4 Gene Linkage and Genetic Mapping

CHAPTER ORGANIZATION

Varieties of maize. [© Katarzyna Citko/ ShutterStock, Inc.]

Genetic mapping means determining the relative positions of genes along a chromosome. It is one of the main experimental tools in genetics. This may seem odd in organisms in which the DNA sequence of the genome has been determined. If every gene in an organism is already sequenced, then what is the point of genetic mapping? The answer is that a gene's sequence does not always reveal its function, nor does a genomic DNA sequence reveal which genes interact in a complex biological process. When a new mutant gene is discovered, the first step in genetic analysis is usually genetic mapping to determine its position in the genome. It is at this point that the genomic sequence, if known, becomes useful, because in some cases the position of the mutant gene coincides with a gene whose sequence suggests a role in the biological process being investigated. For example, in the case of flower color, a new mutation may map to a region containing a gene whose sequence suggests that it encodes an enzyme in anthocyanin synthesis. But the function of a gene is not always revealed by its DNA sequence, and so in some cases, further genetic or molecular analysis is necessary to sort out which one of the genes in a sequenced region corresponds to a mutant gene mapped to that region. In human genetics, genetic mapping is important because it enables genes associated with hereditary diseases, such as those that predispose to breast cancer, to be localized and correlated with the genomic sequence in the region.

4.1 Linked alleles tend to stay together in meiosis.

In meiosis, homologous chromosomes form pairs in prophase I by undergoing synapsis, and the individual members of each pair separate from one another at anaphase I. Genes that are close enough together in the same chromosome might therefore be expected to be transmitted together. Thomas Hunt Morgan examined this issue using two genes present in the X chromosome of *Drosophila*. One was a mutation for white eyes, the other a mutation for miniature wings. Morgan found that the *white* and *miniature* alleles present in each X chromosome of a female do tend to remain together in inheritance, a phenomenon known as **linkage**. Nevertheless, the linkage is incomplete. Some gametes are produced that have different combinations of the *white* and *miniature* alleles than those in the parental chromosomes. The new combinations are produced because homologous chromosomes can exchange segments when they are paired. This process (crossing-over) results in **recombination** of alleles between the homologous chromosomes. The probability of recombination between any two genes serves as a measure of genetic distance between the genes and allows the construction of a **genetic map**, which is a diagram of a chromosome showing the relative positions of the genes. The linear order of genes along a genetic map is consistent with the conclusion that each gene occupies a well-defined position, or **locus**, in the chromosome, with the alleles of a gene in a heterozygote occupying corresponding locations in the pair of homologous chromosomes.

In discussing linked genes, it is necessary to distinguish which alleles are present together in the parental chromosomes. This is done by means of a slash ("/"). The alleles in one chromosome are depicted to the left of the slash, and those in the homologous chromosome are depicted to the right of the slash. For example, in the cross *AA BB* × *aa bb*, the genotype of the doubly heterozygous progeny is denoted *A B*/*a b* because the *A* and *B* alleles were inherited in one parental chromosome and the alleles *a* and *b* were inherited in the other parental chromosome. In this genotype the *A* and *B* alleles are said to be in the **coupling** or ***cis* configuration**; likewise, the *a* and *b* alleles are in coupling. Among the four possible types of gametes, the *A B* and *a b* types are called **parental combinations** because the alleles are in the same configuration as in the parental chromosomes, and the *A b* and *a B* types are called **recombinants** (**FIGURE 4.1**, part A).

Another possible configuration of the *A, a* and *B, b* allele pairs is *A b*/*a B*. In this case the *A* and *B* alleles are said to be in the **repulsion** or ***trans* configuration**. Now the parental and recombinant gametic types are reversed (Figure 4.1, part B). The *A b* and *a B* types are the parental combinations, and the *A B* and *a b* types are the recombinants.

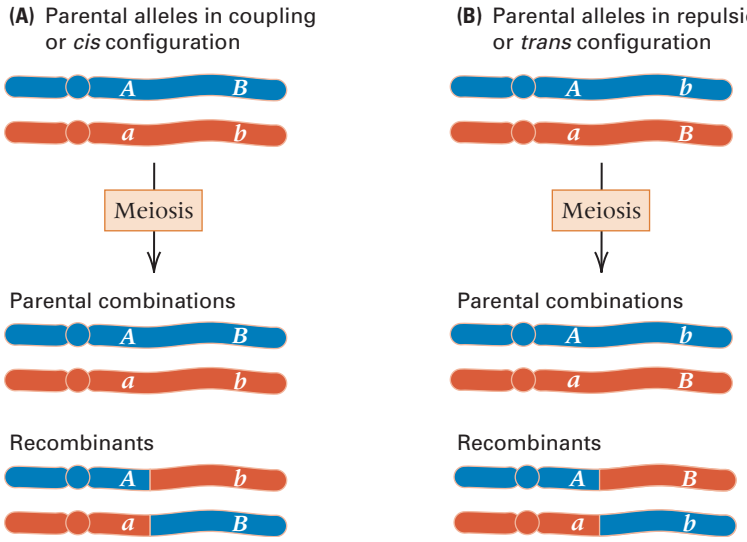


FIGURE 4.1 For any pair of alleles, the gametes produced through meiosis have the alleles either in a parental configuration or in a recombinant configuration. Which types are parental and which recombinant depends on whether the configuration of the alleles in the parent is (A) coupling or (B) repulsion.

The degree of linkage is measured by the frequency of recombination.

In his early experiments with *Drosophila*, Morgan found mutations in each of several X-linked genes that provided ideal materials for studying linkage. One of these genes, with alleles w^+ and w, determines normal red eye color versus white eyes; another such gene, with the alleles m^+ and m, determines whether the size of the wings is normal or miniature. The initial cross is shown as Cross 1 in **FIGURE 4.2**. It was a cross between females with white eyes and normal wings and males with red eyes and miniature wings:

$$\frac{w\ m^+}{w\ m^+}\,♀♀ \times \frac{w^+\ m}{\mathrm{Y}}\,♂♂$$

In this way of writing the genotypes, the horizontal line replaces the slash. Alleles

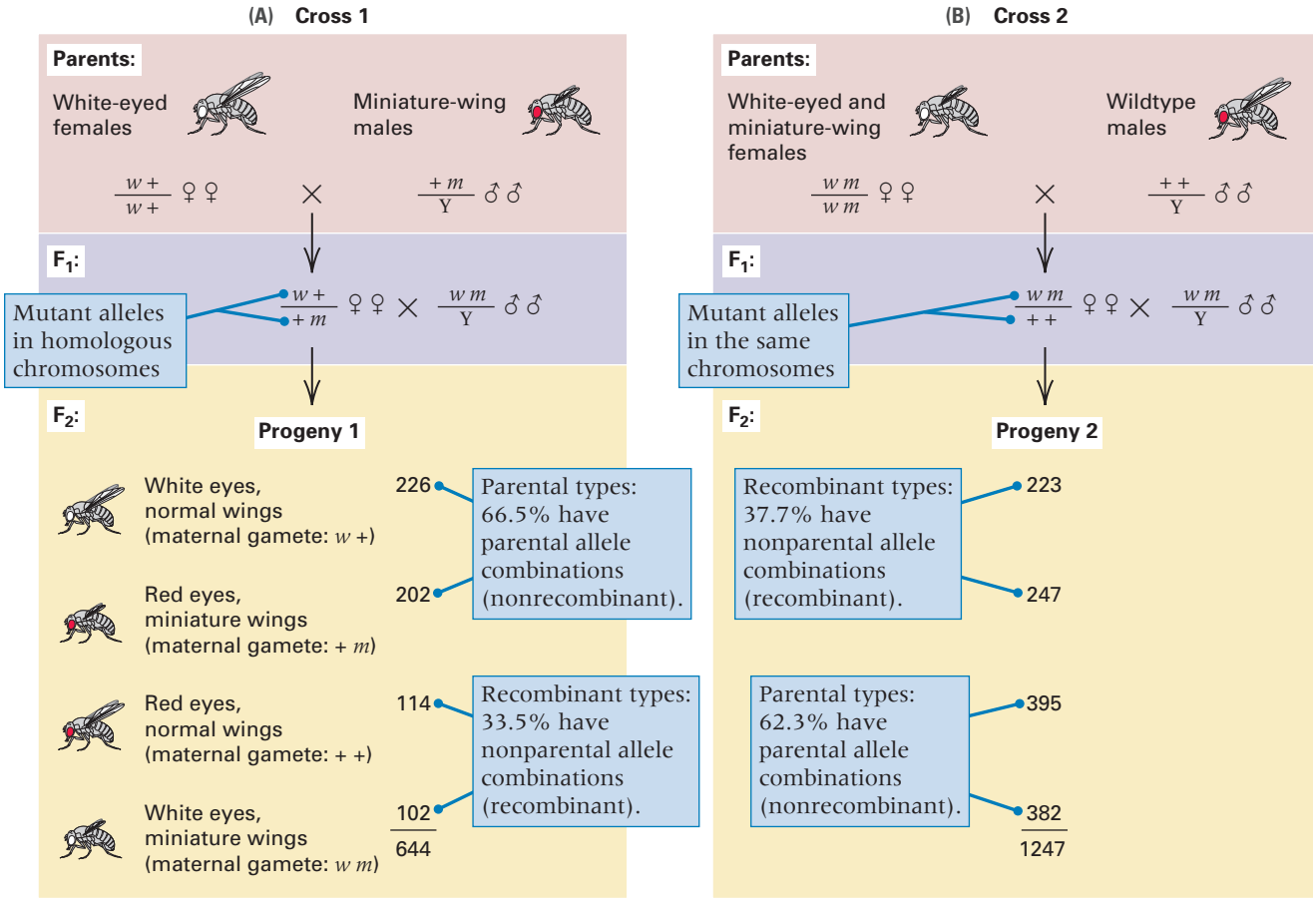


FIGURE 4.2 An experiment demonstrating that the frequency of recombination between two mutant alleles is independent of whether they are present in the same chromosome or in homologous chromosomes. (A) Cross 1 produces F_1 females with the genotype $w\ +/+\ m$, and the $w - m$ recombination frequency is 33.5 percent. (B) Cross 2 produces F_1 females with the genotype $w\ m/+\ +$, and the $w - m$ recombination frequency is 37.7 percent. These values are within the range of variation expected to occur by chance.

written above the line are present in one chromosome, and those written below the line are present in the homologous chromosome. In the females, both X chromosomes carry *w* and m^+. In males, the X chromosome carries the alleles w^+ and *m*. (The Y written below the line denotes the Y chromosome in the male.) Figure 4.2 illustrates a simplified symbolism, commonly used in *Drosophila* genetics, in which a wildtype allele is denoted by a + sign in the appropriate position. The + symbolism is unambiguous because the linked genes in a chromosome are always written in the same order. Using the + notation,

$$\frac{w\ +}{w\ +} \quad \text{means} \quad \frac{w\ m^+}{w\ m^+}$$

and

$$\frac{+\ \ m}{\mathrm{Y}} \quad \text{means} \quad \frac{w^+\ m}{\mathrm{Y}}$$

The resulting F_1 female progeny from Cross 1 have the genotype *w* +/+ *m* (or, equivalently, *w* m^+/w^+ *m*). In this genotype, the w^+ and m^+ alleles are in repulsion. When these females were mated with *w m*/Y males, the offspring denoted as Progeny 1 in Figure 4.2 were obtained. In each class of progeny, the gamete from the female parent is shown in the column at the left, and the gamete from the male parent carries either *w m* or the Y chromosome. The cross is equivalent to a testcross, and so the phenotype of each class of progeny reveals the alleles present in the gamete from the mother.

The results of Cross 1 show a great departure from the 1 : 1 : 1 : 1 ratio of the four male phenotypes that is expected with independent assortment. If genes in the same chromosome tended to remain together in inheritance but were not completely linked, this pattern of deviation might be observed. In this case, the combinations of phenotypic traits in the parents of the original cross (parental phenotypes) were present in 428/644 (66.5 percent) of the F_2 males, and nonparental combinations (recombinant phenotypes) of the traits were present in 216/644 (33.5 percent). The 33.5 percent recombinant X chromosomes is called the **frequency of recombination**, and it should be contrasted with the 50 percent recombination expected with independent assortment.

The recombinant X chromosomes w^+ m^+ and *w m* result from crossing-over in meiosis in F_1 females. In this example, the frequency of recombination between the linked *w* and *m* genes was 33.5 percent. With other pairs of linked genes, the frequency of recombination ranges from near 0 to 50 percent. Even genes in the same chromosome can undergo independent assortment (frequency of recombination equal to 50 percent) if they are sufficiently far apart. This implies the following principle:

KEY CONCEPT

Genes with recombination frequencies smaller than 50 percent are present in the same chromosome (linked). Two genes that undergo independent assortment, indicated by a recombination frequency equal to 50 percent, either are in nonhomologous chromosomes or are located far apart in a single chromosome.

The frequency of recombination is the same for coupling and repulsion heterozygotes.

Morgan also studied progeny from the coupling configuration of the w^+ and m^+ alleles, which results from the mating designated as Cross 2 in Figure 4.2. In this case, the original parents had the genotypes

$$\frac{w\ m}{w\ m}\,♀♀ \times \frac{+\ +}{\mathrm{Y}}\,♂♂$$

The resulting F_1 female progeny from Cross 2 have the genotype *w m*/+ + (equivalently, *w m*/w^+ m^+). In this case the wildtype alleles are in the same chromosome. When these F_1 female progeny were crossed with *w m*/Y males, they yielded the types of progeny tabulated as Progeny 2 in Figure 4.2.

Because the alleles in Cross 2 are in the coupling configuration, the parental-type gametes carry either *w m* or + +, and the recombinant gametes carry either *w* + or + *m*. The types of gametes are the same as those observed in Cross 1, but the parental and recombinant types are opposite. Yet the frequency of recombination is approximately the same: 37.7 percent versus 33.5 percent. The difference is within the range expected to result from random variation from experiment to experiment. The consistent finding of equal recombination frequencies in experiments in which the mutant alleles are in the *trans* or the *cis* configuration leads to the following conclusion:

KEY CONCEPT

Recombination between linked genes takes place with the same frequency whether the alleles of the genes are in the repulsion (*trans*) configuration or in the coupling (*cis*) configuration; it is the same no matter how the alleles are arranged.

The frequency of recombination differs from one gene pair to the next.

The principle that the frequency of recombination depends on the particular pair of genes may be illustrated using the recessive allele *y* of another

X-linked gene in *Drosophila*, which results in yellow body color instead of the usual gray color determined by the y^+ allele. The *yellow body (y)* and *white eye (w)* genes are linked. The frequency of recombination between the genes is as shown in the data in **FIGURE 4.3**. The layout of the crosses is like that in Figure 4.2. In Cross 1, the female has *y* and *w* in the *trans* configuration (+ *w*/*y* +); in Cross 2, the alleles are in the *cis* configuration (*y w*/+ +). The *y* and *w* genes exhibit a much lower frequency of recombination than that observed with *w* and *m* in Figure 4.2. To put it another way, the genes *y* and *w* are more closely linked than are *w* and *m*. In Cross 1, the recombinant progeny are + + and *y w*, and they account for 130/9027 = 1.4 percent of the total. In Cross 2, the recombinant progeny are + *w* and *y* +, and they account for 94/7838 = 1.2 percent of the total. Once again, the parental and recombinant gametes are reversed in Crosses 1 and 2, because the configuration of alleles in the female parent is *trans* in Cross 1 but *cis* in Cross 2, yet the frequency of recombination between the genes is within experimental error.

The results of these and other experiments give support to two general principles of recombination:

- The recombination frequency is a characteristic of a particular pair of genes.
- Recombination frequencies are the same in *cis* (coupling) and *trans* (repulsion) heterozygotes.

Recombination does not occur in *Drosophila* males.

Early experiments in *Drosophila* genetics also indicated that the organism is unusual in that recombination does not take place in males. The absence of recombination in *Drosophila* males means that all alleles located in a particular chromosome show complete linkage in the male. For example, the genes *cn* (cinnabar eyes) and *bw* (brown eyes) are both in chromosome 2, but they are so far apart that in females, they show 50 percent recombination. Because the genes exhibit 50 percent recombination, the cross

$$\frac{cn\ bw}{++}\,♀♀ \times \frac{cn\ bw}{cn\ bw}\,♂♂$$

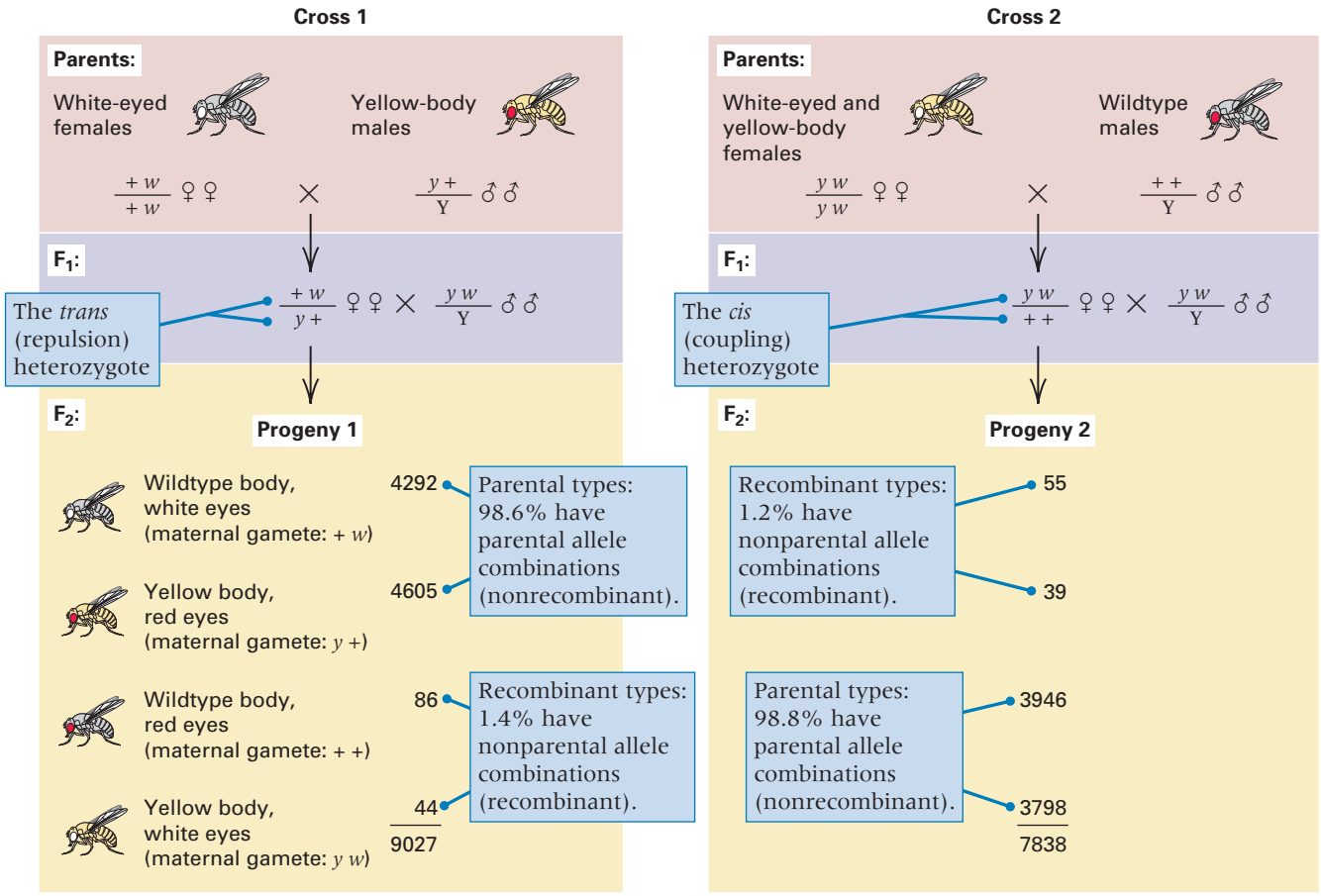


FIGURE 4.3 An experiment demonstrating that the frequency of recombination between two genes depends on the genes. The frequency of recombination between *w* and *y* is much less than that between *w* and *m* in Figure 4.2. The *y*–*w* experiment also confirms the equal frequency of recombination in *trans* and *cis* heterozygous genotypes. (A) The *trans* heterozygous females, + *w*/*y* +, yield 1.4 percent recombination. (B) The *cis* heterozygous females, *y w*/++, yield 1.2 percent recombination.

yields progeny of genotype *cn bw*/*cn bw* and + +/*cn bw* (the nonrecombinant types) as well as *cn* +/*cn bw* and + *bw*/*cn bw* (the recombinant types) in the proportions 1 : 1 : 1 : 1. The outcome of the reciprocal cross is different. Because no crossing-over occurs in males, the reciprocal cross

$$\frac{cn\ \ bw}{cn\ \ bw}\,♀♀ \times \frac{cn\ \ bw}{+\ +}\,♂♂$$

yields progeny only of the nonrecombinant genotypes *cn bw*/*cn bw* and + +/*cn bw* in equal proportions. The absence of recombination in *Drosophila* males is a convenience often exploited in experimental design; as shown in the case of *cn* and *bw*, all the alleles present in any chromosome in a male must be transmitted as a group, without being recombined with alleles present in the homologous chromosome. The absence of crossing-over in *Drosophila* males is atypical; in most other animals and plants, recombination takes place in both sexes, though not necessarily with the same frequency.

4.2 Recombination results from crossing-over between linked alleles.

The linkage of the genes in a chromosome can be represented in the form of a *genetic map*, which shows the linear order of the genes along the chromosome spaced so that the distances between adjacent genes is proportional to the frequency of recombination between them. A genetic map is also called a **linkage map** or a **chromosome map**. The concept of genetic mapping was first developed by Morgan's student Alfred H. Sturtevant in 1913. The early geneticists understood that recombination between genes takes place by an exchange of segments between homologous chromosomes in the process now called crossing-over. Each crossover is manifested physically as a chiasma, or cross-shaped configuration, between homologous chromosomes; chiasmata are observed in prophase I of meiosis. Each chiasma results from the breaking and rejoining of chromatids during meiosis, with the result that there is an exchange of corresponding segments between them. The theory of crossing-over is that each chiasma results in a new association of genetic markers. This process is illustrated in **FIGURE 4.4**. When there is no crossing-over (part A), the alleles present in each homologous chromosome remain in the same combination. When a crossover does take place (part B), the outermost alleles in two of the chromatids are interchanged (recombined).

The unit of distance in a genetic map is called a **map unit**; one map unit is equal to 1 percent recombination. For example, two genes that recombine with a frequency of 3.1 percent are said to be located 3.1 map units apart. One map unit is also called a **centimorgan**, abbreviated cM, in honor of T. H. Morgan. A distance of 3.1 map units therefore equals 3.1 centimorgans and indicates 3.1 percent recombination between the genes. An example is shown in part A of **FIGURE 4.5**, which deals with the *Drosophila* mutants *w* for white eyes and *dm (diminutive)* for small body. The female parent in the testcross is the *trans* heterozygote, but as we have seen, this configuration is equivalent in frequency of recombination to the *cis* heterozygote. Among 1000 progeny there are 31 recombinants. Using this estimate, we can express the genetic distance between *w* and *dm* in four completely equivalent ways:

- As the *frequency of recombination*—in this case 0.031
- As the *percent recombination*, or 3.1 percent
- As the distance in *map units*—in this example, 3.1 map units
- As the distance in *centimorgans*, or 3.1 centimorgans (3.1 cM)

A genetic map based on these data is shown in Figure 4.5, part B. The chromosome is represented as a horizontal line, and each gene is assigned a position on the line according to its genetic distance from other genes. In this example, there are only two genes, *w* and *dm*, and they are separated by a distance of 3.1 centimorgans (3.1 cM), or 3.1 map units. Genetic maps are usually truncated to show only the genes of interest. The full genetic map of the *Drosophila* X chromosome extends considerably farther in both directions than indicated in this figure.

Physically, one map unit corresponds to a length of the chromosome in which, on the average, one crossover is formed in every 50 cells undergoing meiosis. This principle is illustrated in **FIGURE 4.6**. If one meiotic cell in 50 has a crossover, the frequency of *crossing-over* equals 1/50, or 2 percent. Yet the frequency of *recombination* between the genes is 1 percent. The correspondence of 1 percent recombination with 2 percent crossing-over is a little confusing until you consider that a crossover results in two recombinant chromatids and two nonrecombinant chromatids (Figure 4.6). A crossover frequency of 2 percent means that of the 200 chromosomes that result from meiosis in 50 cells, exactly 2 chromosomes (those involved in the crossover) are recombinant for genetic markers spanning the particular chromosome segment. To put the matter in another way, 2 percent crossing-over corresponds to 1 percent recombination because only half of the chromatids in each cell with a crossover are actually recombinant.

In situations in which there are **genetic markers** along the chromosome, such as the *A, a* and *B, b* pairs of alleles in Figure 4.6, recombination between the marker genes takes place only when a crossover occurs *between* the genes. **FIGURE 4.7** illustrates a case in which a crossover takes place between the gene *A* and the centromere, rather than between the genes *A* and *B*. The crossover does result in the physical exchange

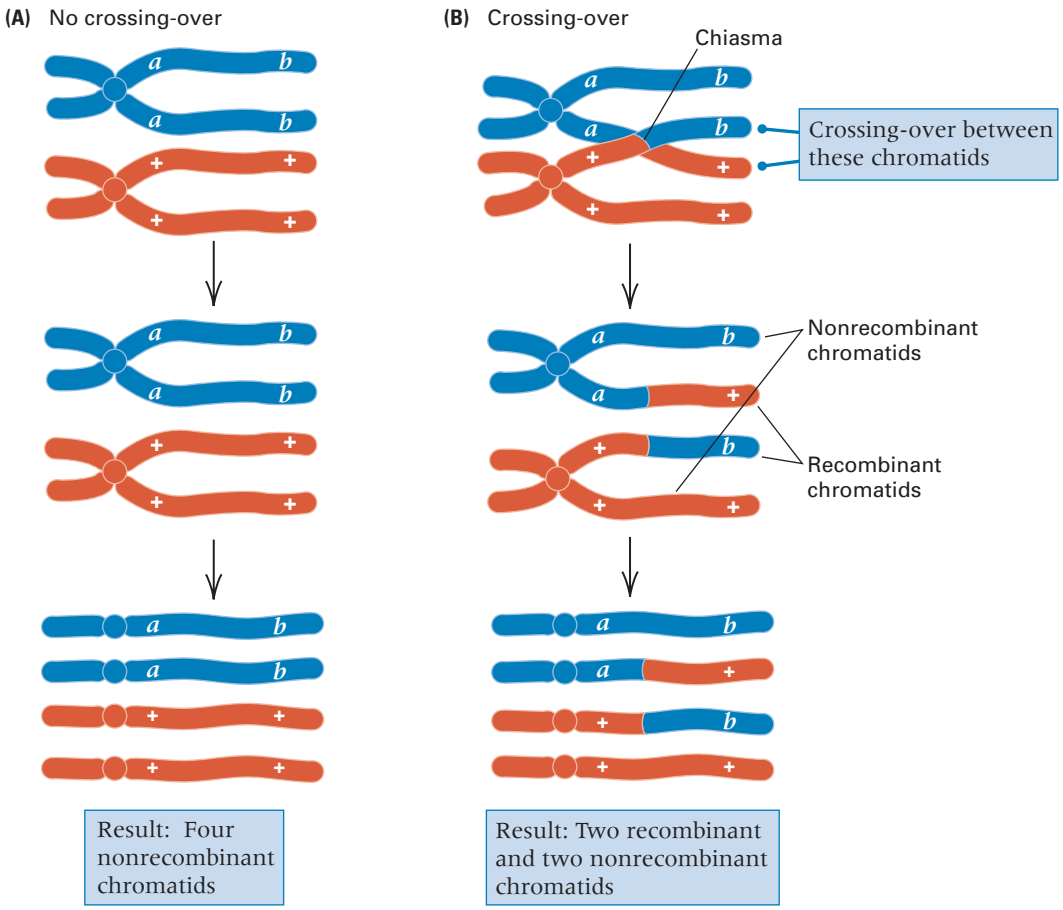


FIGURE 4.4 Diagram illustrating crossing-over between two genes. (A) When there is no crossover between two genes, the alleles are not recombined. (B) When there is a crossover between them, the result is two recombinant and two nonrecombinant products, because the exchange takes place between only two of the four chromatids.

of segments between the innermost chromatids. However, because it is located outside the region between *A* and *B*, all of the resulting gametes must carry either the *A B* or the *a b* allele combination. These are nonrecombinant chromosomes. The presence of the crossover is undetected because it is not in the region between the genetic markers.

In some cases, the region between genetic markers is large enough that two (or even more) crossovers can be formed in a single meiotic cell. One possible configuration for two crossovers is shown in **FIGURE 4.8**. In this example, both crossovers are between the same pair of chromatids. The result is that there is a physical exchange of a segment of chromosome between the marker genes, but the double crossover remains undetected because the markers themselves are not recombined. The absence of recombination results from the fact that the second crossover reverses the effect of the first, insofar as recombination between *A* and *B* is concerned. The resulting chromosomes are either *A B* or *a b*, both of which are nonrecombinant.

Because double crossovers in a region between two genes can remain undetected (this happens when they do not result in recombinant chromosomes), there is an important distinction between the distance between two genes as measured by the recombination frequency and as measured in map units:

- The *map distance* between two genes equals one-half of the average number of crossovers that take place in the region per meiotic cell; it is a measure of crossing-over.
- The *recombination frequency* between two genes indicates how much recombination is actually observed in a particular experiment; it is a measure of recombination.

The difference between map distance and recombination frequency arises because double crossovers that do not yield recombinant gametes, like the one depicted in Figure 4.8, *do* contribute to the map distance but *do not* contribute to the recombination frequency. The distinction is important only when the region in question is large enough for double crossing-over to occur. If the region between the genes is short enough that no more than one crossover can occur in the region in any one meiosis, then map units and recombination frequencies

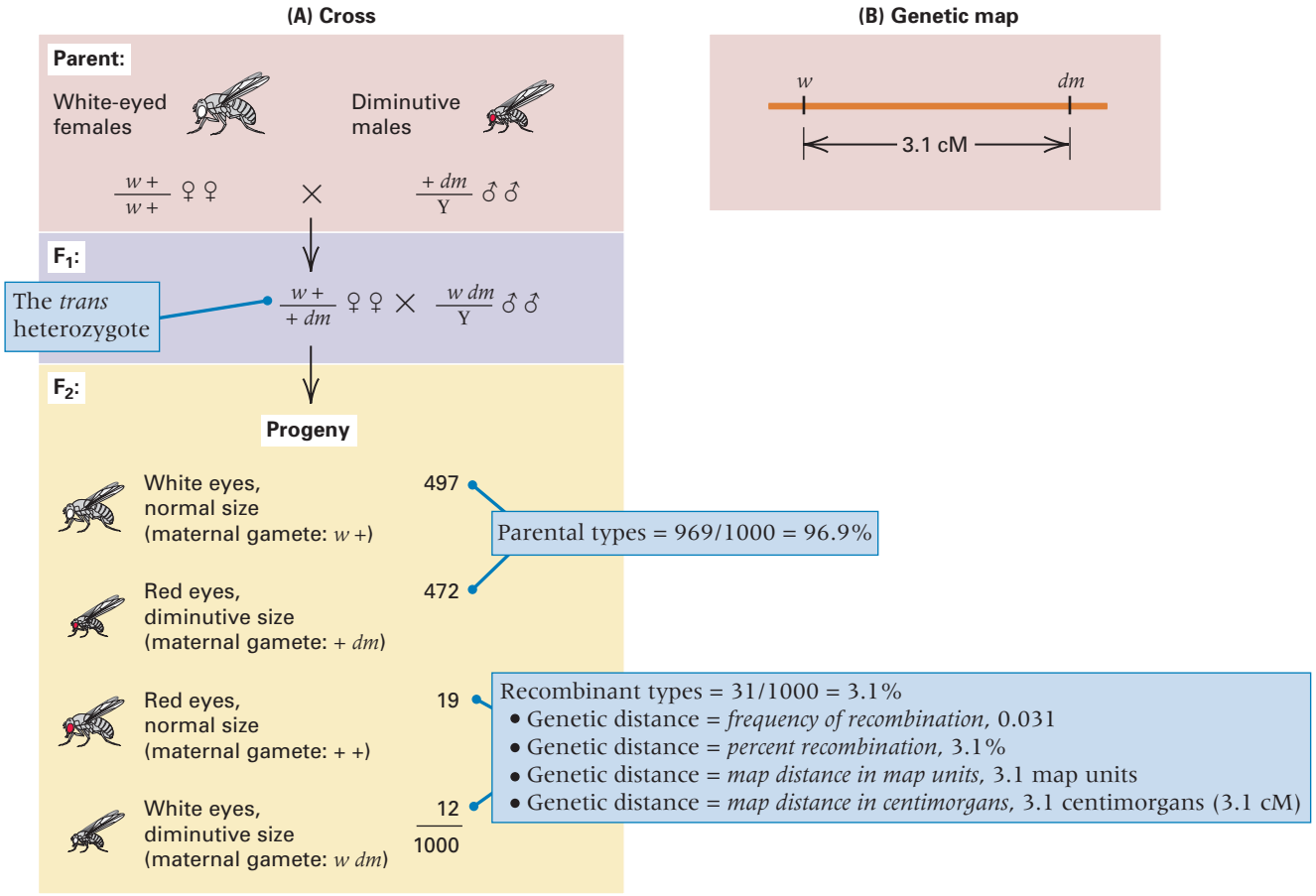


FIGURE 4.5 An experiment illustrating how the frequency of recombination is used to construct a genetic map. (A) There is 3.1 percent recombination between the genes *w* and *dm*. (B) A genetic map with *w* and *dm* positioned 3.1 map units (3.1 centimorgans, cM) apart, corresponding to 3.1 percent recombination. The map distance equals frequency of recombination only when the frequency of recombination is sufficiently small.

are the same (because there are no multiple crossovers that can undo each other). This is the basis for defining a map unit as being equal to 1 percent recombination:

KEY CONCEPT

Over an interval so short that multiple crossovers are precluded (typically yielding 10 percent recombination or less), the map distance equals the recombination frequency because all crossovers result in recombinant gametes.

Furthermore, when adjacent chromosome regions separating linked genes are so short that multiple crossovers are not formed, the recombination frequencies (and hence the map distances) between the genes are additive. This important feature of recombination, as well as the logic used in genetic mapping, is illustrated by the example in **FIGURE 4.9**. The genes are located in the X chromosome of *Drosophila*—*y* for yellow body, *rb* for ruby eye color, and *cv* for shortened wing crossvein. The experimentally measured recombination frequency between genes *y* and *rb* is 7.5 percent, and that between *rb* and *cv* is 6.2 percent. The genetic map might be any one of three possibilities, depending on which gene is in the middle (*y*, *cv*, or *rb*). Map C, which has *y* in the middle, can be excluded because it implies that the recombination frequency between *rb* and *cv* should be greater than that between *rb* and *y*, and this contradicts the observed data. In other words, map C can be excluded because it implies that the frequency of recombination between *y* and *cv* must be negative.

Maps A and B are both consistent with the observed recombination frequencies. They differ in their predictions regarding the recombination frequency between *y* and *cv*. Using the principle of additivity of map distances, the predicted *y*−*cv* map distance in A is 13.7 map units, whereas the predicted *y*−*cv* map distance in B is 1.3 map units. In fact, the observed recombination frequency between *y* and *cv* is 13.3 percent. Map A is therefore correct. However, there are actually two genetic maps corresponding to map A. They differ only in whether *y* is placed at the left or at the right. One map is *y*−*rb*−*cv*, which is the one shown in Figure 4.9;

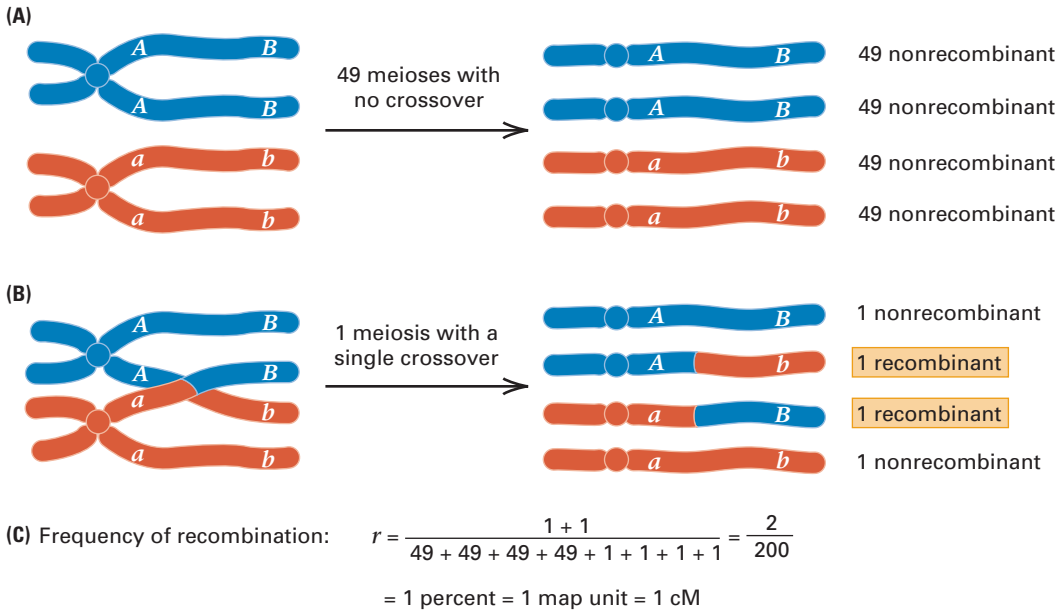


FIGURE 4.6 Diagram of chromosomal configurations in 50 meiotic cells, in which 1 has a crossover between two genes. (A) The 49 cells without a crossover result in 98 *A B* and 98 *a b* chromosomes; these are all nonrecombinant. (B) The cell with a crossover yields chromosomes that are *A B*, *A b*, *a B*, and *a b*, of which the middle two types are recombinant chromosomes. (C) The recombination frequency equals 2/200, or 1 percent, also called 1 map unit or 1 cM. Hence, 1 percent recombination means that 1 meiotic cell in 50 has a crossover in the region between the genes.

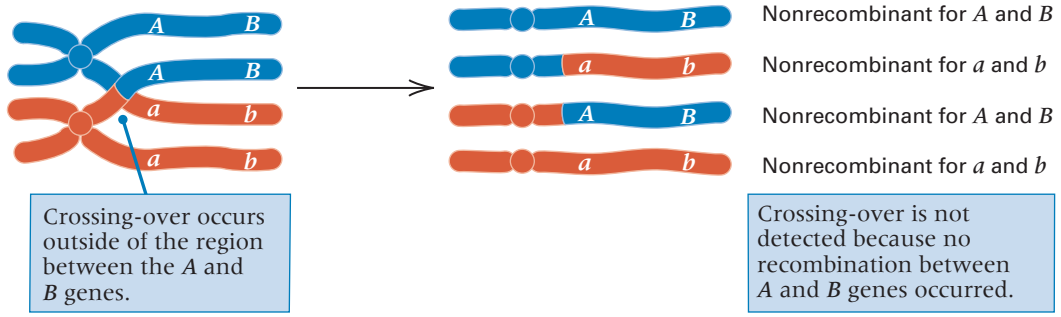


FIGURE 4.7 Crossing-over outside the region between two genes is not detectable through recombination. Although a segment of chromosome is exchanged, the genetic markers outside the region of the crossovers stay in the nonrecombinant configurations, in this case *A B* and *a b*.

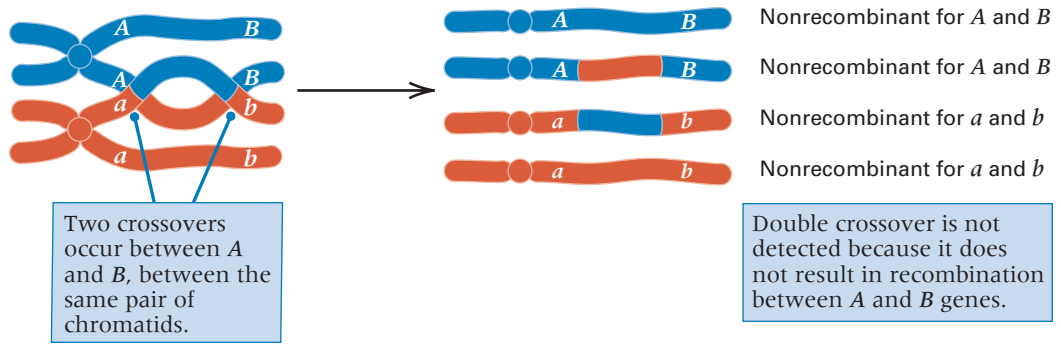


FIGURE 4.8 If two crossovers take place between marker genes *A* and *B*, and both involve the same pair of chromatids, then neither crossover is detected because all of the resulting chromosomes are nonrecombinant *A B* or *a b*.

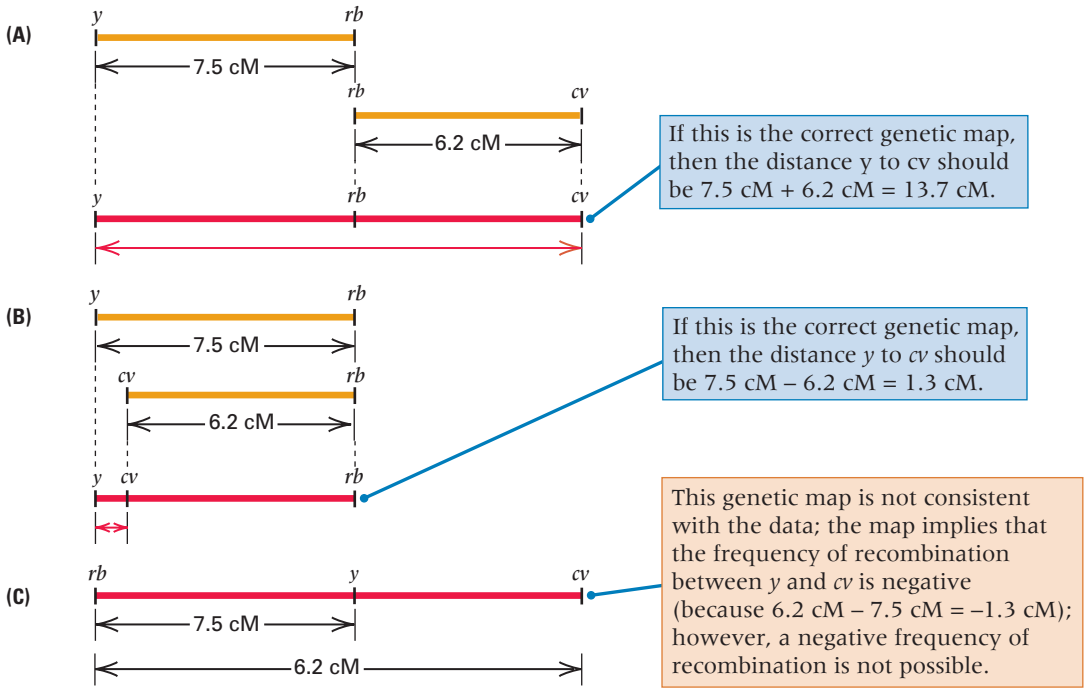
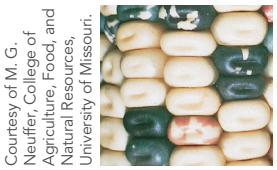


FIGURE 4.9 In *Drosophila*, the genes *y* (yellow body) and *rb* (ruby eyes) have a recombination frequency of 7.5 percent, and *rb* and *cv* (shortened wing crossvein) have a recombination frequency of 6.2 percent. There are three possible genetic maps, depending on whether *rb* is in the middle (part A), *cv* is in the middle (part B), or *y* is in the middle (part C). Map (C) can be excluded because it implies that *rb* and *y* should be closer than *rb* and *cv*, whereas the observed recombination frequency between *rb* and *y* is actually greater than that between *rb* and *cv*. Maps (A) and (B) are compatible with the data given.

the other is *cv*–*rb*–*y*. The two ways of depicting the genetic map are completely equivalent because there is no way of knowing from the recombination data whether *y* or *cv* is closer to the telomere. (Other data indicate that *y* is, in fact, near the telomere.)

A genetic map can be expanded by this type of reasoning to include all of the known genes in a chromosome; these genes constitute a **linkage group**. The number of linkage groups is the same as the haploid number of chromosomes of the species. For example, cultivated corn (*Zea mays*) has ten pairs of chromosomes and ten linkage groups. A partial genetic map of chromosome 10 is shown in **FIGURE 4.10**, along with the dramatic phenotypes caused by some of the mutations. The ears of corn shown in parts C and F demonstrate the result of Mendelian segregation. The ear in part C shows a 3 : 1 segregation of yellow : orange kernels produced by the recessive *orange pericarp-2* (*orp-2*) allele in a cross between two heterozygous genotypes.

The ear in part F shows a 1 : 1 segregation of marbled:white kernels produced by the dominant allele *R1-mb* in a cross between a heterozygous genotype and a homozygous wildtype.



Physical distance is often—but not always—correlated with map distance.

Generally speaking, the greater the physical separation between genes along a chromosome, the greater the map distance between them. Physical distance and genetic map distance are usually correlated, because a greater distance between genetic markers affords a greater chance for a crossover to take place; crossing-over is a physical exchange between the chromatids of paired homologous chromosomes.

On the other hand, the general correlation between physical distance and genetic map distance is by no means absolute. We have already noted that the frequency of recombination between genes may differ in males and females. An unequal frequency of recombination means that the sexes can have different map distances in their genetic maps, although the physical chromosomes of the two sexes are the same and the

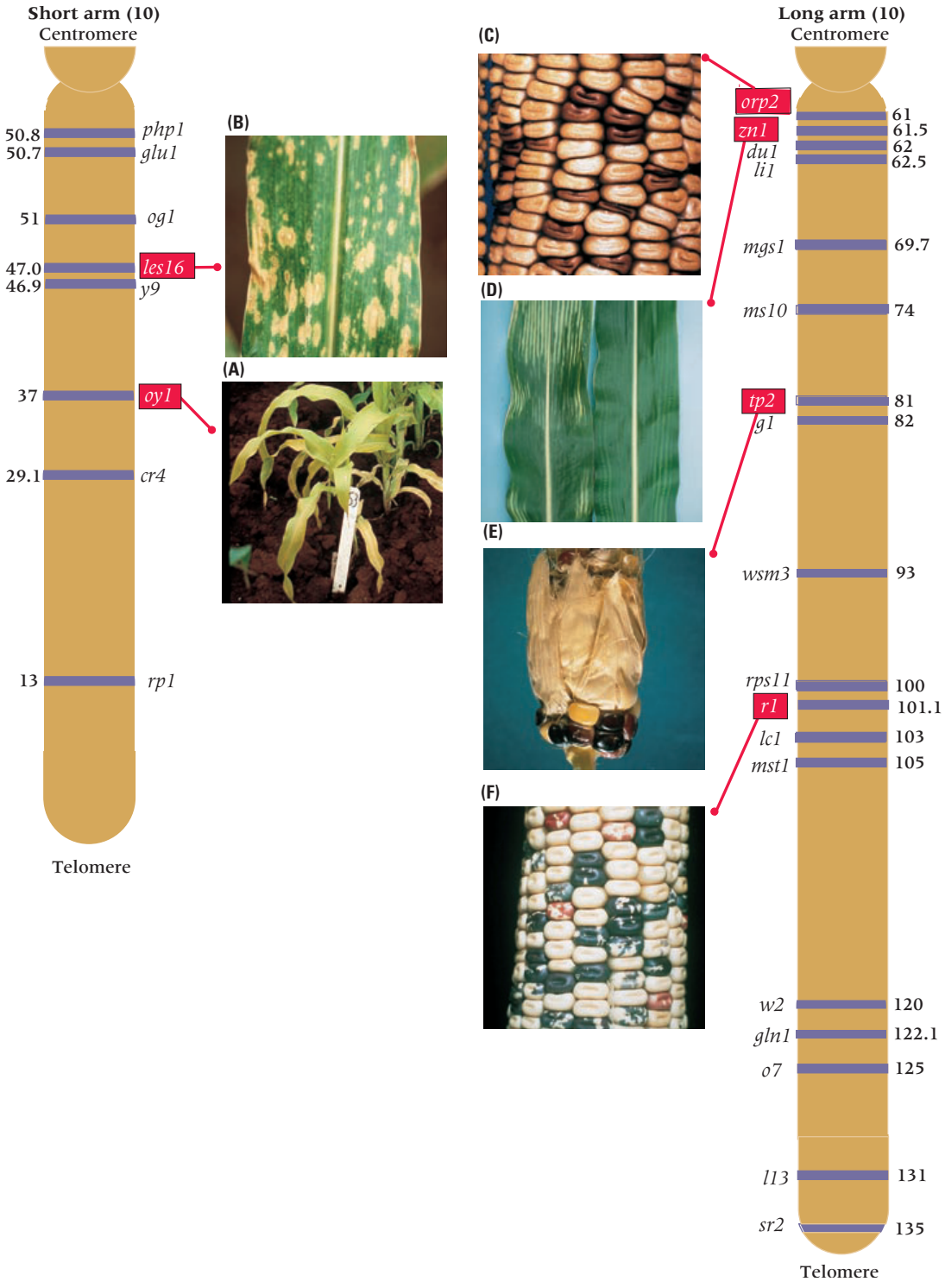


FIGURE 4.10 Genetic map of chromosome 10 of corn, *Zea mays*. The map distance to each gene is given in standard map units (centimorgans) relative to a position 0 for the telomere of the short arm (lower left). (A) Mutations in the gene *oil yellow-1 (oy1)* result in a yellow-green plant. The plant in the foreground is heterozygous for the dominant allele *Oy1*; behind is a normal plant. (B) Mutations in the gene *lesion-16 (les16)* result in many small to medium-sized, irregularly spaced, discolored spots on the leaf blade and sheath. The photograph shows the phenotype of a heterozygote for *Les16*, a dominant allele. (C) The *orp2* allele is a recessive expressed as orange pericarp, a maternal tissue that surrounds the kernels. The ear shows the segregation of *orp2* in a cross between two heterozygous genotypes, yielding a 3 : 1 ratio of yellow : orange seeds. (D) The gene *zn1* is *zebra necrotic-1*, in which dying tissue appears in longitudinal leaf bands. The leaf on the left is homozygous *zn1*, that on the right is wildtype. (E) Mutations in the gene *teopod-2 (tp2)* result in many small, partially podded ears and a simple tassel. An ear from a plant heterozygous for the dominant allele *Tp2* is shown. (F) The mutation *R1-mb* is an allele of the *r1* gene, resulting in red or purple color in the aleurone layer of the seed. Note the marbled color in kernels of an ear segregating for *R1-mb*. [Adapted from an illustration by E. H. Coe. Photos courtesy of M. G. Neuffer, College of Agriculture, Food, and Natural Resources, University of Missouri.]

genes must have the same linear order. For example, because there is no recombination in male *Drosophila*, the map distance between any pair of genes located in the same chromosome, when measured in the male, is 0. (On the other hand, genes on different chromosomes do undergo independent assortment in males.)

The general correlation between physical distance and genetic map distance can even break down in a single chromosome. For example, crossing-over is much less frequent in heterochromatin, which consists primarily of gene-poor regions near the centromeres, than in euchromatin. Consequently, a given length of heterochromatin will appear much shorter in the genetic map than an equal length of euchromatin. In heterochromatic regions, therefore, the genetic map gives a distorted picture of the physical map. An example of such distortion is illustrated in **FIGURE 4.11**, which compares the physical map and the genetic map of chromosome 2 in *Drosophila*. The physical map depicts the appearance of the chromosome in metaphase of mitosis. Two genes near the tips and two near the euchromatin–heterochromatin junction are indicated in the genetic map. The map distances across the euchromatic arms are 54.5 and 49.5 map units, respectively, for a total euchromatic map distance of 104.0 map units. However, the heterochromatin, which constitutes approximately 25 percent of the entire chromosome, has a genetic length in map units of only 3.0 percent. The distorted length of the heterochromatin in the genetic map results from the reduced frequency of crossing-over in the heterochromatin. In spite of the distortion of the genetic map across the heterochromatin, in the regions of euchromatin there is a good correlation between the physical distance between genes and their distance, in map units, in the genetic map.

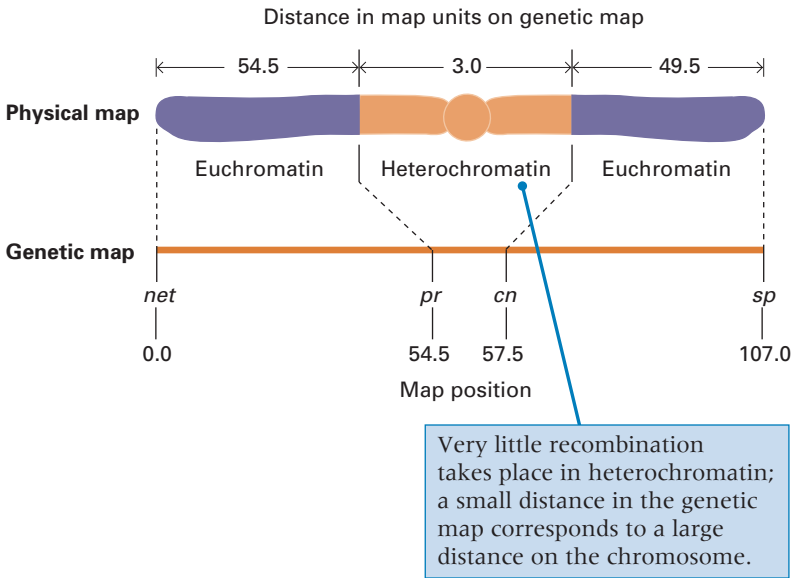


FIGURE 4.11 Chromosome 2 in *Drosophila* as it appears in metaphase of mitosis (physical map, top) and in the genetic map (bottom). The genes *pr* and *cn* are actually in euchromatin but are located near the junction with heterochromatin. The total map length is 54.5 + 49.5 + 3.0 = 107.0 map units. The heterochromatin accounts for 3.0/107.0 = 2.8 percent of the total map length but constitutes approximately 25 percent of the physical length of the metaphase chromosome.

One crossover can undo the effects of another.

When two genes are located far apart along a chromosome, more than one crossover can be formed between them in a single meiosis, and this complicates the interpretation of recombination data. The probability of multiple crossovers increases with the distance between the genes. Multiple crossing-over complicates genetic mapping because map distance is based on the number of physical exchanges that are formed, and some of the multiple exchanges between two genes do not result in recombination of the genes and hence are not detected. As we saw in Figure 4.8, the effect of one crossover can be canceled by another crossover further along the way. If two exchanges between the same two chromatids take place between the genes *A* and *B*, then their net effect will be that all chromosomes are nonrecombinant, either *A B* or *a b*. Two of the products of this meiosis have an interchange of their middle segments, but the chromosomes are not recombinant for the genetic markers and so are genetically indistinguishable from noncrossover chromosomes. The possibility of such canceling events means that the observed recombination value is an *underestimate* of the true exchange frequency and the map distance between the genes. In higher organisms, double crossing-over is effectively precluded in chromosome segments that are sufficiently short, usually about 10 map units or less. Therefore, multiple crossovers that cancel each other's effects can be avoided by using recombination data for closely linked genes to build up genetic linkage maps.

The minimum recombination frequency between two genes is 0. The recombination frequency also has a maximum:

KEY CONCEPT

No matter how far apart two genes may be, the maximum frequency of recombination between any two genes is 50 percent.

Fifty percent recombination is the same value that would be observed if the genes were on nonhomologous chromosomes and assorted independently. The

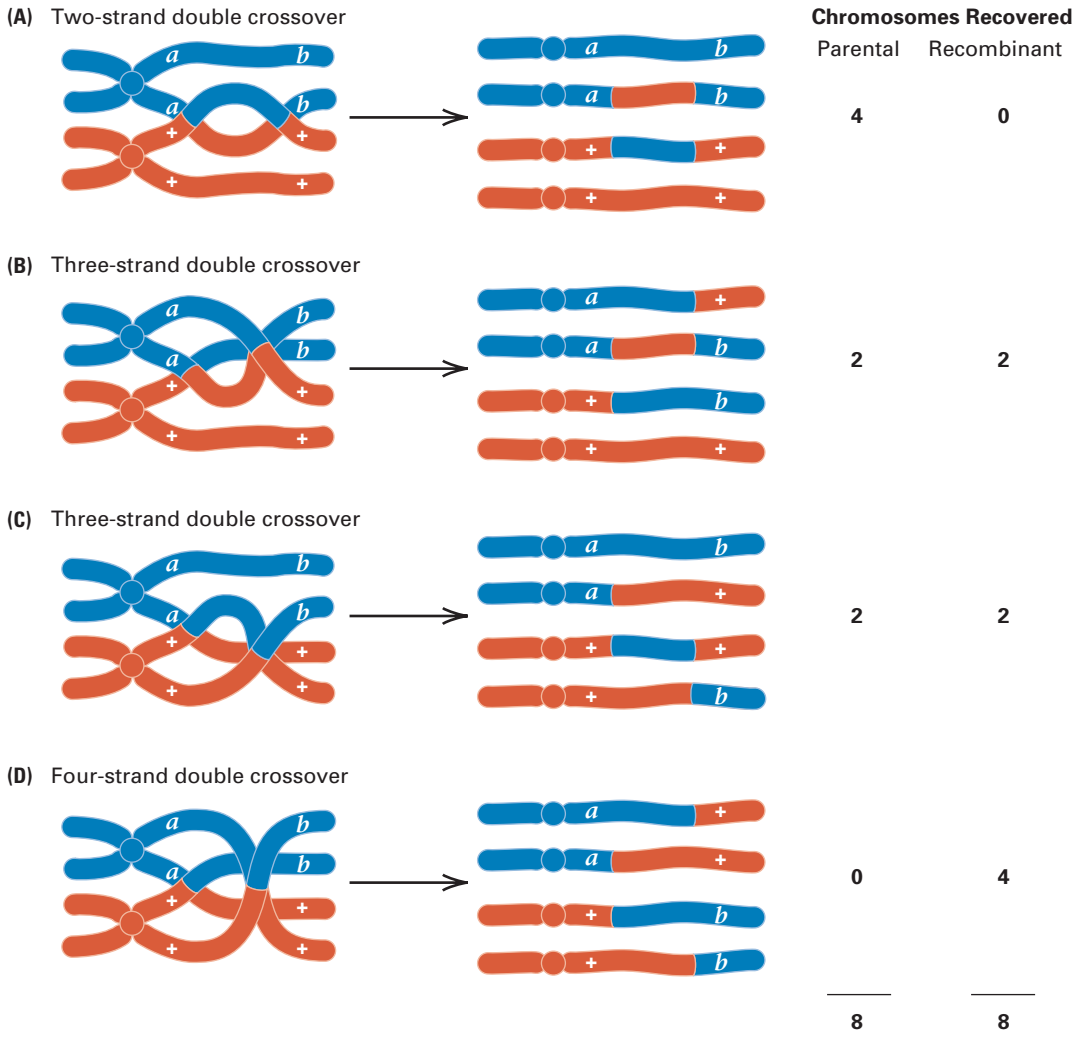


FIGURE 4.12 Diagram showing that the result of two crossovers in the interval between two genes is indistinguishable from independent assortment of the genes, provided that the chromatids participate at random in the crossovers. (A) A two-strand double crossover. (B) and (C) The two types of three-strand double crossovers. (D) A four-strand double crossover.

maximum frequency of recombination is observed when the genes are so far apart in the chromosome that at least one crossover is almost always formed between them. In part B of Figure 4.6, it can be seen that a single exchange in every meiosis would result in half of the products having parental combinations and the other half having recombinant combinations of the genes. The occurrence of two exchanges between two genes has the same effect, as shown in **FIGURE 4.12**. Part A shows a two-strand double crossover, in which the same chromatids participate in both exchanges; no recombination of the marker genes is detectable. When the two exchanges have one chromatid in common (three-strand double crossover, parts B and C), the result is indistinguishable from that of a single exchange; two products with parental combinations and two with recombinant combinations are produced. Note that there are two types of three-strand doubles, depending on which three chromatids participate. The final possibility is that the second exchange connects the chromatids that did not participate in the first exchange (four-strand double crossover, part D), in which case all four products are recombinant.

In most organisms, when double crossovers are formed, the chromatids that take part in the two exchange events are selected at random. In this case, the expected proportions of the three types of double exchanges are 1/4 four-strand doubles, 1/2 three-strand doubles, and 1/4 two-strand doubles. This means that on the average, $(1/4)(0) + (1/2)(2) + (1/4)(4) = 2$ recombinant chromatids will be found among the 4 chromatids produced from meioses with two exchanges between a pair of genes. This is the same proportion obtained with a single exchange between the genes. Moreover, a maximum of 50 percent recombination is obtained for any number of exchanges.

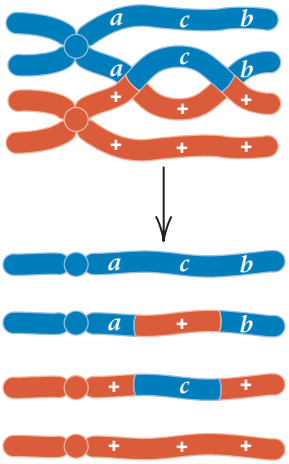


FIGURE 4.13 Diagram showing that two crossovers that occur between the same chromatids and span the middle pair of alleles in a triple heterozygote will result in a reciprocal exchange of the middle pair of alleles between the two chromatids.

Double crossing-over is detectable in recombination experiments that employ **three-point crosses**, which include three pairs of alleles. If a third pair of alleles, c^+ and c, is located between the outermost genetic markers (**FIGURE 4.13**), double exchanges in the region can be detected when the crossovers flank the *c* gene. The two crossovers, which in this example take place between the same pair of chromatids, would result in a reciprocal exchange of the c^+ and c alleles between the chromatids. A three-point cross is an efficient way to obtain recombination data; it is also a simple method for determining the order of the three genes, as we will see in the next section.

4.3 Double crossovers are revealed in three-point crosses.

The data in **TABLE 4.1** result from a testcross in corn with three genes in a single chromosome. The analysis illustrates the approach to interpreting a three-point cross. The recessive alleles of the genes in this cross are *lz* (for lazy or prostrate growth habit), *gl* (for glossy leaf), and *su* (for sugary endosperm), and the multiply heterozygous parent in the cross had the genotype

$$\frac{Lz\ \ Gl\ \ Su}{lz\ \ gl\ \ su}$$

where each symbol with an initial capital letter represents the dominant allele. (The use of this type of symbolism is customary in corn genetics.) The two classes of progeny that inherit noncrossover (parental-type) gametes are therefore the wildtype plants and those with the lazy-glossy-sugary phenotype. The number of progeny in these classes is far larger than the number in any of the crossover classes. Because the frequency of recombination is never greater than 50 percent, the very fact that these progeny are the most numerous indicates that the gametes that gave rise to them have the parental allele configurations, in this case *Lz Gl Su* and *lz gl su*. Using this principle, we could have inferred the genotype of the heterozygous parent even if the

TABLE 4.1

Interpreting a Three-Point Cross

Phenotype of testcross progeny	Genotype of gamete from hybrid parent	Number of progeny
Wildtype	*Lz Gl Su*	286
Lazy	*lz Gl Su*	33
Glossy	*Lz gl Su*	59
Sugary	*Lz Gl su*	4
Lazy, glossy	*lz gl Su*	2
Lazy, sugary	*lz Gl su*	44
Glossy, sugary	*Lz gl su*	40
Lazy, glossy, sugary	*lz gl su*	272
		740

The two most frequent classes identify the nonrecombinant gametes.

The two rarest classes identify the double-recombinant gametes.

These reciprocal classes result from single recombination between one pair of adjacent genes.

These reciprocal classes result from single recombination between the other pair of adjacent genes.

genotype had not been stated. This is a point important enough to state more generally:

KEY CONCEPT

In any genetic cross, no matter how complex, the two most frequent types of gametes with respect to any pair of genes are *nonrecombinant*; these provide the linkage phase (*cis* versus *trans*) of the alleles of the genes in the multiply heterozygous parent.

In mapping experiments, the gene sequence is usually not known. In this example, the order in which the three genes are shown is entirely arbitrary. However, there is an easy way to determine the correct order from three-point data. The gene order can be deduced by identifying the genotypes of the double-crossover gametes produced by the heterozygous parent and comparing these with the nonrecombinant gametes. Because the probability of two simultaneous exchanges is considerably smaller than that of either single exchange, the double-crossover gametes will be the least frequent types. It is clear in Table 4.1 that the classes composed of four plants with the sugary phenotype and two plants with the lazy-glossy phenotype (products of the *Lz Gl su* and *lz gl Su* gametes, respectively) are the least frequent and therefore constitute the double-crossover progeny. Now we apply another principle:

KEY CONCEPT

The effect of double crossing-over is to interchange the members of the *middle* pair of alleles between the chromosomes.

This principle is illustrated in **FIGURE 4.14**. With three genes there are three possible orders, depending on which gene is in the middle. If *gl* were in the middle (part A), the double-recombinant gametes would be *Lz gl Su* and *lz Gl su*, which is inconsistent with the data. Likewise, if *lz* were in the middle (part C), the double-recombinant gametes would be *Gl lz Su* and *gl Lz su*, which is also inconsistent with the data.

The correct order of the genes, *lz*–*su*–*gl*, is given in part B, because in this case, the double-recombinant gametes are *Lz su Gl* and *lz Su gl*, which Table 4.1 indicates is actually the case. Although one can always infer which gene is in the middle by going through all three possibilities, there is a shortcut. Each double-recombinant gamete will always match one of the parental gametes in two of the alleles. In Table 4.1, for example, the double-recombinant gamete *Lz Gl su* matches the parental gamete *Lz Gl Su* except for the allele *su*. Similarly, the double-recombinant gamete *lz gl Su* matches the parental gamete *lz gl su* except for the allele *Su*. The middle gene can be identified because the "odd man out" in the comparisons—in this case, the alleles of *Su*—is always the gene in the middle. The reason is that only the middle pair of alleles is interchanged by double crossing-over.

Taking the correct gene order into account, the genotype of the heterozygous parent in the cross yielding the progeny in Table 4.1 should be written as

$$\frac{Lz\ \ Su\ \ Gl}{lz\ \ su\ \ gl}$$

The consequences of single crossing-over in this genotype are shown in **FIGURE 4.15**. A single crossover in the *lz*–*su* region (part A) yields the reciprocal recombinants *Lz su gl* and *lz Su Gl*, and a single crossover in the *su*–*gl* region (part B) yields the reciprocal recombinants *Lz Su gl* and *lz su Gl*. The consequences of double crossing-over are illustrated in **FIGURE 4.16**. There are four different types of double crossovers: a two-strand double (part A), two types of three-strand doubles (parts B and C), and a four-strand double (part D). These types were illustrated earlier in Figure 4.12, where the main point was that with two genetic markers flanking the crossovers, the occurrence of double

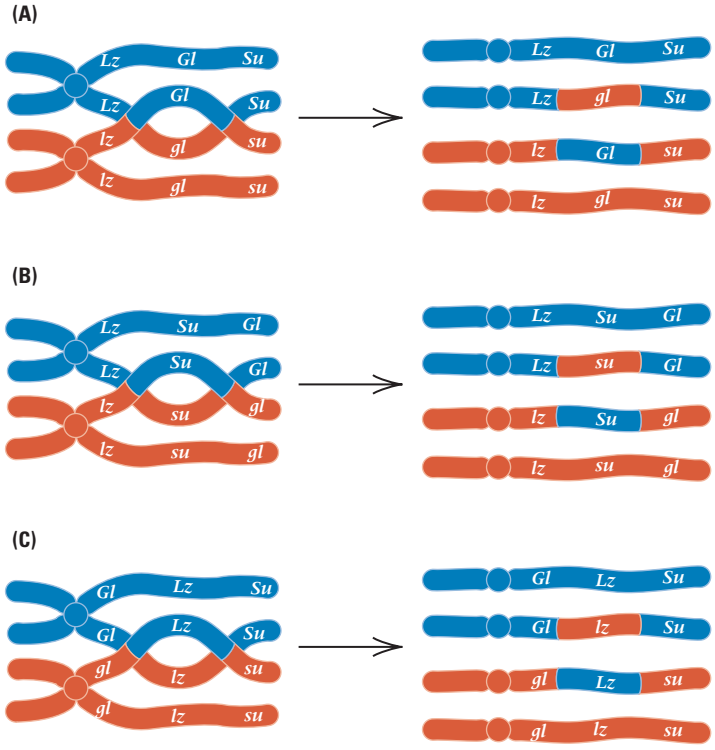


FIGURE 4.14 The order of genes in a three-point testcross may be deduced from the principle that double recombination interchanges the middle pair of alleles. For the genes *Lu*, *Gl*, and *Su*, there are three possible orders (parts A, B, and C), each of which predicts a different pair of gametes as the result of double recombination. Only the order in part B is consistent with the finding that *Lz Gl su* and *lz gl Su* are the double-recombinant gametes.

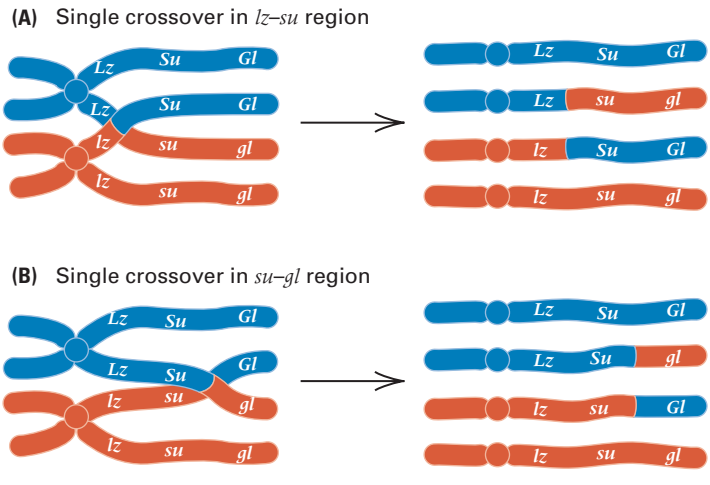


FIGURE 4.15 Result of single crossovers in a triple heterozygote, using the *Lz*–*Su*–*Gl* region as an example. (A) A crossover between *Lz* and *Su* results in two gametes that show recombination between *Lz* and *Su* and two gametes that are nonrecombinant. (B) A crossover between *Su* and *Gl* results in two gametes that show recombination between *Su* and *Gl* and two gametes that are nonrecombinant.

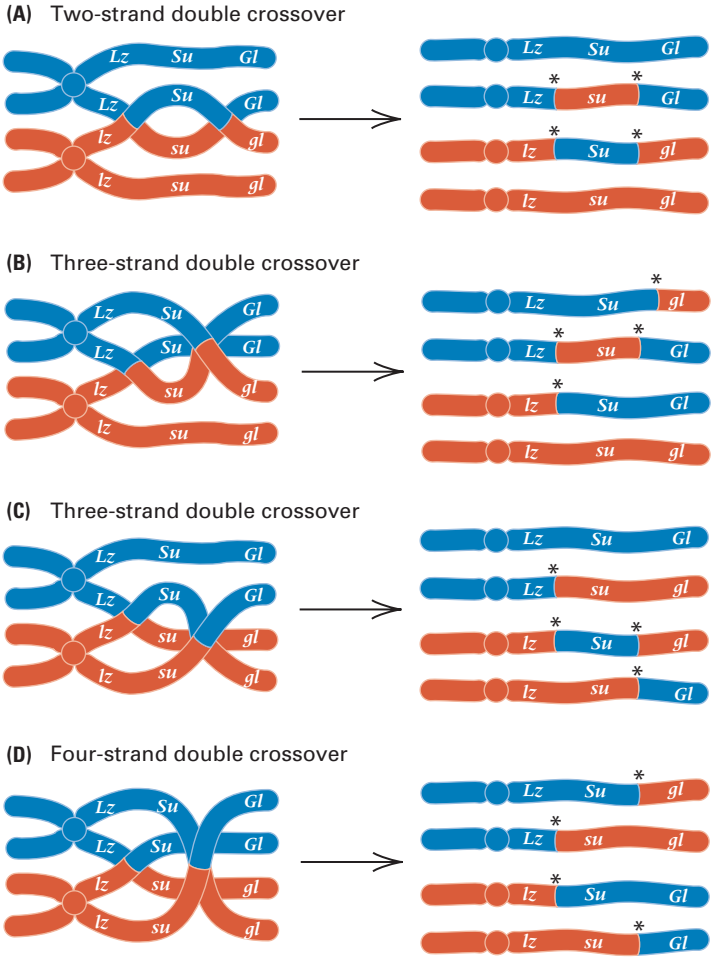


FIGURE 4.16 Result of double crossovers in a triple heterozygote, using the *Lz*–*Su*–*Gl* region as an example. Note that chromosomes showing double recombination derive from the two-strand double crossover (A) or from either type of three-strand double crossover (B and C). The four-strand double crossover (D) results only in single-recombinant chromosomes.

crossovers cannot be detected genetically. The difference in the present case is that, here, the genetic marker *su* is located in the middle between the two crossovers, so some of the double crossovers can be detected genetically. On the right in Figure 4.16, the asterisks mark the sites of crossing-over between nonsister chromatids. In terms of recombination, the result is that

- A two-strand double crossover (part A) yields the reciprocal double-recombinant products *Lz su Gl* and *lz Su gl*.
- One three-strand double crossover (part B) yields the double-recombinant product *Lz su Gl* and two single-recombinant products, *Lz Su gl* and *lz Su Gl*.
- The other three-strand double crossover (part C) yields the double-recombinant product *lz Su gl* and two single-recombinant products, *Lz su gl* and *lz su Gl*.
- The four-strand double crossover (part C) yields reciprocal single recombinants in the *lz*–*su* region, namely *Lz su gl* and *lz Su Gl*, and reciprocal single recombinants in the *su*–*gl* region, namely *Lz Su gl* and *lz su Gl*.

Note that the products of recombination in the three-strand double crossovers (parts B and C) are the reciprocals of each other. Because these two types of double crossovers are equally frequent, the reciprocal products of recombination are expected to appear in equal numbers.

We can now summarize the data in Table 4.1 in a more informative way by writing the genes in the correct order and grouping reciprocal gametic genotypes together. This grouping is shown in **TABLE 4.2**. Note that each class of single recombinants consists of two reciprocal products and that these are found in approximately equal frequencies (40 versus 33 and 59 versus 44). This observation illustrates an important principle:

KEY CONCEPT

The two reciprocal products resulting from any crossover, or any combination of crossovers, are expected to appear in approximately equal frequencies among the progeny.

In calculating the frequency of recombination from the data, remember that the double-recombinant chromosomes result from *two* exchanges, one in each of the chromosome regions defined by the three genes. Therefore, chromosomes that are recombinant between *lz*

and *su* are represented by the following chromosome types:

Lz su gl	40
lz Su Gl	33
Lz su Gl	4
lz Su gl	2
	79

The total implies that 79/740, or 10.7 percent, of the chromosomes recovered in the progeny are recombinant between the *lz* and *su* genes, so the map distance between these genes is 10.7 map units, or 10.7 centimorgans. Similarly, the chromosomes that are recombinant between *su* and *gl* are represented by

Lz Su gl	59
lz su Gl	44
Lz su Gl	4
lz Su gl	2
	109

In this case the recombination frequency between *su* and *gl* is 109/740, or 14.7 percent, so the map distance between these genes is 14.7 map units, or 14.7 centimorgans. The genetic map of the chromosome segment in which the three genes are located is therefore

The most common error in learning how to interpret three-point crosses is to forget to include the double recombinants when calculating the recombination frequency between adjacent genes. You can keep from falling into this trap by remembering that the double-recombinant chromosomes have single recombination in *both* regions.

Interference decreases the chance of multiple crossing-over.

The detection of double crossing-over makes it possible to determine whether exchanges in two different regions of a pair of chromosomes are formed independently of each other. Using the information from the example with corn, we know from the recombination frequencies that the probability of recombination is 0.107 between *lz* and *su* and 0.147 between *su* and *gl*. If recombination is independent in the two regions (which means that the formation of one crossover does not alter the probability of the second crossover), the probability of a single recombination in both regions is the product of these separate probabilities, or $0.107 \times 0.147 = 0.0157$ (1.57 percent). This implies that in a sample of 740 gametes, the expected number of double recombinants would be 740×0.0157, or 11.6, whereas the number actually observed was only 6 (Table 4.2). Such deficiencies in the observed number of double recombinants are common; they reflect a phenomenon called chromosome **interference**, in which a crossover in one region of a chromosome reduces the probability of a second crossover in a nearby region. Over short genetic distances, chromosome interference is nearly complete.

The **coefficient of coincidence** is the observed number of double-recombinant chromosomes divided by the expected number. Its value provides a quantitative measure of the degree of interference, which is defined as

$$i = \text{Interference} = 1 - (\text{Coefficient of coincidence})$$

From the data in the corn example, the coefficient of coincidence is calculated as follows:

- Observed number of double recombinants = 6
- Expected number of double recombinants = $0.107 \times 0.147 \times 740 = 11.6$
- Coefficient of coincidence = $6/11.6 = 0.52$

The 0.52 means that the observed number of double recombinants was only about half of the number expected if crossing-over in the two regions were independent. The value of the interference depends on the distance between the genetic markers and on the species. In some species, the interference increases as the distance between the two outside markers becomes smaller, until a point is reached at which double crossing-over is eliminated; that is, no double recombinants are found,

TABLE 4.2

Comparing Reciprocal Products in a Three-Point Cross

Genotype of gamete from hybrid parent	Number of progeny	Intervals showing recombination
Lz Su Gl *lz su gl*	286 272	
Lz su gl *lz Su Gl*	40 33	*lz*–*su*
Lz Su gl *lz su Gl*	59 44	*su*–*gl*
Lz su Gl *lz Su gl*	4 2	*lz*–*su* + *su*–*gl*
	740	

Total number of recombinants in *lz*–*su* region:
40 + 33 + 4 + 2 = 79

Total number of recombinants in *su*–*gl* region:
59 + 44 + 4 + 2 = 109

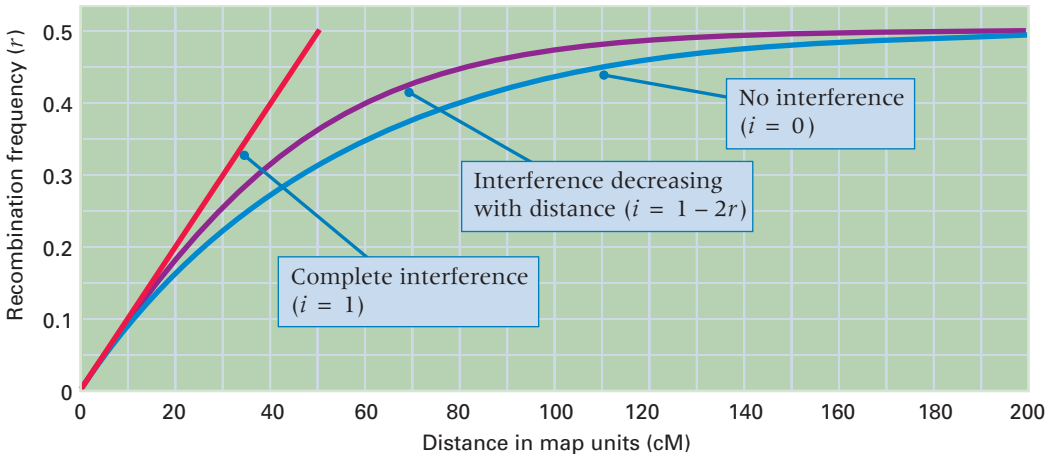


FIGURE 4.17 A mapping function is the relation between genetic map distance across an interval and the observed frequency of recombination across the interval. Map distance is defined as one-half the average number of crossovers converted into a percentage. The three mapping functions correspond to different assumptions about interference, *i*.

and the coefficient of coincidence equals 0 (or, to say the same thing, the interference equals 1). In *Drosophila* this distance is about 10 map units.

The effect of interference on the relationship between genetic map distance and the frequency of recombination is illustrated in **FIGURE 4.17**. Each curve is an example of a *mapping function,* which is the mathematical relation between the genetic distance across an interval in map units (centimorgans) and the observed frequency of recombination across the interval. In other words, a mapping function tells one how to convert a *map distance* between genetic markers into a *recombination frequency* between the markers. As we have seen, when the map distance between the markers is small, the recombination frequency equals the map distance. This principle is reflected in the curves in Figure 4.17 in the region in which the map distance is smaller than about 10 cM. At less than this distance, all of the curves are nearly straight lines, which means that map distance and recombination frequency are equal; 1 map unit equals 1 percent recombination, and 10 map units equals 10 percent recombination. For distances greater than 10 map units, the recombination frequency becomes smaller than the map distance according to the pattern of interference along the chromosome. Each pattern of interference yields a different mapping function, as shown by the three examples in Figure 4.17.

4.4 Polymorphic DNA sequences are used in human genetic mapping.

Until quite recently, mapping genes in human beings was very tedious and slow. Numerous practical obstacles complicated genetic mapping in human pedigrees:

1. Most genes that cause genetic diseases are rare, so they are observed in only a small number of families.
2. Many mutant genes of interest in human genetics are recessive, so they are not detected in heterozygous genotypes.
3. The number of offspring per human family is relatively small, so segregation cannot usually be detected in single sibships.
4. The human geneticist cannot perform testcrosses or backcrosses, because human matings are not dictated by an experimenter.

Human genetics has been revolutionized by the use of techniques for manipulating DNA. These techniques have enabled investigators to carry out genetic

Q A MOMENT TO THINK

Problem: In his pioneering 1913 studies in *Drosophila* that resulted in the first genetic linkage map, A. H. Sturtevant included three genetic markers now known to cover almost the entirety of the euchromatin of the X chromosome. The marker *w* (white eyes) is near the tip, *m* (miniature body) near the middle, and *r* (rudimentary wings) near the centromere. In two-point crosses, Sturtevant obtained recombination frequencies of 0.32 for the interval *w*–*m*, 0.25 for the interval *m*–*r*, and 0.45 for the interval *w*–*r*. He noted that 0.45 is smaller than the sum of 0.32 + 0.25 = 0.57, which is the value expected if the frequencies of recombination over such large distances were additive. He commented that the discrepancy "is probably due to the occurrence of two breaks in the same chromosome, or double crossing-over." From Sturtevant's data, calculate the coincidence and the interference across the region. (The answer can be found at the end of the chapter.)

mapping in human pedigrees primarily by using genetic markers present in the DNA itself, rather than through the phenotypes produced by mutant genes. There are many minor differences in DNA sequence from one person to the next. On the average, the DNA sequences at corresponding positions in any two chromosomes, taken from any two people, differ at approximately one in every thousand base pairs. A genetic difference that is relatively common in a population is called a **polymorphism**. Most polymorphisms in DNA sequence are not associated with any inherited disease or disability; many occur in DNA sequences that do not code for proteins. Nevertheless, each of the polymorphisms serves as a convenient genetic marker, and those genetically linked to genes that cause hereditary diseases are particularly important.

Single-nucleotide polymorphisms (SNPs) are abundant in the human genome.

A **single-nucleotide polymorphism**, or **SNP** (pronounced "snip"), is present at a particular nucleotide site if the DNA molecules in the population often differ in the identity of the nucleotide pair that occupies the site. For example, some DNA molecules may have a T–A base pair at a particular nucleotide site, whereas other DNA molecules in the same population may have a C–G base pair at the same site. This difference constitutes a SNP. The SNP defines two "alleles" for which there could be three genotypes among individuals in the population: homozygous with T–A at the corresponding site in both homologous chromosomes, homozygous with C–G at the corresponding site in both homologous chromosomes, or heterozygous with T–A in one chromosome and C–G in the homologous chromosome. The word *allele* is in quotation marks above because the SNP need not be in a coding sequence, or even in a gene. In the human genome, any two randomly chosen DNA molecules are likely to differ at one SNP site about every 1000 bp in noncoding DNA and at about one SNP site every 3000 bp in protein-coding DNA. In the definition of a SNP, the DNA molecules must differ at the nucleotide site "often." This provision excludes rare genetic variation of the sort found in less than 1 percent of the DNA molecules in a population. The reason for the exclusion is that genetic variants that are too rare are not generally as useful in genetic analysis as the more common variants. SNPs are the most common form of genetic differences among people. About 10 million SNPs have been identified in which the alternative nucleotides are each relatively common in the human population, and about half a million of these are typically used in a search for SNPs that might be associated with complex diseases such as diabetes or high blood pressure.

Identifying the particular nucleotide present at each of a million SNPs is made possible through the use of **DNA microarrays** composed of about 20 million infinitesimal spots on a glass slide about the size of a postage stamp. Each tiny spot contains a unique DNA oligonucleotide sequence present in millions of copies synthesized by microchemistry when the microarray is manufactured. Each oligonucleotide sequence is designed to hybridize specifically with small fragments of genomic DNA that include one or the other of the nucleotide pairs present in a SNP. The microarrays also include numerous controls for each hybridization. The controls consist of oligonucleotides containing deliberate mismatches that are intended to guard against being misled by particular nucleotide sequences that are particularly "sticky" and hybridize too readily with genomic fragments and other sequences that form structures that hybridize poorly or not at all. Such microarrays, sometimes called **SNP chips**, enable the SNP genotype of an individual to be determined with nearly 100 percent accuracy.

The principles behind oligonucleotide hybridization are illustrated in **FIGURE 4.18**. Here the length of each oligonucleotide is seven nucleotides, but in practice this is too short. Typical SNP chips consist of oligonucleotides at least 25 nucleotides in length. Part A shows the two types of DNA duplexes that might form a SNP. In this example, some chromosomes carry a DNA molecule with a T–A base pair at the position shown in red, while the DNA molecule in other chromosomes has a C–G base pair at the corresponding position. Short fragments of genomic DNA are labeled with a fluorescent tag, and then the single strands hybridized with a SNP chip containing the complementary oligonucleotides as well as the numerous controls. The duplex containing the T–A hybridizes only with the two oligonucleotides on the left, and that containing the C–G hybridizes only with the two oligonucleotides on the right.

After the hybridization takes place, the SNP chip is examined with fluorescence microscopy to detect the spots that fluoresce as a result of the tag on the genomic DNA. The possible patterns are shown in part B. Genomic DNA from an individual whose chromosomes contain two copies of the T–A form of the duplex (homozygous TA/TA) causes the two leftmost spots to fluoresce, but the two rightmost spots remain unlabeled. Similarly, genomic DNA from a homozygous CG/CG individual causes the two rightmost spots to fluoresce but not the two on the left. Finally, genomic DNA from a heterozygous TA/CG individual causes fluorescence of all four spots because the TA duplex labels the two leftmost spots and the CG duplex labels the two rightmost spots.

Use of SNP chips or other available technologies for high-throughput genotyping of millions of SNPs in thousands of individuals allows genetic risk factors

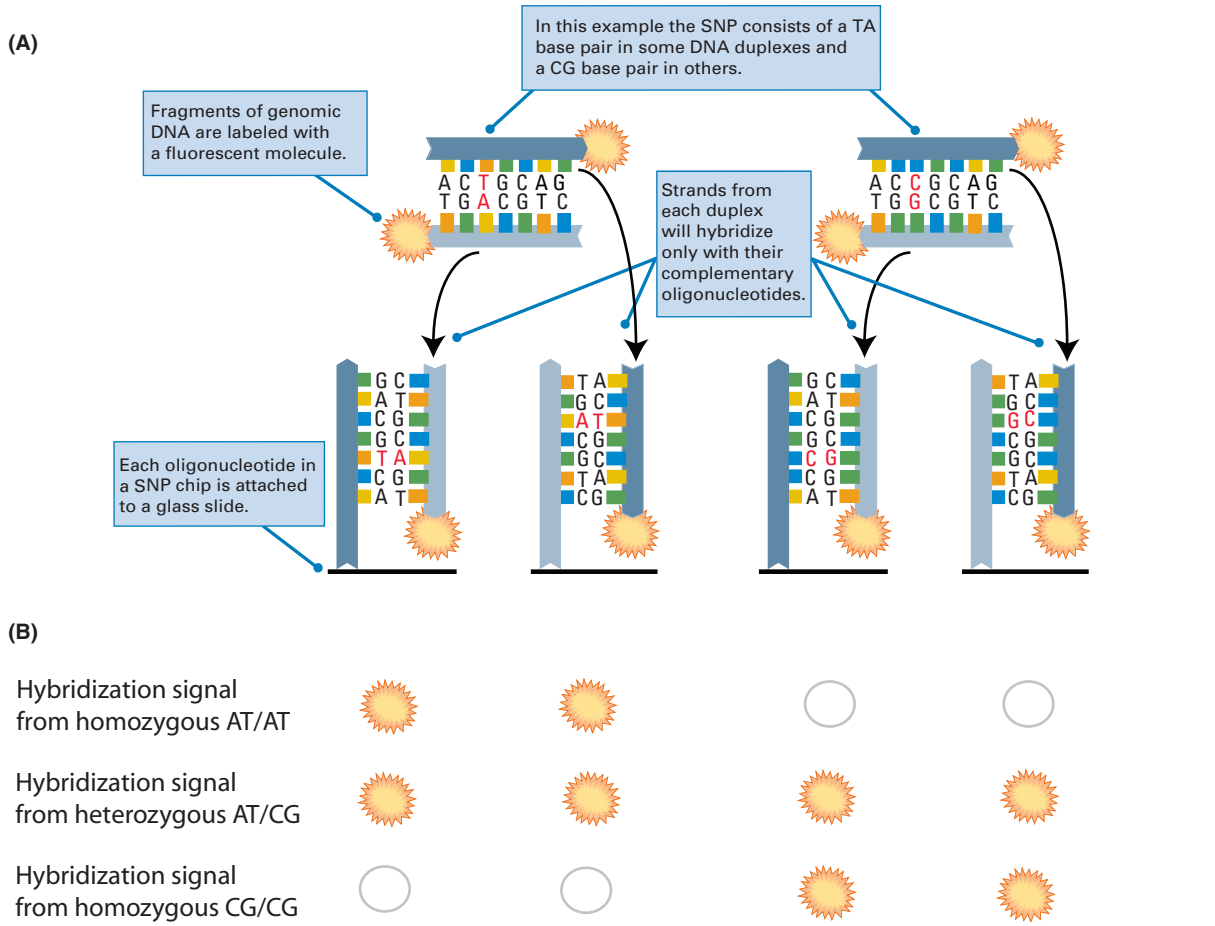


FIGURE 4.18 (A) Oligonucleotides attached to a glass slide in a SNP chip can be used to identify duplex DNA molecules containing base pairs for a SNP; in this example, a TA base pair versus a CG base pair. (B) The SNP genotype of an individual can be determined by hybridization because DNA from genotypes that are homozygous TA/TA, homozygous CG/CG, or heterozygous TA/CG each gives a different pattern of fluorescence.

for disease to be identified. A typical study compares the genotypes of patients with particular diseases with healthy people matched with the patients for such factors as sex, age, and ethnic group. Comparing the SNP genotypes among these groups often reveals which SNPs are in the genome.

SNPs in restriction sites yield restriction fragment length polymorphisms (RFLPs).

Some polymorphisms in DNA sequence are detected by means of a type of enzyme called a **restriction endonuclease**, which cleaves double-stranded DNA molecules wherever a particular, short sequence of bases is present. For example, the restriction enzyme *Eco*RI cleaves DNA wherever the sequence GAATTC appears in either strand, as illustrated in **FIGURE 4.19**. Restriction enzymes are considered in detail in Chapter 6. For present purposes, their significance is related to the fact that a difference in DNA sequence that eliminates a cleavage site can be detected because the region lacking the cleavage site will be cleaved into one larger fragment instead of two smaller ones (**FIGURE 4.20**). More rarely, a mutation in the DNA sequence will create a new site rather than destroy one already present. The main point is that any difference in DNA sequence that alters a cleavage site also changes the length of the DNA fragments produced by cleavage with the corresponding restriction enzyme. The different DNA fragments can be separated by size by an electric field in a supporting gel and detected by various means. Differences in DNA fragment length produced by presence or absence of the cleavage sites in DNA molecules are known as **restriction fragment length polymorphisms (RFLPs)**.

RFLPs are typically formed in one of two ways. A mutation that changes a base sequence may result in loss or gain of a cleavage site that is recognized by the restriction endonuclease in use. **FIGURE 4.21**, part A, gives an example. On the left is shown the relevant region in the homologous DNA molecules in a person who is heterozygous for such a sequence polymorphism. The homologous chromosomes in the person

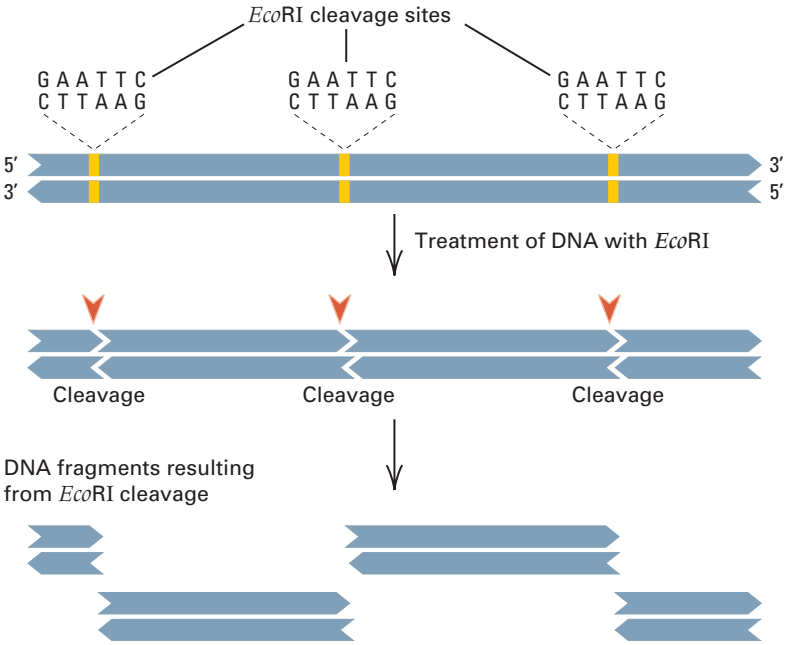


FIGURE 4.19 The restriction enzyme *Eco*RI cleaves double-stranded DNA wherever the sequence 5′-GAATTC-3′ is present. In the example shown here, the DNA molecule contains three *Eco*RI cleavage sites, and it is cleaved at each site, producing a number of fragments.

are distinguished by the letters *a* and *b*. In the region of interest, chromosome *a* contains two cleavage sites and chromosome *b* contains three. On the right is shown the position of the DNA fragments produced by cleavage after separation in an electric field. Each fragment appears as a discrete band in the gel. The fragment from chromosome *a* migrates more slowly than those from chromosome *b* because it is longer, and longer fragments move more slowly through the gel. In this example, DNA from a person heterozygous for the *a* and *b* types of chromosomes (genotype *ab*) would yield three bands in a gel. Similarly, DNA from homozygous *aa* would yield one band, and that from homozygous *bb* would yield two bands.

Simple-sequence repeats (SSRs) often differ in copy number.

Yet another type of DNA polymorphism results from differences in the number of copies of a short DNA sequence that may be repeated many times in tandem at a particular site in a chromosome (Figure 4.21, part B). In a particular chromosome, the tandem repeats may contain any number of copies, typically ranging from ten to a few hundred. When a DNA molecule is cleaved with a restriction endonuclease that cleaves at sites flanking the tandem repeat, the size of the DNA fragment produced is determined by the number of repeats present in the molecule. Figure 4.21, part B, illustrates homologous DNA sequences in a heterozygous person containing one chromosome *a* with two copies of the repeat and another chromosome *b* with five copies of the repeat. When cleaved and separated in a gel, chromosome *a* yields a shorter fragment than that from chromosome *b*, because *a* contains fewer copies of the repeat. A genetic polymorphism resulting from a tandemly repeated short DNA sequence is called a **simple sequence repeat (SSR)**. An example of an SSR is the repeating sequence

5′-...TGTGTGTGTGTG...-3′

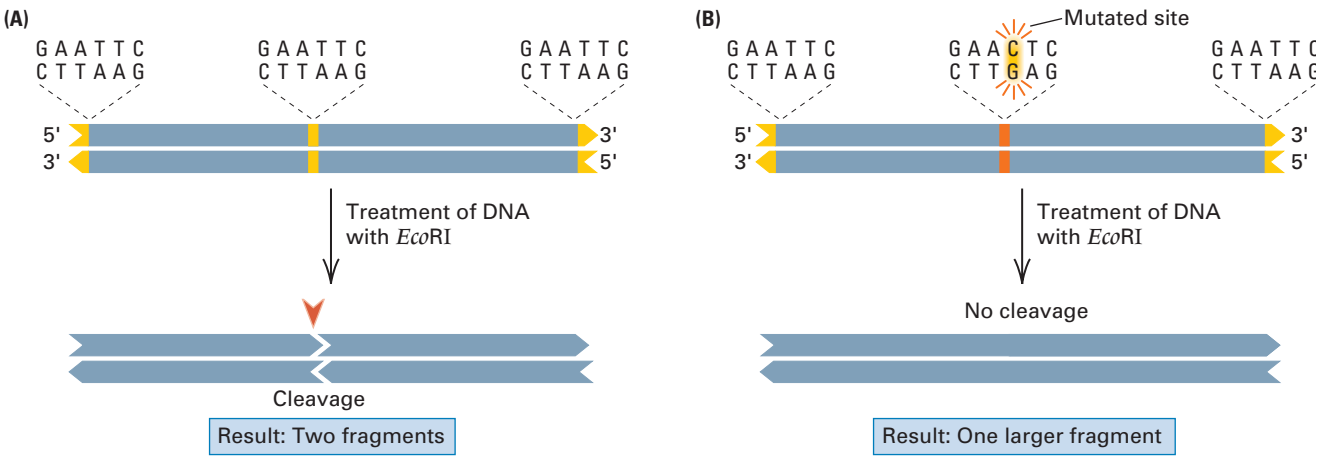


FIGURE 4.20 A minor difference in the DNA sequence of two molecules can be detected if the difference eliminates a restriction site. (A) This molecule contains three restriction sites for *Eco*RI, including one at each end. It is cleaved into two fragments by the enzyme. (B) This molecule has a mutant base sequence in the *Eco*RI site in the middle. It changes 5′-GAATTC-3′ into 5′-GAACTC-3′, which is no longer cleaved by *Eco*RI. Treatment of this molecule with *Eco*RI results in one larger fragment.

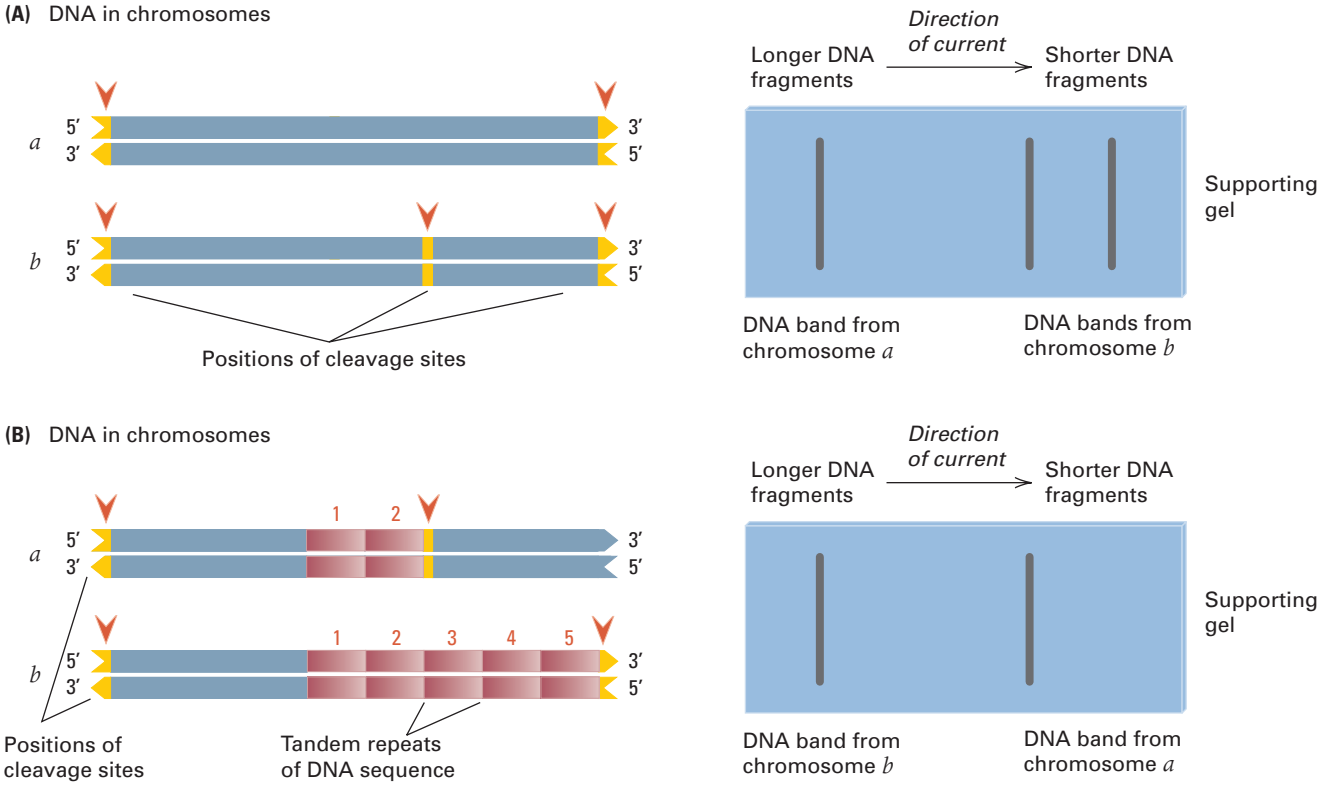


FIGURE 4.21 Two types of genetic variation that are widespread in most natural populations of animals and plants. (A) RFLP (restriction fragment length polymorphism), in which alleles differ in the presence or absence of a cleavage site in the DNA. The different alleles yield different fragment lengths (shown in the gel pattern at the right) when the molecules are cleaved with a restriction enzyme. (B) SSR (simple sequence repeat), in which alleles differ in the number of repeating units present between two cleavage sites.

and the polymorphism consists of differences in the number of TG repeats. A particular "allele" of the SSR is defined by the number of TG repeats it includes.

One source of the utility of SSRs in human genetic mapping is the high density of SSRs across the genome. There is an average of one SSR per 2 kb of human DNA. Some examples are shown in **TABLE 4.3**. The prevalence of different SSRs differs. Some dinucleotide repeats, such as 5′-AC-3′ and 5′-AT-3′, are far more frequent than others, such as 5′-GC-3′. Overall, dinucleotide repeats are much more abundant than trinucleotide repeats. The second source of utility of SSRs in genetic mapping is the large number of alleles that can be present in any human population. The large number of alleles also implies that most people will be heterozygous, and so their DNA will yield two bands upon cleavage with the appropriate restriction endonuclease. Because of their high degree of variation among people, DNA polymorphisms are also widely used in DNA typing in criminal investigations.

In genetic mapping, the phenotype of a person with respect to a DNA polymorphism is a pattern of bands in a gel. As with any other type of genetic marker, the genotype of a person with respect to the polymorphism is inferred, insofar as it is possible, from the phenotype.

TABLE 4.3

Some Simple Sequence Repeats in the Human Genome

SSR repeat unit	Number of SSRs in the human genome
5′-AC-3′	80,330
5′-AT-3′	56,260
5′-AG-3′	23,780
5′-GC-3′	290
5′-AAT-3′	11,890
5′-AAC-3′	7,540
5′-AGG-3′	4,350
5′-AAG-3′	4,060
5′-ATG-3′	2,030
5′-CGG-3′	1,740
5′-ACC-3′	1,160
5′-AGC-3′	870
5′-ACT-3′	580

Data from International Human Genome Sequencing Consortium, *Nature* 409 (2001): 860–921.

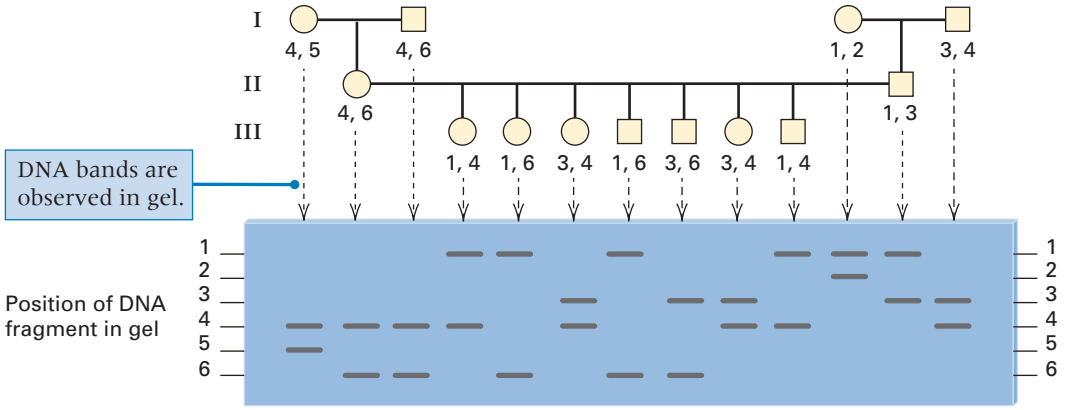


FIGURE 4.22 Human pedigree showing segregation of SSR alleles. Six alleles (1–6) are present in the pedigree, but any one person can have only one allele (if homozygous) or two alleles (if heterozygous).

Linkage between different polymorphic loci is detected through lack of independent assortment of the alleles in pedigrees, and recombination and genetic mapping are carried out using the same principles as apply in other organisms, except that in human beings, because of the small family size, different pedigrees are pooled for analysis. Primarily through the use of DNA polymorphisms, genetic mapping in humans has progressed rapidly.

To give an example of the type of data used in human genetic mapping, a three-generation pedigree of a family segregating for several alleles of an SSR is illustrated in **FIGURE 4.22**. In this example, each of the parents is heterozygous, as are all of the children. Yet every person can be assigned his or her genotype because the SSR alleles are codominant. At present, DNA polymorphisms are the principal types of genetic markers used in genetic mapping in human pedigrees. Such polymorphisms are prevalent, are located in virtually all regions of the chromosome set, and have multiple alleles and so yield a high proportion of heterozygous genotypes. Furthermore, only a small amount of biological material is needed to perform the necessary tests.

One feature of the human genetic map is that there is about 60 percent more recombination in females than in males, so the female and male genetic maps differ in length. The female map is about 4400 cM, the male map about 2700 cM. Averaged over both sexes, the length of the human genetic map for all 23 pairs of chromosomes is about 3500 cM. Because the total DNA content per haploid set of human chromosomes is 3200 million base pairs, there is, very roughly, 1 cM per million base pairs in the human genome.

Gene dosage can differ owing to copy-number variation (CNV).

In addition to the small-scale variation in copy number represented by such genomic features as simple-sequence repeats (SSRs), a substantial portion of the human genome can be duplicated or deleted in much larger but still submicroscopic chunks ranging from 1 kb to 1 Mb (Mb stands for megabase pairs, or one million base pairs). This type of variation is known as **copy-number variation (CNV)**. The extra or missing copies of the genome in CNVs can be detected by means of hybridization with oligonucleotides in DNA microarrays. Since each spot on the microarray consists of millions of identical copies of a particular oligonucleotide sequence, the number of these that undergo hybridization depends on how many copies of the complementary sequence are present in genomic DNA. A typical region is present in two copies (one inherited from the mother and the other from the father). If an individual has an extra copy of the region, the ratio of hybridization and, therefore, fluorescence intensity is 3 : 2, while if an individual has a missing copy, the ratio of hybridization and, therefore, fluorescence intensity is 1 : 2. These differences can readily be detected with DNA microarrays. Moreoever, since CNVs are relatively large, a microarray typically has many different oligonucleotides that are complementary to sequences at intervals across the CNV; hence, the CNV results in a increase or decrease in signal intensity of all the oligonucleotides included in the CNV. Current SNP chips also include about a million oligonucleotide probes designed to detect known CNVs.

CNVs by definition exceed 1 kb in size, but many are much larger. In one study of about 300 individuals with their ancestry in Africa, Europe, or Asia, approximately 1500 CNVs were discovered by hybridization with microarrays. These averaged 200 to 300 kb in length. In the aggregate, the CNVs included 300 to 450 million base pairs, or 10 to 15 percent of the nucleotides in the entire genome. Many of the CNVs were located in regions near known mutant genes associated with hereditary diseases. CNVs in alpha and beta hemoglobin

THE HUMAN CONNECTION

Starch Contrast

George H. Perry[1,2], Nathaniel J. Dominy[3], Katrina G. Claw[1], Arthur S. Lee[2], Heike Fiegler[4], Richard Redon[4], John Werner[1], and 6 other authors 2007

[1]Arizona State University, Tempe, AZ;
[2]Brigham and Women's Hospital, Boston, MA;
[3]University of California, Santa Cruz, CA;
[4]The Wellcome Trust Sanger Institute, Hinxton UK.

Diet and the Evolution of Human Amylase Gene Copy Number Variation

We favor a model in which *AMY1* copy number has been subject to positive or directional selection in at least some high-starch populations . . .

Evolution of increased enzyme activity can occur through regulatory mutations that increase transcription of a single gene or through increases in gene copy number. This study reports a strong correlation between amount of starch in the diets of human populations and the number of copies of a gene encoding salivary amylase, a starch-degrading enzyme. Humans are not alone among primates in having evolved higher amylase activity. A group of Old World Monkeys called cercopithecines, which includes macaques and mangabeys, produces even more salivary amylase than humans. Cercopithecines are unique among primates in storing starchy foods, such as the seeds of unripe fruits, in a cheek pouch, and it is a plausible hypothesis that the increased amylase facilitates digestion of the starch. It is not known whether the increased amylase production in cercopithecines is due to copy-number variation or to some other mechanism.

Hominin evolution is characterized by significant dietary shifts, facilitated in part by the development of stone tool technology, the control of fire and, most recently, the domestication of plants and animals. . . . Starch, for instance, has become an increasingly prominent component of the human diet, particularly among agricultural societies. . . . A distinction can be made between "high-starch" populations for which starchy food resources comprise a substantial portion of the diet and "low-starch" populations with traditional diets that incorporate relatively few starchy foods [but] instead emphasize proteinaceous resources (for example, meats and blood) and simple saccharides (for example, from fruit, honey, and milk). . . . We estimated [salivary amylase gene] *AMY1* copy number in three high-starch and four low-starch population samples. . . . Notably, the proportion of individuals from the combined high-starch sample with at least six *AMY1* copies (70 percent) was nearly two times greater than that for low-starch populations (37 percent). . . . Diet more strongly predicts *AMY1* copy number than geographic proximity. . . . We favor a model in which *AMY1* copy number has been subject to positive or directional selection in at least some high-starch populations but has evolved neutrally (that is, through genetic drift) in low-starch populations. . . . Comparisons with other great apes suggest that *AMY1* copy number was probably gained in the human lineage. . . . The initial human-specific increase in *AMY1* copy number may have been coincident with a dietary shift early in hominin evolutionary history. For example, it is hypothesized that starch-rich plant underground storage organs such as bulbs, corms and tubers were a critical food resource for early hominins.

Source: Science 328: 710–722.

genes are known to be associated with resistance to malaria, and CNVs in an HIV-1 receptor gene *CCL3* are associated with resistance to AIDS. The prevalence of CNVs has prompted the inclusion of CNV oligonucleotides onto DNA microarrays to be able to assess what effects CNVs may have on the risk of complex diseases such as autism or Alzheimer's disease.

Copy-number variation has helped human populations adapt to a high-starch diet.

Increased copy number is sometimes beneficial, with the result that natural selection operates to gradually increase copy number throughout an entire population of organisms. An excellent example in human evolution

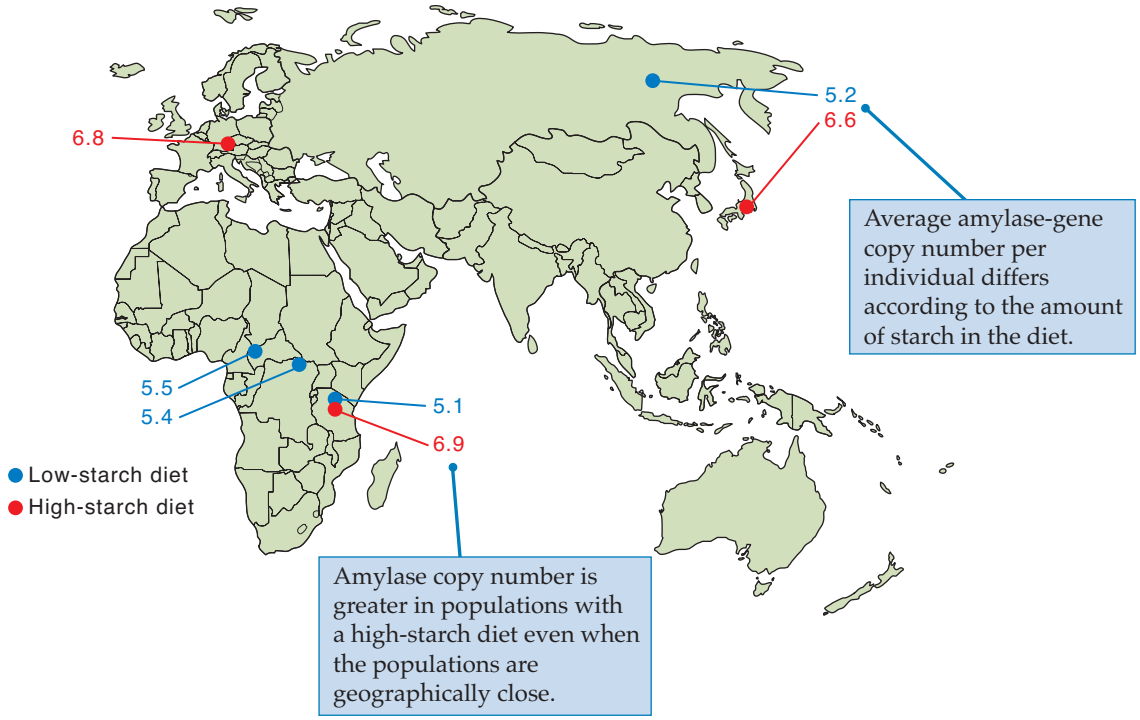


FIGURE 4.23 Amylase copy number varies with the amount of starch in human diets. [Data from G. H. Perry, *et al. Nat. Genet.* 39 (2007): 1256–1260.]

is copy-number variation in the gene for the starch-degrading enzyme amylase, which is produced in the salivary gland and pancreas.

As anatomically modern humans left their original homes in Southern Africa about 200,000 years ago and migrated first north to populate all of Africa and then, about 60,000 years ago, into the Middle East and then other parts of the world, the migrating populations had to cope with different climates, soil types, water availability, and native plants and animals. Some of the novel conditions were accommodated by changes in diet. Traditional hunter-gatherer populations in humid climates such as tropical forests typically eat a diet rich in meat or fish and relatively low in starch. Pastoralists, who follow herds like reindeer or bison as they migrate during the year, also have diets rich in meat and relatively low in starch. In contrast, hunter-gatherer populations in arid climates rely more on roots and tubers for food and increase their intake of starch. Populations that switch to agriculture also have a dramatic increase in dietary starch because cultivated staple foods like wheat, rice, potatoes, and corn are rich in starch.

One consequence of copy-number variation is that the amount of protein produced from any gene included in the region is usually proportional to the copy number. Increased gene copy number usually implies increased protein level. In human populations with high-starch diets, it might be expected that increased amylase is beneficial because more calories can be absorbed as the starch is digested. In human populations, increased amylase results from copy-number variation. The results of one study of amylase gene copy number are shown in **FIGURE 4.23**. Seven populations were assayed for amylase gene copy number, four with low-starch diets (blue) and three with high-starch diets (red). On average, individuals in populations with a low-starch diet had an average of 5.4 copies of the amylase gene, whereas those with a high-starch diet had an average of 6.7 copies of the gene. The comparisons exclude geography as an explanation for the difference. Note, for example, the contrasting populations in Tanzania in East Africa at the lower right and the contrasting populations in Siberia and Japan at the upper right. Our nearest primate relatives, chimpanzees and gorillas, have only one copy of the amylase gene; therefore, amylase gene copy number apparently has increased generally in human evolution, perhaps beginning at a time when our ancestors left the humid forests for the arid savannahs and began to eat more starchy roots and tubers.

4.5 Tetrads contain all four products of meiosis.

In some species of fungi, each meiotic tetrad is contained in a sac-like structure, called an **ascus**, and can be recovered as an intact group. Each product of meiosis is included in a reproductive cell called an **ascospore**, and all of the ascospores formed from one meiotic cell remain together in the ascus **FIGURE 4.24**. The advantage of these organisms for the study of recombination is the

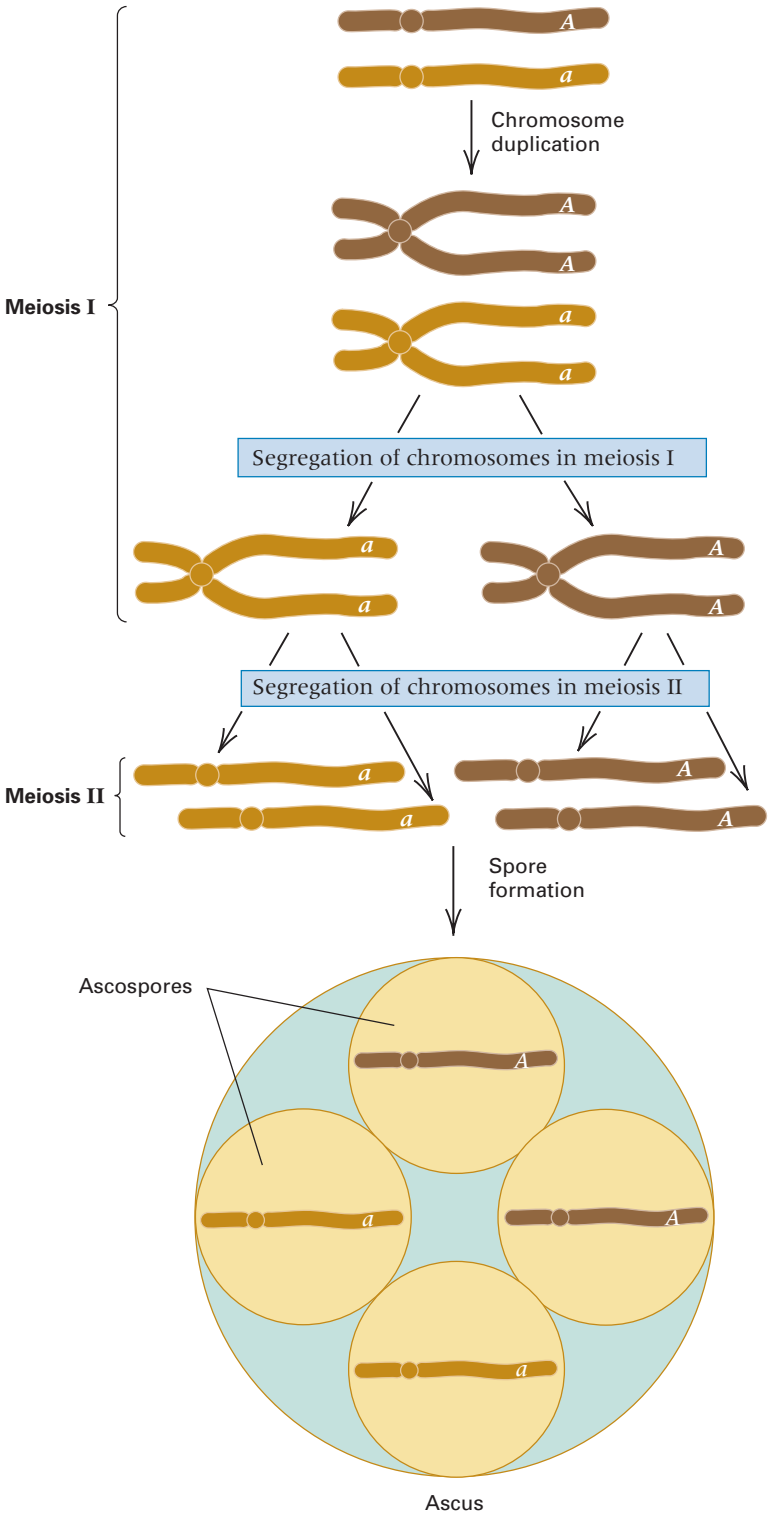


FIGURE 4.24 Formation of an ascus containing all of the four products of a single meiosis. Each ascospore present in the ascus is a reproductive cell formed from one of the products of meiosis.

potential for analyzing all of the products from each meiotic division. For example, one can see immediately from the diagram in Figure 4.24 that a tetrad containing the products of a single meiosis in a heterozygous *Aa* organism contains 2 *A* ascospores and 2 *a* ascospores. The 2 *A* : 2 *a* segregation means that the Mendelian ratio of 1 : 1 is realized in the products of each individual meiotic division and is not merely an average over a large number of meioses.

Two other features of ascus-producing organisms are especially useful for genetic analysis: (1) They are haploid, so dominance is not a complicating factor because the genotype is expressed directly in the phenotype. (2) They produce very large numbers of progeny, making it possible to detect rare events and to estimate their frequencies accurately. Furthermore, the life cycles of the organisms tend to be short. The only diploid stage is the zygote, which undergoes meiosis soon after it is formed; the resulting haploid meiotic products (which form the ascospores) germinate to regenerate the vegetative stage **FIGURE 4.25**. In most of the organisms, the meiotic products, or their derivatives, are not arranged in any particular order in the ascus. However, bread molds of the genus *Neurospora* and related organisms have the useful characteristic that the meiotic products are arranged in a definite order directly related to the planes of the meiotic divisions. In *Neurospora*, each of the four products of meiosis also undergoes a mitotic division, with the result that each member of the tetrad yields a *pair* of genetically identical ascospores. We will examine the ordered system after first looking at unordered tetrads.

Unordered tetrads have no relation to the geometry of meiosis.

In the tetrads when two pairs of alleles are segregating, three patterns of segregation are possible. For example, in the cross *A B* × *a b*, the three types of tetrads are

(*AB*) (*AB*) (*ab*) (*ab*) referred to as **parental ditype**, or **PD**. Only two genotypes are represented, and their alleles have the same combinations found in the parents.

(*Ab*) (*Ab*) (*aB*) (*aB*) referred to as **nonparental ditype**, or **NPD**. Only two genotypes are represented, but their alleles have nonparental combinations.

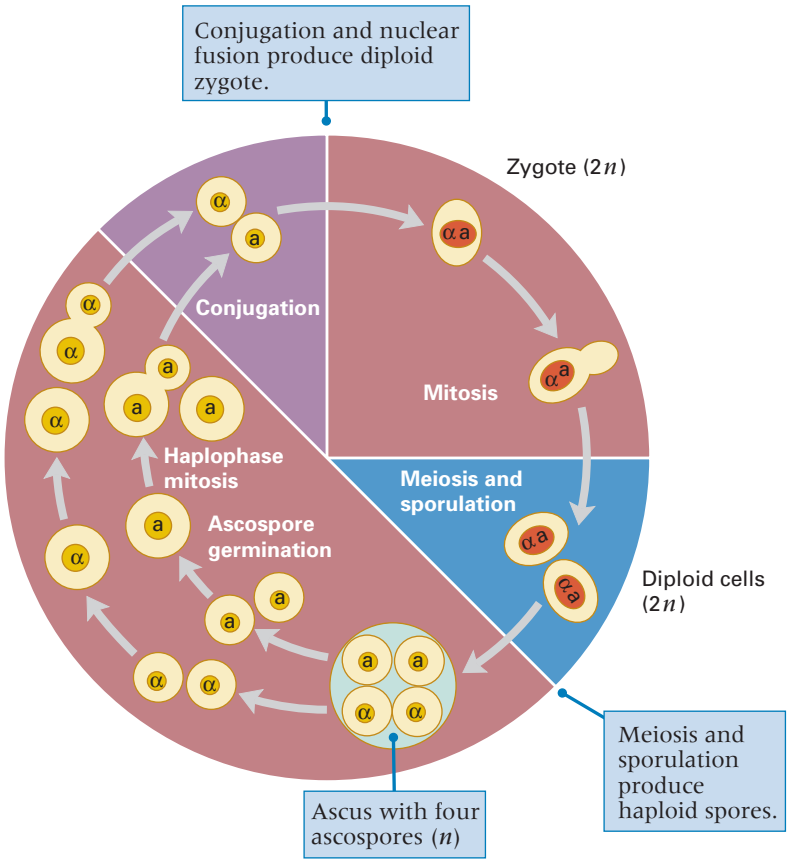


FIGURE 4.25 Life cycle of the yeast *Saccharomyces cerevisiae*. Mating type is determined by the alleles a and α. Both haploid and diploid cells normally multiply by mitosis (budding). Depletion of nutrients in the growth medium induces meiosis and sporulation of cells in the diploid state. Diploid nuclei are shown in red, haploid nuclei in yellow.

(AB) (Ab) (aB) (ab) referred to as **tetratype**, or **TT**. All four of the possible genotypes are present.

Tetratype tetrads demonstrate that crossing-over takes place at the four-strand stage of meiosis and is reciprocal.

We noted earlier that tetrads from heterozygous organisms regularly contain 2 *A* and 2 *a* ascospores, which implies that Mendelian segregation takes place in each meiosis. The existence of tetratype tetrads for linked genes demonstrates two features about crossing-over that we have assumed, so far without proof.

1. The exchange of segments between parental chromatids takes place in the first meiotic prophase, *after the chromosomes have duplicated*. Tetratype tetrads demonstrate this assertion because only two of the four products of meiosis show recombination. This would not be possible unless crossing-over took place at the four-strand stage.
2. The exchange process consists of the breaking and rejoining of the two chromatids, resulting in the *reciprocal* exchange of equal and corresponding segments. Tetratype tetrads demonstrate this point because they contain the reciprocal products (*A b* and *a B* if the parental alleles were in coupling, *A B* and *a b* if they were in repulsion).

Tetrad analysis affords a convenient test for linkage.

Tetrad analysis is an effective way to determine whether two genes are linked, because

KEY CONCEPT

When genes are *unlinked*, the parental ditype tetrads and the nonparental ditype tetrads are expected in equal frequencies (PD = NPD).

The reason for the equality PD = NPD for unlinked genes is shown in part A of **FIGURE 4.26**, where the two pairs of alleles *A, a* and *B, b* are located in different chromosomes. In the absence of crossing-over between

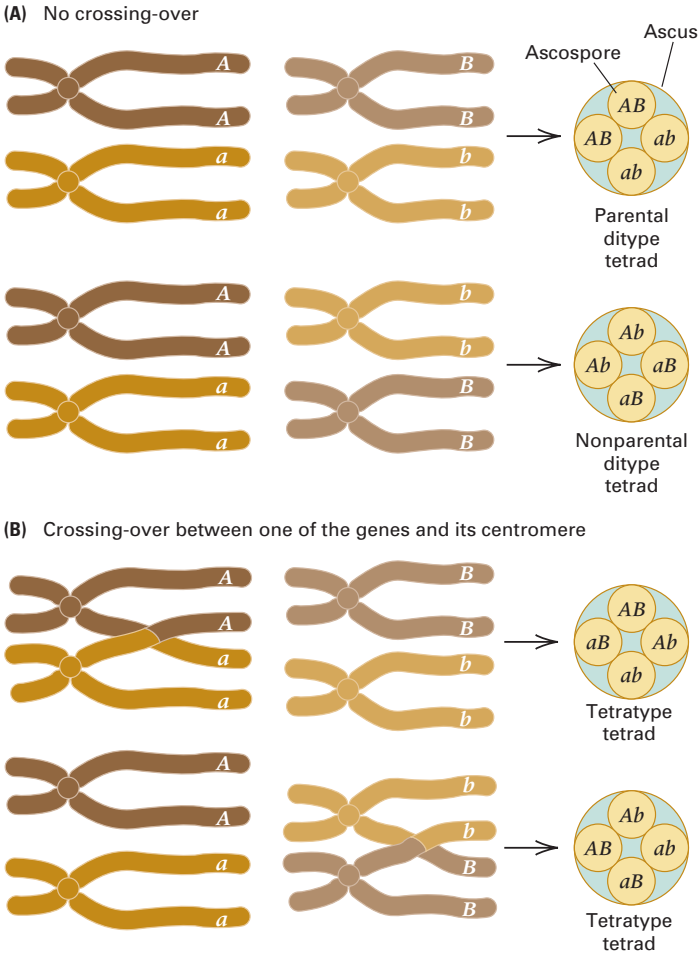


FIGURE 4.26 Types of unordered asci produced with two genes in different chromosomes. (A) In the absence of a crossover, random arrangement of chromosome pairs at metaphase I results in two different combinations of chromatids, one yielding PD tetrads and the other NPD tetrads. (B) When a crossover takes place between one gene and its centromere, the two chromosome arrangements yield TT tetrads. If both genes are closely linked to their centromeres (so that crossing-over is rare), few TT tetrads are produced.

either gene and its centromere, the two chromosomal configurations are equally likely at metaphase I, and so PD = NPD. When there is a crossover between either gene and its centromere (Figure 4.26, part B), a tetratype tetrad results, but this does not change the fact that PD = NPD.

In contrast, when genes are linked, parental ditypes are far more frequent than nonparental ditypes. To see why, assume that the genes are linked and consider the events required for the production of the three types of tetrads. **FIGURE 4.27** shows that when no crossing-over takes place between the genes, a PD tetrad is formed. Single crossover between the genes results in a TT tetrad. The formation of a two-strand, three-strand, or four-strand double crossover results in a PD, TT, or NPD tetrad, respectively. With linked genes, meiotic cells with no crossovers always outnumber those with four-strand double crossovers. Therefore,

KEY CONCEPT

Linkage is indicated when nonparental ditype tetrads appear with a much lower frequency than parental ditype tetrads (NPD << PD).

The relative frequencies of the different types of tetrads can be used to determine the map distance between two linked genes. The simplest case is one in which the genes are sufficiently close that double and higher levels of crossing-over can be neglected. In this case, tetratype tetrads arise only from meiotic cells in which a single crossover occurs between the genes (Figure 4.27, part A and part B). As we saw in Figure 4.6, the genetic map distance across an interval is defined as one-half the proportion of cells with a crossover in the interval, so the map distance implied by the tetrads is given by

$$\text{Map distance} = \frac{1}{2} \times \frac{\text{Number of tetratype tetrads}}{\text{Total number of tetrads}} \times 100 \quad (4.1)$$

To take a specific example, suppose 100 tetrads are analyzed from the cross *A B* × *a b*, and the result is that 91 are PD and 9 TT. The finding that NPD << PD means that the genes are linked, and the fact that NPD = 0 means that the genes are so closely linked that double crossing-over does not occur between them. The map distance between *A* and *B* is calculated as follows:

$$\text{Map distance} = \frac{1}{2} \times \frac{9}{100} \times 100 = 4.5\,\text{cM}$$

We must emphasize that Equation (4.1) is valid only when NPD = 0, so that interference across the region prevents the occurrence of double crossing-over. When double crossovers do take place in the interval, then NPD ≠ 0, and the formula for map distance has to be modified to take the double crossovers into account.

The mapping procedure using tetrads differs from that presented earlier in the chapter in that the map distance is not calculated directly from the number of recombinant and nonrecombinant chromatids. Instead, the map distance is calculated directly from the tetrads and the inferred crossovers that give rise to each type of tetrad. However, it is not necessary to carry out a full tetrad analysis for estimating linkage. The alternative is to examine spores chosen at random after allowing the tetrads to break open and disseminate their spores. This procedure is called *random-spore analysis*, and the linkage relationships are determined exactly as described earlier for *Drosophila* and corn. In particular,

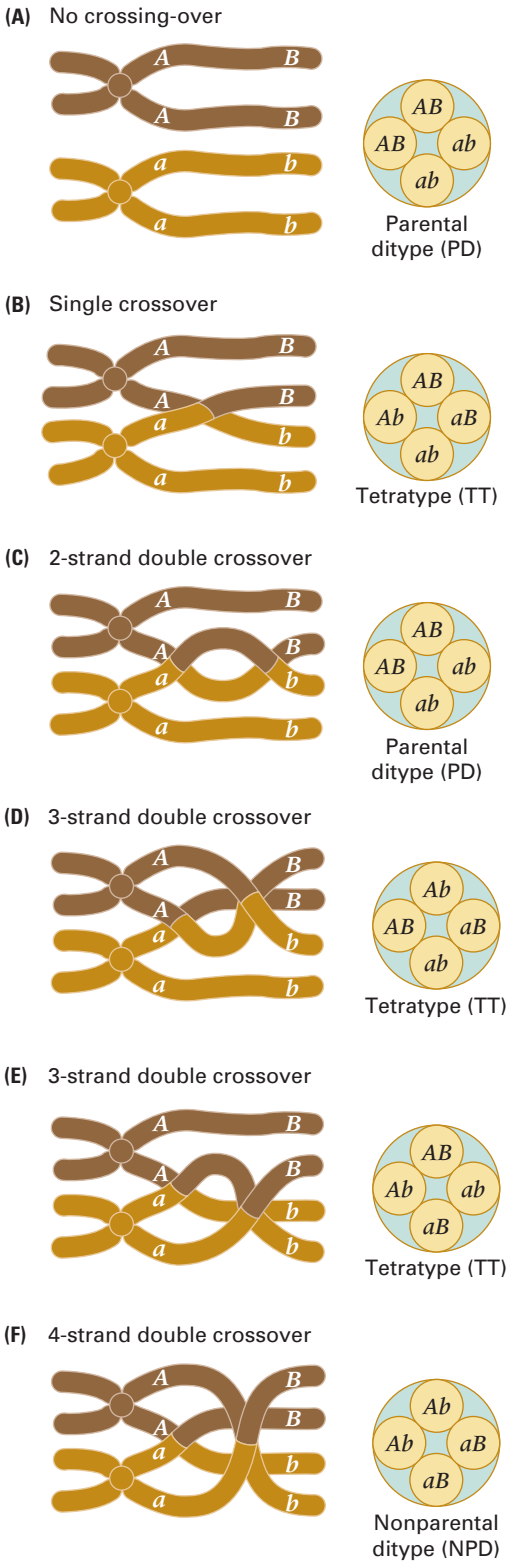


FIGURE 4.27 Types of tetrads produced with two linked genes. (A) In the absence of a crossover, a PD tetrad is produced. (B) With a single crossover between the genes, a TT tetrad is produced. (C–F) Among the four possible types of double crossovers between the genes, only the four-strand double crossover in part F yields an NPD tetrad.

the frequency of recombination equals the number of spores that are recombinant for the genetic markers divided by the total number of spores.

The geometry of meiosis is revealed in ordered tetrads.

In the bread mold *Neurospora crassa*, the products of meiosis are contained in an *ordered* array of ascospores (**FIGURE 4.28**). A zygote nucleus, contained in a sac-like ascus, undergoes meiosis almost immediately after it is formed. The four nuclei produced by meiosis are in a linear, ordered sequence in the ascus, and each of them undergoes a mitotic division to form two genetically identical and adjacent ascospores. Each mature ascus contains eight ascospores arranged in four pairs, each pair derived from one of the products of meiosis. The ascospores can be removed one by one from an ascus and each germinated in a culture tube to determine its genotype.

Ordered asci also can be classified as PD, NPD, or TT with respect to two pairs of alleles, which makes it possible to assess the degree of linkage between the genes. The fact that the arrangement of meiotic products is ordered also makes it possible to determine the recombination frequency between any particular gene and its centromere. The logic of the mapping technique is based on the feature of meiosis shown in **FIGURE 4.29**.

KEY CONCEPT

Homologous centromeres of parental chromosomes separate at the first meiotic division; the centromeres of sister chromatids separate at the second meiotic division.

Thus, in the absence of crossing-over between a gene and its centromere, the alleles of the gene (for example, *A* and *a*) must separate in the first meiotic division; this separation is called **first-division segregation**. If, instead, a crossover is formed between the gene and its centromere, the *A* and *a* alleles do not become separated until the second meiotic division; this separation is called **second-division segregation**. The distinction between first-division and second-division segregation is shown in Figure 4.29. As shown in part A, only two possible arrangements of the products of meiosis can yield first-division segregation—*A A a a* or *a a A A*. However, four patterns of second-division segregation are possible because of the random arrangement of homologous chromosomes at metaphase I and of the chromatids at metaphase II. These four arrangements, which are shown in part B, are

A a A a, *a A a A*, *A a a A*, and *a A A a*

The percentage of asci with second-division segregation patterns for a gene can be used to map the gene with respect to its centromere. For example, let us assume that 30 percent of a sample of asci from a cross have a

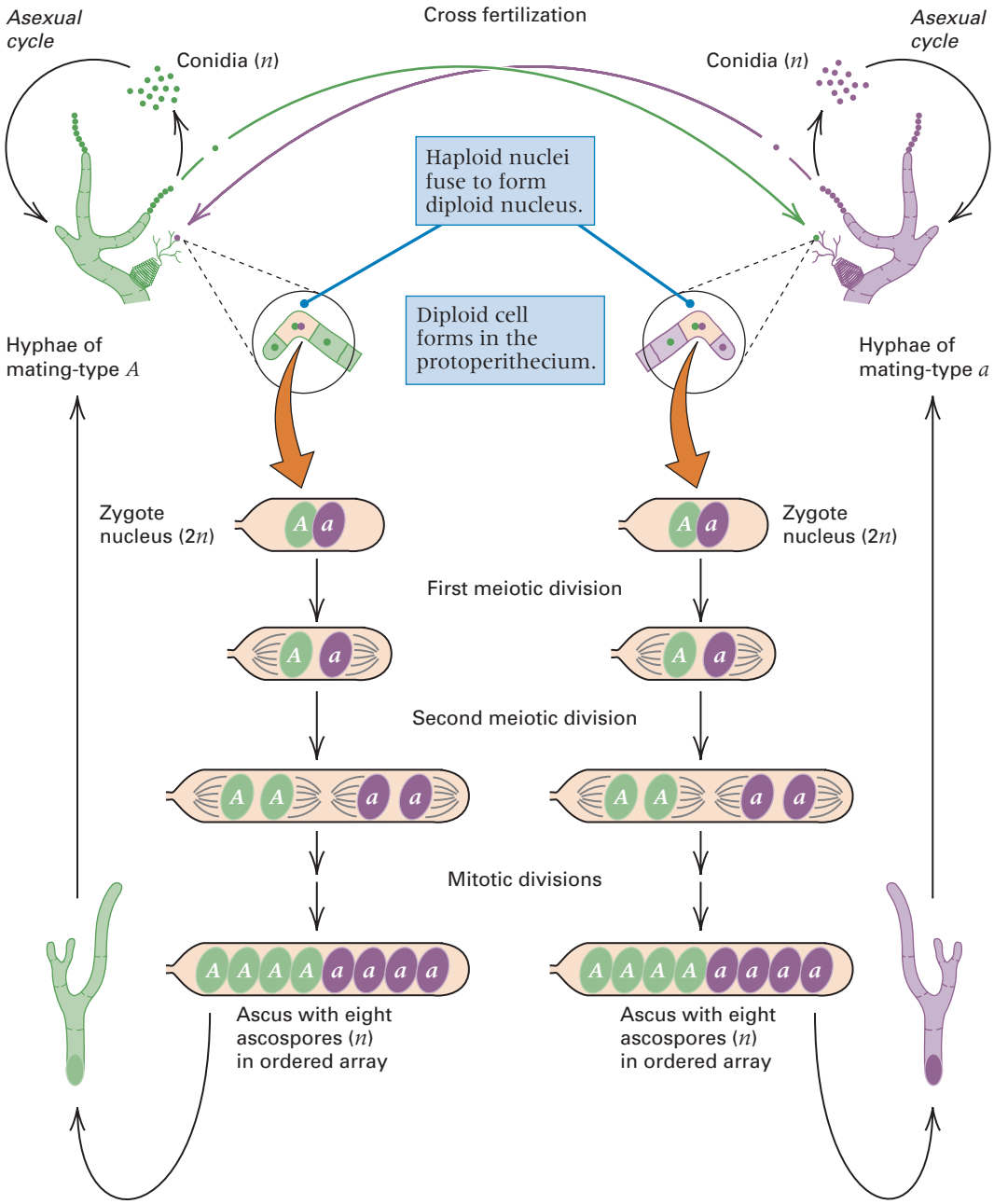


FIGURE 4.28 The life cycle of *Neurospora crassa*. The vegetative body consists of partly segmented filaments called hyphae. Conidia are asexual spores that function in the fertilization of organisms of the opposite mating type. A protoperithecium develops into a structure in which numerous cells undergo meiosis.

second-division segregation pattern for the *A* and *a* alleles. This means that 30 percent of the cells undergoing meiosis had a crossover between the *A* gene and its centromere. Because the map distance between two genes is, by definition, equal to one-half times the proportion of cells with a crossover between the genes, the map distance between a gene and its centromere is given by the equation

$$\text{Map distance} = \frac{1}{2} \times \frac{\text{Number of asci with second division segregation}}{\text{Total number of asci}} \times 100 \qquad (4.2)$$

Equation (4.2) is valid as long as the gene is close enough to the centromere for us to neglect multiple crossovers. Reliable linkage values are best determined for genes that are near the centromere. The location of more distant genes is then accomplished by mapping these genes relative to genes nearer the centromere.

If a gene is far from its centromere, crossing-over between the gene and its centromere will be so frequent that the *A* and *a* alleles become randomized with respect to the four chromatids. The result is that the six possible spore arrangements shown in Figure 4.29 are

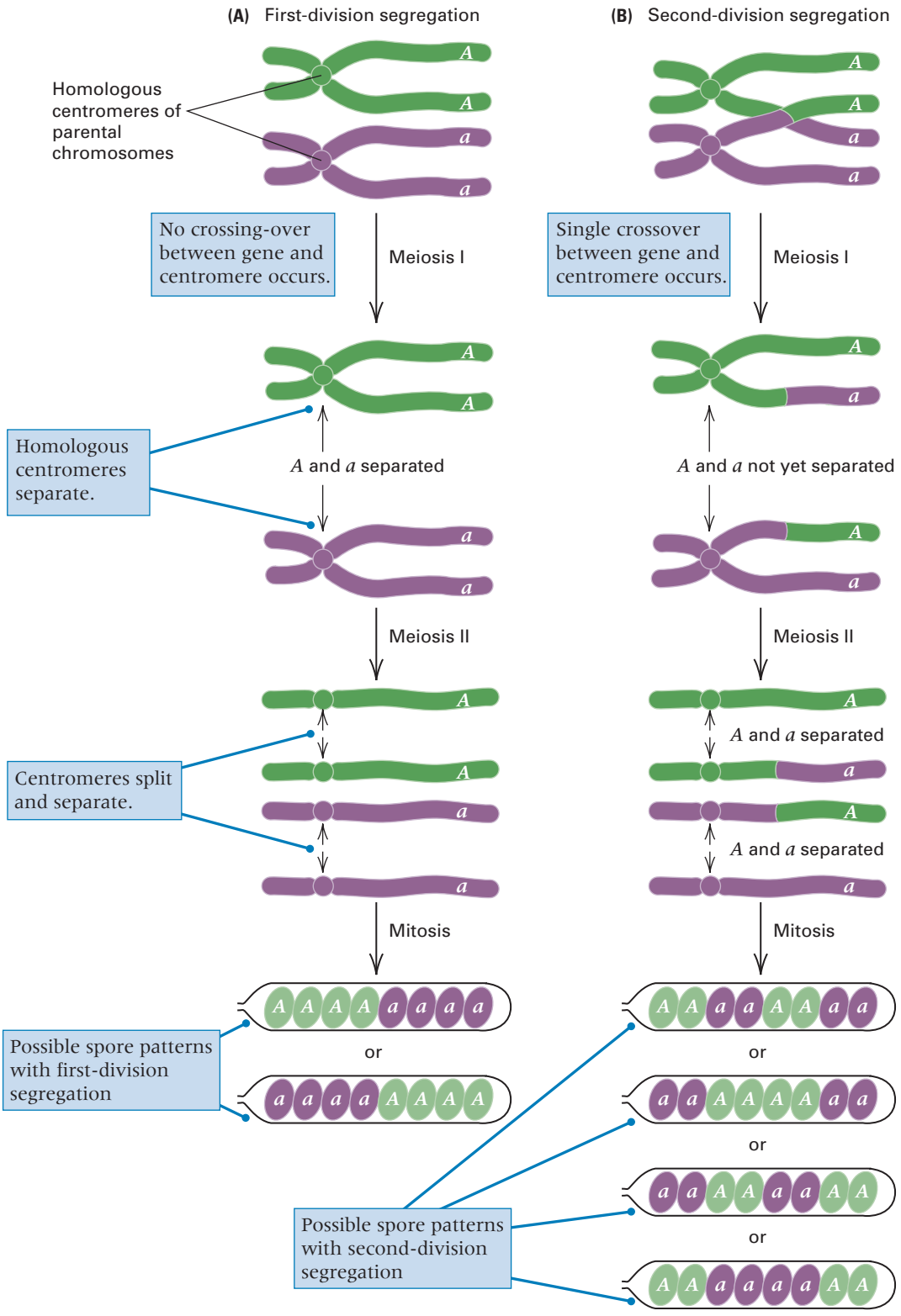


FIGURE 4.29 First- and second-division segregation in *Neurospora*. (A) First-division segregation patterns are found in the ascus when a crossover between the gene and centromere does not take place. The alleles separate (segregate) in meiosis I. Two spore patterns are possible, depending on the orientation of the pair of chromosomes on the first-division spindle. The orientation shown results in the pattern in the upper ascus. (B) Second-division segregation patterns are found in the ascus when a crossover between the gene and the centromere delays separation of *A* from *a* until meiosis II. Four patterns of spores are possible, depending on the orientation of the pair of chromosomes on the first-division spindle and that of the chromatids of each chromosome on the second-division spindle. The orientation shown results in the pattern in the top ascus.

all equally frequent. Therefore, when the chromatids participating in each cross over are chosen at random,

KEY CONCEPT

The maximum frequency of second-division segregation asci is 2/3.

Gene conversion suggests a molecular mechanism of recombination.

Genetic recombination may be regarded as a process of breakage and repair between two DNA molecules. In eukaryotes, the process takes place early in meiosis after each molecule has replicated, and with respect to genetic markers, it results in two molecules of the parental type and two recombinants. For genetic studies of recombination, fungi such as yeast or *Neurospora* are particularly useful, because all four products of any meiosis can be recovered in a four-spore (yeast) or eight-spore (*Neurospora*) ascus. As we have noted, most asci from heterozygous *Aa* diploids contain ratios of

2 *A* : 2 *a* in four-spored asci, or
4 *A* : 4 *a* in eight-spored asci

demonstrating normal Mendelian segregation. Occasionally, however, aberrant ratios are also found, such as

3 *A* : 1 *a* or 1 *A* : 3 *a* in four-spored asci, and
5 *A* : 3 *a* or 3 *A* : 5 *a* in eight-spored asci

Different types of aberrant ratios can also occur. The aberrant asci are said to result from **gene conversion** because it appears as if one allele has "converted" the other allele into a form like itself. Gene conversion is frequently accompanied by recombination between genetic markers on either side of the conversion event, even when the flanking markers are tightly linked. This

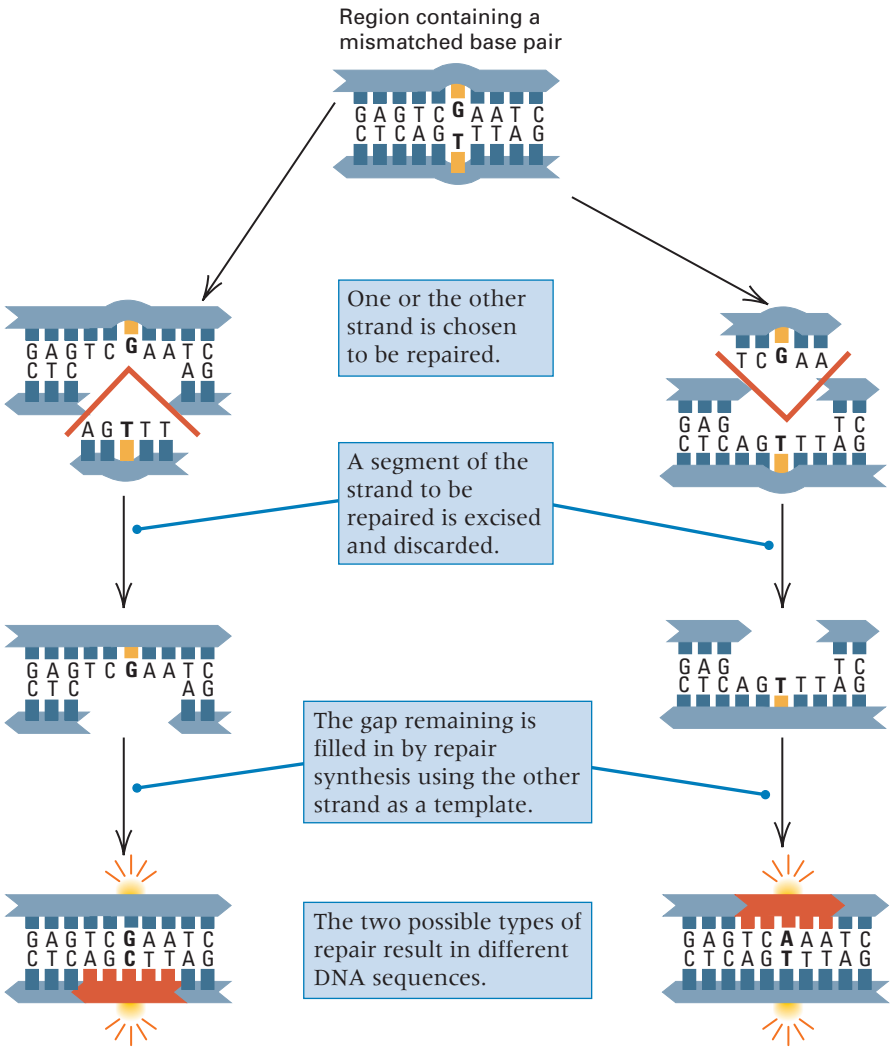


FIGURE 4.30 Mismatch repair consists of the excision of a segment of a DNA strand containing a base mismatch followed by repair synthesis. Either strand can be excised and corrected. In this example, the G–T mismatch is corrected to either G–C (left) or A–T (right).

implies that gene conversion can be one consequence of the recombination process.

Gene conversion results from a normal DNA repair process in the cell known as **mismatch repair**. In this process, an enzyme recognizes any base pair in a DNA duplex in which the paired bases are mismatched—for example, G paired with T, or A paired with C. When such a mismatch is found in a molecule of duplex DNA, a small segment of one strand is excised and replaced with a new segment synthesized using the remaining strand as a template. In this manner the mismatched base pair is replaced. **FIGURE 4.30** shows an example in which a mismatched G−T pair is being repaired. The strand that is excised could be either the strand containing T or the one containing G, and the newly synthesized (repaired) segment, shown in red, would contain either a C or an A, respectively. The two possible products of repair differ in DNA sequence.

The role of mismatch repair in gene conversion is illustrated in **FIGURE 4.31**. The pair of DNA duplexes across the top represents the DNA molecules of two alleles in a cell undergoing meiosis. One duplex contains a G−C base pair highlighted in color; this corresponds to the *A* allele. The other duplex contains an A−T base pair at the same position, which corresponds to the *a* allele. In the process of recombination, the participating DNA duplexes can exchange pairing partners. The result is shown in the second row. The exchange of pairing partners creates a **heteroduplex** region in which any bases that are not identical in the parental duplexes become mismatched. In this example, one heteroduplex contains a G−T base pair and the other an A−C base pair. At this point, the mismatch repair system comes into play and corrects the mismatches. Each mismatch can be repaired in either of two ways, so there are four possible ways in which the mismatches can be repaired. One type of repair results in gene conversion of *a* to *A*, another results in gene conversion of *A* to *a*, and the remaining two restore the sequences of the original duplexes and so do not result in gene conversion.

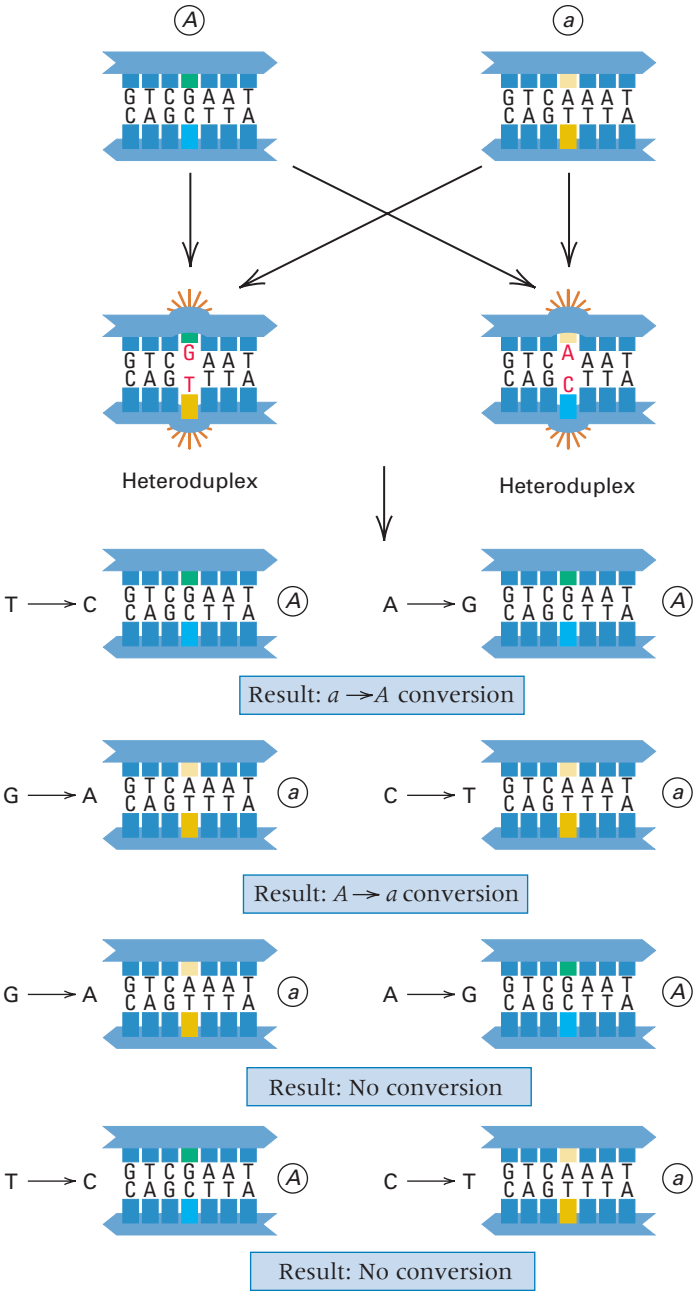


FIGURE 4.31 Mismatch repair resulting in gene conversion. Only a small part of the heteroduplex region is shown.

4.6 Recombination is initiated by a double-stranded break in DNA.

In prophase I of meiosis, chiasmata are the physical manifestations of crossing-over between DNA molecules. These structures bridge between pairs of sister chromatids in a bivalent and are important in the proper alignment of the bivalent at the metaphase plate in preparation for anaphase I. Bivalents that lack chiasmata to help hold them together are prone to undergo nondisjunction.

The crossovers needed for the chiasmata to form are initiated by programmed double-stranded breaks in DNA (**FIGURE 4.32**, part A). In a double-stranded break, the size of the gap is usuallly increased by nuclease digestion of the broken ends, with greater degradation of the 5′ ends leaving overhanging 3′ ends as shown in the illustration. These gaps are repaired using the unbroken homologous DNA molecule as a template, but in meiosis the repair process can result in crossovers that yield chiasmata between nonsister chromatids. These crossovers are also the physical basis of what is observed genetically as recombination.

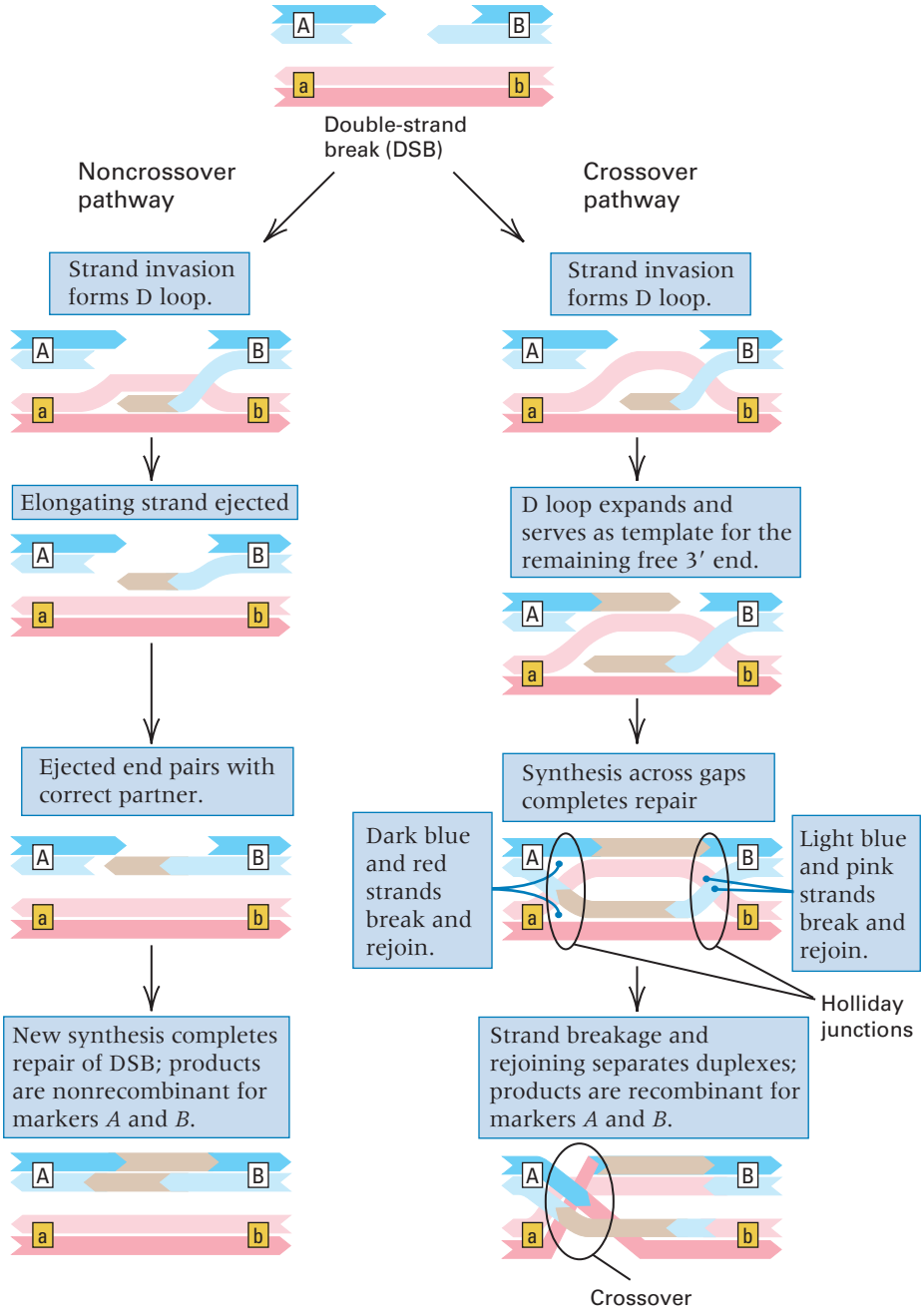


FIGURE 4.32 (A) Double-strand break in a duplex DNA molecule with overhanging 3′ ends facing the gap. Repair of the break makes use of the nonbroken homologous duplex. (B) Repair pathway that does not result in crossing-over, although the heteroduplex regions can undergo gene conversion. (C) Repair pathway that does result in crossing-over, also with possible gene conversion in heteroduplex regions. [Adapted from D. K. Bishop and D. Zickler, *Cell* 117 (2004): 9–15.]

A double-stranded break does not necessarily result in a crossover, however. Repair of the double-stranded break by the noncrossover pathway is illustrated in Figure 4.32, part B. The first step in repair is that a broken 3′ end invades the homologous unbroken DNA duplex, forming a short heteroduplex region with one strand and a looped-out region of the other strand called a **D loop.** (Specific proteins are required to mediate strand invasion; in *E. coli* the strand-invasion protein is known as RecA.) In the illustration, the heteroduplex region is the region where the light blue strand is paired with the red strand. Because it is a heteroduplex, any base-pair mismatches in this region could be corrected by mismatch repair in such a way as to result in gene conversion. Such heteroduplex regions are typically only a few hundred base pairs in length. They are much shorter than a gene and vastly shorter than a chromosome, and so gene conversions are rare events except for short regions very near the site of a double-stranded break.

At one end of the heteroduplex, the free 3′ end of the broken DNA strand is extended (brown), but after a time it is ejected from the template, and the strands of the unbroken duplex are able to come together again. At this point, the extension of the 3′ end is long enough that pairing can take place with the complementary strand in the broken duplex. At the same time, this pairing provides a template for the 3′ end of the other broken strand. Extension of the 3′ ends across the remaining gaps completes the repair of the double-stranded break. Note that although gene conversion can occur in the noncrossover pathway, the resulting duplex DNA molecules are nonrecombinant.

The crossover pathway for repairing a double-stranded break is illustrated in Figure 4.32, part C. Again invasion of the unbroken duplex forms a D loop and a short heteroduplex region in which gene conversion can occur. As in the noncrossover pathway, the free 3′ end of the broken DNA strand is extended (brown), but in this case it continues until it displaces the partner strand (pink) of the template strand (red). The displaced strand can then serve as a template for the elongation of the 3′ end of the other broken strand. Eventually, the extensions of the broken strands become long enough that they can be attached to the broken 5′ ends. This completes the repair of the double-stranded break, but note that the resulting structure includes two places where the strands have exchanged pairing partners. Each of the structures where pairing partners are switched is called a **Holliday junction**, named after Robin Holliday, who first predicted that such structures would be involved in recombination.

The problem with Holliday junctions is that they are places where DNA strands from different duplex molecules are interconnected. How the strands are interconnected is shown for the DNA double helices in **FIGURE 4.33**, part A. Resolution of the Holliday junctions is necessary for the DNA molecules to become free of one another. This requires breakage and rejoining of one pair of DNA strands at each Holliday junction. The breakage and rejoining is an enzymatic function carried out by an enzyme called the **Holliday junction-resolving enzyme**.

Parts B and C in Figure 4.33 show two ways in which the Holliday structures can be resolved. Breakage and rejoining of the strands indicated by the red arrows results in a crossover at the site of the left-hand Holliday junction, whereas breakage and rejoining of the strands indicated by the blue arrows results in a crossover at the site of the right-hand Holliday junction. In both cases, the resulting DNA molecules have

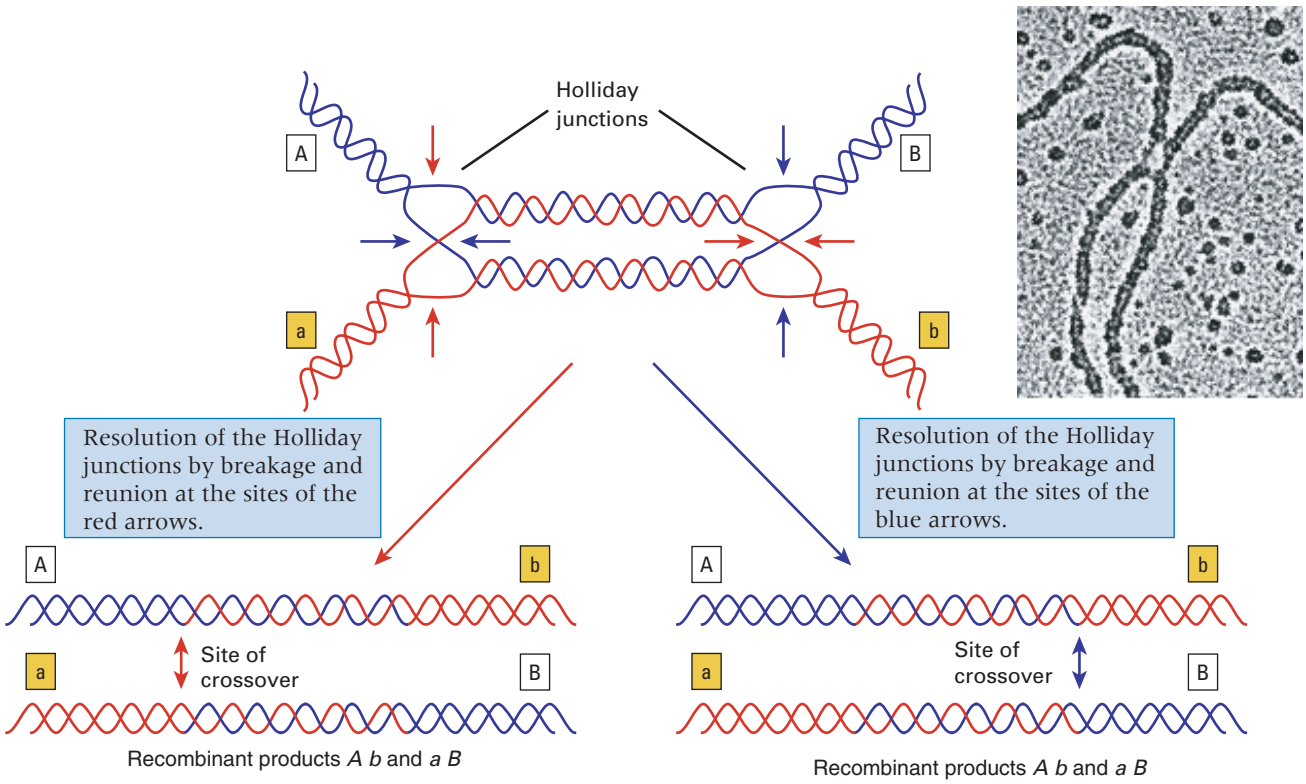


FIGURE 4.33 (A) Two Holliday junctions in a pair of DNA molecules undergoing recombination (the parental chromosomes are AB and ab); (B and C) two modes of resolution depending on which strands are broken and rejoined. Part D is an electron micrograph showing a single Holliday junction between a pair of DNA molecules. [Illustration modified from B. Alberts. *Essential Cell Biology*. Garland Science, 1997. Illustration reproduced with permission of Huntington Potter, Johnnie B. Byrd, Sr., Alzheimer's Center & Research Institute.]

a crossover that yields reciprocal recombinant *A b* and *a B* products. (In principle, resolution could also take place at the red arrows in one Holliday junction and the blue arrows in the other, but these resolutions result in noncrossover products. It is unclear how often these noncrossover types of resolution take place.)

Recombination tends to take place at preferred positions in the genome.

In some organisms, including humans and other mammals, programmed double-stranded DNA breaks are much more likely to occur at certain positions in the genome than others. Crossovers resulting in recombination are much more likely to occur at these positions, which are referred to as **hotspots of recombination**. The human genome contains about 30 thousand hotspots of recombination, spaced an average of 100 kb apart. These hotspots are often located in the spaces between genes, and they differ greatly from one to the next in the likelihood of a double-stranded break. One particular protein has been implicated in about 40 percent of hotspots across the human genome. The protein is known as PRDM9, and it is known to bind with double-stranded DNA at sites that match or nearly match the 13-mer sequence

5′-CCGCCGTMWCCWC-3′

where M means A or C, and W means T or A. The protein does not cause the double-stranded break directly. PRDM9 is actually a methyl transferase that attaches methyl ($-CH_3$) groups to the amino acid lysine of an abundant protein called histone H3 that, together with other histone proteins, binds with DNA to form a chain of small DNA-protein beads known as *nucleosomes*. DNA is usually packaged in the form of nucleosomes containing histone H3, and hence wherever the target 13-mer of PRDM9 occurs, nucleosomes containing histone H3 are also present. Evidently, methylation of histone H3 by PRDM9 predispposes the DNA to undergo a double-stranded break, but the molecular details of break formation are still unknown. More than 30 alleles of the gene encoding PRDM9 have been identified, which differ in their propensity to recognize the 13-mer target and to methylate histone H3. Variation in the alleles encoding PRDM9 accounts for the differing efficiency of PRDM9-associated hotspots.

Although recombination tends to be initiated by means of double-stranded breaks at hotspots of recombination, it should be emphasized that there are so many hotspots relative to number of crossover events that the sites of recombination in any particular meiotic cell show a great deal of randomness. In the human genome, for example, there are about 30 thousand hotspots of recombination and about 60 crossover events per meiosis. Roughly speaking, each crossover could take place at any of about 500 hotspots in the vicinity. The relatively small number of crossovers per meiosis means that crossovers occur essentially at random sites chosen from among the very large number of hotspots that occur across the genome.

CHAPTER SUMMARY

- Genes that are located in the same chromosome and that do not show independent assortment are said to be linked.
- The alleles of linked genes present together in the same chromosome tend to be inherited as a group.
- Crossing-over between homologous chromosomes results in recombination, which breaks up combinations of linked alleles.
- The frequency of recombination serves as a measure of distance between linked genes along a chromosome, providing a genetic map of the relative positions of the genes.
- The map distance between genes in a genetic map is related to the frequency of crossing-over between the genes in meiosis.
- Physical distance along a chromosome is often, but not always, correlated with map distance.
- Variations in DNA sequence among individuals (polymorphisms) serve as genetic markers along the genome that are used for genetic mapping, tracing the genetic ancestry of individuals, and many other purposes.
- Among the most useful types of DNA polymorphisms are single-nucleotide polymorphisms (SNPs), restriction fragment length polymorphisms (RFLPs), simple-sequence repeats (SSRs), and copy-number polymorphisms (CNPs).
- Tetrads are sensitive indicators of linkage because they include all the products of meiosis.
- At the DNA level, recombination is initiated by a double-stranded break in a DNA molecule. Use of the homologous DNA molecule as a template for repair can result in a crossover, in which both strands of the participating DNA molecules are broken and rejoined.

LEARNING OUTCOMES

- Given a genetic map with two genes, predict the kinds and relative frequencies of gametes that would be produced by an individual of a specified genotype.
- Analyze the results of a genetic cross with three linked genes to deduce the genotypes of the parents, the order of the genes along the chromosome, the map distances between the genes, and the degree of interference between crossovers.
- Distinguish between a single-nucleotide polymorphism (SNP), a restriction fragment length polymorphism (RFLP), and copy-number variation (CNV). Describe one experimental method by which each type of polymorphism can be detected.
- Explain how linkage is detected between two genes in an organism with unordered tetrads and estimate the distance in the genetic map between the genes.
- Explain how linkage between a gene and its centromere is detected in an organism with ordered tetrads and estimate the map distance between the gene and its centromere.

ISSUES AND IDEAS

- Distinguish between genetic recombination and genetic complementation. Is it possible for two mutant genes to show complementation but not recombination? Is it possible for two mutant genes to show recombination but not complementation?
- In genetic analysis, why is it important to know the position of a gene along a chromosome?
- What is the maximum frequency of recombination between two genes? Is there a maximum map distance between two genes?
- Why is the frequency of recombination over a long interval of a chromosome always smaller than the map distance over the same interval?
- What is meant by the term *chromosome interference*?
- In human genetics, why are molecular variations in DNA sequence, rather than phenotypes such as eye color or blood-group differences, used for genetic analysis?
- In genetic analysis, what is so special about the ability to examine tetrads in certain fungi?
- Explain how tetratype tetrads demonstrate that recombination takes place at the four-strand stage of meiosis and is reciprocal.
- Explain why the observation PD >> NPD with respect to tetrads is a sensitive indicator of linkage.

SOLUTIONS: STEP BY STEP

Problem 1

In *Drosophila*, the recessive mutant allele *spineless* (*ss*) results in thin bristles, *cinnabar* (*cn*) results in bright red eyes, and *ebony* (*e*) results in a black body color. A cross is carried out between females of genotype *ss cn e* / + + + and males of genotype *ss cn e* / *ss cn e*. In this type of symbolism, each + denotes the nonmutant allele of the gene written at the corresponding position. From this cross, the following 1000 progeny were obtained:

ss cn e / *ss cn e*	241
ss + *e* / *ss cn e*	223
+ *cn* + / *ss cn e*	202
+ + + / *ss cn e*	212
ss cn + / *ss cn e*	25
ss + + / *ss cn e*	31
+ *cn e* / *ss cn e*	31
+ + *e* / *ss cn e*	35

Determine which, if any, of the genes are linked, and for those that are linked, estimate the frequency of recombination between the genes.

Solution. To determine which, if any, genes are linked, consider the mutants in pairs, and sum the data to find the total number of parental types and recombinant types for each pair. For *ss* and *cn*, the parental types are 241 + 212 + 25 + 35 = 513 and the recombinant types are 223 + 202 + 31 + 31 = 487. These numbers are close enough to 500 : 500 that one may infer that ss and *cn* are unlinked. Similarly, for *cn* and *e*, the parental types sum to 515 and the recombinant types to 485, hence *cn* and *e* are unlinked. For *ss* and *e*, however, the parental types sum to 878 and the recombinant types to 122, and this result implies that *ss* and *e* are linked. The estimated frequency of recombination is 122/1000 = 0.122 or 12.2 percent.

Problem 2

In *Drosophila*, the recessive mutant allele *cinnabar (cn)* results in bright red eyes, *curved* (*c*) results in curved wings, and *plexus* (*px*) results in extra wing veins. All three genes are linked. In a cross between *cn c px* / + + + females and *cn c px* / *cn c px* males, the following progeny were counted:

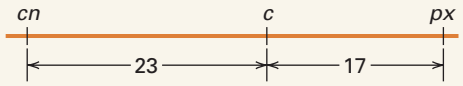
cn c px / *cn c px*	296
cn c + / *cn c px*	63
cn + + / *cn c px*	119
cn + *px* / *cn c px*	10
+ *c px* / *cn c px*	86
+ *c* + / *cn c px*	15
+ + + / *cn c px*	329
+ + *px* / *cn c px*	82
Total	1000

(a) What is the frequency of recombination between *cn* and *c*?
(b) What is the frequency of recombination between *c* and *px*?
(c) What is the frequency of recombination between *cn* and *px*?
(d) Why is the frequency of recombination between *cn* and *px* smaller than the sum of that between *cn* and *c* and that between *c* and *px*?
(e) What is the coefficient of coincidence across this region? What is the value of the interference?
(f) Draw a genetic map of the region, showing the locations of *cn*, *c*, and *px* and the map distances between the genes.

Solution. Do not try to hurry through linkage problems! You will be rewarded by taking time to organize the information in the optimal manner. First, group the progeny types into reciprocal pairs—*cn c px* with + + +, *cn c* + with + + *px*, and so forth—and make a new list organized as shown here. (Ignore the *cn c px* chromosome from the father because it contributes no information about recombination.)

cn c px	296	625
+ + +	329	
cn c +	63	145
+ + *px*	82	
cn + +	119	205
+ *c px*	86	
cn + *px*	10	25
+ *c* +	15	
Total		1000

In this tabulation, a space has been inserted between the pairs of reciprocal products in order to keep the groups separate. The number next to each brace is the total number of chromosomes in the group. The most numerous group of reciprocal chromosomes (in this case, *cn c px* and + + +) consists of the nonrecombinants, and the least numerous group of reciprocal chromosomes (in this case, *cn* + *px* and + *c* +) consists of the double recombinants. Rearrange the order of the groups, if necessary, so that the nonrecombinants are at the top of the list and the double recombinants are at the bottom. (In the present example, rearrangement is not necessary.) At this point, also make sure that the order of the genes is correct as given, by comparing the genotypes of the double recombinants with those of the nonrecombinants. If the gene order is correct, then it will require two recombination events (one in each interval) to derive the double-recombinant chromosomes from the nonrecombinants. If this is not the case, rearrange the order of the genes. (The "odd man out" in comparing the double recombinants with the nonrecombinants is always the gene in the middle.) In this particular example, the gene order is correct as given. Finally, with this preliminary bookkeeping done, we can proceed to tackle the questions. **(a)** The frequency of recombination between *cn* and *c* is given by the totals of all classes of progeny showing recombination in the *cn* − *c* interval, in this case (205 + 25)/1000 = 0.23. **(b)** The frequency of recombination between *c* and *px* equals (145 + 25)/1000 = 0.17. **(c)** The frequency of recombination between *cn* and *px* equals (145 + 205)/1000 = 0.35. (Note that the double recombinants are not included in this total, because the double recombinants are not recombined for *cn* and *px*; their allele combinations for *cn* and *px* are the same as in the parents.) **(d)** The frequency of recombination between *cn* and *px* (0.35) is smaller than the sum of that between *cn* and *c* and that between *c* and *px* (0.23 + 0.17 = 0.40) because of double crossovers. **(e)** The coefficient of coincidence equals the observed number of double recombinants divided by the expected number. The observed number is 25 and the expected number is 0.23 × 0.17 × 1000 = 39.1; the coefficient of coincidence therefore equals 25/39.1 = 0.64. The interference equals 1 − coefficient of coincidence, and so the interference equals 1 − 0.64 = 0.36. **(f)** The genetic map is shown in the accompanying diagram. The distances are in map units (centimorgans). However, the map distances of 23 and 17 map units are based on the 23 percent and 17 percent recombination observed between *cn* and *c* and between *c* and *px*, respectively; the actual distances in map units are probably a little greater than these estimates because of a small amount of double recombination within each of the intervals.



Problem 3

In *Neurospora*, the gene *arg12* encodes the enzyme needed to convert ornithine to citrulline in the pathway of arginine biosynthesis. The gene was discovered in the experiments of Beadle and Tatum discussed in Chapter 1. In a cross between *arg12* and *ARG12* strains, where *arg12* denotes the mutant allele and *ARG12* the nonmutant allele, two-thirds of the resulting asci show second-division segregation. What does this observation imply about the map distance between *arg12* and the centromere?

Solution. Two-thirds is the maximum proportion of second-division segregation that can occur. This value is observed for any gene that is so far from the centromere that one or more crossovers are almost certain to take place between the gene and the centromere. Hence, we can deduce that *arg12* is at least 50 map units from the centromere. This map distance is a minimum, and the true map distance could be greater.

CONCEPTS IN ACTION: PROBLEMS FOR SOLUTION

4.1 A double heterozygote has the repulsion configuration *A b* / *a B* of two linked genes that have a frequency of recombination of 0.20. If a randomly chosen gamete carries *A*, what is the probability that it also carries *B*?

4.2 What gametes, and in what frequencies, are produced by a female *Drosophila* of genotype *A B* / *a b* when the genes are present in the same chromosome and the frequency of recombination between the genes is 8 percent? What gametes, and in what frequencies, are produced by a male of the same genotype?

4.3 A cell undergoing meiosis in an organism with unordered tetrads undergoes a double crossover between two markers. If the ratio of 2-strand : 3-strand : 4-strand doubles is 1 : 2 : 1, what is the ratio of PD : TT : NPD tetrads? (PD stands for parental ditype tetrad, TT for tetratype, and NPD for nonparental ditype.)

4.4 A coefficient of coincidence of 0.36 implies which one or more of the following statements are true:

(a) The frequency of double crossovers was 36 percent.
(b) The frequency of double crossovers was 36 percent of the number that would be expected if there were no interference.
(c) There were 0.36 times as many single crossovers as double crossovers.
(d) There were 0.36 times as many single crossovers in one region as there were in an adjacent region.
(e) There were 0.36 times as many parental as recombinant progeny.

4.5 A gene in *Neurospora*, a fungus with ordered tetrads, shows 10% second-division segregation. What is the map distance between the gene and the centromere?

4.6 In *Drosophila pseudoobscura*, the eye-color mutation *purple (pr)* and the wing mutation *crossveinless (cv)* are located in chromosome 3 at a distance of 18 map units. What phenotypes, and in what proportions, would you expect in the progeny from the mating of pr^+ cv^+ / *pr cv* females with *pr cv* / *pr cv* males?

4.7 Construct a genetic map of a chromosome from the following recombination frequencies between individual pairs of genes: *r*–*c*, 10; *c*–*p*, 12; *p*–*r*, 3; *s*–*c*, 16; *s*–*r*, 8. You will discover that the distances are not strictly additive. Why aren't they?

4.8 A *Drosophila* cross is carried out with a female that is heterozygous for both the *y* (yellow body) and *bb* (bobbed bristles) mutations. Both genes are located in the X chromosome. Among 200 male progeny, there were 49 wildtype for both traits, 51 with yellow body, 41 with bobbed bristles, and 59 mutant for both genes. Do these genes show evidence for linkage? [Note: The appropriate chi-square test is a test for a 1 : 1 ratio of parental : recombinant gametes.]

4.9 Two genes in chromosome 7 of corn are identified by the recessive alleles *gl* (glossy), determining glossy leaves, and *ra* (ramosa), determining branching of ears. When a plant heterozygous for each of these alleles was crossed with a homozygous recessive plant, the progeny consisted of the following genotypes with the numbers of each indicated:

Gl ra/gl ra	98	*gl Ra/gl ra*	91
Gl Ra/gl ra	7	*gl ra/gl ra*	4

Calculate the frequency of recombination between these genes.

4.10 In the yellow-fever mosquito, *Aedes aegypti*, a dominant gene *DDT* for DDT resistance (DDT is dichlorodiphenyltrichloroethane, a long-lasting insecticide) and a dominant gene *Dl* for Dieldrin resistance (Dieldrin is another long-lasting insecticide) are known to be in the same chromosome. A cross was carried out between a DDT-resistant strain and a Dieldrin-resistant strain, and female progeny resistant to both insecticides were testcrossed with wildtype males. Among the progeny, 99 were resistant to both insecticides, 88 were resistant to DDT only, 89 were resistant to Dieldrin only, and 106 were sensitive to both insecticides.

(a) Are *DDT* and *Dl* alleles of the same gene? How can you tell?
(b) Are *DDT* and *Dl* linked?
(c) What can you deduce about the genetic positions of *DDT* and *Dl* along the chromosome?

4.11 Two pure-breeding strains of mice are crossed to produce F_1 mice that are heterozygous for three linked genes with alleles *Aa*, *Bb*, and *Dd*. Numerous triply heterozygous F_1 mice are testcrossed, and the genotypes and numbers of the resulting progeny are as follows:

A–	*B*–	*D*–	10
A–	*B*–	*dd*	350
A–	*bb*	*D*–	100
A–	*bb*	*dd*	40
aa	*B*–	*D*–	60
aa	*B*–	*dd*	120
aa	*bb*	*D*–	320
Total			1000

(a) Which gene is in the middle?
(b) Specify the genotype of the F_1 triple heterozygote as completely as possible, with the genes in the correct order and the correct alleles on each chromosome.
(c) Which two genes are closest together?
(d) What is the map distance between the two closest genes? Assume that interference is complete between these two genes.
(e) If interference is not complete, how will the true map distance differ from the value you calculated in part **(d)**? Briefly explain why this is so.

4.12 In corn, the genes *v* (virescent seedlings), *pr* (red aleurone), and *bm* (brown midrib) are all on chromosome 5, but not necessarily in the order given. The cross

$$v^+\ pr\ bm/v\ pr^+\ bm^+ \times v\ pr\ bm/v\ pr\ bm$$

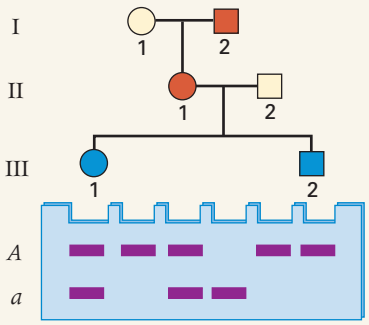
produces 1000 progeny with the following phenotypes:

v^+	pr	bm	226
v	pr^+	bm^+	229
v^+	pr	bm^+	153
v	pr^+	bm	185
v^+	pr^+	bm	59
v	pr	bm^+	71
v^+	pr^+	bm^+	36
v	pr	bm	41

(a) Determine the gene order, the recombination frequencies between adjacent genes, the coefficient of coincidence, and the interference.
(b) Explain why, in this example, the recombination frequencies are not good estimates of map distance.

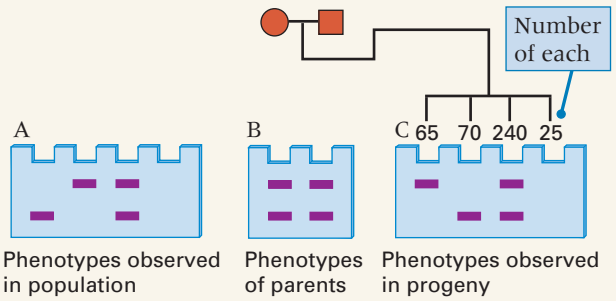
4.13 The male I-2 in the accompanying pedigree is affected with Huntington disease, a type of neuromuscular degeneration caused by a rare autosomal dominant mutation *HD* with complete penetrance. The wildtype, nonmutant allele is denoted *hd*. The woman II-1 is also affected. The RFLP alleles *A* and *a* yielding the bands in the gel are linked to the Huntington locus with a recombination frequency of 10 percent.

(a) Is the genotype of II-1 *HD A/hd a* or is it *HD a/hd A*?
(b) Given the pattern of bands in the gel, what is the probability that III-1 will be affected?
(c) Given the pattern of bands in the gel, what is the probability that III-2 will be affected?



4.14 A human geneticist discovers the molecular variation in DNA sequence illustrated in the accompanying diagrams of electrophoresis gels. In the human population as a whole, she finds any of four phenotypes, shown in panel A. She believes that this may be a simple genetic polymorphism with three alleles, like the ABO blood groups. There are two alleles that yield DNA fragments of different sizes, fast (*F*) or slow (*S*) migration, and a "null" allele (*O*) in which the DNA fragment is deleted. The genotypes in panel A would therefore be, from left to right, *FF* or *FO*, *SS* or *SO*, *FS*, and *OO*. In the population as a whole, the putative *OO* genotype is extremely common, and the *FS* genotype is quite rare. To investigate this hypothesis further, the geneticist studies offspring of matings between parents who have the putative FS genotype (panel B). The types of progeny, and their numbers, are shown in panel C.

(a) What result would be expected from the three-allele hypothesis?
(b) Are the observed data consistent with this result? Why or why not?
(c) Suggest a genetic hypothesis that can explain the data in panel C.
(d) Are the data consistent with your hypothesis?



4.15 The following classes and frequencies of ordered tetrads were obtained from the cross $a^+\ b^+ \times a\ b$ in *Neurospora*. (Only one member of each pair of spores is shown.) What is the order of the genes in relation to the centromere?

Spore pair				Number of asci
1-2	3-4	5-6	7-8	
$a^+\ b^+$	$a^+\ b^+$	$a\ b$	$a\ b$	1766
$a^+\ b^+$	$a\ b$	$a^+\ b^+$	$a\ b$	220
$a^+\ b^+$	$a\ b^+$	$a^+\ b$	$a\ b$	14

4.16 A portion of the linkage map of chromosome 2 in the tomato is illustrated here. The oblate phenotype is a flattened fruit, the peach phenotype is hairy fruit (like a peach), and compound influorescence means clustered flowers.

Among 1000 gametes produced by a plant of genotype *o ci* +/+ + *p*, what types of gametes would be expected, and what number would be expected of each? Assume that the chromosome interference across this region is 80 percent but that interference within each region is complete.

4.17 The yeast *Saccharomyces cerevisiae* has unordered tetrads. In a cross carried out to study the linkage relationships among three genes, the tetrads in the accompanying table were obtained. The cross was between a strain of genotype + *b c* and one of genotype *a* + +.

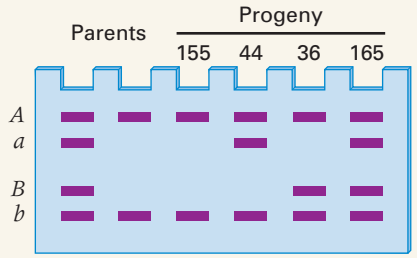
(a) From these data determine which, if any, of the genes are linked.
(b) For any linked genes, determine the map distances.

Tetrad type	Genotypes of spores in tetrads				Number of tetrads
1	*a* + +	*a* + +	+ *b c*	+ *b c*	132
2	*a b* +	*a b* +	+ + *c*	+ + *c*	124
3	*a* + +	*a* + *c*	+ *b* +	+ *b c*	64
4	*a b* +	*a b c*	+ + +	+ + *c*	80
Total					400

4.18 A small portion of the genetic map of *Neurospora crassa* chromosome VI is illustrated here. The *cys-1* mutation blocks cysteine synthesis, and the *pan-2* mutation blocks pantothenic acid synthesis. Assuming complete chromosome interference, determine the expected frequencies of the following types of asci in a cross of *cys-1 pan-2* × *CYS-1 PAN-2*.

(a) First-division segregation of *cys-1* and first-division segregation of *pan-2*.
(b) First-division segregation of *cys-1* and second-division segregation of *pan-2*.
(c) Second-division segregation of *cys-1* and first-division segregation of *pan-2*.
(d) Second-division segregation of *cys-1* and second-division segregation of *pan-2*.
(e) Parental ditype, tetratype, and nonparental ditype tetrads.

4.19 The accompanying gel diagram shows the positions of DNA bands associated with the *A, a* and *B, b* allele pairs for two linked genes. On the left are the phenotypes of the parents, and on the right are the phenotypes of the progeny and the number of each observed. Is the linkage phase of *A* and *B* in the doubly heterozygous parent coupling or repulsion? What is the frequency of recombination between these genes?



4.20 Janet is performing a test cross to determine the linkage relationships between three *Drosophila* genes, *dpy, unc* and *dor*. Her entire grade depends on getting this right! She testcrosses females heterozygous for the recessive alleles:

dpy (dumpy body) / dpy^+ (normal body)
unc (uncoordinated) / unc^+ (coordinated)
dor (deep orange eye) / dor^+ (red eye).

The cross yields the following results:

normal body, red eye, coordinated	75
normal body, red eye, uncoordinated	348
normal body, deep orange eye, uncoordinated	96
dumpy body, red eye, coordinated	110
dumpy body, deep orange eye, coordinated	306
dumpy body, deep orange eye, uncoordinated	65

(a) What is the genotype of the F_1 heterozygous female?
(b) Construct a map of the region indicating the order of the genes and the distances (in map units) between them.
(c) What is the interference in this region?

A MOMENT TO THINK

Answer to Problem: The coincidence is equal to the observed number of double crossovers divided by the expected number, but the observed and expected frequencies (proportions) can be used as well, because in the conversion to frequencies, both numerator and denominator are divided by the same number. Set $r_1 = 0.32$ and $r_2 = 0.25$. We know that r_1 includes the proportion of gametes with single recombination in region 1 (call this s_1) plus the proportion of gametes with double recombination (call this d). Hence $r_1 = s_1 + d$. Similarly, $r_2 = s_2 + d$, where s_2 is the proportion of gametes with single recombination in region 2. In the two-point cross between *w* and *r*, the observed frequency of recombination, 0.45, equals $s_1 + s_2$ because none of the double recombinants are detected. Now we can solve for d because $r_1 + r_2 = (s_1 + d) + (s_2 + d) = s_1 + s_2 + 2d = 0.57$, whereas $s_1 + s_2 = 0.45$. Subtracting, we obtain $2d = 0.57 - 0.45 = 0.12$, or $d = 0.06$. This is the "observed" (inferred in this case) frequency of double recombinants. The expected frequency of double recombinants equals $r_1 \times r_2 = 0.32 \times 0.25 = 0.08$. Thus the coincidence across the region is $0.06/0.08 = 0.75$, and the interference equals $1 - 0.75 = 0.25$. In other words, there is about a 25% deficit in double recombinants from the frequency that would be expected with independence.

GENETICS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetics information on the Internet. To explore these sites, visit the Jones & Bartlett companion site to accompany *Essential Genetics: A Genomic Perspective, Sixth Edition*, at http://biology.jbpub.com/Hartl/EssentialGenetics.

There you will find a chapter-by-chapter list of highlighted keywords. When you select one of the keywords, you will be linked to a Web site containing information related to that keyword.