Chapter 10: Morgan

Thomas Hunt Morgan: Introduction

Thomas Hunt Morgan was a famous geneticist who, in the initial years of the 20th century, studied *Drosophila*, the fruit fly, in his lab at Columbia University in New York City. Morgan’s choice of *Drosophila* was both fortuitous and prescient not only for his own historic findings but also for introducing a model organism that has evolved to become a major workhorse in the science of genetics. These tiny flies reproduce quickly and leave a large number of progeny. While Mendel had to wait months to plant and harvest two generations of peas, Morgan could study have a dozen generations in the same time. Moreover, *Drosophila* have only three chromosomes.

Two of Morgan’s many findings stand out. Despite all the complicated looping of the DNA around chromosomal proteins, Morgan found that the genes on a chromosome have a remarkable statistical property—namely, statistically genes appear as if are linearly arranged along the chromosome. Thus, one can draw a schematic of a chromosome as a single strait line, with the genes in linear order along that straight line, even though the

---

1 This part of the chapter contains some advanced material. For the undergraduate student it is important to know the following: (1) who Morgan was; (2) what recombination is; (3) that recombination is a function of physical distance; (4) the definition of a linkage study; (5) the meaning of the recombination fraction, θ; and (6) the meaning of θ = 0 and θ = .5. It is also important to inspect the pedigrees and read the text in order to arrive at some appreciation for linkage analysis. You will not be given data to compute θ, a lod score, or the probability of offspring from a mating using linkage.
actual physical construction of the chromosome is a series of looped and folded DNA.

Figure 11.1 provides an example of the linear arrangements of genes on chromosomes.

Morgan’s second finding was no less important. He discovered that chromosomes can recombine and exchange genetic material.

[Insert Figure 10.1 about here]

**Recombination**

*Recombination or crossing over*, as it also called, refers to the fact that in the genesis of a sperm or egg, the maternal chromosome pairs with its counterpart paternal chromosome and exchanges genetic material. We have already discussed recombination in Chapter 2 under the topic of meiosis. Here, we will deal with the statistical implications of crossing over. The process is diagrammed in Figure 10.2.

[Insert Figure 10.2 about here]

The probability that a recombination event occurs between two loci is a function of the distance between the two loci. The alleles at two loci that are far apart on a chromosome are more likely to recombine than the alleles for two that are close together on the chromosome. For example, in Figure 10.1, it is more likely that loci A and D will have a recombination event between them that it is that loci A and B will involve a recombination. To same the same thing in a different way, the alleles for two genes that are physically close together on a chromosome are more likely to be inherited as a unit than the alleles for two genes that are far apart on the same chromosome.
The facts that (1) genes are arranged linearly on chromosomes and (2) recombination is a function of the physical distance between two genes permit human geneticists to locate genes for disorders and traits among the 23 human chromosomes. The procedure for doing this is called linkage analysis, a statistical technique for tracing the within-pedigree cosegregation of one or more genetic markers with a trait. Careful wording has been used to construct this definition, so let us take it apart, phrase by phrase, starting from the end and working backwards. The trait in human linkage analysis is usually a disorder. (The special case in which the trait is a score on a continuous variable will be discussed later in the text). The markers are genes whose location on a particular chromosome are already known. The term cosegregation refers to the fact that the allele for the trait and the allele(s) for the marker(s) tend to be transmitted together because they are physically close together on the same chromosome. Finally, the within-pedigree phrase denotes that the cosegregation of the marker allele(s) and the trait allele takes place within a pedigree.

The pedigree depicted in Figure 10.3 illustrates the principles of linkage between a marker locus and a gene for a rare dominant disorder. For the disorder, allele D causes the disorder while allele d is the normal allele. The two marker alleles are denoted as A and a. The grandfather in this pedigree has haplotype AD/ad. This means that grandfather has allele A on the same chromosome that contains allele D and, conversely, allele a on the chromosome with d. Grandmother is haplotype ad/ad. In generation II, the two offspring that received grandfather’s allele A (i.e., persons II.2 and II.7) also have the
disorder because they inherited the AD chromosome from their father. The two aa offspring in generation II (i.e., II.4 and II.5) are unaffected because they inherited the ad chromosome from their father.

The grandchildren are consistent with linkage. Granddaughter III.2 received the AD chromosome from her father which was transmitted to him by his father. Her marker genotype is AA because she inherited the other A from her mother. Her brother, III.1, deserves some mention. He has marker genotype Aa, the same marker genotype as his affected father, his affected grandfather, and his affected aunt (II.7). Why is he not affected? The reason is that he received his A allele from his unaffected mother, not from his father. His father gave the ad chromosome that he inherited from his mother (I.2). Note that two other grandchildren also have marker genotype Aa but are unaffected (III.3 and III.6). Once again, this is due to their inheritance of allele A from unaffected parents, II.3 and II.6, respectively. Grandson III.7 is affected because he inherited the AD chromosome from his mother.

It takes time and considerable mental concentration to trace alleles through pedigrees, so do not despair if you find the description of this pedigree somewhat confusing. To assist you in decoding it, examine Figure 10.4 which gives a color-coded schematic for the transmission of the chromosomes. The red chromosome is the one that causes the disorder because it has allele D, so everyone in the pedigree with a red chromosome will have the disorder. Grandfather’s other chromosome is blue, both of
grandmother’s chromosomes are yellow, and the chromosomes of people who marry into the pedigree are white. Notice how all the unaffected grandchildren with genotype $Aa$ inherited the $A$ allele from someone who married into the pedigree.

[Insert Figure 10.4 about here]

**Advanced Topics**

**Estimating recombination**

This pedigree in Figure 10.4 is idealized in the sense that no recombination has taken place between the $A$ locus and the $D$ locus. The situation becomes more complicated when the trait locus and the marker are far enough apart that recombination can occur. This situation is illustrated by the pedigree in Figure 10.5. Note how in this pedigree the allele for the disorder, $D$, is on the same chromosome as allele $a$ whereas in the previous pedigree, $D$ was located on the same chromosome as marker allele $A$.

Consequently affected offspring, with the expectation of II.9, have haplotype $aD/ad$. This is the key point about linkage—in some pedigrees the disease allele, $D$, will be associated with marker allele $A$ while in other pedigrees it will be associated with marker allele $a$. The central feature is that within any single pedigree, the disease allele will consistently cosegregate with the same marker allele. This is what is meant by the phrase *within-pedigree cosegregation* in the definition of linkage.
Offspring II.9 in Figure 10.5 deserves comment. This person is a termed a *recombinant* because the chromosome that he/she inherited from parent I.1 has undergone a recombination between the marker and the disease locus. Instead of receiving either a whole red chromosome or a whole blue chromosome, this person has been given the top portion of the blue chromosome recombined with the lower portion of the red chromosome.

Recombination is measured by the recombination fraction customarily denoted by $\theta$ (lowercase Greek theta). Many people working in the field of genetics define $\theta$ as the probability of a recombination between two loci, and even though that definition is not quite correct, we shall use it here. The lowest possible value for $\theta$ is 0 and its upper limit is 0.5. When $\theta = 0$, there is no recombination between the marker and the trait locus. In practice, this would mean one of two things: (1) either the maker is actually the trait locus, or (2) the two loci are so close together that they rarely recombine and the estimate of $\theta$ rounds off to 0. When $\theta = 0.5$, then the loci are not linked. The marker and trait loci are either on entirely different chromosomes or they are so far apart on the same chromosome that they act as independently assorting loci.

---

2 A *haplotype* is defined as the alleles on a chromosome. If a person is haplotype $A_b/aB$, then that person has one chromosome with alleles $A$ and $b$ on it and a second chromosome with alleles $a$ and $B$ on it.

3 Strictly, $\theta$ is a conditional probability and answers the following question: given that one inherits the allele at locus 1 from the maternal chromosome, what is the probability that one will inherit the allele at locus 2 from the paternal chromosome?
Conceptually, it is quite easy to estimate \( \theta \)--simply count the number of recombinants among offspring and divide by the total number of offspring. For example, the pedigree in Figure 10.5 has one recombinant offspring and nine total offspring, so the estimate of \( \theta \) is \( \frac{1}{9} \) or .11. In practice, however, most geneticists use sophisticated computer algorithms that simultaneously estimate \( \theta \) and test for linkage.

Our old friend the Punnett rectangle can be used to find the expected frequency of offspring genotypes in the presence of linkage. The logic remains the same—one parent’s gametes and their probabilities form the rows of the rectangle while the other parent’s games and probabilities form the columns. The only trick is that the probabilities of the gametes are a function of \( \theta \).

As an example, consider a father who has haplotype \( \text{AD/ad} \) and a mother with haplotype \( \text{ad/ad} \). All of mother’s gametes will be genotype \( \text{ad} \). Father, however, may have one of four different gametes, given by the rows in Table 10.16. The probability that father gives gamete \( \text{AD} \) equals the probability of two independent events—that probability father gives allele \( \text{A} \) (versus allele \( \text{a} \)) at the marker (which equals \( \frac{1}{2} \)) and the probability that a recombination does not occur between the marker and trait locus which equals the quantity \( 1 - \theta \). Mathematically, the probability father transmits \( \text{AD} \) is the product of these two probabilities or \( \frac{1}{2}(1 - \theta) \).

[Insert Table 10.1 about here]
Similarly, the probability that father’s gamete has genotype $Ad$ is the product of the probabilities of two independent events—the probability that father transmits $A$ at the marker (again, $1/2$) and the probability that a recombination occurs between the marker and the trait locus (or $\theta$ in this case). Consequently, the probability of transmitting gamete $Ad$ equals $1/2\theta$. Similar logic will give the probability for father’s gametes $aD$ and $ad$. It is not necessary to understand the probability theory that goes into these calculations. The important point is that the logic of the Punnett rectangle can be applied to the case of linkage just as it can be applied to the case of two independently assorting loci.

Table 10.2 gives numerical estimates of the expected frequency among offspring as a function of the recombination fraction, $\theta$. When $\theta = 0$, father can give only two types of gametes, $AD$ or $ad$. As $\theta$ increases, the proportion of recombinants (i.e., $Ad$ and $aD$) to nonrecombinants (i.e., $AD$ and $ad$) increased. When $\theta$ reaches its upper limit of .5, then all four types of gametes are equally probable. Compare this column in Table 10.2 with the Punnett rectangle in Table 9.4 used to illustrate independent assortment in Mendel’s peas. It is clear that when $\theta = .50$, the two loci assort independently.

[Insert Table 10.2 about here]

**Detecting Linkage**

Dramatic advances in genetics have given us over a thousand different marker loci scattered over the 23 human chromosomes. When geneticists have a genetic disorder but
do not know where the gene is, they gather a large number of families in which the

disorder runs and obtain some biological specimen (usually blood, but sometimes cheek

scrapings) from each member. DNA is extracted from the biological specimen and then
genotyped on a large number of the marker loci. Ideally, marker loci are selected so that
they are evenly spaced but cover each and every chromosome. Sophisticated

mathematical procedures, implemented in equally sophisticated computer algorithms, are
then used to test for linkage through the genome. This type of procedure is called a whole

genome scan.

Although we treated linkage as the study of a single marker with a trait locus, in

practice geneticists prefer to examine several linked markers simultaneously, a procedure

known as multipoint linkage. Multipoint linkage has the major statistical advantage of

being able to detect linkage more powerfully than single point linkage (the analysis of

only a single marker). The actual mathematics of multipoint linkage are too complicated

for us to explore here, but once again, the Punnett rectangle could still be used to calculate

the expected frequency of offspring.

There are two generic types of statistical techniques used in linkage analysis,

parametric linkage analysis and nonparametric linkage analysis. Parametric linkage

analysis uses statistical procedures to estimate \( \theta \) and sometimes other quantities. An

important term in parametric linkage analysis is the \textit{lod score}. The term lod is specific to

genetics and refers to the common logarithm (i.e., the base 10 logarithm) of the odds for

linkage. The odds for linkage equal the ratio of two probabilities. The numerator is the
probability of observing the data given that $\theta$ is some value less than .50 (i.e., the marker and the trait loci are linked) and the denominator is probability of observing the data given that $\theta = .50$ (i.e., the marker and the trait loci are not linked). It is not necessary for us to learn how to compute lod scores, but it is helpful to go through an example.

Figure 10.6 depicts a pedigree for a dominant disorder. We illustrate the calculation of the lod score by assuming that the affected father has haplotype $AD/ad$ and that $\theta$ equals .10. We begin the calculating by finding the probability of each offspring under the hypothesis that $\theta = .10$. We can use Table 10.1 for this purpose. The probability of II.1 is .45, the probability of II.2 is .45, the probability of II.3 is .45, the probability of II.4 is .05, and the probabilities of II.5 and II.6 are both .45. We next multiply all these probabilities together to arrive at the probability of observing this family under the hypothesis of linkage with $\theta = .10$. This quantity is $0.05(.45)^5 = 0.0009226$.

We next calculate the probability of the family under the hypothesis that $\theta = .50$. Again, start by finding the probability for each offspring under the hypothesis that $\theta = .50$. From Table 10.1, this probability will be equal for all offspring and is .25. By multiplying .25 by itself six times, we have the probability of this family when $\theta = .50$-- $0.25^6 = 0.000241$.

The lod score is then the base 10 log of the ratio of these two probabilities, or

$$\text{lod} = \log_{10} \left( \frac{0.0009226}{0.0002441} \right) = 0.577.$$
In traditional linkage analysis of Mendelian traits, a lod score of 3 or more is required in order to be confident of linkage. The present lod score is much less than 3, so we would not take this pedigree as sufficient evidence to claim linkage between the A locus and the disorder. In reality, linkage studies involve studying many pedigrees and lod scores are added over pedigrees.

Parametric linkage and lod scores are suitable for single gene disorders. For complex disorders, like all of psychopathology, many geneticists prefer to use the nonparametric approach. The advantage of the nonparametric techniques is that it is not necessary to make assumptions about the mode of inheritance for the disorder, something that we quite frankly do not know about in the case of psychopathology. The disadvantage of nonparametric approaches is that they are less powerful than the parametric techniques.

It is not possible to survey all the nonparametric techniques here. Instead, we will illustrate one of them, the affected sib-pair method. Here, the geneticist gathers data on a large number of sibships to locate those that have at least two members of a sibship who are affected with the disorder. These affected sib pairs are then genotyped at the marker locus, and the sib pairs are placed into one of two mutually exclusive categories based on their genotypes at the marker. The first category includes all sib pairs who have the same genotype at the marker; we can term these the marker-concordant pairs. The second

---

4 In practice, the parents of the sibs are also genotyped. I omit this complication to make the logic of the design easier to understand.
category is for the *marker-discordant* pairs, i.e., those sib pairs who have different genotypes at the marker.

If the marker is not linked to the gene for the disorder, then we should expect an equal number in both categories\(^5\). However, if the marker is linked to the disease locus, then there should be more marker-concordant pairs than marker-discordant pairs. Hence, one simply performs a statistical test to see whether the number of marker-concordant pairs is significantly larger than the number of marker-discordant pairs.

---

\(^5\) Strictly speaking, when there is no linkage, the ratio of marker concordant to marker discordant pairs is a complicated function of the frequencies of the marker alleles. The example in the text assumes that there are a very large number of alleles so that the frequency of any single allele is always quite small.
Table 10.1. Expected frequency of children from a mating where father is haplotype AD/ad and mother is haplotype ad/ad.

<table>
<thead>
<tr>
<th>Father’s Gametes</th>
<th>Recombinant?</th>
<th>Probability</th>
<th>Mother’s gametes: ad</th>
<th>Probability = 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>No</td>
<td>1/2(1 - θ)</td>
<td>AD/ad</td>
<td>1/2(1 - θ)</td>
</tr>
<tr>
<td>Ad</td>
<td>Yes</td>
<td>1/20</td>
<td>Ad/ad</td>
<td>1/20</td>
</tr>
<tr>
<td>AD</td>
<td>Yes</td>
<td>1/20</td>
<td>aD/ad</td>
<td>1/20</td>
</tr>
<tr>
<td>Ad</td>
<td>No</td>
<td>1/2(1 - θ)</td>
<td>ad/ad</td>
<td>1/2(1 - θ)</td>
</tr>
</tbody>
</table>
Table 10.2. Expected frequency of children from a mating where father is haplotype AD/ad and mother is haplotype ad/ad as a function of the recombination fraction, \( \theta \).

<table>
<thead>
<tr>
<th>Offspring Haplotype:</th>
<th>Recombinant?</th>
<th>Probability</th>
<th>( \theta = )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>.10</td>
</tr>
<tr>
<td>AD/ad</td>
<td>No</td>
<td>1/2(1 - ( \theta ))</td>
<td>.50</td>
</tr>
<tr>
<td>Ad/ad</td>
<td>Yes</td>
<td>1/2 ( \theta )</td>
<td>.00</td>
</tr>
<tr>
<td>aD/ad</td>
<td>Yes</td>
<td>1/2 ( \theta )</td>
<td>.00</td>
</tr>
<tr>
<td>ad/ad</td>
<td>No</td>
<td>1/2 (1 - ( \theta ))</td>
<td>.50</td>
</tr>
</tbody>
</table>
Figure 10.1. An example of homologous chromosomes. Each chromosome has the same ordering of genes but may have different alleles at a locus.
Figure 10.2. Recombination (crossing over). Homologous chromosomes pair up and exchange genetic material.
Figure 10.3. An example of a pedigree demonstrating linkage between a dominant disorder and a marker. The disease allele is on the grandfather’s chromosome that carries the A marker allele.
Figure 10.4. The same pedigree in Figure 11.3 but showing the chromosomes of the individuals. All those with the $AD$ chromosome develop the disorder.
Figure 10.5. Example of a recombinant (person II.9).
Figure 10.6. A pedigree used to calculate a lod score.