

LECTURE 5: LINKAGE AND GENETIC MAPPING

Reading: Ch. 5, p. 113-131

Problems: Ch. 5, solved problems I, II; 5-2, 5-4, 5-5, 5.7 – 5.9, 5-12, 5-16a; 5-17 – 5-19, 5-21; 5-22a-e; 5-23

The dihybrid crosses that we've considered up to this point are those segregating for genes on different chromosomes. What genotypes might we expect to see if the genes are located on the same chromosome?

Genes on the same chromosome assort together more often than not

In dihybrid crosses, departures from a 1:1:1:1 ratio of F1 gametes indicates that the two genes are on the same chromosome (linked). Among the F2 progeny, there are more **parental types** than **recombinant types**.

Consider two *Drosophila* genes linked to the X-chromosome, the eye color gene *white* and a body color gene *yellow*. The wild-type *white* allele is w^+ (red eyes); the recessive mutant allele is w (white eyes). The wild-type allele is y^+ (brown bodies); the recessive mutant allele is y (yellow bodies).

Cross:

P: white, brown female x red, yellow male
 $w^+ y^+ / w^+ y^+$ x $w^+ y / Y$

F1: red, brown female x white brown male
 $w^+ y / w^+ y^+$ x $w^+ y^+ / Y$

F2 males:

white, brown males	$w^+ y^+ / Y$	Parental type	4484
red, yellow males	$w^+ y / Y$	Parental type	4413
red, brown males	$w^+ y^+ / Y$	Recombinant type	76
white, yellow males	$w^+ y / Y$	Recombinant type	53

If the white and yellow genes sort independently of one another, than the F1 female would produce 4 kinds of gametes with equal probability ($w^+ y^+$, $w^+ y$, $w^+ y^+$, $w^+ y$). We can reveal the ratio of her gametes by observing only the male progeny, as sons receive their only X from their mother. As you can see from the data, the four kinds of gametes do not occur with equal frequency. Instead **parental types** vastly outnumber **recombinant types** (99% to 1%).

Bottom line: When two genes assort independently, the numbers of parental and recombinant F2 progeny are equal. Two genes are considered **linked** when the number of F2 progeny with parental genotypes exceeds the number of F2 progeny with recombinant genotypes.

The **recombination frequency (RF)**, the percentage of total progeny that are recombinant) depends upon the gene pair under consideration. Linked genes have a recombination frequency of less than 50%. The example we use above indicates tight linkage (the genes are close together), whereas other gene pairs give different percentages. A RF of 1% indicates tight linkage, whereas a RF closer to 50% would indicate that the genes lie farther apart.

Autosomal traits also exhibit linkage; using the testcross

Transmission of sex-linked traits is easier to follow, since the genotypes of the F1 female chromosomes can be determined by examining her sons (they receive an X from their mother and a Y from their father). Autosomal traits are a bit harder to follow because all progeny receive two copies of autosomal genes. This is where the **test cross** comes in handy!

In this cross, we look at an autosomal gene for body color ($b+$ is brown and recessive b is black) and a gene for wing shape ($c+$ is straight and recessive c is curved).

P: black, straight female x brown, curved male
 $b\ c+ / b\ c+$ x $b+ c / b+ c$

F1: brown, straight progeny
 All $b\ c+ / b+ c$

Test cross: brown, straight F1 female x black, curved male
 $b\ c+ / b+ c$ x $b\ c / b\ c$

Test cross progeny:

black, straight	$b\ c+ / b\ c$	Parental type	2934
brown, curved	$b+ c / b\ c$	Parental type	2768
black, curved	$b\ c / b\ c$	Recombinant type	871
brown, straight	$b+ c+ / b\ c$	Recombinant type	846

The test cross shows that these autosomal genes exhibit linkage; they do not assort independently. Instead the parental types are transmitted together > 50% of the time.

Chi Square Test (p. 117-120); Read about this test and do the problems!

What if linkage is not very tight, and the percentage of recombinant classes approaches 50%? How can we determine statistically if the two genes are truly linked? The Chi Square Test is a probability test that measures the "goodness of fit" between experimental and predicted results.

We test whether the data is consistent with the hypothesis that the genes are not linked, as this hypothesis gives a precise prediction: unlinked genes will assort independently, giving half parental and half recombinant progeny. If the chi square test shows that the observed data are different from that expected from independent assortment, we can reject the null hypothesis and assume the two genes are linked. $\chi^2 = \text{Sum of: } \frac{(\# \text{ observed} - \# \text{ expected})^2}{\# \text{ expected}}$

$$\chi^2 = \text{Sum of: } \frac{(\# \text{ observed} - \# \text{ expected})^2}{\# \text{ expected}}$$

For the example above, the total number of progeny is 7419; if the two genes are not linked, we would expect 25% progeny in each class, or 1855.

$$\chi^2 = (2934 - 1855)^2 / 1855 + (2768 - 1855)^2 / 1855 + (871 - 1855)^2 / 1855 + (846 - 1855)^2 / 1855 \\ = 628 + 494 + 522 + 549 = 2193$$

p value < 0.001 (use Table 5.1 on page 120 of your book)

Use the **chi square value** together with the **degrees of freedom** (measure of the number of independently varying parameters in the experiment; with 4 (N) classes, there are 3 degrees of freedom) to determine a p value. A p value is the probability that a particular set of observed experimental results represents a chance deviation from the values predicted by a particular hypothesis. A low p value suggests that the data showing deviation from the values predicted by the hypothesis are significant enough to reject the null hypothesis. A p value of 0.001 means that it is highly unlikely (1 chance in a 1000) that the observed distribution of phenotypes arose from two unlinked genes.

Recombination results when crossing over during meiosis separates linked genes

Reciprocal exchanges between homologous chromosomes are the physical basis of recombination. We will cover the actual mechanism of recombination in another lecture.

Using chromosomes that had cytologically visible abnormalities, Creighton and McClintock working with maize, and Stern, working with *Drosophila*, showed that recombination depends upon the physical exchange of equal parts between maternal and paternal chromosomes during meiosis.

Recombination frequencies for pairs of genes reflect the distances between them along a chromosome.

Since genes are arranged linearly on chromosomes, Morgan proposed that different gene pairs exhibited different linkage rates because the closer together two genes are the less likely they will be separated by a recombination event. That is the probability of a crossover occurring between two genes increases with the distance separating them. Sturtevant, an undergraduate in Morgan's lab, suggested that recombination frequency could be used to gauge the physical distance between two genes: 1% RF = 1 cM = 1 map unit. Recombination frequency = # recombinants/total progeny x 100.

Experimental recombination frequencies between two genes are never greater than 50%.

Recombinants among the F2 progeny are never in the majority. Genes on different chromosomes yield 50% recombination frequency because of independent assortment. Genes that lie far apart on the same chromosome also show 50%. The only way to tell for sure whether the two genes are on the same chromosome is to show definite linkage with other genes that lie in between them. How do we do that? By mapping.

131	vg+ b pr+	recombinant for b relative to vg and pr
118	vg b+ pr	"
13	vg b pr+	recombinant for pr relative to vg and b
<u>9</u>	vg+ b+ pr	"
4197		

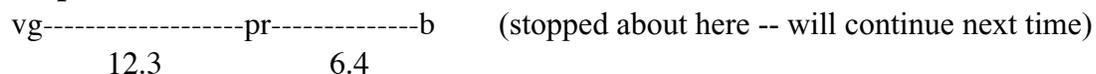
distance between vg and b is: $(252 + 241 + 131 + 118) / 4197 \times 100 = 17.7$ m.u. (***** this is not correct - see below!**)

distance between vg and pr is $(252 + 241 + 13 + 9) / 4197 \times 100 = 12.3$ m.u.

distance between b and pr is $(131 + 118 + 13 + 9) / 4197 \times 100 = 6.4$ m.u.

The recombination frequencies show that vg and b are separated by the largest distance and therefore are the outside genes. The two smallest classes represent the double crossover events, in which pr alleles are recombined relative to vg and b. The arrangement of alleles in double recombinants indicates the relative order of the three genes, because in gametes formed by double crossovers, the gene whose alleles have recombined relative to the parental configuration must be the one in the middle.

Map:



But, again, the distance separating the outside genes vg and b (17.7 map units) is smaller than the sum of the two intervening distances ($12.3 + 6.4 = 18.7$). Why? Because the distance between vg and b that we calculated does not account for all the recombination events (double crossovers).

Our three-point cross data lets us adjust the recombination frequency. How? We add the recombination frequency of the double crossovers twice, since each individual in the double crossover groups is a result of two exchanges between vg and b. DOUBLE CROSSOVERS are needed to generate gametes in which the middle gene is recombined relative to the two flanking genes. Thus when we calculate the genetic distance between the two outside markers, vg and b, we must add in the double crossovers twice.

***** CORRECTION TO ABOVE TO ACCOUNT FOR DOUBLE CROSSOVERS:**

distance between vg and b is $(252 + 241 + 131 + 118 + 13 + 13 + 9 + 9) / 4197 \times 100 = 18.7$ m.u.

*******We'll cover the rest below at the beginning of the next lecture*******

A few additional points about mapping:

- Mapping reveals the relative order of genes, not the actual physical distance.
- The most accurate maps are made by summing the genetic distances of genes lying close together (small intervals). One needs to connect genes that lie far apart through the genes that lie between them.

- **INTERFERENCE**: A measure of the independence of crossovers from each other. (That is, does a crossover in one region affect the likelihood of a crossover in an adjacent region?)

Calculating interference: First of all, what is the probability of double crossovers occurring? Consider our example of vg, pr, and b linkage. We can calculate the probability of a double crossover using the Law of the Product rule. As long as a crossover in one region does not affect the probability of a crossover in another region, the probability of a double crossover is simply the product of their separate probabilities.

Probability of single crossover between vg and pr = 0.123 (corresponds to 12.3 map units)

Probability of single crossover between pr and b = 0.064 (corresponds to 6.4 map units)

Probability of both is $0.123 \times 0.064 = 0.0079$ or 0.79%

But the number actually observed was different: $([13 + 9]/4197) \times 100 = 0.52\%$, meaning that the two crossovers are not occurring independently, but instead a crossover in one region reduces the likelihood of a crossover in an adjacent region.

Coefficient of coincidence = frequency observed / frequency expected = $0.52 / 0.79 = 0.66$
(thus, in our example, only 66% of the expected DCOs took place).

Interference = $1 - \text{coefficient of coincidence} = 1 - 0.66 = 0.34$.

If interference is zero, this means that the frequency of double crossovers is as expected and that a crossover in one region occurs independently of a crossover in an adjacent region. If interference is 1, this means that interference is complete and that no double crossovers are observed because a crossover in one region completely blocks a crossover in an adjacent region.