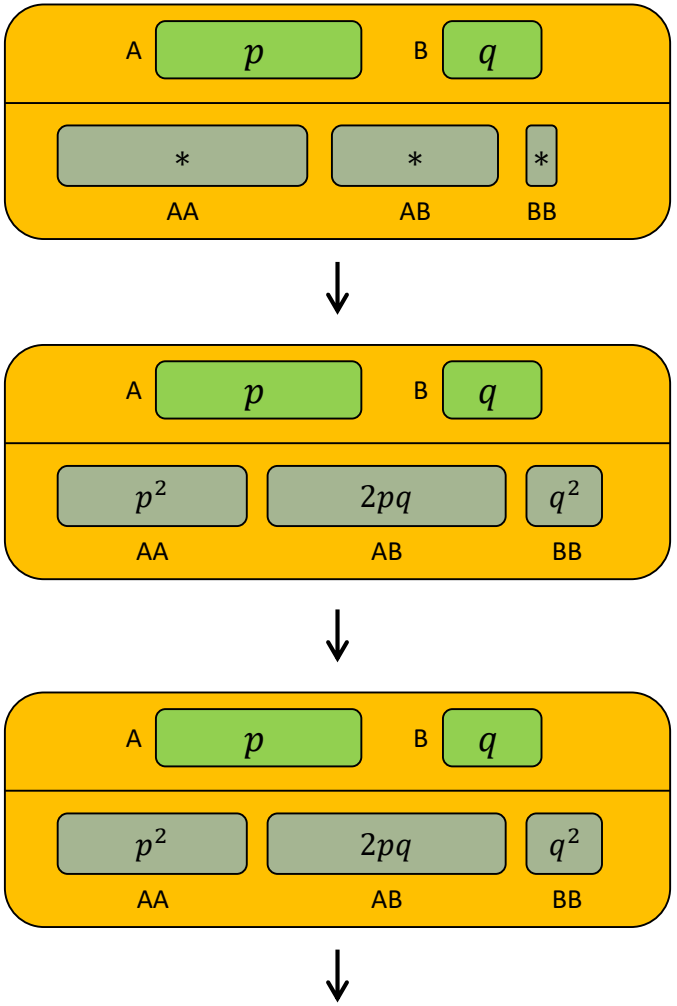
Hardy (1908): Shows

«... *using a little mathemathics of the multiplication table kind*»:

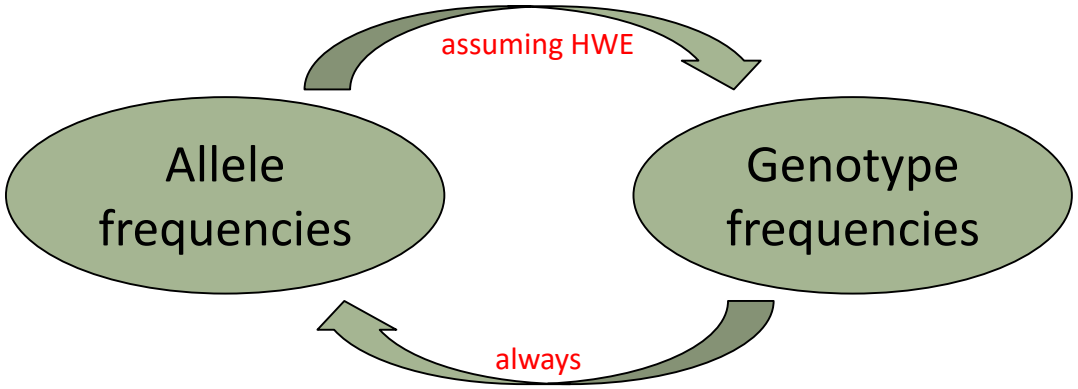
- allele freqs are unchanged from generation to generation
- after 1 generation the genotype freqs stay unchanged

$$P(AA) = p^2$$
$$P(AB) = 2pq$$
$$P(BB) = q^2$$

HW equilibrium



$$p_{AA} = p^2$$
$$p_{AB} = 2pq$$
$$p_{BB} = q^2$$

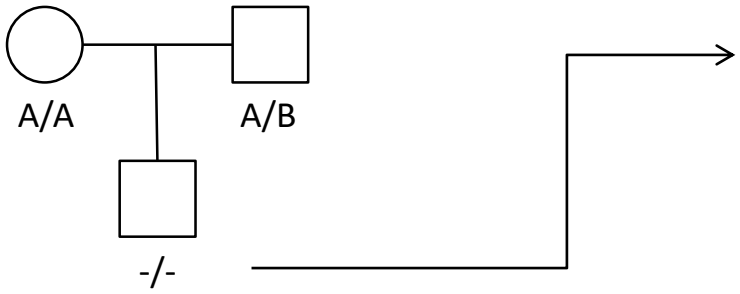


$$p = p_{AA} + 0.5\, p_{AB}$$
$$q = p_{BB} + 0.5\, p_{AB}$$

Ingredient 2: Transition probabilities

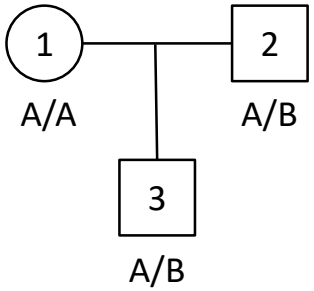
$P(g_{child} \mid g_{parents})$

- Easy - directly from Mendel's laws!



parents \ child	A/A	AB	BB
AA × AA	1	0	0
AA × AB	0.5	0.5	0
AA × BB	0	1	0
AB × AA	0.5	0.5	0
AB × AB	0.25	0.5	0.25
AB × BB	0	0.5	0.5
BB × AA	0	1	0
BB × AB	0	0.5	0.5
BB × BB	0	0	1

Example



$$L = P(g_1, g_2, g_3)$$

$$= P(g_1) \cdot P(g_2) \cdot P(g_3 \mid g_1, g_2)$$

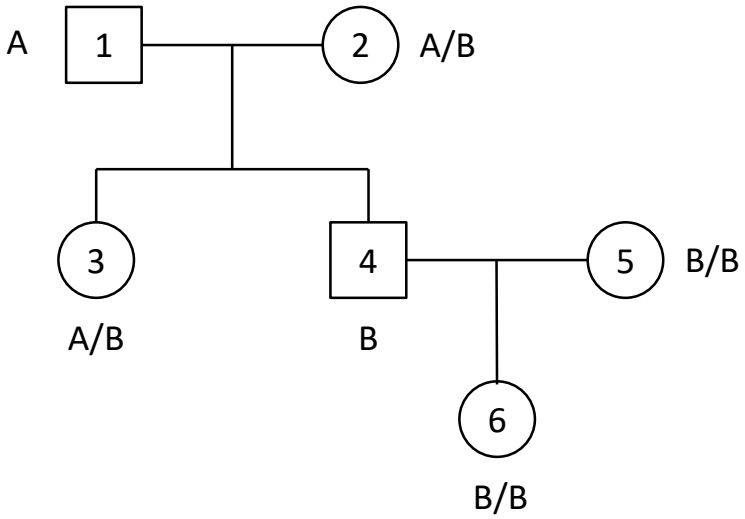
$$= P(AA) \cdot P(AB) \cdot P(AB \mid \text{parents} = AA \times AB)$$

$$= p^2 \cdot 2pq \cdot 0.5$$

$$= p^3 q$$

assuming HWE!

Example on X



$$L = P(genotypes \mid pedigree, p, q)$$

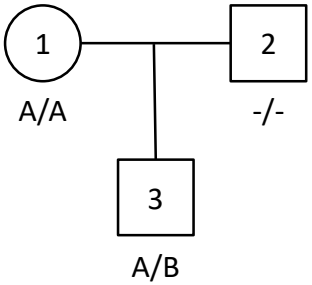
$$= \overset{1}{p} \cdot \overset{2}{2pq} \cdot \overset{3}{0.5} \cdot \overset{4}{0.5} \cdot \overset{5}{q^2} \cdot \overset{6}{1}$$

contribution from each individual

$$= 0.5\, p^2\, q^3$$

Ingredient 3: How to deal with untyped individuals

Solution: Sum of all possible genotypes for the untyped



$$P(g_1, g_3) = \sum_{g_2} P(g_1, g_2, g_3) = \sum_{g_2} P(g_1) \cdot P(g_2) \cdot P(g_3 | g_1, g_2)$$

$$= P(AA) \cdot P(AA) \cdot P(AB | AA \times AA) \; + \; P(AA) \cdot P(AB) \cdot P(AB | AA \times AB) \; + \; P(AA) \cdot P(BB) \cdot P(AB | AA \times BB)$$

$$= \; p^2 \cdot p^2 \cdot 0 \qquad + \; p^2 \cdot 2pq \cdot 0.5 \qquad + \; p^2 \cdot q^2 \cdot 1$$

$$= p^3 q + p^2 q^2 = p^2 q (p + q) = p^2 q$$

Pedigree likelihood: General formula

- Given:
  - pedigree with *n* individuals
  - *k* members are genotyped: $g_1, g_2, \ldots, g_k$

- Then:

$$P(g_1, \ldots, g_k) = \sum_{G_1} \sum_{G_2} \cdots \sum_{G_n} \overbrace{P(g_1) \cdots P(g_j)}^{\text{founders}} \cdot \overbrace{P(g_{j+1}|par) \cdots P(g_n|par)}^{\text{non-founders}}$$

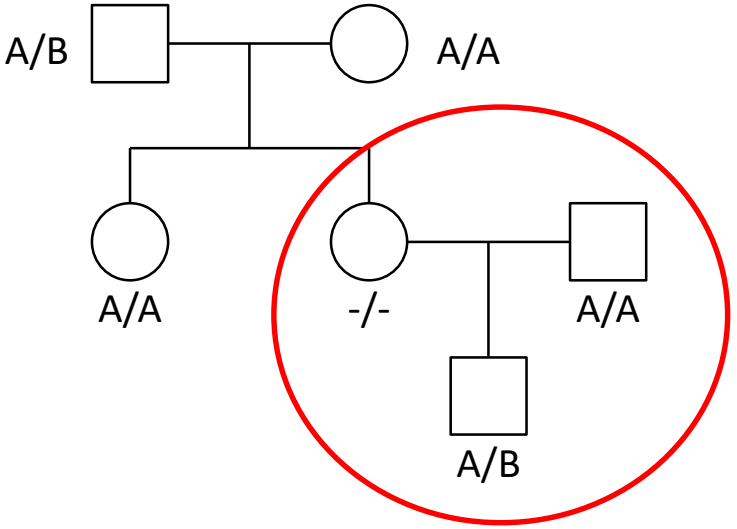
G_i = set of possible genotypes for individual *i*

- If everyone is typed: Only one term → easy

- Number of terms grows exponentially in #(untyped)
  - but clever algorithms exist!

Computer algorithms for pedigree likelihoods

- Elston-Stewart algorithm
  - a *peeling* algorithm
  - linear in pedigree size!



- Lander-Green
  - based on inheritance vectors
  - hidden Markov model
  - best choice with many *linked* markers
  - small/medium pedigrees only

Software for pedigree likelihoods

- R/ped suite
  - Elston-Stewart
  - mutations, theta correction, ++
- Familias
  - GUI for forensic applications
  - Elston-Stewart
  - handles mutations, theta correction, ++
- MERLIN
  - command line program
  - Lander-Green
  - gold standard for cases with dense SNP markers (but not too large pedigrees)
  - not mutations, not theta correction
- Other (old) software still in use
  - FastLink
  - Allegro
  - +++

Your turn: Exercises!