Genetic linkage & mapping

What it means for genes to be linked. How to determine recombination frequency for a pair of genes.

Key points:

When genes are found on different chromosomes or far apart on the same chromosome, they assort independently and are said to be unlinked. When genes are close together on the same chromosome, they are said to be linked. That means the alleles, or gene versions, already together on one chromosome will be inherited as a unit more frequently than not.

We can see if two genes are linked, and how tightly, by using data from genetic crosses to calculate the recombination frequency. By finding recombination frequencies for many gene pairs, we can make linkage maps that show the order and relative distances of the genes on the chromosome.

Introduction

In general, organisms have a lot more genes than chromosomes. For instance, we humans have roughly 19,000 genes on 23chromosomes (present in two sets). Similarly, the humble fruit fly—a favourite subject of study for geneticists—has around 13,000 genes on4chromosomes (also present in two sets).

The consequence? Each gene isn't going to get its own chromosome. In fact, not even close! Quite a few genes are going to be lined up in a row on each chromosome, and some of them are going to be squished very close together .

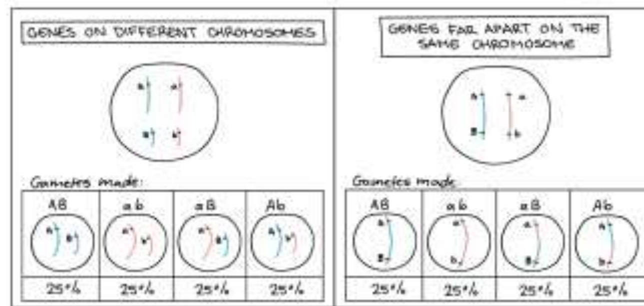
Does this affect how genes are inherited? In some cases, the answer is yes. Genes that are sufficiently close together on a chromosome will tend to "stick together," and the versions (alleles) of those genes that are together on a chromosome will tend to be inherited as a pair more often than not.

This phenomenon is called genetic linkage. When genes are linked, genetic crosses involving those genes will lead to ratios of gametes (egg and sperm) and offspring types that are not what we'd predict from Mendel' s law of independent assortment. Let's take a closer look at why this is the case.

What is genetic linkage?

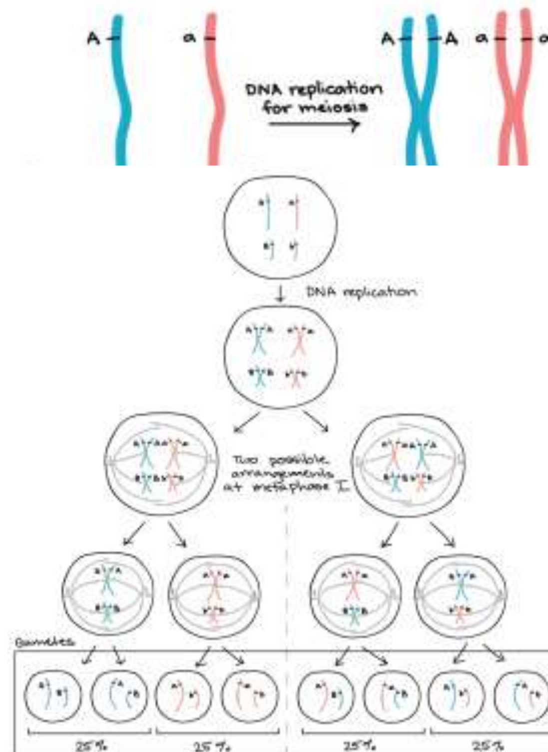
When genes are on separate chromosomes, or very far apart on the same chromosomes, they assort

independently. That is, when the genes go into gametes, the allele received for one gene doesn't affect the allele received for the other . In a double heterozygous organism (AaBb), this results in the formation of all 4possible types of gametes with equal, or 25%, frequency .



Why is this the case?

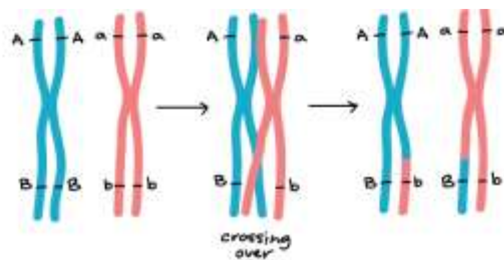
Genes on separate chromosomes assort independently because of the random orientation of homologous chromosome pairs during meiosis. Homologous chromosomes are paired chromosomes that carry the same genes, but may have different alleles of those genes. One member of each homologous pair comes from an organism's mom, the other from its dad.



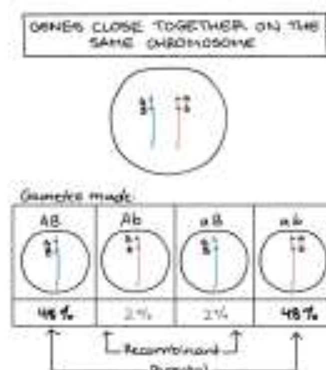
As illustrated in the diagram below, the homologues of each pair separate in the first stage of meiosis. In this process, which side the "dad" and "mom" chromosomes of each pair go to is random.

When we are following two genes, this results in four types of gametes that are produced with equal frequency.

When genes are on the same chromosome but very far apart, they assort independently due to crossing over (homologous recombination). This is a process that happens at the very beginning of meiosis, in which homologous chromosomes randomly exchange matching fragments. Crossing over can put new alleles together in combination on the same chromosome, causing them to go into the same gamete. When genes are far apart, crossing over happens often enough that all types of gametes are produced with 25% frequency.



When genes are very close together on the same chromosome, crossing over still occurs, but the outcome (in terms of gamete types produced) is different. Instead of assorting independently, the genes tend to "stick together" during meiosis. That is, the alleles of the genes that are already together on a chromosome will tend to be passed as a unit to gametes. In this case, the genes are linked. For example, two linked genes might behave like this:

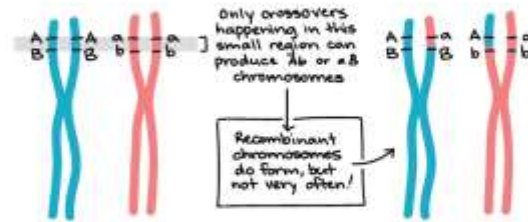


Now, we see gamete types that are present in very unequal proportions. The common types of gametes contain parental configurations of alleles—that is, the ones that were already together on the chromosome in the organism before meiosis (i.e. on the chromosome it got from its parents). The rare types of gametes contain recombinant configurations of alleles, that is, ones that can only form if a recombination event (crossover) occurs in between the genes.

Why are the recombinant gamete types rare? The basic reason is that crossovers between two genes that are close together are not very common.

Crossovers during meiosis happen at more or less random positions along the chromosome, so the

frequency of crossovers between two genes depends on the distance between them. A very short distance is, effectively, a very small "target" for crossover events, meaning that few such events will take place (as compared to the number of events between two further-apart genes).



Thanks to this relationship, we can use the frequency of recombination events between two genes (i.e., their degree of genetic linkage) to estimate their relative distance apart on the chromosome. Two very close together genes will have very few recombination events and be tightly linked, while two genes that are slightly further apart will have more recombination events and be less tightly linked. In the next section, we'll see how to calculate the recombination frequency between two genes, using information from genetic crosses.

Finding recombination frequency

Let's suppose we are interested in seeing whether two genes in the fruit fly (Drosophila) are linked to each other, and if so, how tightly linked they are. In our example, the genes are:

The purple gene, with a dominant pr allele that specifies normal, red eyes and a recessive b allele that specifies purple eyes.

The vestigial gene, with a dominant vg allele that specifies normal, long wings and a recessive vg allele that specifies short, "vestigial" wings.

If we want to measure recombination frequency between these genes, we first need to construct a fly in which we can observe recombination. That is, we need to make a fly that is not just heterozygous for both genes, but where we know exactly which genes are together on the chromosome. To do so, we can start by crossing two homozygous flies as shown below:

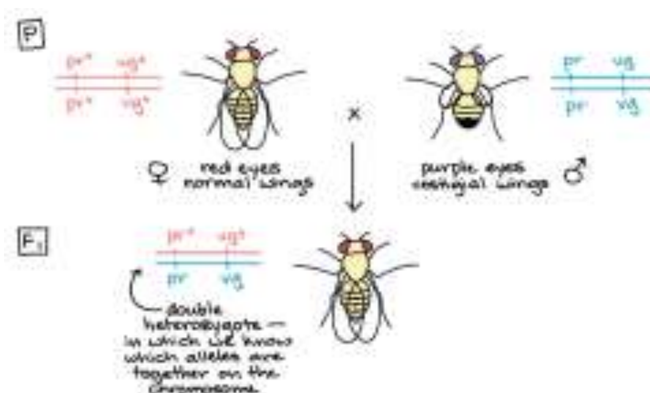


Image modified from "Drosophila melanogaster," by Madboy74 (CC0/public domain).

This cross gives us exactly what we need to observe recombination: a fly that's heterozygous for the purple and vestigial genes, in which we know clearly which alleles are together on a single chromosome.

Now, we need a way to "see" recombination events.

The most direct approach would be to look into the gametes made by the heterozygous fly and see what alleles they had on their chromosomes. Practically, though, it's much simpler to use those gametes in a cross and see what the offspring look like!

To do so, we can cross a double heterozygous fly with a tester, a fly that's homozygous recessive for all the genes of interest (in this case, the pr and vg alleles).

The purpose of using a tester is to ensure that the alleles provided by the non-tester parent fully determine the phenotype, or appearance, of the offspring. When we cross our fly of interest to a tester, we can directly "read" the genotype of each gamete from the physical appearance of the offspring.

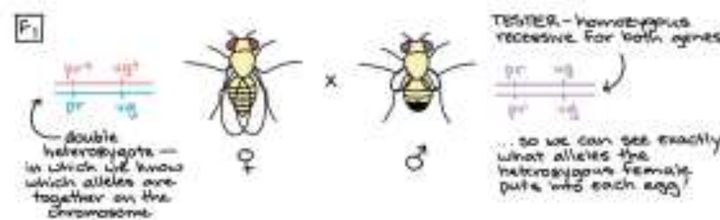


Image modified from "Drosophila melanogaster," by Madboy74 (CC0/public domain).

Below, we can see a modified Punnet square showing the results of the cross between our double Heterozygous fly and the tester fly. Four different types of eggs are produced by a double heterozygous female fly, each of which combines with a sperm from the male tester fly. Four different phenotypic (appearance-based) classes of offspring are produced in this cross, each corresponding to a particular gamete from the female parent:

The four classes of offspring are not produced in equal numbers, which tells us that the purple and

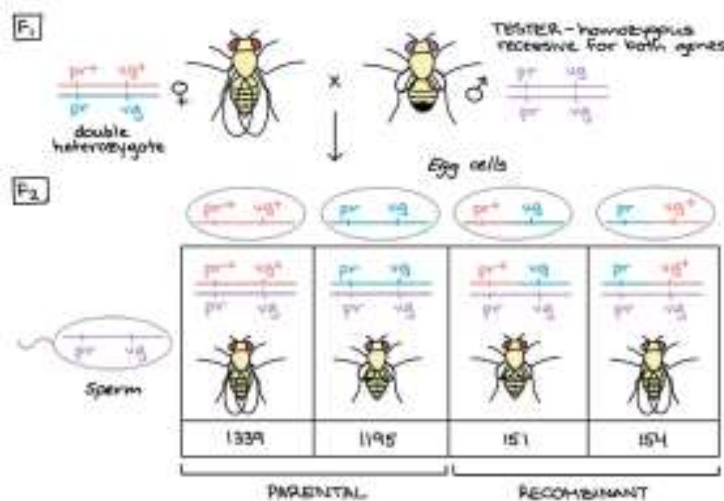


Image modified from "Drosophila melanogaster," by Madboy74 (CC0/public domain).

Vestigial genes are linked. As we expect for linked genes, the parental chromosome configurations are over-represented in the offspring, while the recombinant chromosome configurations are underrepresented. To measure linkage quantitatively, we can calculate the recombination frequency (RF) between the purple and vestigial genes:

$$\text{Recombination frequency (RF)} = \frac{\text{Recombinants}}{\text{Total offspring}}$$

In our case, the recombinant progeny classes are the red-eyed, vestigial-winged flies and the purple-eyed, long-winged flies. We can identify these flies as the recombinant classes for two reasons: one, we know from the series of crosses we performed that they must have inherited a chromosome from their mother that had undergone a recombination event; and two, they are the underrepresented classes (relative to the overrepresented, parental classes).

So, for the cross above, we can write our equation as follows:

$$\text{RF} = \frac{151 + 154}{1339 + 1195 + 151 + 154} \times 100\% = 10.7\%$$

The recombination frequency between the purple and vestigial genes is 10.7%.

Recombination frequency and linkage maps

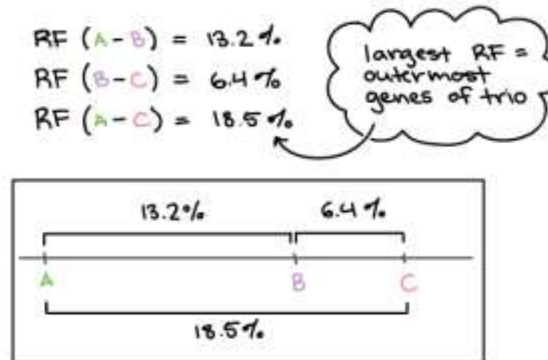
What is the benefit of calculating recombination frequency? One way that recombination frequencies have been used historically is to build linkage maps, chromosomal maps based on recombination frequencies. In fact, studying linkage helped early geneticists establish that chromosomes were in fact linear, and that each gene had its own specific place on a chromosome.

Recombination frequency is not a direct measure of how physically far apart genes are on chromosomes. However, it provides an estimate or approximation of physical distance. So, we can say that a pair of genes with a larger recombination frequency are likely farther apart, while a pair with a smaller recombination frequency are likely closer together.

Importantly, recombination frequency "maxes out" at 50% (which corresponds to genes being unlinked, or assorting independently). That is, 50% is the largest recombination frequency we'll ever directly measure between genes. So, if we want to figure out the map distance between genes further apart than this, we must do so by adding the recombination frequencies of multiple pairs of genes, "building up" a map that extends between the two distant genes.

Comparison of recombination frequencies can also be used to figure out the order of genes on a

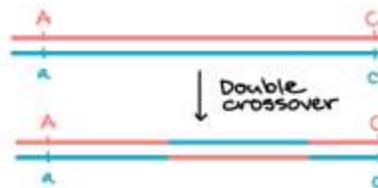
chromosome. For example, let's suppose we have three genes, A, B, and C, and we want to know their order on the chromosome (ABC? ACB? CAB?) If we look at recombination frequencies among all three possible pairs of genes (AC, AB, BC), we can figure out which genes lie furthest apart, and which other gene lies in the middle. Specifically, the pair of genes with the largest recombination frequency must flank the third gene:



Recombination frequencies are based on those for fly genes *v*, *cv*, and *ct*, as given in D. C Bergmann[4].

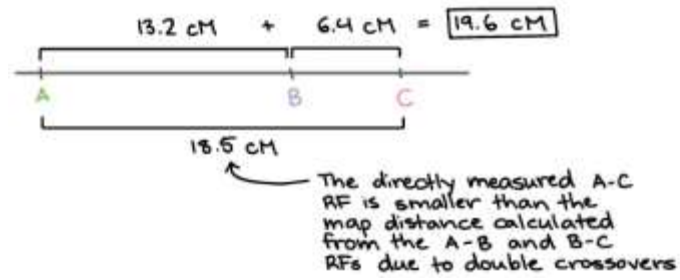
By doing this type of analysis with more and more genes (e.g., adding in genes D, E, and F and figuring out their relationships to A, B, and C) we can build up linkage maps of entire chromosomes. In linkage maps, you may see distances expressed as centimorgans or map units rather than recombination frequencies. Luckily, there's a direct relationship among these values: a 1% recombination frequency is equivalent to 1 centimorgan or 1 map unit.

Is map distance always the same as recombination frequency? Sometimes, the directly measured recombination frequency between two genes is not the most accurate measure of their map distance. That's because, in addition to the single crossovers we've discussed in this article, double crossovers (two separate crossovers between the two genes) can also occur:



Double crossovers are "invisible" if we're only monitoring two genes, in that they put the original two genes back on the same chromosome (but with a swapped-out bit in the middle). For example, the double crossover shown above wouldn't be detectable if we were just looking at genes A and C, since these genes end up back in their original configuration.

Because of this, double crossovers are not counted in the directly measured recombination frequency resulting a slight underestimate of the actual number of recombination events. This is why, in the example below, the recombination frequency directly measured between A and C is a bit smaller than the sum of the recombination frequencies between A-Band B-C. When B is included, double crossovers between A and C can be detected and accounted for.



By measuring recombination frequencies for closer together gene pairs and adding them up, we can minimize "invisible" double crossovers and get more accurate map distances.