

LECTURE 4: LINKAGE AND GENETIC MAPPING**Reading for this and next lecture:** Ch. 5, p 123-141**Problems for this and next lecture:** Ch. 5, solved problems I, II; #2 – 5, 7 – 9, 12, 14, 15, 20, 21, 23, 24 and 27

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%).

How do we know that that there isn't just a preference for inheritance of a particular combination of alleles (e.g., w with y^+ and w^+ with y) rather than a preference for the parental combination of alleles? Perform a cross that begins with different combinations of alleles in the parents!

P: red, brown female x white, yellow male

$$w^+ y^+ / w^+ y^+ \quad \times \quad w y / Y$$

$$\text{F1: red, brown female} \quad \times \quad \text{red brown male}$$

$$w^+ y^+ / w y \quad \times \quad w^+ y^+ / Y$$

F2 males:

red, brown males	$w^+ y^+ / Y$	Parental type	}	99%
white, yellow males	$w y / Y$	Parental type		
red, yellow males	$w^+ y / Y$	Recombinant type	}	1%
white, brown males	$w y^+ / Y$	Recombinant type		

In this case, the parental types are different combinations of alleles than those in the previous cross, and again, the parental types dominate over the recombinants. The assignment of “parental” versus “recombinant” to progeny from an F1 mother depends upon the set of alleles that she received from each of her parents.

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 easy 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).

$$\text{P: black, straight female} \quad \times \quad \text{brown, curved male}$$

$$b c^+ / b c^+ \quad \times \quad b^+ c / b^+ c$$

$$\text{F1: brown, straight progeny}$$

$$\text{All } b c^+ / b^+ c$$

$$\text{Test cross: brown, straight F1 female} \quad \times \quad \text{black, curved male}$$

$$b c^+ / b^+ c \quad \times \quad b c / b c$$

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. Linkage between two genes is never 100% - if you look at enough progeny, you will find a recombinant!

Could you determine the extent of linkage between b and c genes in the following crosses?

- (1) black, straight female (b c+ / b c) x black, curved male (b c / b c)?

Answer: No, alleles at the body color gene are not segregating in the cross

- (2) black, curved female (b c / b c) x brown, straight male (b c+ / b+ c)

Answer: Even though the genotypes for body color and wing shape of the individuals are exactly same testcross as the one we originally did to demonstrate linkage of these two genes, this cross would not give the same results. This is because there is no crossing over between homologous chromosomes in male flies. Only parental combinations of alleles would be observed among the progeny of this cross!

Chi Square Test (p. 127-131); Read about this test and do the problems!

What if linkage is not very tight, and the percentage of recombinant classes approaches 50%? Or if our sample size is small and we are not sure if deviations from the expected ratios reflect linkage or just chance? 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 "null 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. (The hypothesis that the two genes are linked does not give a precise prediction, because we would not know whether to predict that the genes are tightly linked, barely linked, or somewhere in between.) If the chi square test shows that the observed data are different from that expected from independent assortment (the null hypothesis), we can reject the null hypothesis and assume the two genes are linked.

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

Just looking at the Chi-Square equation, one can see that the chi-square value represents a measure of the deviation of the observed results from the expected results – the greater the value, the greater the deviation, the less likely that the differences observed can be explained by chance

alone. You should also see that this probability test also takes into account that the sample size is an important component of statistical significance.

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.

$$\begin{aligned}\chi^2 &= (2934 - 1855)^2 / 1855 + (2768 - 1855)^2 / 1855 + (871 - 1855)^2 / 1855 + (846 - 1855)^2 / 1855 \\ &= 628 + 494 + 522 + 549 = 2193\end{aligned}$$

p value < 0.001 (use Table 5.1 on page 129 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.

You could also do this problem with null hypothesis that parental: recombinant classes would be 1:1. With only two classes, you would have 1 degree of freedom. Chapter 5, problem #6, illustrates why you might choose one null hypothesis over another.

Next time, we'll talk about how to set up crosses to determine the positions of genes relative to each other along a chromosome and to gauge the distance between them.