

Mapping a Molecular Marker in *Drosophila*

INTRODUCTION

In this lab activity, you will study the inheritance of a molecular marker in *Drosophila melanogaster*. Specifically, you will conduct an experiment to demonstrate that the marker is linked to a particular gene for a visible characteristic in *Drosophila*. This experiment demonstrates one of the strategies that is often used in the early stages of genome characterization in humans and other organisms. In this experiment, you will study the inheritance of the marker and that of two genes associated with physical mutations, *black* body and *Bar* eye. Your goal will be to demonstrate that the marker is linked to the *black* body gene and to estimate from the data how closely the two loci are linked.

STRATEGIES FOR GENOME CHARACTERIZATION

Information to characterize an organism's genome can be obtained in various ways. One of the most obvious is by sequencing some or all of the organism's DNA. However, this step is generally preceded by other steps that provide more general information about genome organization. For example, since the function of most DNA from a given organism is often unknown, it is generally necessary to begin by locating specific genes.

One way to locate specific genes is to use molecular markers to identify these genes. Various kinds of molecular markers can be utilized in different ways. Some are very specific, such as nucleic acid probes. These probes recognize and bind with unique DNA sequences. However, to generate DNA probes, some sequence information for the gene in question is required. This is possible if, for example, the gene of interest has already been identified and sequenced in a closely related species. In this case, it is likely that the gene sequences would be similar enough that the sequence data from one species could be used to design a probe to locate the gene in the other organism.

However, it is also possible to find genes for which no sequence information is yet available. This can be done by using markers that are based on random genetic polymorphisms (polymorphisms are variations in length or other attributes). Such an approach was used to locate several human genes associated with genetic disorders during the early stages of the Human Genome Project. (The gene associated with cystic fibrosis was discovered this way; the story of the discovery of this gene can be found at <http://www.hhmi.org/genetictrail/a100.html>.) Some polymorphisms used for markers are based on the presence or absence of particular restriction enzyme sites. These markers provide characteristic "genetic fingerprints" for individuals because restriction enzymes cut the DNA into different-sized fragments.

Another type of molecular marker is based on polymorphisms that are due to actual variation in the length of a particular region of DNA (not just the presence or absence of a restriction site). You will be using such a marker — a *microsatellite sequence* — in this lab activity. In microsatellite

sequences, the length variations are due to the presence of small nucleotide sequences that are repeated several to many times. The length variations can be detected by using a polymerase chain reaction (PCR) to amplify the microsatellite sequence (if you are unfamiliar with PCR, please review the description presented in Lab Activity 6, entitled “Designing PCR Primers”). The PCR primers used to amplify these sequences flank the regions where the repeats occur. If fewer repeats are present, the resulting PCR product is shorter; if more repeats are present, it is longer.

Although the significance of most microsatellite variations is unknown, some that occur in humans are associated with genetic diseases, such as Huntington’s disease. Many other microsatellites in humans and other organisms appear to have no effect on health or viability. However, they are of interest to geneticists because they can be used as molecular markers. For example, when geneticists search for the location of genes associated with diseases, they begin by screening members of families that carry the disease, looking for patterns associated with large numbers of molecular markers. What they hope to find is a pattern that will indicate that one or more of the markers is linked to the disease-causing form of the gene.

To illustrate, consider the example of a family that carries a hypothetical dominant condition (see Figure 8.1). The condition is exhibited by the father and three out of seven siblings. In this case, the father carries a molecular marker (the 8 kb fragment in Figure 8.1) that is also present in the three affected siblings. Significantly, the mother lacks this marker, and so do the unaffected siblings. In this example, it seems quite likely that the marker is physically associated with the gene that causes the disease. Further confirmation could be obtained by finding a similar pattern among families of any affected brothers or sisters of the affected father in Figure 8.1. The situation illustrated in Figure 8.1 suggests that the marker is very closely linked to the gene in question (because all affected individuals carry the marker). In most cases, a linked marker would occur more frequently with a given phenotype than apart from it, but it would not always be associated with it because of genetic recombination.

Most genetic analyses of this type would be more complex than the situation described in Figure 8.1. For one thing, it is usually necessary to screen large numbers of markers before finding one that shows a pattern suggesting linkage. However, despite the complexity, this approach often pays off, yielding a marker that is physically linked to a gene of interest. If the marker has been well characterized, its location on a particular chromosome may already be known. In this case, the location of the gene of interest can be narrowed to a single chromosome. Furthermore, whether the location of the marker is known or not, the marker itself can be used as a nucleotide probe. Such a probe could then be used to screen a genomic DNA library, which could lead to the isolation of the actual gene of interest itself.

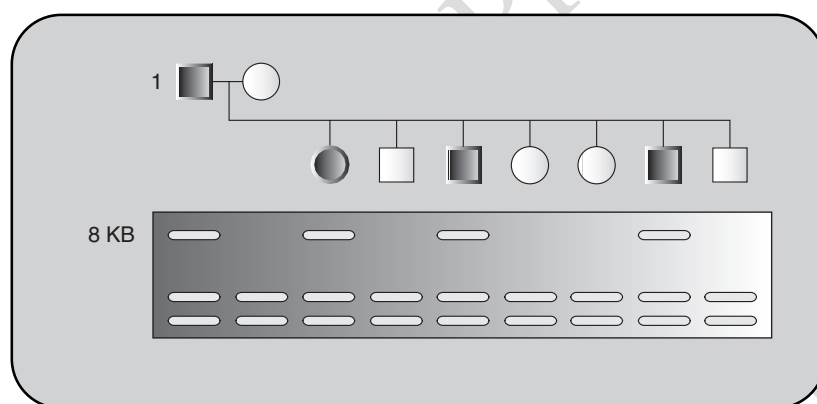


Figure 8.1 – The upper portion of the figure shows a pedigree for a family carrying a gene for a hypothetical dominant condition. As indicated by the shaded symbols, the father and three of his children are affected by the condition. The bottom half of the figure shows the electrophoresis patterns for a molecular marker (the 8 kb fragment) and other nonspecific bands for each individual in the pedigree.

BACKGROUND INFORMATION FOR THE LAB PROCEDURE

Since it is somewhat difficult to conduct genetic experiments with human subjects, in this lab, you will work with a molecular marker present in *Drosophila*. Furthermore, for the purpose of this lab activity, it is not feasible to conduct many experiments using a large number of randomly selected molecular markers and different strains of fruit flies. Therefore, you will be given a specific cross to analyze, and you will know ahead of time that the molecular marker is linked to the *black* body gene. However, it is still important that you understand how this type of analysis works to comprehend how it shows that the marker and the *black* gene are linked. In addition, you will carry the analysis one step further, by using the data you obtain to determine how many map units separate the marker and the *black* body gene.

Your class will be conducting the molecular analysis with the parents and offspring from a test cross involving the two strains carrying different versions of the molecular marker (see the details given later). The test cross will utilize F1 female flies that are heterozygous for the genes associated with *black* and *Bar* (and also heterozygous for the molecular marker those strains carry — discussed later) and male flies that are homozygous for the *black* mutation (and also homozygous for one form of the molecular marker).

Your instructor may choose to give you the parental flies (*black* body and *Bar* eye) and have you conduct all the steps needed to prepare the cross. Alternatively, you may be given the F1 flies and the black flies and instructed to test-cross them, or you may simply be given the test-cross progeny to analyze. The first and second alternatives require more time (at least 4 to 5 weeks for the first alternative) and a great amount of care to obtain meaningful results. If your instructor chooses to have you conduct all or some of the steps necessary to obtain the resulting progeny for the cross, you should consult a separate exercise in this manual (Lab Activity 3, entitled “Introduction to *Drosophila* and Conducting Crosses”) to learn appropriate techniques for mating fruit flies. Since obtaining meaningful results will require the collection of large amounts of data, this experiment should be conducted as a joint effort by an entire lab section or genetics class.

The Strains and the Genotypes

The two strains of flies that serve as original parental strains in this cross are each homozygous for a physically observable phenotype as well as a molecular marker (distinguishable by PCR and electrophoresis — discussed later).

The *black* Parents

The *black* parental strain is homozygous for the recessive allele that produces a black body color. Since this is a recessive phenotype, the symbol for the black allele is “b,” and black fruit flies have the genotype “bb.” Instead of using a capital *B* to indicate the dominant form of this gene (which produces normal, brownish, body color), *Drosophila* geneticists use the symbol + to indicate all normal (wild-type) alleles. Using this system, the three possible genotypes for this character are ++, +b (which both exhibit the dominant, wild-type phenotype), and bb (which exhibits the recessive, black phenotype). Using this system, you can tell just by looking at the three possible genotypes that the phenotype being studied is a recessive one (because a lowercase *b* is used as the symbol for black).

The *black* parental strain is also homozygous for the form of the molecular marker that produces the “short” (about 132 base pairs) PCR product. We will symbolize the genotype for this trait in the black flies as “SS” (flies that are homozygous for the longer form of the marker will be symbolized as “LL,” and heterozygotes will be symbolized as “LS”). You can’t tell by looking at one of these black

parental flies that it has the genotype SS. The only way you can tell is by extracting its DNA, amplifying it with PCR, and then running the PCR product on an electrophoresis gel. DNA from a fly with the genotype SS will yield a PCR product that will produce only one band on a gel. DNA from an LL fly will yield a product that will produce one band also, but the band will be closer to the top of the gel (because it is a longer PCR product), and DNA from an LS fly will yield two bands.

The *Bar* Eye Parents

The allele associated with *Bar* eye is an X-linked “semidominant” allele. Because it’s on the X chromosome, only females can have two copies of the mutant allele (BB). Such females have extremely narrow eyes that are about one-third the width of the eyes of wild-type flies. Females with the genotype +B have slightly wider eyes, but they are still narrower than those of wild-type flies (++). Males can only have two genotypes for this gene, since it is an X-linked gene. The two genotypes can be symbolized as +Y and BY (where “Y” indicates the Y chromosome); such flies have normal eyes and somewhat narrowed eyes, respectively.

The *Bar* eye parental flies are homozygous for the “long” form (about 165 base pairs) of the molecular marker, having the genotype LL. Remember that “S” and “L” are alternative forms of the same marker. Also remember that the microsatellite marker is at a different locus from either *Bar* or *black* (although it is linked to *black*).

The Test Cross

This cross is conducted by using virgin females that are heterozygous for all three loci (F1 flies; genotype +B, +b, LS) and male flies from the *black* parental strain (+Y, bb, SS). Only females are used as the heterozygous parent because male *Drosophila* do not exhibit genetic recombination. To estimate the number of map units that separate the *black* locus from the molecular marker, it is necessary to determine the percentage recombination that occurs between these two loci.

Expected Results: Testing the Null Hypothesis

To analyze the results of this cross, you will need to assess each fly from the test cross progeny for two characteristics. These characteristics include the visible one, which is body color, and the molecular characteristic, which is number, and types of PCR products for the molecular marker (LL, SS, or LS). This means that you will be observing each fly and recording its body color, then you will be extracting its DNA, preparing a PCR reaction, and characterizing its “molecular genotype” using gel electrophoresis. The results that you should expect in this case will be observations that support the hypothesis that the molecular marker is, in fact, linked to the gene for *black* body (you will also be quantifying the apparent amount of genetic recombination).

The simplest way to infer that linkage exists between two genetic loci is to disprove the null hypothesis, which, in this case, is that the two loci are not linked. If you can clearly identify what results the null hypothesis would predict, it becomes easy to identify results that would disprove the null hypothesis (implying that the two genes are linked). To help you visualize this relationship, consider Figure 8.2. This figure illustrates one possible interpretation of the test cross you will analyze. It is based on the null hypothesis that the gene for *black* body and the locus for the molecular marker are unlinked (realize that this null hypothesis is, in fact, incorrect).

Now, to consider the implications of the null hypothesis and to understand the kinds of data you will expect to observe if the null hypothesis is incorrect (as expected), answer the following questions:

Q1 Consider the first category of test-cross offspring shown in Figure 8.2 (+b, LS).

- A. What would be the physical phenotype of these flies?
- B. If PCR was conducted with the DNA of one of these flies using the primers for the molecular marker, what would be the appearance of the bands on an electrophoresis gel with the PCR products?

Q2 If Figure 8.2 was accurate (i.e., if the null hypothesis was correct), (a) what proportion of the test-cross progeny would be black flies that are heterozygous for the molecular marker? (b) What proportion would be flies with normal body color, which are homozygous for one form of the molecular marker?

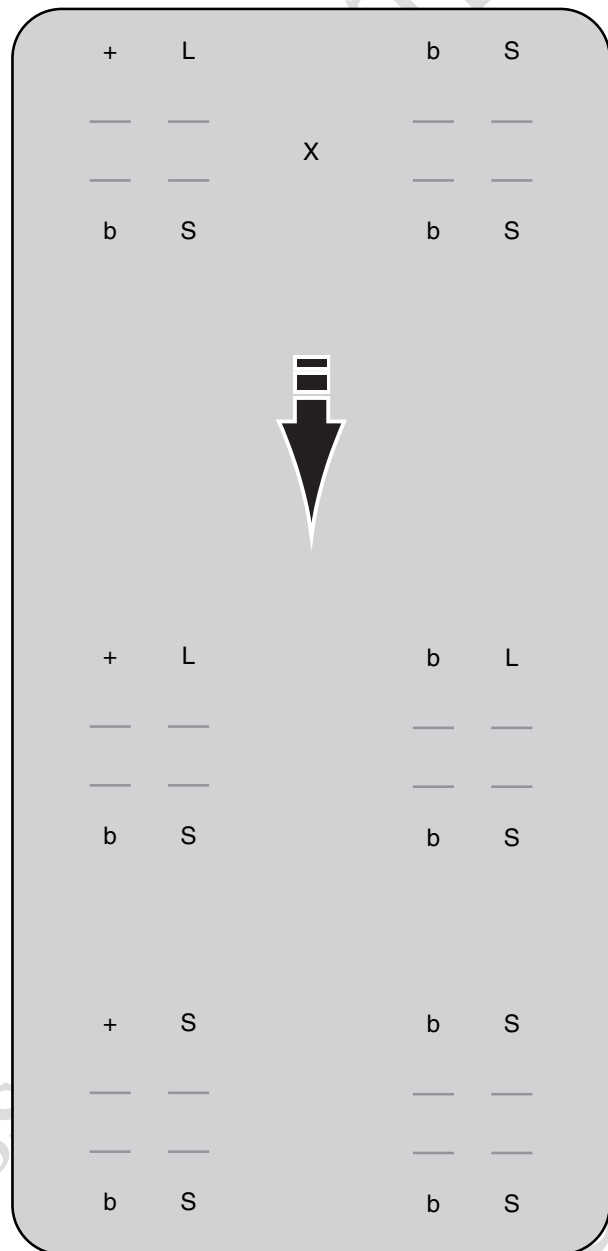


Figure 8.2 – An illustration of results of a test cross for the cross involving the black gene and a molecular marker. The figure illustrates how the alleles would assort if the null hypothesis were correct (i.e., if the gene for *black* and the locus of the molecular marker were unlinked). Each horizontal line represents a single chromosome.

- Q3** If the null hypothesis is incorrect (as expected), the same four genotypes should be present in test-cross progeny, as shown in Figure 8.2; however, the proportion of each type of progeny will differ (and the arrangement of the genes on the chromosomes will be different). Describe in general terms how the proportion of flies in each genotypic category will be different assuming the null hypothesis is incorrect (describe how the number of flies *in each category* would be expected to change). *Note:* It is important to remember that the parents of the heterozygous female flies in the test cross had the following genotypes: bb , SS , and $++$, LL .
- Q4** Assuming that the data you collect do, in fact, indicate that the *black* gene and the locus for the molecular marker are linked, how can you estimate the number of map units that separate these two loci? Write out a formula (using expected categories from the test-cross progeny) indicating how you will estimate this value.

PROCEDURE

Either you will be given flies representing the progeny of the test cross described earlier, or your instructor will have you generate these progeny via the appropriate crosses. Together with the rest of your class, you should analyze as many of the test-cross progeny as possible (it should be possible for each student to process several flies, if necessary). Each fly should be given a code number that will identify it in your records and will also be used to designate its DNA and the resulting PCR product. The instructions for Steps 1 through 3 are given here for individual flies.

Step 1: Observe the anesthetized fly using a dissecting microscope. Record a code number for this fly and record its body color.

Step 2: Grinding individual flies in a buffer solution

- a. Label a microcentrifuge tube with the appropriate code number representing the fly, and place the anesthetized fly in the tube.

- b. Add 50 μl of homogenization buffer to the tube, and grind the fly with a plastic pestle as completely as possible. Use care when grinding and when removing the pestle to leave as much fly homogenate in the bottom of the tube as possible.
- c. Incubate the tube for 25 minutes at approximately 30°C. This can be accomplished by placing the tube in a floating tube “rack” in a 30°C water bath.
- d. Incubate the tube for 1 to 2 minutes at about 95°C using a similar procedure. Use care when removing the tubes from the hot water.
- e. If a microcentrifuge is available, briefly spin the homogenized fly so that remaining tissues form a pellet at the bottom of the tube. Keep this solution on ice during subsequent steps.

Step 3: Preparing and amplifying PCR reactions for each fly

- a. Label a PCR tube with the appropriate code and transfer 2 μl of Fly DNA solution into it.
- b. Add 23 μl of PCR “master mix.” The master mix contains all the components needed for a PCR reaction, except the Fly DNA (added in Step a). The final PCR reaction contains the following: 10 mM Tris-HCl (pH 8.3), 0.75 units of *Taq* polymerase, 50 mM KCl, 1.5 mM MgCl_2 , 0.2 mM dNTP mix, and 50 pmoles of each primer.
- c. If your thermal cycler requires oil (your instructor will advise you of this), add 1 to 2 drops of mineral oil to the side of the PCR tube. It will form a layer over the aqueous PCR mixture.
- d. If you added oil, burst-spin the PCR tube in a microcentrifuge to separate the aqueous PCR mix completely from the oil. *Caution:* Place PCR tubes within the “larger” microcentrifuge tubes *with the microcentrifuge tube caps cut off* within the microcentrifuge; otherwise, the PCR tubes will not remain in the rotor. Spin the tubes for several seconds.
- e. Your instructor will set up the thermal cycler to conduct PCR.

Step 4: Conducting electrophoresis of PCR samples (this will be done on a different day from the preceding procedures)

- a. Add 5 μl of loading dye to the aqueous portion of each PCR sample. Burst-spin as done in Step 3 to separate the oil from the dyed aqueous solution if necessary.
- b. Set your micropipette for the volume indicated by your instructor. Carefully submerge the tip of your micropipette below the oil layer (if present) and slowly depress the plunger, expelling small air bubbles into the dyed aqueous solution. Then *slowly* release the plunger, aspirating the measured amount of the dyed PCR solution. Raise the tip above the oil layer and move the tip around the upper wall of the tube to remove excess oil adhering to the outside of the tip.
- c. Add your sample to one well in the gel (note which well[s] you load), by submerging the tip below the buffer solution and directly into the opening of a well. Do not expel the solution until you are sure that your tip is inserted into a well, and do not insert your tip too far into the well (be careful to avoid puncturing the bottom of the gel with your pipette tip). When you expel the sample into the well, you should be able to see the colored solution settling into the well and filling it. One student should also load one lane on each gel with a molecular marker for size determination.

- d. After the gel is loaded, your instructor will set the voltage.
- e. The gel should be allowed to run until the dye front is close to the opposite end of the gel (unless a somewhat long gel is used). The gel should be viewed immediately following electrophoresis.

Step 5: Characterizing the bands present on gels — View the gel(s) on an ultraviolet light box or use a gel documentation system to photograph the gel (wear gloves when handling the gel — it contains a potent mutagen, ethidium bromide). If you view the gel directly (rather than viewing an image of the gel), *wear appropriate plastic eye protection*. Since prolonged exposure to UV light may also harm the skin, students may wish to wear a plastic face shield and appropriate clothing to cover exposed skin.

The pattern in each sample lane should represent either a “long/long” homozygote (one band closer to the top of the gel), a “long/short” heterozygote (two bands), or a “short/short” homozygote (one band closer to the bottom of the gel). To verify that appropriately sized bands were produced by PCR, you should compare the sample bands to the DNA in the molecular marker lane. Depending on the efficiency of your DNA isolation technique, there may be one or more lighter bands in addition to the band(s) of expected size; however, in most cases, one or two bright bands should be obvious.

To accomplish the analysis of the data from this experiment, you must determine and tabulate the body color and the molecular genotype (LL, LS, or SS) of each fly. You should do this in a large group with the assistance of your instructor. Each student will then analyze the entire data set and answer the questions that accompany this lab activity.

Q5 Indicate the number of flies in each of the four genotypic categories among the test-cross progeny (i.e., +b, LS; +b, SS; bb, LS; bb, SS).

Q6 How many map units appear to separate the locus for the *black* gene from the locus for the molecular marker? (Show the calculations you used to arrive at this number.)