

Calculating LOD Scores for Human Pedigrees

Genetics 466

John Doebley

Humans pose a challenge for linkage mapping because of their small family sizes, our inability to make controlled crosses with the critters, and their long generation time. A method for working around these difficulties was devised by J. B. S. Haldane and C. A. Smith in the 1940's and this method has seen an explosion in its usage since the discovery of molecular markers in the 1980's. Haldane and Smith devised a means of calculating the probability of linkage between two genes, often a disease gene and a marker gene, by considering the ratio of the probability of observing a specific pedigree assuming linkage to the probability of observing that pedigree assuming non-linkage (independence) between the two genes.

Consider the following pedigree in which two genes **A** and **B** are segregating in a small family.

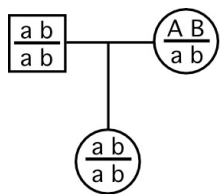


Figure 1

The linkage phase is known, **A** and **B** are in coupling phase, although the recombination frequency (r) between **A** and **B** is unknown. What Haldane and Smith did was to calculate the *odds* ratio for linkage vs. independence. They asked: what is the probability of obtaining the observed pedigree given some value of r relative to the probability of obtaining that same pedigree given independence.

$$(1) \quad odds = \frac{P(\text{pedigree} | r)}{P(\text{pedigree} | \text{unlinked})}$$

When this ratio is large (> 1), the data favor linkage, and when it is small (< 1), the data favor independence.

For the pedigree (Fig. 1), we need to calculate the probabilities of the child inheriting a recombinant vs. nonrecombinant gamete from her mother given some

value of r and given independence. There is a probability of 0.5 that the girl will inherit a nonrecombinant gamete (either **AB** or **ab**) as shown in the pedigree if the genes are unlinked and so the denominator of the odds ratio is 0.5. If **A** and **B** are linked, the probability that the girl will inherit a nonrecombinant gamete (either **AB** or **ab**) is 1 minus the fraction of recombinants (r). If r is 0, then the child can only inherit a nonrecombinant, if r is 0.5, then she has a 50:50 chance. Thus, the odds ratio for this pedigree is

$$(2) \quad odds = \frac{1-r}{0.5}$$

Now consider a second family in which the same two genes are segregating.

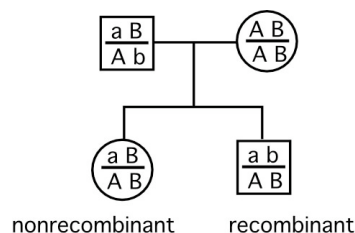


Figure 2

The father is the heteroallelic parent and the linkage phase is reversed relative to Fig. 1. The daughter received a nonrecombinant gamete from her father, which has a probability of $(1-r)$ if the genes are linked. The son received a recombinant gamete, which has a probability of r or the fraction of recombinants. Thus, the probability of this pedigree given some value for r is $(1-r)r$. Since the probabilities of recombinant and nonrecombinant gametes are the same (0.5) if **A** and **B** are unlinked, then the probability of this pedigree given independence is 0.5 for the daughter times 0.5 for the son or $(0.5)^2$. Thus, the odds ratio for the pedigree is

$$(3) \quad odds = \frac{(1-r)r}{0.5^2}$$

We can combine the evidence from the two pedigrees. An important point is that we cannot simply treat these two families as one large

experiment since the linkage phase differs between them. For the first pedigree, **a** and **b** are coupled, while in the second pedigree, **a** and **B** are coupled. However, we can multiply the odds across families to get the combined odds in the same way that we would multiply the probabilities for any two independent events. Thus, the combined odds for both pedigrees is

$$(4) \quad \text{odds} = \frac{r(1-r)^2}{0.5^3}.$$

We can solve equation (4) to find the value of r that maximizes the odds ratio. We can do this by iteration, that is simply solving equation (4) with different values of r such as 0, 0.01, 0.02, 0.03 0.5. Otherwise, we can find the first and second derivatives of equation (4) which tell that the odds ratio is maximized when $r=0.333$. Plugging $r=0.333$ back into equation (4) gives us an odds ratio of 1.18, indicating that linkage is slightly favored over non-linkage. As a matter of convenience, the normal procedure is to take the \log_{10} of the odds ratio (LOD), which in our case would be 0.074. The LOD score will be >0 when linkage is favored and <0 when non-linkage is favored. Because the LOD scores for different families are exponents, you can add them together to get a cumulative LOD score. By convention, human geneticists consider LODs > 3.0 as an indication of convincing support for linkage. This represents odds of 1000:1 in favor of linkage. It may seem overly conservative to require such a large LOD score; however, for reasons that go beyond the scope of this course, it has been determined to be an appropriate significance threshold for this type of test.

For the pedigrees above, the linkage phase of the two genes was known; often however phase is not known and one needs to calculate a LOD score in light of this uncertainty. Figure 3 shows another pedigree for which phase is unknown.

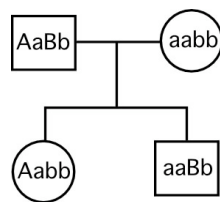


Figure 3

Here the uncertainty is with the linkage phase for the father. There are two possibilities: coupling (**AB/ab**) and repulsion (**Ab/aB**). We would like to calculate $P(\text{pedigree} | r)$, which I will designate $P(\text{PR})$. We can partition $P(\text{PR})$ into two parts depending on whether the genes are in coupling (C) or repulsion (R) in the father:

$$(5) \quad P(\text{PR}) = P(\text{PR}|C)P(C) + P(\text{PR}|R)P(R).$$

If the father's phase is coupling, then both children received recombinant gametes in which case the probability of the pedigree given r [or $P(\text{PR}|C)$] is r^2 . If the father's phase is repulsion, then both children received nonrecombinant gametes in which case the probability of the pedigree given r [or $P(\text{PR}|R)$] is $(1-r)^2$. We assume that the probabilities of coupling and repulsion are the same or that $P(C)=P(R)=1/2$. Plugging these expressions and values into equation (5) gives us

$$(6) \quad P(\text{pedigree} | r) = \frac{r^2}{2} + \frac{(1-r)^2}{2}.$$

Finally, $P(\text{pedigree} | \text{unlinked})$ for the pedigree (Fig. 3) is not affected by phase and is $(1/2)^2$ and so the odds ratio is

$$(7) \quad \text{odds} = \frac{\frac{r^2}{2} + \frac{(1-r)^2}{2}}{(1/2)^2}.$$

Solving this equation to find the value of r that maximizes the odds ratio gives $r = 0.234$ and a LOD of 0.149. Since we are working with the same two genes as in the first two pedigrees, we can sum the LODs ($0.149+0.074=0.223$). Obviously, this combined LOD score is a long way from the value of 3.0 needed to declare linkage, so you can see that a researcher would need to study a large number of families. For this reason, researchers often work with ethnic groups like the Amish who have large families and among whom recessive hereditary disorders are more frequent than in the general population because of a higher instance of marriage among relatives (inbreeding).

When calculating LOD scores from pedigrees, we also need to be able to determine the number of informative individuals or gametes in the pedigree. Let's look at one more pedigree (Fig. 4). In this pedigree, we are looking at an autosomal dominant disease gene segregating along with the MN blood group gene. An informative gamete is one for which we can tell whether or not a recombination has

occurred during the meiosis in which it was formed. Because individual 1 is homozygous MM and individual 2 does not have the disease, the gametes that formed their two children are not informative. Individual 4 is heterozygous for both the disease and blood type genes. The gametes that led to each of his three children are informative. There are no other informative gamete in this pedigree. Thus, three is the total number of informative gametes represented in Fig. 4.

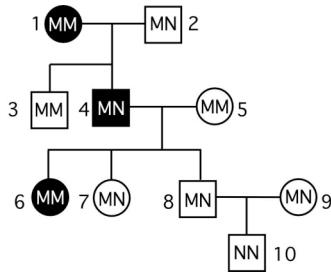


Figure 4

It should be obvious to you that this type of analysis can become rather complicated very rapidly with larger pedigrees, marker genes with multiple alleles, recessive disease genes, incomplete penetrance, variable expressivity, dominant marker genes, uncertain linkage phase, etc. Rather than discourage human geneticists, I think these complications functioned to attract some very bright people to tackle the problems. Their efforts contributed to the identification of many human disease susceptibility genes for conditions such as breast cancer, Alzheimer's disease, manic depression, schizophrenia, cystic fibrosis, diabetes and more. If you have managed to read this far, drop me an email and let me know.