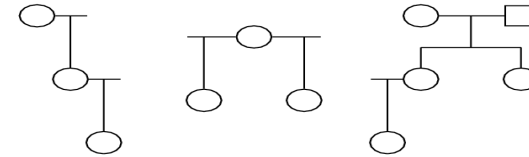


### 3.5 TWO-LOCUS KINSHIP AND GENE IBD

#### 3.5.1 Gene identity at two linked loci

- The simplest example is for the three unilineal ( $K_2=0$ ) pairwise relationships of grandmother-granddaughter (G), half-sisters (H), and aunt-niece (N).
- Each of these relationships has  $\mathbf{k} = (1/2, 1/2, 0)$ , and hence they are indistinguishable on the basis of data at independently segregating loci, i.e. unlinked loci.
- For such relationships, gene identity at two linked loci is summarized by  $K_{11}(\rho) = P(\text{share 1 gene IBD at each of 2 loci at recombination } \rho)$ .
- For the three relationships G, H, and N, we have:
  - G:  $K_{11}(\rho) = (1/2)(1 - \rho)$
  - H:  $K_{11}(\rho) = (1/2)(\rho^2 + (1 - \rho)^2) = (1/2)R$  say
  - N:  $K_{11}(\rho) = (1/2)((1 - \rho)R + \rho/2)$ .
- Thus the relationships are identifiable on the basis of data at two linked loci ( $0 < \rho < 1/2$ ), but not on the basis of data at unlinked loci.
- All the three relationships have  $K_{11}(0) = 1/2$  and  $K_{11}(\rho) = 1/4$ .

#### 3.5.2 Two-locus kinship

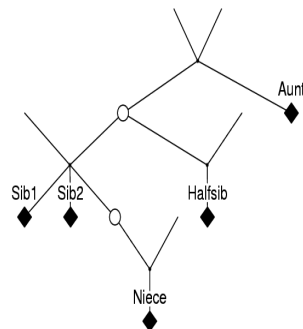


- Note that, although  $K_{11}(\rho)$  is sufficient to specify pairwise genotype and phenotype distributions, it does not determine the two-locus kinship of the individuals, unlike at a single locus where  $\psi = (K_1 + 2K_2)/4$ .
- The shared genes at the two loci may be on the same haplotype in the individual, or on different ones. In fact, in H they are necessarily on the same (maternal) haplotype in the two half-sibs, while in G they may be on either haplotype of the grandmother. For N, for the first term of  $K_{11}$  they are on the same haplotype in the aunt, while the last term corresponds to the case where the genes at the two loci are on two different haplotypes in the aunt.
- In fact, G and H have the same two-locus kinship,  $(1/8)(1 - \rho)^2 R$ .

#### 3.5.3 An example of three related individuals

- Suppose multilocus genetic marker data are available on a pair of sibs, and on a third related individual, who may be an aunt, niece, or half-sister of the pair.
- The gene IBD states at one locus for the 6 genes of the three individuals are the same for niece and half-sib, but different for the aunt.
- For example, the state of 6 distinct genes has probability  $(1/4) \times (1/4) = 1/16$  for the aunt, but 0 for niece and half-sib.
- To distinguish these alternatives we must consider loci jointly (linked), and the three individuals jointly (not pairwise):
 

Individuals:	Pairwise	Joint	
Loci unlinked	$H \equiv N \equiv A$	$H \equiv N$	
Loci linked	$N \equiv A$	$H, N, A$	identifiable



### 3.6 HOMOZYGOSITY MAPPING

#### 3.6.1 Likelihoods for linkage

- Homozygosity mapping was used to map many rare recessive traits. It is no longer important as a practical method but is a unique case where we can see the relationship between IBD and linkage likelihoods.
- Suppose the frequency of the recessive disease allele is  $q$ , and at the marker locus alleles  $A_u$  have frequencies  $p_u$ .
- Suppose that the affected individual has inbreeding coefficient  $f$ , and probability  $f_2(\rho)$  of carrying genes IBD at both of two loci between which the recombination frequency is  $\rho$ .

ibd state	ibd probability	trait data	marker $A_u A_u$	marker $A_u A_v$
I I	$f_2(\rho)$	prob: $q$	prob: $p_u$	prob: 0
I N	$f - f_2(\rho)$	prob: $q$	prob: $p_u^2$	prob: $2 p_u p_v$
N I	$f - f_2(\rho)$	prob: $q^2$	prob: $p_u$	prob: 0
N N	$1 - 2f + f_2(\rho)$	prob: $q^2$	prob: $p_u^2$	prob: $2 p_u p_v$

## Homozygosity mapping likelihoods

- If the individual has marker phenotype  $AuAv$ , he cannot be ibd at marker locus, and likelihood ratio is
 
$$\begin{aligned} L(\rho)/L(\rho = 1/2) &= P(\text{data}; \rho) / P(\text{data}; \rho = 1/2) \\ &= (2 \text{ pu pv} (q (f-f2(\rho)) + q^2 (1-2f+f2(\rho))) / (2 \text{ pu pv} (q (f-f^2) + q^2(1-f)^2)) \\ &= ((f-f2(\rho)) + q(1-2f+f2(\rho))) / ((1-f)(f + q(1-f))). \end{aligned}$$
- This is decreasing in  $f2(\rho)$ , hence increasing in  $\rho$ : MLE  $\rho^* = 1/2$ .
- If the individual has homozygous marker phenotype  $Au Au$  the likelihood ratio is
 
$$\begin{aligned} L(\rho) / L(\rho = 1/2) &= P(\text{data}; \rho) / P(\text{data}; \rho = 1/2) \\ &= (q \text{ pu} f2(\rho) + q^2 \text{ pu} (f-f2(\rho)) + q \text{ pu}^2 (f-f2(\rho)) + q^2 \text{ pu}^2 (1-2f+f2(\rho))) \\ &\quad / (q \text{ pu} f^2 + q^2 \text{ pu} f(1-f) + q \text{ pu}^2 f(1-f) + q^2 \text{ pu}^2 (1-f)^2) \\ &= (f2(\rho) + (q + \text{pu}) (f-f2(\rho)) + q \text{ pu} (1-2f+f2(\rho))) \\ &\quad / (f^2 + (q + \text{pu})f(1-f) + q \text{ pu} (1-f)^2). \end{aligned}$$
- The coefficient of  $f2(\rho)$  is  $(1-q)(1-\text{pu}) > 0$ .  $f2(\rho)$  decreases as  $\rho$  increases, so the MLE is  $\rho^* = 0$
- Also  $f2(0) = f$ , and
 
$$L(\rho = 0) / L(\rho = 1/2) = (f + (1-f) q \text{ pu}) / ((f + (1-f) q) (f + (1-f) \text{pu})).$$
- This is always  $\geq 1$ , and increases as  $q$  or  $\text{pu} \rightarrow 0$ . When  $q=0$ ,
 
$$L(\rho = 0) / L(\rho = 1/2) = 1 / (f + (1-f) \text{pu}); \text{ lod} = -\log(f + (1-f)\text{pu})$$

## 3.6.3 Allelic association and fine-scale mapping

- A small number of unrelated affected individuals all homozygous at the same polymorphic marker locus provides strong evidence for linkage for a rare recessive trait.
- For example one may sample the affected offspring of cousin marriages.
- The evidence is even stronger if the affected individuals are homozygous for the same marker allele, since this shows allelic association (LD) between the recessive trait allele and this marker.
- This suggests that the "unrelated" affected individuals are in fact related, perhaps many generations ago, and the trait allele they share is has been maintained in LD with the marker over those generations: see section 3.1

## 3.6.2 Lod scores and elods

- Log-likelihoods are additive over unrelated pedigrees  $i$ . The base-10 lod score is  $\text{lod}(\rho) = \sum_i \log_{10} Li(\rho)/Li(\rho = 1/2)$  where  $Li(\cdot)$  is the likelihood contributed by pedigree  $i$ .
- The maximized lod score is  $\max_{\{0 \leq \rho \leq 1/2\}} (\text{lod}(\rho))$ .
- Combining over pedigrees, the MLE  $\rho^*$  may be neither 0 nor 1/2. Then  $f2(\rho)$  is also relevant in determining the lod score, not only  $f$ .
- Again, a useful measure of information for linkage analysis is the elod;  $\text{elod}(\rho) = E(\text{lod}(\rho); \rho)$ . The elod is additive over independent pedigrees.
- For simplicity consider  $\rho = 0$  and  $q \approx 0$ . Then each affected individual has prob  $\approx 1$  of having two ibd genes at the disease locus, and hence also at the marker ( $\rho = 0$ ), and so has prob  $\text{pu}$  of being  $AuAu$ . Hence each contributes  $-\sum_u \text{pu} \log(f + (1-f) \text{pu})$  to the elod.
- For example, for the affected offspring of first-cousin marriages ( $f=1/16$ ), and a polymorphic marker locus (for example,  $\text{pu}=0.1$  for each of 10 alleles) the value is  $-\log((1/16) + (15/16)^*(1/10)) = \log(6.4) \approx 0.8$ .
- This is a base-10 elod: about 4 individuals would give a lod score of 3.