How far back can we go? Examples and conclusions.

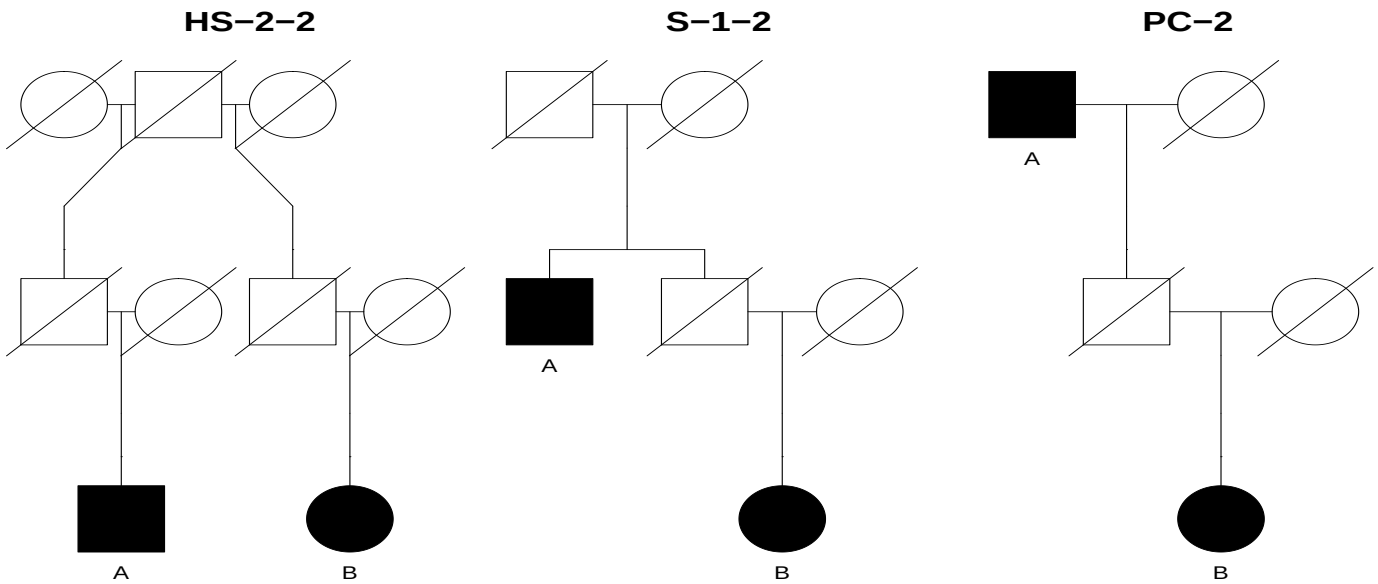
Presentation at Graphical Models and Genetics Applications Workshop, Warwick, April 15-17

FEST: R-package for computing likelihoods for hypothesized family relations

- Linked *autosomal* SNP markers
- Restrict attention to *pairwise* relations: Likelihoods and posterior probabilities for specific alternative family relations.
- Provides a front-end to MERLIN which do the actual likelihood computations. (Lander-Green algorithm)
- May specify any possible family relation under certain constraints (no inbreeding loops, very distant family relations cause computational problems)

Extended pairwise family relationships

- Half-sib relations. *One* common ancestor. (Jefferson example)
  HS-n_1-n_2: n_1 generations between common ancestor and A, n_2 generations between the ancestor and B.
- Sibling relations. *Two* common ancestors.
  S-n_1-n_2: n_1 generations between common ancestors and A, n_2 generations between the ancestors and B.
- Parent-child relations. Person A is the ancestor of person B.
  PC-n: n generations between A and B.



IBD probabilities for different numbers of IBD alleles and pairwise family relationships

IBD	HS-n_1-n_2	S-n_1-n_2	PC-n
0	$1-\left(\frac{1}{2}\right)^{n_1+n_2-1}$	$1-\left(\frac{1}{2}\right)^{n_1+n_2-2}-\frac{1}{4}\mathbf{1}(n_1=n_2=1)$	$1-\left(\frac{1}{2}\right)^{n-1}$
1	$\left(\frac{1}{2}\right)^{n_1+n_2-1}$	$\left(\frac{1}{2}\right)^{n_1+n_2-2}-\frac{1}{2}\mathbf{1}(n_1=n_2=1)$	$\left(\frac{1}{2}\right)^{n-1}$
2	0	$\frac{1}{4}\mathbf{1}(n_1=n_2=1)$	0

- Note that for HS-1-1, S-1-2 and PC-2 the probabilities are 0.5, 0.5 and 0 for IBD = 0, 1 and respectively (example shown by Thore).

Example: Checking family relationships prior to linkage or association studies

- Linkage study on autosomal dominant epilepsy
- Data from 3722 SNPs on 22 autosomal chromosomes
- Posterior probabilities computed for each supposed each true (corresponding to columns) and hypothesised family relation.
- Flat prior

relation	S-1-1	HS-1-1	S-2-2	S-3-3	Unrelated
S-1-1	1.000	0.000	0.000	0.000	0.000
HS-1-1	0.000	1.000	0.000	0.000	0.000
S-2-2	0.000	0.000	1.000	0.000	0.000
S-3-3	0.000	0.000	0.001	0.974	0.025

Example: Checking relationships between founders

- A recessive disease. Are the founders related?
- Data from 3439 SNPs on 22 autosomal chromosomes
- Posterior probabilities computed for each founder pair and hypothesised family relationship.
- Flat prior

	S-2-2	S-3-3	HS-2-2	HS-3-3	Unrelated
Founder pair 1	0.000	0.001	0.000	0.034	0.965
Founder pair 2	0.000	0.001	0.000	0.028	0.972
Founder pair 3	0.987	0.000	0.013	0.000	0.000

Example. How far back can you go from a genome-wide scan?

- Simulate data on a pair of individuals for six extended half-sib pedigrees
- Compute average posterior probabilities of that true relationship versus the hypothesis that they are unrelated.
- Based on Affymetrix 500K frequency data and genetic map data.

Posterior probabilities	HS-1-1	HS-2-2	HS-3-3	HS-4-4	HS-5-5	HS-6-6
IBD prob. share 1 allele	1/2	1/8	1/32	1/128	1/512	1/2024
22 unlinked markers	0.633	0.514	0.501	0.500	0.500	0.500
220 linked markers	0.943	0.594	0.507	0.500	0.500	0.500
2200 linked markers	1.000	0.942	0.617	0.514	0.502	0.501
22 000 linked markers	1.000	1.000	0.950	0.706	0.558	0.509
Affymetrix 500 K	1.000	1.000	1.000	0.848	0.642	0.531

Example (cont): How far back can you go from a genome-wide scan?

- A second simulation study used the same Affymetrix data and genetic map, but now comparison with closer alternatives are considered.
- Specifically, data are simulated for each of the HS-1-1, HS-2-2, ..., HS-5-5 relationships and compared against all five options in addition to the unrelated case.

True family relation	HS-1-1	HS-2-2	HS-3-3	HS-4-4	HS-5-5	Unrel
HS-1-1	1.000	0.000	0.000	0.000	0.000	0.000
HS-2-2	0.000	0.952	0.048	0.000	0.000	0.000
HS-3-3	0.000	0.039	0.750	0.182	0.029	0.000
HS-4-4	0.000	0.000	0.161	0.450	0.287	0.102
HS-5-5	0.000	0.000	0.024	0.275	0.376	0.325
Unrelated	0.000	0.000	0.002	0.090	0.327	0.581

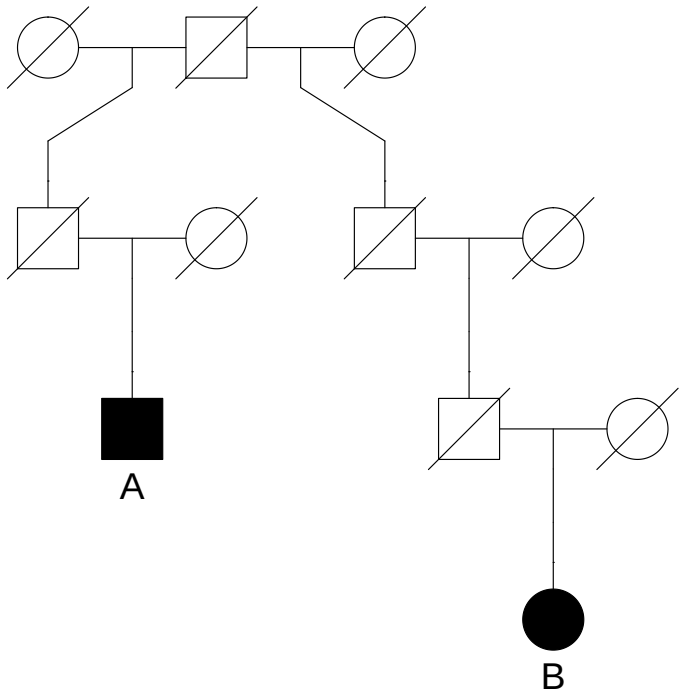
Conclusions: How far back can you go from a genome-wide scan?

If we consider extended half-sib relationships versus the unrelated alternative, previous example shows

- HS-3-3 can be clearly distinguished from unrelated. (number of meiosis $= 6$)
- HS-4-4 can be determined with reasonable certainty. (number of meiosis $= 8$)
- More distant relationships, corresponding to number of meiosis > 8, seem to be outside the scale of what is possible with 500K markers.

Graphs and pedigrees:

Pedigrees are often represented graphically but not necessarily as graphs. A standard representation of a pedigree (squares for males, circles for females):



Graphs representations of pedigrees

- Many possible graph representations of pedigrees: relationship graphs, marriage node graphs, segregation network, ...
- Pedigree graphs are DAGs. Have natural ordering whereby parents precede offspring.
- Pedigrees can have undirected cycles due to inbreeding.

Bayesian network and pedigrees

A pedigree may be represented in form of a Bayesian network that is a DAG satisfying the directed Markov property:

$$p(\mathbf{x}) = \prod_{i=1}^{K} p(x_i|\mathbf{x}_{pa(i)})$$

where $pa(i)$ are the parents of node i. Bayesian network representations of a pedigree:

- Segregation network: contains all details of the inheritance of the two alleles from the parents to the offspring.
- Genotype network: Each node corresponds to an individual. The Markov property holds under the Mendelian inheritance model. The joint probability distribution of the genotype configuration of the pedigree may then be written as

$$p(\mathbf{g}) = \prod_{\text{founders } i} p(g_i) \prod_{\text{non-founders } j} p(g_j|g_{M_j}, g_{F_j})$$