15 The Dynamic Genome: Transposable Elements

Kernels on an ear of corn. The spotted kernels on this ear of corn result from the interaction of a mobile genetic element (a transposable element) with a corn gene whose product is required for pigmentation. [*Cliff Weil and Susan Wessler.*]

A boy is born with a disease that makes his immune system ineffective. Diagnostic testing determines that he has a recessive genetic disorder called SCID (severe combined immunodeficiency disease), more commonly known as *bubble-boy disease*. This disease is caused by a mutation in the gene encoding the blood enzyme adenosine deaminase (ADA). As a result of the loss of this enzyme, the precursor cells that give rise to one of the cell types of the immune system are missing. Because this boy has no ability to fight infection, he has to live in a completely isolated and sterile environment—that is, a bubble in which the air is filtered for sterility (Figure 15-1). No pharmaceutical or other conventional therapy is available to treat this disease. Giving the boy a tissue transplant containing the precursor cells from another person would not work in the vast majority of cases, because a precise tissue match between donor and patient is extremely rare. Consequently, the donor cells would end up creating an immune response against the boy's own tissues (graft-versus-host disease).

In the past two decades, techniques have been developed that offer the possibility of a different kind of transplantation therapy—**gene therapy**—that could help people with SCID and other incurable diseases. In regard to SCID, a normal ADA gene is "transplanted" into cells of a patient's immune system, thereby permitting these cells to survive and function normally. In the earliest human gene-therapy trials, scientists modified a type of virus called a retrovirus in the laboratory ("engineered") so that it could insert itself and a normal ADA gene into chromosomes of the immune cells taken from patients with SCID. In this chapter, you will see that retroviruses have many biological properties in common with a type of mobile element called a *retrotransposon,* which is present in our genome and the genomes of most eukaryotes. Lessons learned about the behavior

KEY QUESTIONS

- Why were transposable elements first discovered genetically in maize but first isolated molecularly from *E. coli*?
- How do transposable elements participate in the spread of antibiotic-resistant bacteria?
- What are the two major mechanisms used by elements to transpose?
- How can humans survive, given that as much as 50 percent of the human genome is derived from transposable elements?
- Why can the host repress the spread of some transposable elements but not others?

OUTLINE

A boy with SCID

Figure 15-1 A patient with SCID must live in a protective bubble. [*UPI/Bettmann/CORBIS.*]

of retrotransposons and other mobile elements from model organisms such as yeast are sources of valuable insights into the design of a new generation of biological vectors for human gene therapy.

Starting in the 1930s, genetic studies of maize yielded results that greatly upset the classical genetic picture of genes residing only at fixed loci on the main chromosomes. The research literature began to carry reports suggesting that certain genetic elements present in the main chromosomes can somehow move from one location to another. These findings were viewed with skepticism for many years, but it is now clear that such mobile elements are widespread in nature.

A variety of colorful names (some of which help to describe their respective properties) have been applied to these genetic elements: controlling elements, jumping genes, mobile genes, mobile elements, and transposons. Here we use the terms *transposable elements* and *mobile elements,* which embrace the entire family of types. Transposable elements can move to new positions within the same chromosome or even to a different chromosome. They have been detected genetically in model organisms such as *E. coli,* maize, yeast, *C. elegans,* and *Drosophila* through the mutations that they produce when they insert into and inactivate genes.

DNA sequencing of genomes from a variety of microbes, plants, and animals indicates that transposable elements exist in virtually all organisms. Surprisingly, they are by far the largest component of the human genome, accounting for almost 50 percent of our chromosomes. Despite their abundance, the normal genetic role of these elements is not known with certainty.

In their studies, scientists are able to exploit the ability of transposable elements to insert into new sites in the genome. Transposable elements engineered in the test tube are valuable tools, both in prokaryotes and in eukaryotes, for genetic mapping, creating mutants, cloning genes, and even producing transgenic organisms. Let us reconstruct some of the steps in the evolution of our present understanding of transposable elements. In doing so, we will uncover the principles guiding these fascinating genetic units.

The *Ds* transposable element helps cause breakage

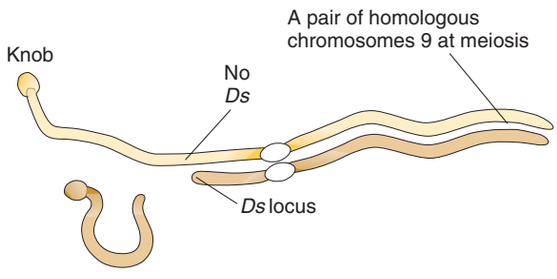


Figure 15-2 Chromosome 9 of corn breaks at the *Ds* locus, where the *Ds* transposable element has inserted.

15.1 Discovery of Transposable Elements in Maize

McClintock's experiments: the *Ds* element

In the 1940s, Barbara McClintock made an astonishing discovery while studying the colored kernels of so-called Indian corn, known as maize (see the Model Organism box on the next page). Maize has 10 chromosomes, numbered from largest (1) to smallest (10). While analyzing the breakage of maize chromosomes, McClintock noticed some unusual phenomena. She found that, in one strain of maize, chromosome 9 broke very frequently and at one particular site, or locus (Figure 15-2). Breakage of the chromosome at this locus, she determined, was due to the presence of two genetic factors. One factor that she called ***Ds*** (for ***Dissociation***) was located at the site of the break. Another, unlinked genetic factor was required to "activate" the breakage of chromosome 9 at the *Ds* locus. Thus, McClintock called this second factor ***Ac*** (for ***Activator***).

McClintock began to suspect that *Ac* and *Ds* were actually mobile genetic elements when she found it impossible to map *Ac.* In some plants, it mapped to one position; in other plants of the same line, it mapped to different positions.

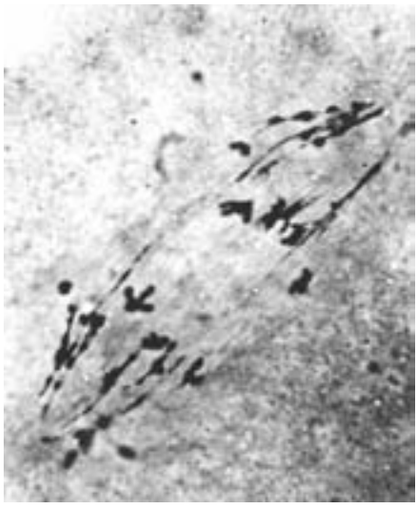
Model Organism *Maize*

Maize, also known as corn, is actually *Zea mays,* a member of the grass family. Grasses—also including rice, wheat, and barley—are the most important source of calories for humanity. Maize was domesticated from the wild grass teosinte by Native Americans in Mexico and Central America and was first introduced to Europe by Columbus on his return from the New World.

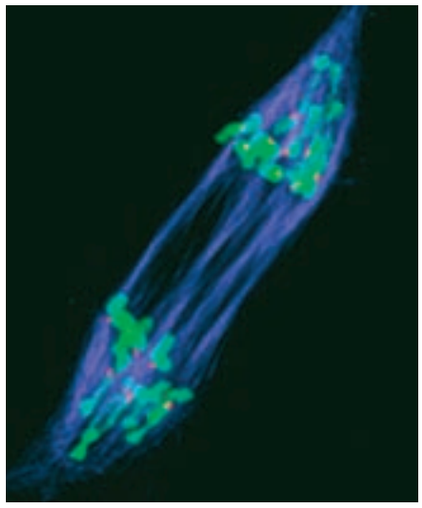
In the 1920s, Rollins A. Emerson set up a laboratory at Cornell University to study the genetics of corn traits, including kernel color, which were ideal for genetic analysis. In addition, the physical separation of male and female flowers into the tassel and ear, respectively, made controlled genetic crosses easy to accomplish. Among the outstanding geneticists attracted to the Emerson laboratory were Marcus Rhoades, Barbara McClintock, and George Beadle (see Chapter 6). Before the advent of molecular biology and the rise of microorganisms as model organisms, geneticists performed microscopic analyses of chromosomes and related their behavior to the segregation of traits. The large pachytene chromosomes of maize and the salivary-gland chromosomes of *Drosophila* made them the organisms of choice for cytogenetic analyses. The results of these early studies led to an understanding of chromosome behavior during meiosis and mitosis, including such events as recombination and the consequences of chromosome breakage such as inversions, translocations, and duplication.

The maize laboratory of Rollins A. Emerson at Cornell University, 1929. Standing from left to right: Charles Burnham, Marcus Rhoades, Rollins Emerson, and Barbara McClintock. Kneeling is George Beadle. Both McClintock and Beadle were awarded a Nobel Prize. [*Courtesy of the Department of Plant Breeding, Cornell University.*]

Maize still serves as a model genetic organism. Molecular biologists continue to exploit its beautiful pachytene chromosomes with new antibody probes (see photograph *b* below) and have used its wealth of genetically well-characterized transposable elements as tools to identify and isolate important genes.



(a)



(b)

Analysis of maize chromosomes, then and now. Maize chromosomes are large and easily visualized by light microscopy. (a) An image from Marcus Rhoades (1952). (b) This image is comparable to that in part *a* except that the spindle is shown in blue (stained with antibodies to tubulin), the centromeres are shown in red (stained with antibodies to a centromere-associated protein), and the chromosomes are shown in green. [*(a) From M. M. Rhoades, "Preferential Segregation in Maize," in J. W. Gowen, ed., Heterosis, pp. 66–80. Iowa State College Press, 1952. (b) From R. K. Dawe, L. Reed, H.-G. Yu, M. G. Muszynski, and E. N. Hiatt, "A Maize Homolog of Mammalian CENPC Is a Constitutive Component of the Inner Kinetochore," Plant Cell 11, 1999, 1227–1238.*]

As if this variable mapping were not enough of a curiosity, rare kernels with dramatically different phenotypes could be derived from the original strain that had frequent breaks in chromosome 9. One such phenotype was a rare colorless kernel containing pigmented spots.

Figure 15-3 compares the phenotype of the chromosome-breaking strain with the phenotype of one of these derivative strains. For the chromosome-breaking strain, a chromosome that breaks at or near *Ds* loses its end containing wild-type alleles of the *C, Sh,* and *Wx* genes. In the example shown in Figure 15-3a, a break occurred in a single cell, which divided mitotically to produce the large sector of mutant tissue (*c sh wx*). Breakage can happen many times in a single kernel, but each sector of tissue will display the loss of expression of all three genes. In contrast, each new derivative affected the expression of only a single gene. One derivative that affected the expression of only the pigment gene *C* is shown in Figure 15-3b. In this example, pigmented spots appeared on a colorless kernel background. Although the expression of *C* was altered in this strange way, the

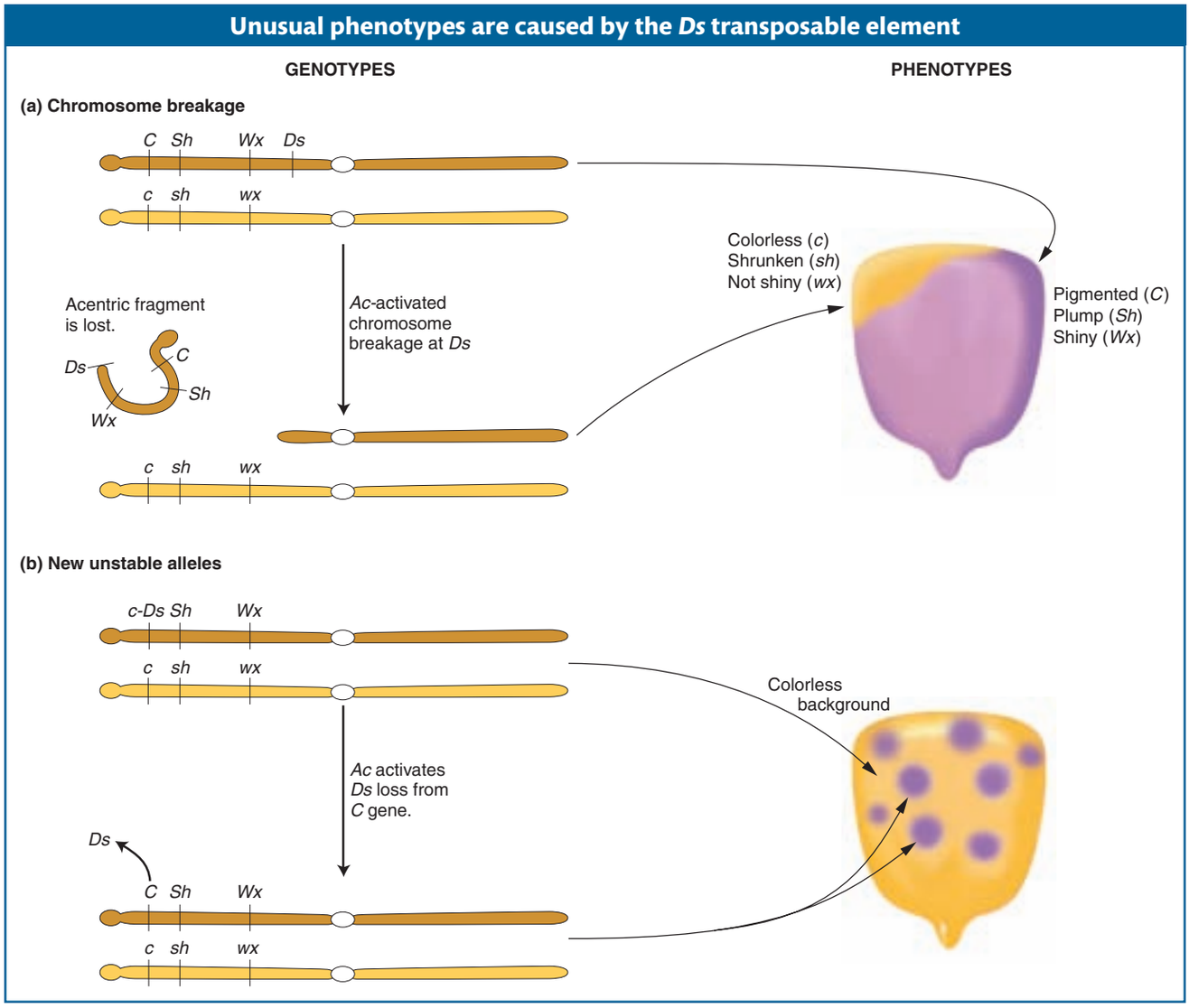


Figure 15-3 New phenotypes in corn are produced through the movement of the *Ds* transposable element on chromosome 9. (a) A chromosome fragment is lost through breakage at the *Ds* locus. Recessive alleles on the homologous chromosome are expressed, producing the colorless sector in the kernel. (b) Insertion of *Ds* in the *C* gene (top) creates colorless corn-kernel cells. Excision of *Ds* from the *C* gene through the action of *Ac* in cells and their mitotic descendants allows color to be expressed again, producing the spotted phenotype.

expression of *Sh* and *Wx* was normal and chromosome 9 no longer sustained frequent breaks.

To explain the new derivatives, McClintock hypothesized that *Ds* had moved from a site near the centromere into the *C* gene located close to the telomeric end. In its new location, *Ds* prevents the expression of *C.* The inactivation of the *C* gene explains the colorless parts of the kernel, but what explains the appearance of the pigmented spots? The spotted kernel is an example of an **unstable phenotype.** McClintock concluded that such unstable phenotypes resulted from the movement or transposition of *Ds* away from the *C* gene. That is, the kernel begins development with a *C* gene that has been mutated by the insertion of *Ds.* However, in some cells of the kernel, *Ds* leaves the *C* gene, allowing the mutant phenotype to revert to wild type and produce pigment in the original cell and in all its mitotic descendants. There are big spots of color when *Ds* leaves the *C* gene early in kernel development (because there are more mitotic descendants), whereas there are small spots when *Ds* leaves the *C* gene later in kernel development. Unstable mutant phenotypes that revert to wild type are a clue to the participation of mobile elements.

Autonomous and nonautonomous elements

What is the relation between *Ac* and *Ds*? How do they interact with genes and chromosomes to produce these interesting and unusual phenotypes? These questions were answered by further genetic analysis. Interactions between *Ds, Ac,* and the pigment gene *C* are used as an example in Figure 15-4. There, *Ds* is shown as a piece of DNA that has inactivated the *C* gene by inserting into its coding region. The allele carrying the insert is called *c-mutable(Ds)* or *c-m(Ds)* for short. A strain with *c-m(Ds)* and no *Ac* has colorless kernels because *Ds* cannot move; it is stuck in the *C* gene. A strain with *c-m(Ds)* and *Ac* has spotted kernels because *Ac* activates *Ds* in some cells to leave the *C* gene, thereby restoring gene function. The leaving element is said to **excise** from the chromosome or **transpose.**

Other strains were isolated in which the *Ac* element itself had inserted into the *C* gene [called *c-m(Ac)*]. Unlike the *c-m(Ds)* allele, which is unstable only when *Ac* is in the genome, *c-m(Ac)* is always unstable. Furthermore, McClintock found that, on rare occasions, an allele of the *Ac* type could be transformed into an allele of the *Ds* type. This transformation was due to the spontaneous generation of a *Ds* element from the inserted *Ac* element. In other words, *Ds* is, in all likelihood, an incomplete, mutated version of *Ac* itself.

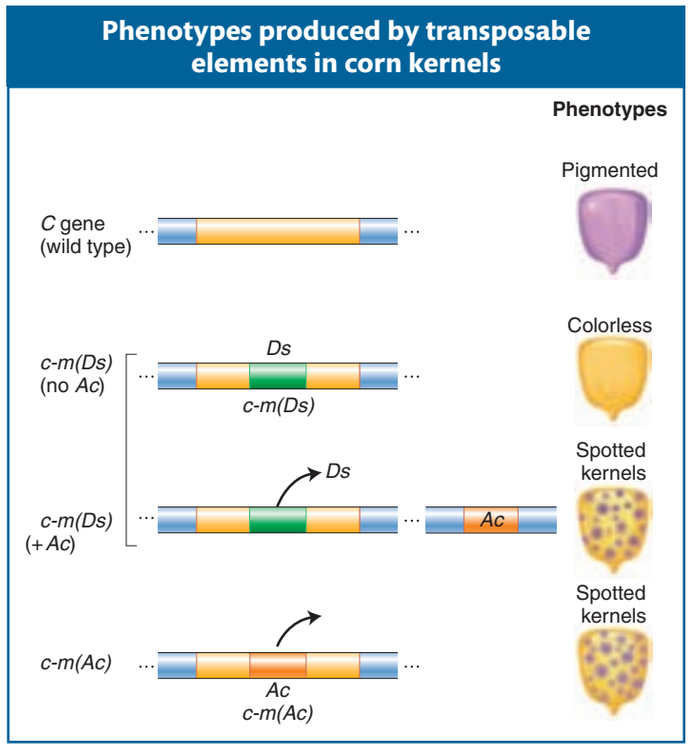


Figure 15-4 Kernel spotting is controlled by the insertion and excision of *Ds* or *Ac* elements in the *C* gene controlling pigment.

Several systems like *Ac/Ds* were found by McClintock and other geneticists working with maize. Two other systems are *Dotted* [(*Dt*), discovered by Marcus Rhoades] and *Suppressor/mutator* [(*Spm*), independently discovered by McClintock and Peter Peterson, who called it *Enhancer/Inhibitor* (*En/In*)]. In addition, as you will see in the sections that follow, elements with similar genetic behavior have been isolated from bacteria, plants, and animals.

The common genetic behavior of these elements led geneticists to propose new categories for all the elements. *Ac* and elements with similar genetic properties are now called **autonomous elements** because they require no other elements for their mobility. Similarly, *Ds* and elements with similar genetic properties are called **nonautonomous elements.** An element *family* is composed of one or more

Transposable elements at work in a rose

Figure 15-5 Mosaicism is caused by the excision of transposable elements in roses. The insertion of a transposable element disrupts pigment production, resulting in white flowers. The excision of the transposable element restores pigment production, resulting in red floral-tissue sectors. [*Photo courtesy of Susan Wessler.*]

autonomous elements and the nonautonomous members that can be mobilized. Autonomous elements encode the information necessary for their own movement and for the movement of unlinked nonautonomous elements in the genome. Because nonautonomous elements do not encode the functions necessary for their own movement, they cannot move unless an autonomous element from their family is present somewhere else in the genome.

Figure 15-5 shows an example of the effects of transposons in a rose.

Message Transposable elements in maize can inactivate a gene in which they reside, cause chromosome breaks, and transpose to new locations within the genome. Autonomous elements can perform these functions unaided; nonautonomous elements can transpose only with the help of an autonomous element elsewhere in the genome.

Transposable elements: only in maize?

Although geneticists accepted McClintock's discovery of transposable elements in maize, many were reluctant to consider the possibility that similar elements resided in the genomes of other organisms. Their existence in all organisms would imply that genomes are inherently unstable and dynamic. This view was inconsistent with the fact that the genetic maps of members of the same species were the same. After all, if genes can be genetically mapped to a precise chromosomal location, doesn't this mapping indicate that they are not moving around?

Because McClintock was a highly respected geneticist, her results were explained by saying that maize is not a natural organism: it is a crop plant that is the product of human selection and domestication. This view was held by some until the 1960s, when the first transposable elements were isolated from the *E. coli* genome and studied at the DNA-sequence level. Transposable elements were subsequently isolated from the genomes of many organisms, including *Drosophila* and yeast. When it became apparent that transposable elements are a significant component of the genomes of most and perhaps all organisms, Barbara McClintock was recognized for her seminal discovery by being awarded the 1983 Nobel Prize in Medicine or Physiology.

15.2 Transposable Elements in Prokaryotes

The genetic discovery of transposable elements led to many questions about what such elements might look like at the DNA-sequence level and how they are able to move from one site to another in the genome. Did all organisms have them? Did all elements look alike or were there different classes of transposable elements? If there were many classes of elements, could they coexist in one genome? Did the number of transposable elements in the genome vary from species to species? The molecular nature of transposable genetic elements was first understood in bacteria. Therefore, we will continue this story by examining the original studies performed with prokaryotes.

There are two broad types of transposable elements in bacteria:

- Short sequences called *IS elements* that can move themselves to new positions but do not carry genes other than those needed for their movement.

- Longer sequences called *transposons* that not only carry the genes they need for their movement but also carry other genes.

Bacterial insertion sequences

Insertion sequences, or **insertion-sequence (IS) elements,** are segments of bacterial DNA that can move from one position on a chromosome to a different position on the same chromosome or on a different chromosome. When IS elements appear in the middle of genes, they interrupt the coding sequence and inactivate the expression of that gene. Owing to their size and in some cases the presence of transcription- and translation-termination signals in the IS element, IS elements can also block the expression of other genes in the same operon if those genes are downstream of the insertion site. IS elements were first found in *E. coli* in the *gal* operon—a cluster of three genes taking part in the metabolism of the sugar galactose.

Identification of discrete IS elements Several *E. coli gal*$^-$ mutants were found to contain large insertions of DNA into the *gal* operon. This finding led naturally to the next question: Are the segments of DNA that insert into genes merely random DNA fragments or are they distinct genetic entities? The answer to this question came from the results of hybridization experiments showing that many different insertion mutations are caused by a small set of insertion sequences. These experiments are performed with the use of λd*gal* phages that contain the *gal*$^-$ operon from several independently isolated *gal* mutant strains. Individual phage particles from the strains are isolated, and their DNA is used to synthesize radioactive RNA in vitro. Certain fragments of this RNA are found to hybridize with the DNA from other *gal*$^-$ mutations containing large DNA insertions but not with wild-type DNA. These results were interpreted to mean that independently isolated *gal* mutants contain the same extra piece of DNA. These particular RNA fragments also hybridize to DNA from other mutants containing IS insertions in other genes, showing that the same bit of DNA can insert in different places in the bacterial chromosome.

Structure of IS elements On the basis of their patterns of cross-hybridization, a number of distinct IS elements have been identified. One sequence, termed IS1, is the 800-bp segment identified in *gal.* Another sequence, termed IS2, is 1350 bp long. Although IS elements differ in DNA sequence, they have several features in common. For example, all IS elements encode a protein, called a **transposase,** which is an enzyme required for the movement of IS elements from one site in the chromosome to another. In addition, all IS elements begin and end with short inverted repeat sequences that are required for their mobility. The transposition of IS elements and other mobile genetic elements will be considered later in the chapter.

The genome of the standard wild-type *E. coli* is rich in IS elements: it contains eight copies of IS1, five copies of IS2, and copies of other less well-studied IS types. Because IS elements are regions of identical sequence, they are sites where crossovers may take place. For example, recombination between the F-factor plasmid and the *E. coli* chromosome to form *Hfr* strains is the result of a single crossover between an IS element located on the plasmid and an IS element located on the chromosome (see Chapter 5, Figure 5-18). Because there are multiple IS elements, the F factor can insert at multiple sites.

Message The bacterial genome contains segments of DNA, termed IS elements, that can move from one position on the chromosome to a different position on the same chromosome or on a different chromosome.

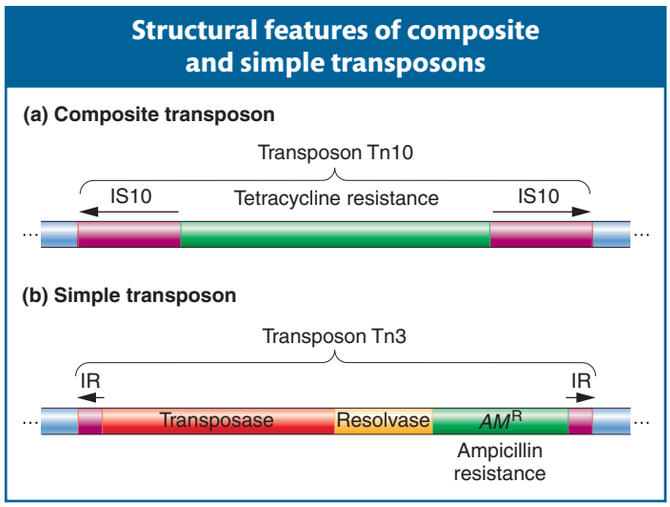


Figure 15-6 (a) Tn10, an example of a composite transposon. The IS elements are inserted in opposite orientation and form inverted repeats (IRs). Each IS element carries a transposase, but only one is usually functional. (b) Tn3, an example of a simple transposon. Short inverted repeats contain no transposase. Instead, simple transposons encode their own transposase. The resolvase is a protein that promotes recombination and resolves the cointegrates (see Figure 15-9).

Prokaryotic transposons

In Chapter 5, you learned about **R factors,** which are plasmids carrying genes that encode resistance to several antibiotics. These R factors (for resistance) are transferred rapidly on cell conjugation, much like the F factor in *E. coli.*

The R factors proved to be just the first of many similar F-like factors to be discovered. R factors have been found to carry many different kinds of genes in bacteria. In particular, R factors pick up genes conferring resistance to different antibiotics. How do they acquire their new genetic abilities? It turns out that the drug-resistance genes reside on a mobile genetic element called a **transposon** (**Tn**). There are two types of bacterial transposons. **Composite transposons** contain a variety of genes that reside between two nearly identical IS elements that are oriented in opposite direction (Figure 15-6a) and, as such, form what is called an **inverted repeat** sequence. Transposase encoded by one of the two IS elements is necessary to catalyze the movement of the entire transposon. An example of a composite transposon is Tn10, shown in Figure 15-6a. Tn10 carries a gene that confers resistance to the antibiotic tetracycline and is flanked by two IS10 elements in opposite orientation. The IS elements that make up composite transposons are not capable of transposing on their own without the rest of the transposon, because of mutations in their inverted repeats.

Simple transposons also consist of bacterial genes flanked by inverted repeat sequences, but these sequences are short (<50 bp) and do not encode the transposase enzyme that is necessary for transposition. Thus, their mobility is not due to an association with IS elements. Instead, simple transposons encode their own transposase in the region between the inverted repeat sequences in addition to carrying bacterial genes. An example of a simple transposon is Tn3, shown in Figure 15-6b.

To review, IS elements are short mobile sequences that encode only those proteins necessary for their mobility. Composite transposons and simple transposons contain additional genes that confer new functions to bacterial cells. Whether composite or simple, transposons are usually just called transposons, and different transposons are designated Tn1, Tn2, Tn505, and so forth.

A transposon can jump from a plasmid to a bacterial chromosome or from one plasmid to another plasmid. In this manner, multiple-drug-resistant plasmids are generated. Figure 15-7 is a composite diagram of an R factor, indicating the various places at which transposons can be located. We next consider the question of how such **transposition** or mobilization events occur.

Message Transposons were originally detected as mobile genetic elements that confer drug resistance. Many of these elements consist of IS elements flanking a gene that encodes drug resistance. This organization promotes the spread of drug-resistant bacteria by facilitating movement of the resistance gene from the chromosome of a resistant bacterium to a plasmid that can be conjugated into another (susceptible) bacterial strain.

Mechanism of transposition

As already stated, the movement of a transposable element depends on the action of a transposase. This enzyme plays key roles in the two stages of transposition: excision (leaving) from the original location and inserting into the new location.

Excision from the original location Most transposable elements in prokaryotes (and in eukaryotes) employ one of two mechanisms of transposition, called

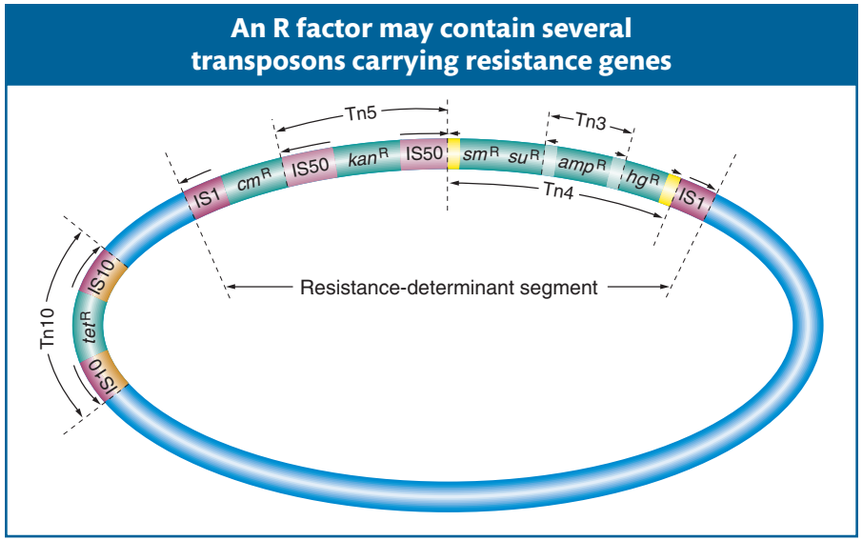


Figure 15-7 A schematic map of a plasmid with several insertions of simple and composite transposons carrying resistance genes. Plasmid sequences are in blue. Genes encoding resistance to the antibiotics tetracycline (*tet*^{R}), kanamycin (*kan*^{R}), streptomycin (*sm*^{R}), sulfonamide (*su*^{R}), and ampicillin (*amp*^{R}) and to mercury (*hg*^{R}) are shown. The resistant-determinant segment can move as a cluster of resistance genes. Tn3 is within Tn4. Each transposon can be transferred independently. [*Simplified from S. N. Cohen and J. A. Shapiro, "Transposable Genetic Elements." Copyright 1980 by Scientific American, Inc. All rights reserved.*]

replicative and **conservative** (nonreplicative), as illustrated in Figure 15-8. In the replicative pathway (as shown for Tn3), a new copy of the transposable element is generated in the transposition event. The results of the transposition are that one copy appears at the new site and one copy remains at the old site. In the conservative pathway (as shown for Tn10), there is no replication. Instead, the element is excised from the chromosome or plasmid and is integrated into the new site. The conservative pathway is also called **"cut and paste."**

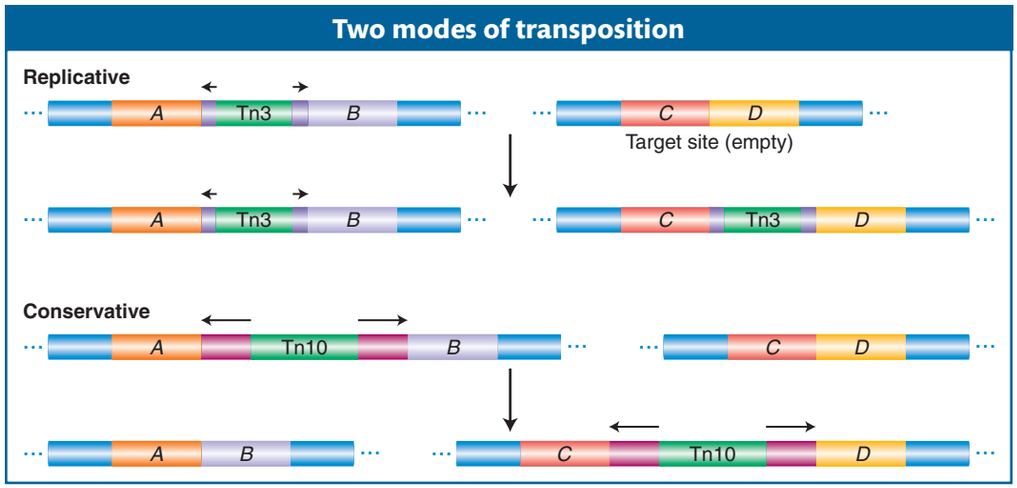


Figure 15-8 Mobile-element transposition may be either replicative or conservative. See text for details. [*Adapted with permission from Nature Reviews: Genetics 1, no. 2, p. 138, Fig. 3, November 2000, "Mobile Elements and the Human Genome," E. T. Luning Prak and H. H. Kazazian, Jr. Copyright 2000 by Macmillan Magazines Ltd.*]

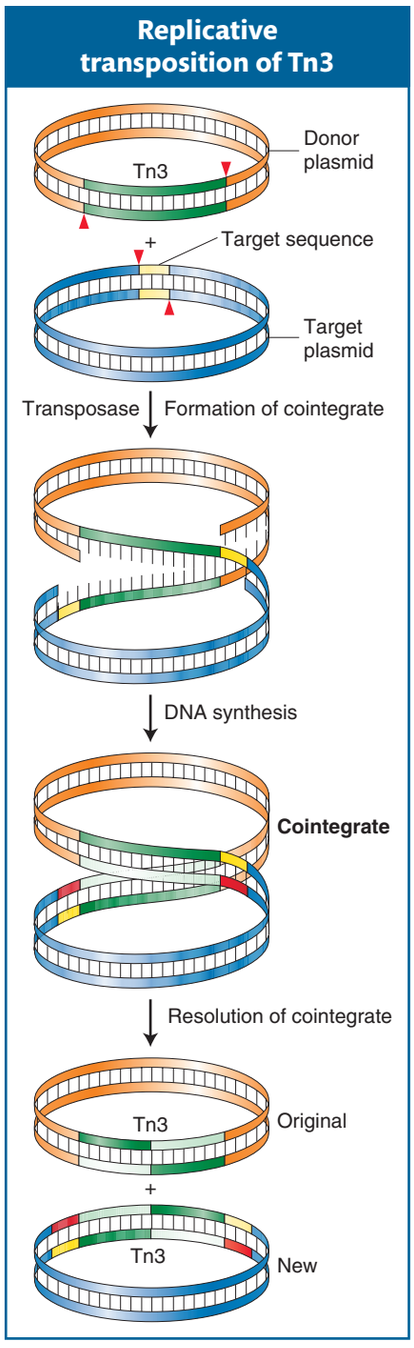


Figure 15-9 Replicative transposition of Tn3 takes place through a cointegrate intermediate. [*After Robert J. Brooker, Genetics: Analysis and Principles, Fig. 18-14. Benjamin-Cummings, 1999.*]

GENETICS PORTAL **ANIMATED ART: Replicative transposition**

Replicative transposition Because this mechanism is a bit complicated, it will be described here in more detail. As Figure 15-8 illustrates, one copy of Tn3 is produced from an initial single copy, yielding two copies of Tn3 altogether.

Figure 15-9 shows the details of the intermediates in the transposition of Tn3 from one plasmid (the donor) to another plasmid (the target). During transposition, the donor and recipient plasmids are temporarily fused together to form a double plasmid. The formation of this intermediate is catalyzed by Tn3-encoded transposase, which makes single-strand cuts at the two ends of Tn3 and staggered cuts at the target sequence and joins the free ends together, forming a fused circle called a **cointegrate.** The transposable element is duplicated in the fusion event. The cointegrate then resolves by a recombination-like event that turns a cointegrate into two smaller circles, leaving one copy of the transposable element in each plasmid. The result is that one copy remains at the original location of the element, whereas the other is integrated at a new genomic position.

Conservative transposition Some transposons, such as Tn10, excise from the chromosome and integrate into the target DNA. In these cases, the DNA of the element is not replicated, and the element is lost from the site of the original chromosome (see Figure 15-8). Like replicative transposition, this reaction is initiated by the element-encoded transposase, which cuts at the ends of the transposon. However, in contrast with replicative transposition, the transposase

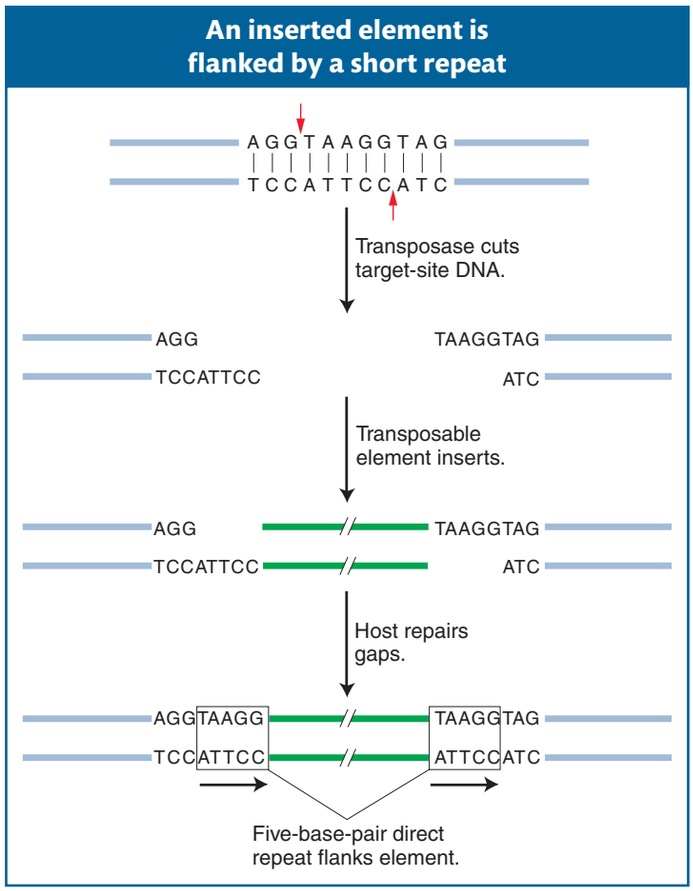


Figure 15-10 A short sequence of DNA is duplicated at the transposon insertion site. The recipient DNA is cleaved at staggered sites (a 5-bp staggered cut is shown), leading to the production of two copies of the five-base-pair sequence flanking the inserted element.

cuts the element out of the donor site. It then makes a staggered cut at a target site and inserts the element into the target site. We will revisit this mechanism in more detail in a discussion of the transposition of eukaryotic transposable elements, including the *Ac*/*Ds* family of maize.

Insertion into a new location We have seen that transposase continues to play an important role in insertion. In one of the first steps of insertion, the transposase makes a staggered cut in the target-site DNA (not unlike the staggered breaks catalyzed by restriction endonucleases in the sugar–phosphate backbone of DNA). Figure 15-10 shows the steps in the insertion of a generic transposable element. In this case, the transposase makes a five-base-pair staggered cut. The transposable element inserts between the staggered ends, and the host DNA repair machinery (see Chapter 16) fills in the gap opposite each single-strand overhang by using the bases in the overhang as a template. There are now two duplicate sequences, each five base pairs in length, at the sites of the former overhangs. These sequences are called a **target-site duplication.** Virtually all transposable elements (in both prokaryotes and eukaryotes) are flanked by a target-site duplication, indicating that all use a mechanism of insertion similar to that shown in Figure 15-10. What differs is the length of the duplication; a particular type of transposable element has a characteristic length for its target-site duplication—as small as two base pairs for some elements. It is important to keep in mind that the transposable elements have *inverted repeats* at their ends and that the inverted repeats are flanked by the target-site duplication—which is a *direct repeat*.

Message In prokaryotes, transposition occurs by at least two different pathways. Some transposable elements can replicate a copy of the element into a target site, leaving one copy behind at the original site. In other cases, transposition consists of the direct excision of the element and its reinsertion into a new site.

15.3 Transposable Elements in Eukaryotes

Although transposable elements were first discovered in maize, the first eukaryotic elements to be molecularly characterized were isolated from mutant yeast and *Drosophila* genes. Eukaryotic transposable elements fall into two classes: class 1 retrotransposons and class 2 DNA transposons. The first class to be isolated, the retrotransposons, are not at all like the prokaryotic IS elements and transposable elements.

Class I: retrotransposons

The laboratory of Gerry Fink was among the first to use yeast as a model organism to study eukaryotic gene regulation. Through the years, he and his colleagues isolated thousands of mutations in the *HIS4* gene, which encodes one of the enzymes in the pathway leading to the synthesis of the amino acid histidine.

They isolated more than 1500 spontaneous *HIS4* mutants and found that 2 of them had an unstable mutant phenotype. The unstable mutants (called pseudorevertants) were more than 1000 times as likely to revert to a phenotype that was similar to wild type as the other *HIS4* mutants. Symbolically, we say that these unstable mutants reverted from His^- to His^+ (wild types have a superscript plus sign, whereas mutants have a superscript minus sign). Like the *E. coli* gal^- mutants, these yeast mutants were found to harbor a large DNA insertion in the *HIS4* gene. The insertion turned out to be very similar to one of a group of transposable elements already characterized in yeast, called the ***Ty* elements.** There are, in fact, about 35 copies of the inserted element, called *Ty1*, in the yeast genome.

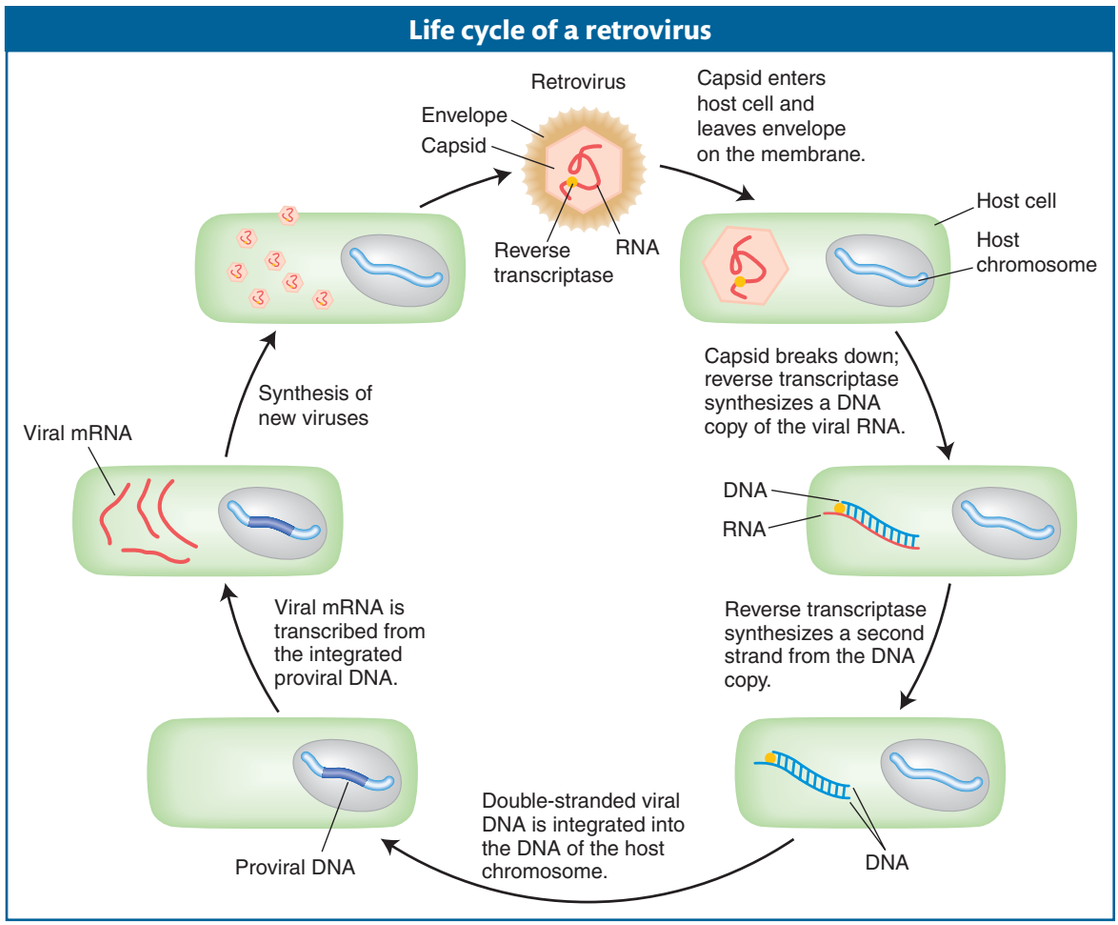


Figure 15-11 The retrovirus RNA genome undergoes reverse transcription into double-stranded DNA inside the host cell.

Cloning of the elements from these mutant alleles led to the surprising discovery that the insertions did not look at all like bacterial IS elements or transposons. Instead, they resembled a well-characterized class of animal viruses called retroviruses. A **retrovirus** is a single-stranded RNA virus that employs a double-stranded DNA intermediate for replication. The RNA is copied into DNA by the enzyme **reverse transcriptase.** The double-stranded DNA is integrated into host chromosomes, from which it is transcribed to produce the RNA viral genome and proteins that form new viral particles. When integrated into host chromosomes as double-stranded DNA, the double-stranded DNA copy of the retroviral genome is called a **provirus.** The life cycle of a typical retrovirus is shown in Figure 15-11. Some retroviruses, such as mouse mammary tumor virus (MMTV) and Rous sarcoma virus (RSV), are responsible for the induction of cancerous tumors. This happens because they insert randomly into the genome and, in the process, may insert next to a gene whose altered expression leads to cancer.

Figure 15-12 shows the similarity in structure and gene content of a retrovirus and the *Ty1* element isolated from the *HIS4* mutants. Both are flanked by **long terminal repeat (LTR)** sequences that are several hundred base pairs long. Both contain the genes *gag* and *pol.*

Retroviruses encode at least three proteins that take part in viral replication: the products of the *gag, pol,* and *env* genes. The *gag*-encoded protein has a role in the maturation of the RNA genome, *pol* encodes the all-important reverse tran-

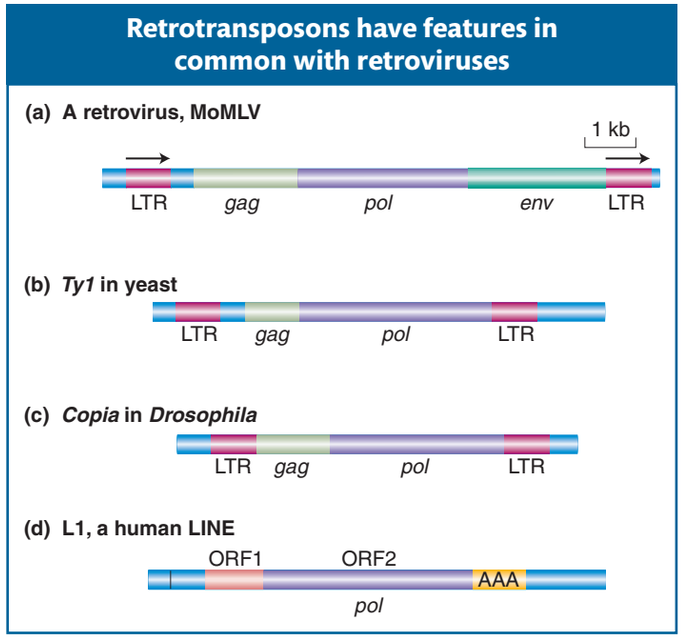


Figure 15-12 Structural comparison of a retrovirus with retrotransposons found in eukaryotic genomes. (a) A retrovirus, Moloney murine leukemia virus (MoMLV), of mice. (b) A retrotransposon, *Ty1,* in yeast. (c) A retrotransposon, *copia,* in *Drosophila.* (d) A long interspersed element (LINE) in humans. Abbreviations: LTR, long terminal repeat; ORF, open reading frame.

scriptase, and *env* encodes the structural protein that surrounds the virus. This protein is necessary for the virus to leave the cell to infect other cells. Interestingly, *Ty1* elements have genes related to *gag* and *pol* but not *env.* These features led to the hypothesis that, like retroviruses, *Ty1* elements are transcribed into RNA transcripts that are copied into double-stranded DNA by the reverse transcriptase. However, unlike retroviruses, *Ty1* elements cannot leave the cell, because they do not encode *env.* Instead, the double-stranded DNA copies are inserted back into the genome of the same cell. These steps are diagrammed in Figure 15-13.

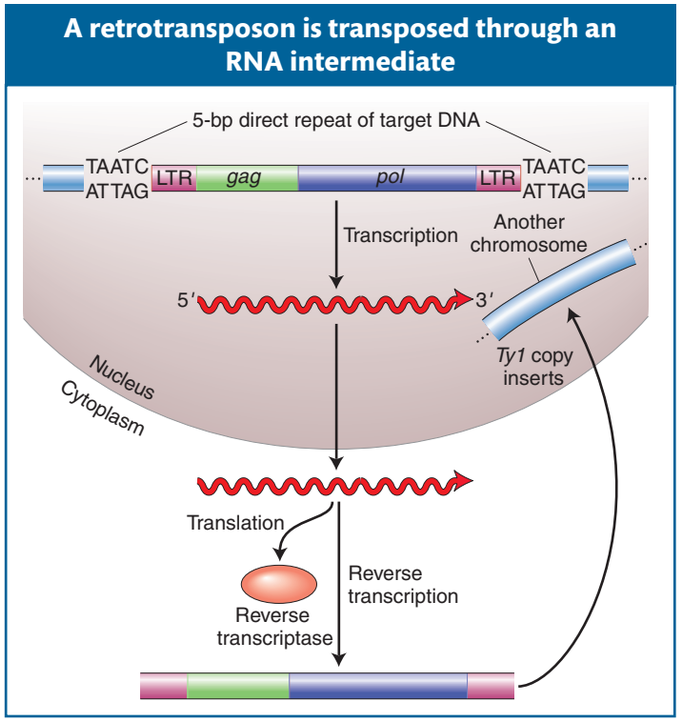


Figure 15-13 An RNA transcript from the retrotransposon undergoes reverse transcription into DNA, by a reverse transcriptase encoded by the retrotransposon. The DNA copy is inserted at a new location in the genome.

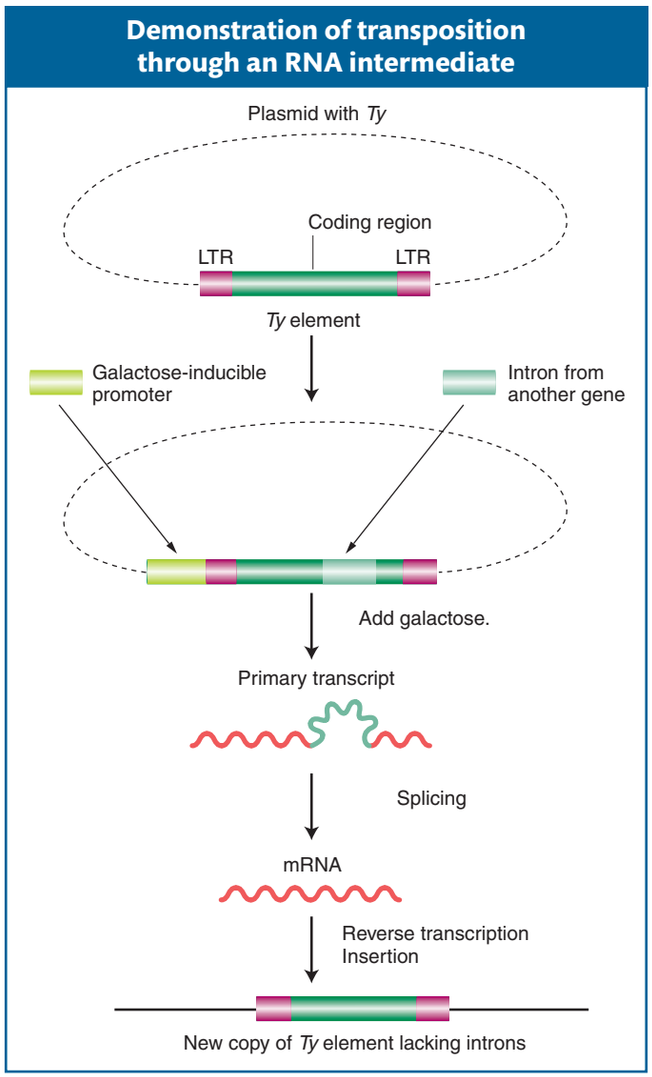


Figure 15-14 A *Ty* element is altered by adding an intron and a promoter that can be activated by the addition of galactose. The intron sequences are spliced before reverse transcription. [*After H. Lodish, D. Baltimore, A. Berk, S. L. Zipursky, P. Matsudaira, and J. Darnell, Molecular Cell Biology, 3rd ed., p. 332. Copyright 1995 by Scientific American Books.*]

In 1985, David Garfinkel, Jef Boeke, and Gerald Fink showed that, like retroviruses, *Ty* elements do in fact transpose through an RNA intermediate. Figure 15-14 diagrams their experimental design. They began by altering a yeast *Ty1* element, cloned on a plasmid. First, near one end of an element, they inserted a promoter that can be activated by the addition of galactose to the medium. Second, they introduced an intron from another yeast gene into the coding region of the *Ty* transposon.

The addition of galactose greatly increases the frequency of transposition of the altered *Ty* element. This increased frequency suggests the participation of RNA, because galactose stimulates the transcription of *Ty* DNA into RNA, beginning at the galactose-sensitive promoter. The key experimental result, however, is the fate of the transposed *Ty* DNA. The researchers found that the intron had been removed from the transposed *Ty* DNA. Because introns are spliced only in the course of RNA processing (see Chapter 8), the transposed *Ty* DNA must have been copied from an RNA intermediate. The conclusion was that RNA is transcribed from the original *Ty* element and spliced. The spliced mRNA undergoes reverse transcription back into double-stranded DNA, which is then integrated into the yeast chromosome. Transposable elements that employ reverse transcriptase to transpose through an RNA intermediate are termed **retrotransposons.** They are also known as **class 1 transposable elements.** Retrotransposons such as *Ty1* that have *long terminal repeats* at their ends are called **LTR-retrotransposons**, and the mechanism they use to transpose is called **"copy and paste"** to distinguish them from the cut-and-paste mechanism that characterizes most DNA transposable elements.

Several spontaneous mutations isolated through the years in *Drosophila* also were shown to contain retrotransposon insertions. The ***copia*-like elements** of *Drosophila* are structurally similar to *Ty1* elements and appear at 10 to 100 positions in the *Drosophila* genome (see Figure 15-12c). Certain classic *Drosophila* mutations result from the insertion of *copia*-like and other elements. For example, the *white-apricot* (w^a) mutation for eye color is caused by the insertion of an element of the *copia* family into the *white* locus. The insertion of LTR-retrotransposons into plant genes (including maize) also has been shown to contribute to spontaneous mutations in this kingdom.

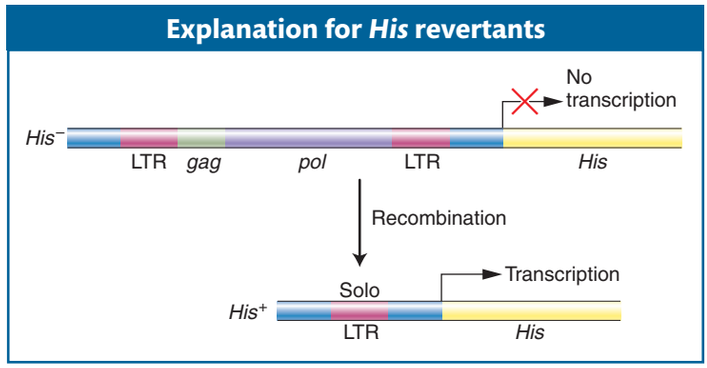


Figure 15-15 *His*⁺ revertants contain a solo LTR in the *His* promoter.

Before we leave retrotransposons (we will return to them later in this chapter), there is one question that needs to be answered. Recall that the first LTR-retrotransposon was discovered in an unstable *His*$^-$ strain of yeast that reverted frequently to *His*$^+$. However, we have just seen that LTR retrotransposons, unlike most DNA transposable elements, do not excise when they transpose. What, then, is responsible for this allele's ~1000-fold increase in reversion frequency when compared to other *His*$^-$ alleles? The answer is shown in Figure 15-15, which shows that the *Ty1* element in the *His*$^-$ allele is located in the promoter region of the *His* gene, where it prevents gene transcription. In contrast, the revertants contain a single copy of the LTR, called a **solo LTR.** This much smaller insertion does not interfere with the transcription of the *His* gene. The solo LTR is the product of recombination between the identical LTRs, which results in the deletion of the rest of the element (see Chapters 4 and 16 for more on recombination). Solo LTRs are a very common feature in the genomes of virtually all eukaryotes, indicating the importance of this process. The sequenced yeast genome contains more than fivefold as many solo LTRs as complete *Ty1* elements.

Message Transposable elements that transpose through RNA intermediates predominate in eukaryotes. Retrotransposons, also known as class 1 elements, encode a reverse transcriptase that produces a double-stranded DNA copy (from an RNA intermediate) that is capable of integrating at a new position in the genome.

Class 2: DNA transposons

Some mobile elements found in eukaryotes appear to transpose by mechanisms similar to those in bacteria. As illustrated in Figure 15-8 for IS elements and transposons, the entity that inserts into a new position in the genome is either the element itself or a copy of the element. Elements that transpose in this manner are called **class 2 elements,** or **DNA transposons.** The first transposable elements discovered by McClintock in maize are now known to be DNA transposons. However, the first DNA transposons to be molecularly characterized were the *P* elements in *Drosophila.*

***P* elements** Of all the transposable elements in *Drosophila,* the most intriguing and useful to the geneticist are the ***P* elements.** The full-size *P* element resembles the simple transposons of bacteria in that its ends are short (31-bp) inverted repeats and it encodes a single protein—the transposase that is responsible for its mobilization (Figure 15-16). The *P* elements vary in size, ranging from 0.5 to 2.9 kb in length. This size difference is due to the presence of many defective *P* elements from which parts of the middle of the element—encoding the transposase gene—have been deleted.

P elements were discovered by Margaret Kidwell, who was studying **hybrid dysgenesis**—a phenomenon that occurs when females from laboratory strains of

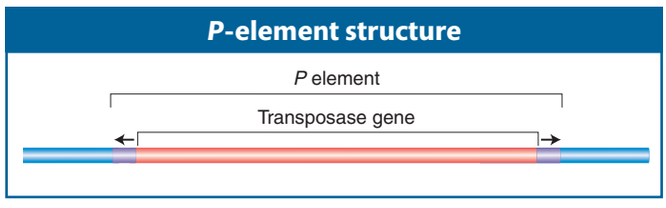


Figure 15-16 DNA sequence analysis of the 2.9-kb *P* element reveals a gene that encodes transposase. A perfect 31-bp inverted repeat resides at each of the element's termini. [*From G. Robin, in J. A. Shapiro, ed., Mobile Genetic Elements, pp. 329–361. Academic Press, 1983.*]

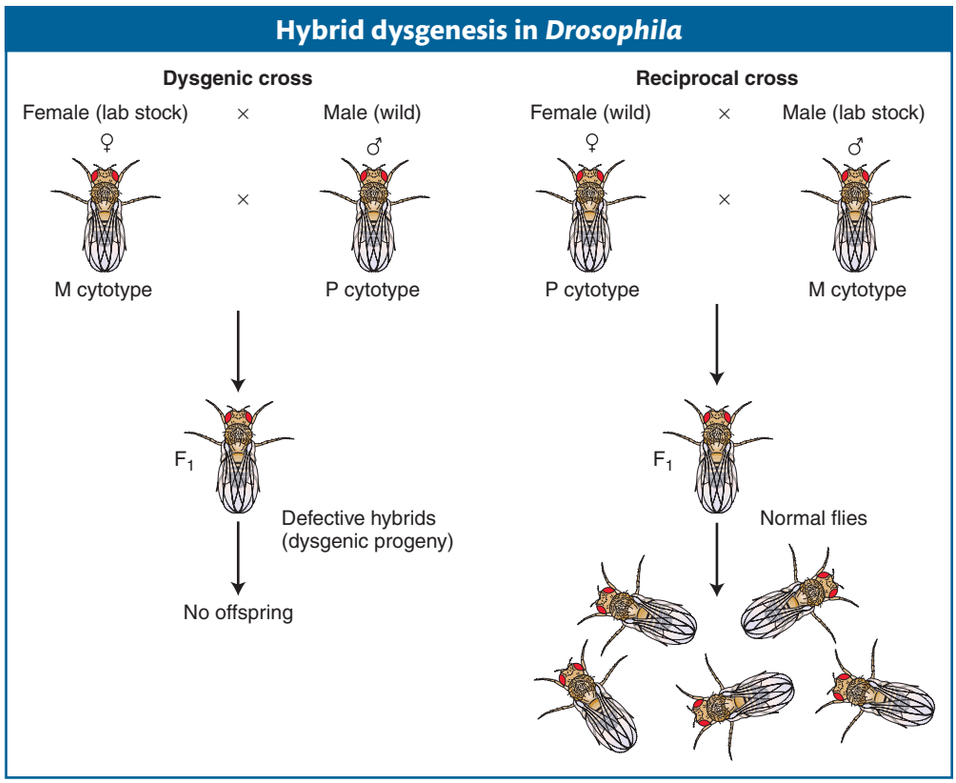


Figure 15-17 In hybrid dysgenesis, a cross between a female from laboratory stock and a wild male yields defective progeny. See text for details.

D. melanogaster are mated with males derived from natural populations. In such crosses, the laboratory stocks are said to possess an **M cytotype** (cell type), and the natural stocks are said to possess a **P cytotype.** In a cross of M (female) × P (male), the progeny show a range of surprising phenotypes that are manifested in the germ line, including sterility, a high mutation rate, and a high frequency of chromosomal aberration and nondisjunction (Figure 15-17). These hybrid progeny are *dysgenic,* or biologically deficient (hence, the expression *hybrid dysgenesis*). Interestingly, the reciprocal cross, P (female) × M (male), produces no dysgenic offspring. An important observation is that a large percentage of the dysgenically induced mutations are unstable; that is, they revert to wild type or to other mutant alleles at very high frequencies. This instability is generally restricted to the germ line of an individual fly possessing an M cytotype by a mechanism explained below.

The unstable *Drosophila* mutants had similarities to the unstable maize mutants characterized by McClintock. Investigators hypothesized that the dysgenic mutations are caused by the insertion of transposable elements into specific genes, thereby rendering them inactive. According to this view, reversion would usually result from the excision of these inserted sequences. This hypothesis has been critically tested by isolating unstable dysgenic mutations at the eye-color locus *white.* Most of the mutations were found to be caused by the insertion of a transposable element into the *white*$^+$ gene. The element, called the *P element,* was found to be present in from 30 to 50 copies per genome in P strains but to be completely absent in M strains.

Why do *P* elements not cause trouble in P strains? The simple answer is that *P*-element transposition is repressed in P strains. At first it was thought that repression was due to a polypeptide repressor that was in P but not M strains. This model is no longer favored. Instead, geneticists now think that all the trans-

posase genes in *P* elements are silenced in P strains. The genes are activated in the F_1 generation as shown in Figure 15-18. Gene silencing has been discussed previously (see Chapters 8 and 12) and will be revisited at the end of this chapter. For some reason, most laboratory strains have no *P* elements, and consequently the silencing mechanism is not activated. In hybrids from the cross M (female, no *P* elements) × P (male, *P* elements), the *P* elements in the newly formed zygote are in a silencing-free environment. The *P* elements derived from the male genome can now transpose throughout the diploid genome, causing a variety of damage as they insert into genes and cause mutations. These molecular events are expressed as the various manifestations of hybrid dysgenesis. On the other hand, P (female) × M (male) crosses do not result in dysgenesis, because, presumably, the egg cytoplasm contains the components required for silencing the *P*-element transposase.

An intriguing question remains unanswered: Why do laboratory strains lack *P* elements, whereas strains in the wild have *P* elements? One hypothesis is that most of the current laboratory strains descended from the original isolates taken from the wild by Morgan and his students almost a century ago. At some point between the capture of those original strains and the present, *P* elements spread

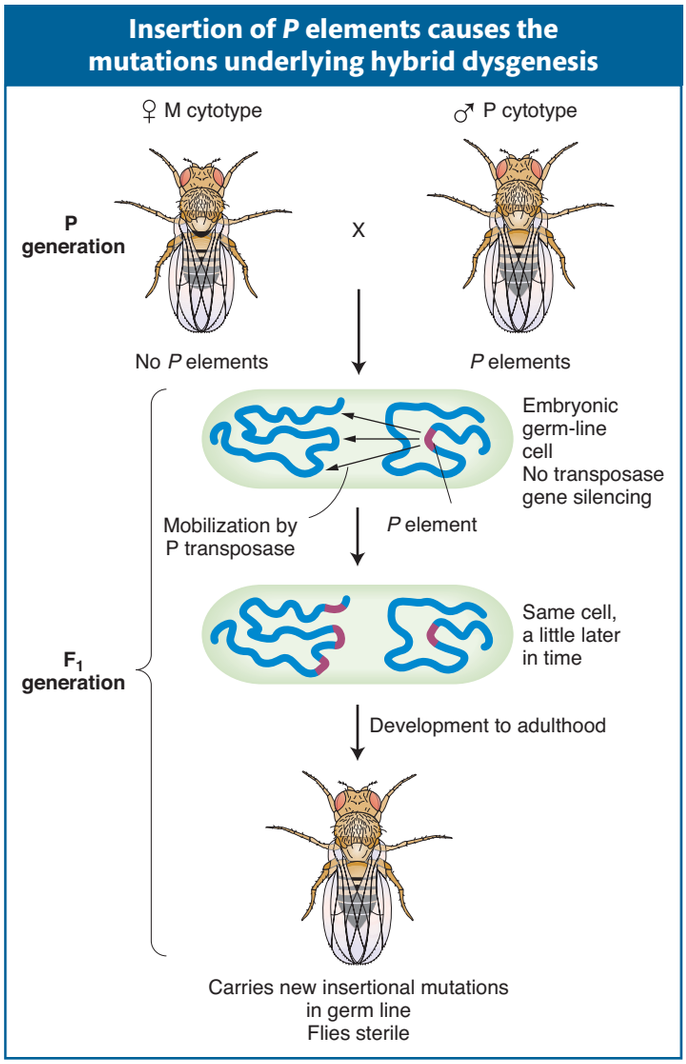


Figure 15-18 Molecular events underlying hybrid dysgenesis. Crosses of male *Drosophila* bearing P transposase with female *Drosophila* that do not have functional *P* elements produce mutations in the germ line of F_1 progeny caused by *P*-element insertions. *P* elements are able to move, causing mutations, because the egg does not silence the transposase gene.

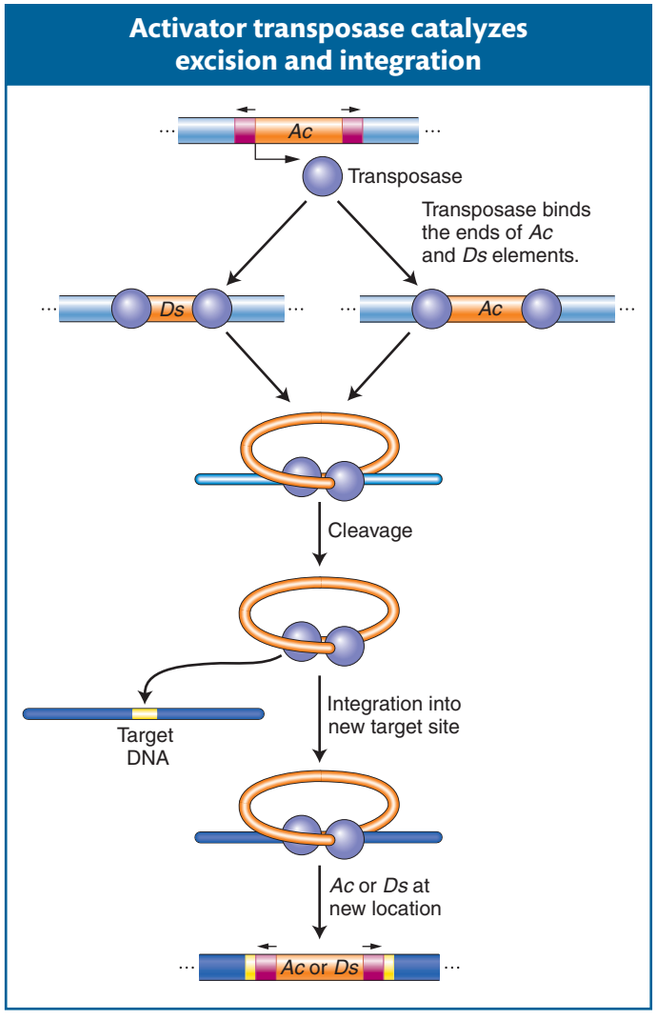


Figure 15-19 The *Ac* element in maize encodes a transposase that binds its own ends or those of a *Ds* element, excising the element, cleaving the target site, and allowing the element to insert elsewhere in the genome.

through natural populations but not through laboratory strains. This difference was not noticed until wild strains were again captured and mated with laboratory strains.

Although the exact scenario of how *P* elements have spread throughout wild populations is not clear, what is clear is that transposable elements can spread rapidly from a few individual members of a population. In this regard, the spread of *P* elements resembles the spread of transposons carrying resistance genes to formerly susceptible bacterial populations.

Maize transposable elements revisited Although the causative agent responsible for unstable mutants was first shown genetically to be transposable elements in maize, it was almost 50 years before the maize *Ac* and *Ds* elements were isolated and shown to be related to DNA transposons in bacteria and in other eukaryotes. Like the *P* element of *Drosophila, Ac* has terminal inverted repeats and encodes a single protein, the transposase. The nonautonomous *Ds* element does not encode transposase and thus cannot transpose on its own. When *Ac* is in the genome, its transposase can bind to both ends of *Ac* or *Ds* elements and promote their transposition (Figure 15-19).

As noted earlier in the chapter, *Ac* and *Ds* are members of a single transposon family, and there are other families of transposable elements in maize. Each family contains an autonomous element encoding a transposase that can move elements in the same family but cannot move elements in other families, because the transposase can bind only to the ends of family members.

Although some organisms such as yeast have no DNA transposons, elements structurally similar to the *P* and *Ac* elements have been isolated from many plant and animal species.

Message The first known transposable elements in maize are DNA transposons that structurally resemble DNA transposons in bacteria and other eukaryotes. DNA transposons encode a transposase that cuts the transposon from the chromosome and catalyzes its reinsertion at other chromosomal locations.

Utility of DNA transposons for gene discovery

Quite apart from their interest as a genetic phenomenon, DNA transposons have become major tools used by geneticists working with a variety of organisms. Their mobility has been exploited to tag genes for cloning and to insert transgenes. The *P* element in *Drosophila* provides one of the best examples of how geneticists exploit the properties of transposable elements in eukaryotes.

Using *P* elements to tag genes for cloning *P* elements can be used to create mutations by insertion, to mark the position of genes, and to facilitate the cloning of genes. *P* elements inserted into genes in vivo disrupt genes at random, creating mutants with different phenotypes. Fruit flies with interesting mutant phenotypes can be selected for cloning of the mutant gene, which is marked by the presence of the *P* element, a method termed **transposon tagging.** After the interrupted gene has been cloned, fragments from the mutant gene can be used as a probe to isolate the wild-type gene.

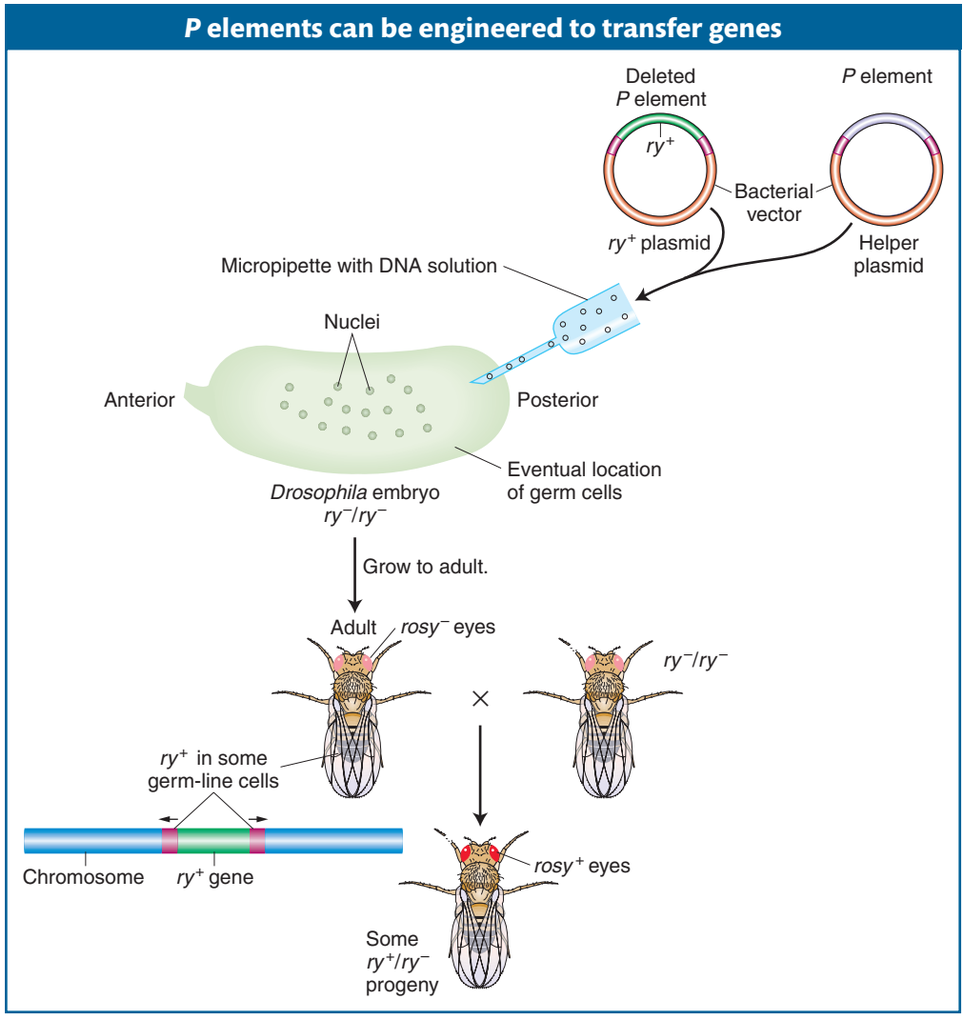


Figure 15-20 *P*-element-mediated gene transfer in *Drosophila*. The *rosy*$^+$ (*ry*$^+$) eye-color gene is engineered into a deleted *P* element carried on a bacterial vector. At the same time, a helper plasmid bearing an intact *P* element is used. Both are injected into an *ry*$^-$ embryo, where *ry*$^+$ transposes with the *P* element into the chromosomes of the germ-line cells.

Using *P* elements to insert genes Gerald Rubin and Allan Spradling showed that *P*-element DNA can be an effective vehicle for transferring donor genes into the germ line of a recipient fly. They devised the following experimental procedure (Figure 15-20). Suppose the goal is to transfer the allele *ry*$^+$, which confers a characteristic eye color, into the fly genome. The recipient genotype is homozygous for the *rosy* (*ry*$^-$) mutation. From this strain, embryos are collected at the completion of about nine nuclear divisions. At this stage, the embryo is one multinucleate cell, and the nuclei destined to form the germ cells are clustered at one end. (*P* elements mobilize only in germ-line cells.) Two types of DNA are injected into embryos of this type. The first is a bacterial plasmid carrying a defective *P* element into which the *ry*$^+$ gene has been inserted. The defective *P* element resembles the maize *Ds* element in that it does not encode transposase but still has the ends that bind transposase and allow transposition. This deleted element is not able to transpose, and so, as mentioned earlier, a helper plasmid encoding transposase also is injected. Flies developing from these embryos are phenotypically still *rosy* mutants, but their offspring include a large proportion of *ry*$^+$ flies. In situ hybridization confirmed that the *ry*$^+$ gene, together with the deleted *P* element, was inserted into one of several distinct chromosome locations. None appeared exactly at the normal locus of the *rosy* gene. These new *ry*$^+$ genes are found to be inherited in a stable, Mendelian fashion.

Because the *P* element can transpose only in *Drosophila,* these applications are restricted to these flies. In contrast, the maize *Ac* element is able to transpose after its introduction into the genomes of many plant species, including the mustard weed *Arabidopsis,* lettuce, carrot, rice, and barley. Like *P* elements, *Ac* has been engineered by geneticists for use in gene isolation by transposon tagging. In this way, *Ac,* the first transposable element discovered by Barbara McClintock, serves as an important tool of plant geneticists more than 50 years later.

Message DNA transposons have been modified and used by scientists in two important ways: (1) to make mutants that can be identified molecularly by the presence of a transposon tag and (2) as vectors that can introduce foreign genes into a chromosome.

15.4 The Dynamic Genome: More Transposable Elements Than Ever Imagined

As you have seen, transposable elements were first discovered with the use of genetic approaches. In these studies, the elements made their presence known when they transposed into a gene or were sites of chromosome breakage or rearrangement. After the DNA of transposable elements was isolated from unstable mutations, scientists could use that DNA as molecular probes to determine if there were more related copies in the genome. In all cases, at least several copies of the element were always present in the genome and, in some cases, as many as several hundred.

Scientists wondered about the prevalence of transposable elements in genomes. Were there other transposable elements in the genome that remained unknown because they had not caused a mutation that could be studied in the laboratory? Were there transposable elements in the vast majority of organisms that were not amenable to genetic analysis? Asked another way, do organisms without mutations induced by transposable elements nonetheless have transposable elements in their genomes? These questions are reminiscent of the question, If a tree falls in the forest, does it make a sound if no one is listening?

Large genomes are largely transposable elements

Long before the advent of DNA-sequencing projects, scientists using a variety of biochemical techniques discovered that DNA content (called ***C*-value**) varied dramatically in eukaryotes and did not correlate with biological complexity. For example, the genomes of salamanders are 20 times as large as the human genome, whereas the genome of barley is more than 10 times as large as the genome of rice, a related grass. The lack of correlation between genome size and the biological complexity of an organism is known as the ***C*-value paradox.**

Barley and rice are both cereal grasses, and, as such, their gene content should be similar. However, if genes are a relatively constant component of the genomes of multicellular organisms, what is responsible for the additional DNA in the larger genomes? On the basis of the results of additional experiments, scientists were able to determine that DNA sequences that are repeated thousands, even hundreds of thousands, of times make up a large fraction of eukaryotic genomes and that some genomes contain much more repetitive DNA than others.

Thanks to many recent projects to sequence the genomes of a wide variety of taxa (including *Drosophila,* humans, the mouse, *Arabidopsis,* and rice), we now know that there are many classes of repetitive sequences in the genomes of higher organisms and that some are similar to the DNA transposons and retrotransposons shown to be responsible for mutations in plants, yeast, and insects. Most

Types of transposable elements in the human genome

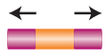
Element	Transposition	Structure	Length	Copy number	Fraction of genome
LINEs	Autonomous		1–5 kb	20,000–40,000	21%
SINEs	Nonautonomous		100–300 bp	1,500,000	13%
DNA transposons	Autonomous		2–3 kb	300,000	3%
	Nonautonomous		80–3000 bp		

Figure 15-21 Several general classes of transposable elements are found in the human genome. [*Reprinted by permission from Nature 409, 880 (15 February 2001), "Initial Sequencing and Analysis of the Human Genome," The International Human Genome Sequencing Consortium. Copyright 2001 by Macmillan Magazines Ltd.*]

remarkably, these sequences make up most of the DNA in the genomes of multicellular eukaryotes.

Rather than correlating with gene content, genome size frequently correlates with the amount of DNA in the genome that is derived from transposable elements. Organisms with big genomes have lots of sequences that resemble transposable elements, whereas organisms with small genomes have many fewer. Two examples, one from the human genome and the other from a comparison of the grass genomes, illustrate this point. The structural features of the transposable elements that are found in human genomes are summarized in Figure 15-21 and will be referred to in the next section.

Message The *C*-value paradox is the lack of correlation between genome size and biological complexity. Genes make up only a small proportion of the genomes of multicellular organisms. Genome size usually corresponds to the amount of transposable-element sequences rather than to gene content.

Transposable elements in the human genome

Almost half of the human genome is derived from transposable elements. The vast majority of these transposable elements are two types of retrotransposons called **long interspersed elements,** or **LINEs,** and **short interspersed elements,** or **SINEs** (see Figure 15-21). LINEs move like a retrotransposon with the help of an element-encoded reverse transcriptase but lack some structural features of retrovirus-like elements, including LTRs (see Figure 15-12d). SINEs can be best described as nonautonomous LINEs, because they have the structural features of LINEs but do not encode their own reverse transcriptase. Presumably, they are mobilized by reverse transcriptase enzymes that are encoded by LINEs residing in the genome.

The most abundant SINE in humans is called ***Alu,*** so named because it contains a target site for the Alu restriction enzyme. The human genome contains more than 1 million whole and partial *Alu* sequences, scattered between genes and within introns. These *Alu* sequences make up more than 10 percent of the human genome. The full *Alu* sequence is about 200 nucleotides long and bears remarkable resemblance to 7SL RNA, an RNA that is part of a complex by which newly synthesized polypeptides are secreted through the endoplasmic reticulum. Presumably, the *Alu* sequences originated as reverse transcripts of these RNA molecules.

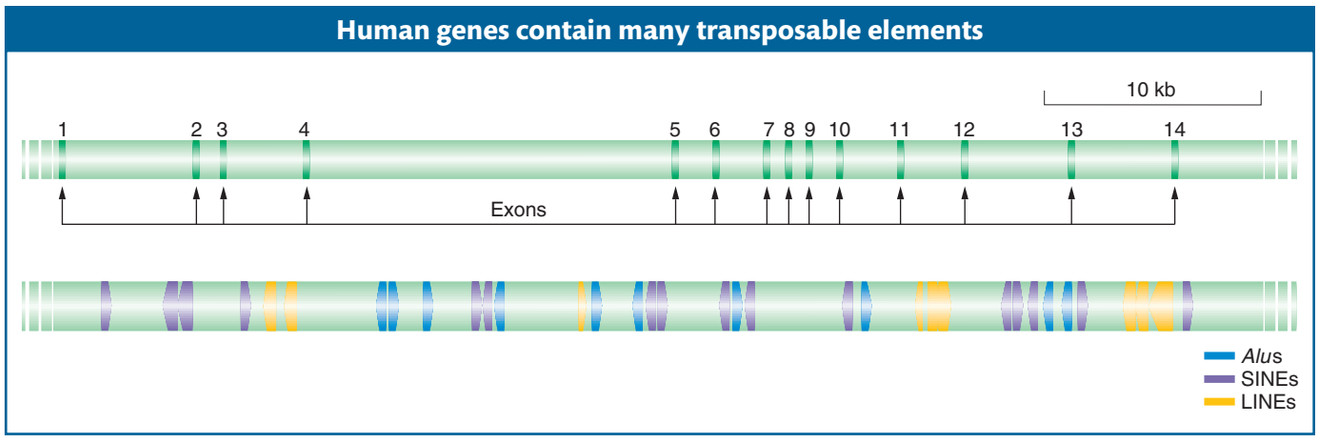


Figure 15-22 Numerous repetitive elements are found in the human gene (*HGO*) encoding homogentisate 1,2-dioxygenase, the enzyme whose deficiency causes alkaptonuria. The upper row diagrams the positions of the *HGO* exons. The locations of *Alu*s (blue), other SINEs (purple), and LINEs (yellow) in the *HGO* sequence are indicated in the lower row. [*After B. Granadino, D. Beltrán-Valero de Bernabé, J. M. Fernández-Cañón, M. A. Peñalva, and S. Rodríguez de Córdoba, "The Human Homogentisate 1,2-dioxygenase (HGO) Gene," Genomics 43, 1997, 115.*]

There is about 20 times as much DNA in the human genome derived from transposable elements as there is DNA encoding all human proteins. Figure 15-22 illustrates the number and diversity of transposable elements present in the human genome, using as an example the positions of individual *Alu*s, other SINEs, and LINEs in the vicinity of a typical human gene.

The human genome seems to be typical for a multicellular organism in the abundance and distribution of transposable elements. Thus, an obvious question is, How do plants and animals survive and thrive with so many insertions in genes and so much mobile DNA in the genome? First, with regard to gene function, all of the elements shown in Figure 15-22 are inserted into introns. Thus, the mRNA produced by this gene will not include any sequences from transposable elements, because they will have been spliced out of the pre-mRNA with the surrounding intron. Presumably, transposable elements insert into both exons and introns, but only the insertions into introns will remain in the population because they are less likely to cause a deleterious mutation. Insertions into exons are said to be subjected to **negative selection.** Second, humans, as well as all other multicellular organisms, can survive with so much mobile DNA in the genome because the vast majority is inactive and cannot move or increase in copy number. Most transposable-element sequences in a genome are relics that have accumulated inactivating mutations through evolutionary time. Others are still capable of movement but are rendered inactive by host regulatory mechanisms (see Section 15.5). There are, however, a few active LINEs and *Alu*s that have managed to escape host control and have inserted into important genes, causing several human diseases. Three separate insertions of LINEs have disrupted the factor VIII gene, causing hemophilia A. At least 11 *Alu* insertions into human genes have been shown to cause several diseases, including hemophilia B (in the factor IX gene), neurofibromatosis (in the *NF1* gene), and breast cancer (in the *BRCA2* gene).

The overall frequency of spontaneous mutation due to the insertion of class 2 elements in humans is quite low, accounting for less than 0.2 percent (1 in 500) of all characterized spontaneous mutations. Surprisingly, retrotransposon insertions account for about 10 percent of spontaneous mutations in another mammal, the mouse. The approximately 50-fold increase in this type of mutation in the mouse most likely corresponds to the much higher activity of these elements in the mouse genome than in the human genome.

Message Transposable elements compose the largest fraction of the human genome, with LINEs and SINEs being the most abundant. The vast majority of transposable elements are ancient relics that can no longer move or increase their copy number. A few elements remain active, and their movement into genes can cause disease.

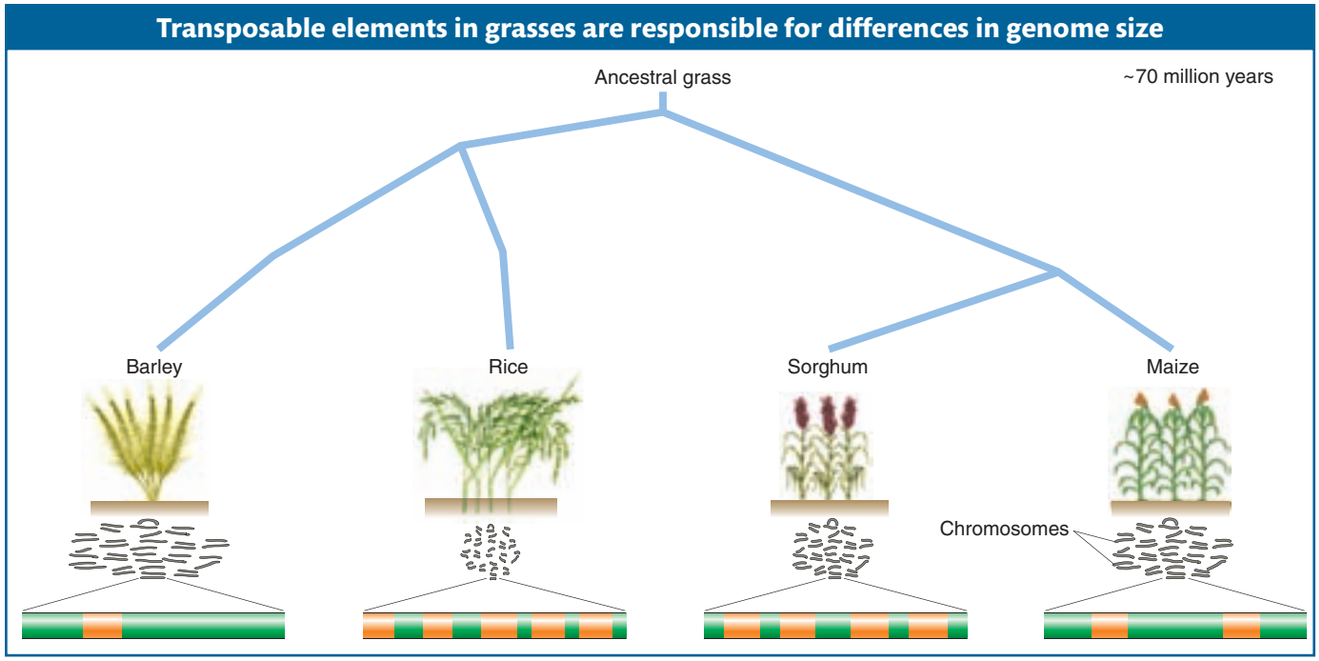


Figure 15-23 The grasses, including barley, rice, sorghum, and maize, diverged from a common ancestor about 70 million years ago. Since that time, the transposable elements have accumulated to different levels in each species. Chromosomes are larger in maize and barley, whose genomes contain large amounts of LTR retrotransposons. Green in the partial genome at the bottom represents a cluster of transposons, whereas orange represents genes.

The grasses: LTR retrotransposons thrive in large genomes

As already mentioned, the *C*-value paradox is the lack of correlation between genome size and biological complexity. How can organisms have very similar gene content but differ dramatically in the size of their genomes? This situation has been investigated in the cereal grasses. Differences in the genome sizes of these grasses have been shown to correlate primarily with the number of one class of elements, the LTR retrotransposons. The cereal grasses are evolutionary relatives that have arisen from a common ancestor in the past 70 million years. As such, their genomes are still very similar with respect to gene content and organization (called **synteny;** see Chapter 14), and regions can be compared directly. These comparisons reveal that linked genes in the small rice genome are physically closer together than are the same genes in the larger maize and barley genomes. In the maize and barley genomes, genes are separated by large clusters of retrotransposons (Figure 15-23).

Safe havens

The abundance of transposable elements in the genomes of multicellular organisms led some investigators to postulate that successful transposable elements (those that are able to attain very high copy numbers) have evolved mechanisms to prevent harm to their hosts by not inserting into host genes. Instead, successful transposable elements insert into so-called **safe havens** in the genome. For the grasses, a safe haven for new insertions appears to be into other retrotransposons. Another safe haven is the heterochromatin of centromeres, where there are very few genes but lots of repetitive DNA (see Chapter 12 for more on heterochromatin). Many classes of transposable elements in both plant and animal species tend to insert into the centric heterochromatin.

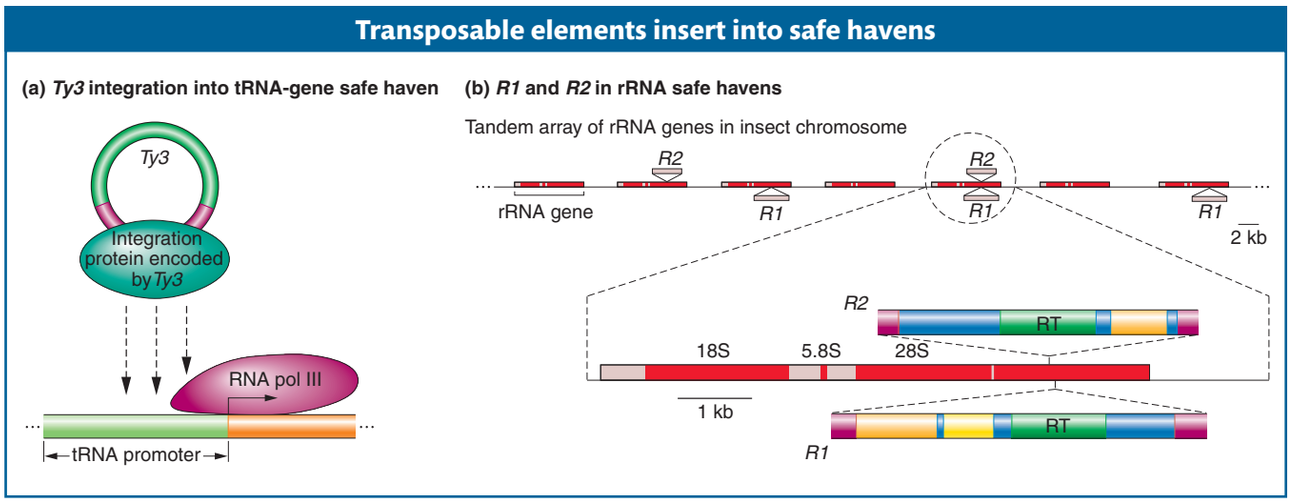


Figure 15-24 Some transposable elements are targeted to specific safe havens. (a) The yeast *Ty3* retrotransposon inserts into the promoter region of transfer RNA genes. (b) The *Drosophila R1* and *R2* non-LTR retrotransposons (LINEs) insert into the genes encoding ribosomal RNA that are found in long tandem arrays on the chromosome. Only the reverse transcriptase (RT) genes of *R1* and *R2* are noted. [*(a) Inspired by D. F. Voytas and J. D. Boeke, "Ty1 and Ty5 of Saccharomyces cerevisiae," in N. L. Craig et al., eds., Mobile DNA II, Chap. 26, Fig. 15, p. 652. ASM Press, 2002. (b) After T. H. Eickbush, "R2 and Related Site-Specific Non-Long Terminal Inverted Repeat Retrotransposons," in N. L. Craig et al., eds., Mobile DNA II, Chap. 34, Fig. 1, p. 814. ASM Press, 2002.*]

Safe havens in small genomes: targeted insertions In contrast with the genomes of multicellular eukaryotes, the genome of unicellular yeast is very compact, with closely spaced genes and very few introns. With almost 70 percent of its genome as exons, there is a high probability that new insertions of transposable elements will disrupt a coding sequence. Yet, as we have seen earlier in this chapter, the yeast genome supports a collection of LTR-retrotransposons called *Ty* elements.

How are transposable elements able to spread to new sites in genomes with few safe havens? Investigators have identified hundreds of *Ty* elements in the sequenced yeast genome and have determined that they are not randomly distributed. Instead, each family of *Ty* elements inserts into a particular genomic region. For example, the *Ty3* family inserts almost exclusively near but not in tRNA genes, at sites where they do not interfere with the production of tRNAs and, presumably, do not harm their hosts. *Ty* elements have evolved a mechanism that allows them to insert into particular regions of the genome: *Ty* proteins necessary for integration interact with specific yeast proteins bound to genomic DNA. *Ty3* proteins, for example, recognize and bind to subunits of the RNA polymerase complex that have assembled at tRNA promoters (Figure 15-24a).

The ability of some transposons to insert preferentially into certain sequences or genomic regions is called **targeting.** A remarkable example of targeting is illustrated by the *R1* and *R2* elements of arthropods, including *Drosophila. R1* and *R2* are LINEs (see Figure 15-21) that insert only into the genes that produce ribosomal RNA. In arthropods, several hundred rRNA genes are organized in tandem arrays (Figure 15-24b). With so many genes encoding the same product, the host tolerates insertion into a subset. However, too many insertions of *R1* and *R2* have been shown to decrease insect viability, presumably by interfering with ribosome assembly.

Gene therapy revisited This chapter began with a description of a recessive genetic disorder called SCID (severe combined immunodeficiency disease). The immune systems of persons afflicted with SCID are severely compromised owing to a mutation in a gene encoding the enzyme adenosine deaminase. To correct this genetic defect, bone-marrow cells from SCID patients were collected and treated with a retrovirus vector containing a good ADA gene. The transformed cells were then infused back into the patients. The immune systems of most of the patients showed significant improvement. However, the therapy had a very serious side effect: two of the patients developed leukemia. In both patients, the retroviral vector had inserted (integrated) near a cellular gene whose aberrant

expression is associated with leukemia. A likely scenario is that insertion of the retroviral vector near the cellular gene altered its expression and, indirectly, caused the leukemia.

Clearly, the serious risk associated with this form of gene therapy might be greatly improved if doctors were able to control where the retroviral vector integrates into the human genome. We have already seen that there are many similarities between LTR retrotransposons and retroviruses. It is hoped that, by understanding *Ty* targeting in yeast, we can learn how to construct retroviral vectors that insert themselves and their transgene cargo into safe havens in the human genome.

Message A successful transposable element increases copy number without harming its host. One way in which an element safely increases copy number is to target new insertions into safe havens, regions of the genome where there are few genes.

15.5 Epigenetic Regulation of Transposable Elements by the Host

Genetic analysis is a very powerful tool used to dissect complex biological processes through the isolation of mutants and, ultimately, mutant genes (for example, see Chapter 13 on development). Many laboratories around the world are using genetic analyses to identify the host genes responsible for repressing the movement of transposable elements and, in this way, maintaining the stability of the genome.

The repression of transposable elements was first investigated in the laboratory of Ron Plasterk in the late 1990s, using the model organism *C. elegans* (a nematode; see the Model Organism box in Chapter 13). This story starts with the observation of a striking difference between the mobility of a transposable element called *Tc*1 in two different cell types of the model organism. *Tc*1 is a DNA transposon that, like the *Ac* element of maize, can lead to an unstable mutant phenotype when it excises from a gene with a visible phenotype (see Figure 15-4). There are 32 *Tc*1 elements in the sequenced genome of the common laboratory strain called Bristol. Significantly, *Tc*1 transposes in somatic but not germ-line cells. That observation suggested to Plasterk that transposition is repressed in the germ line by the host. Evidently, germ-line repression results from the silencing of the transposase genes of all 32 *Tc*1 copies in germ-line cells. Can you propose an explanation for why it makes biological sense for a host to repress transposition in the germ line and not in the soma?

Plasterk and his co-workers set out to identify *C. elegans* genes responsible for silencing the transposase gene. They began with a *C. elegans* strain that had *Tc*1 inserted in the *unc-22* gene (designated *unc-22/Tc*1; Figure 15-25). This was the same gene that was silenced in the experiment of Fire and Mello that led to their sharing the Nobel Prize (see Chapter 8). Whereas wild-type *C. elegans* glides smoothly on the surface of the agar in a petri dish, worms with the mutant *unc-22/Tc*1 gene have a twitching movement that can be easily observed with a microscope. Because *Tc*1 cannot normally transpose in the germ line, it remains inserted in the *unc-22*

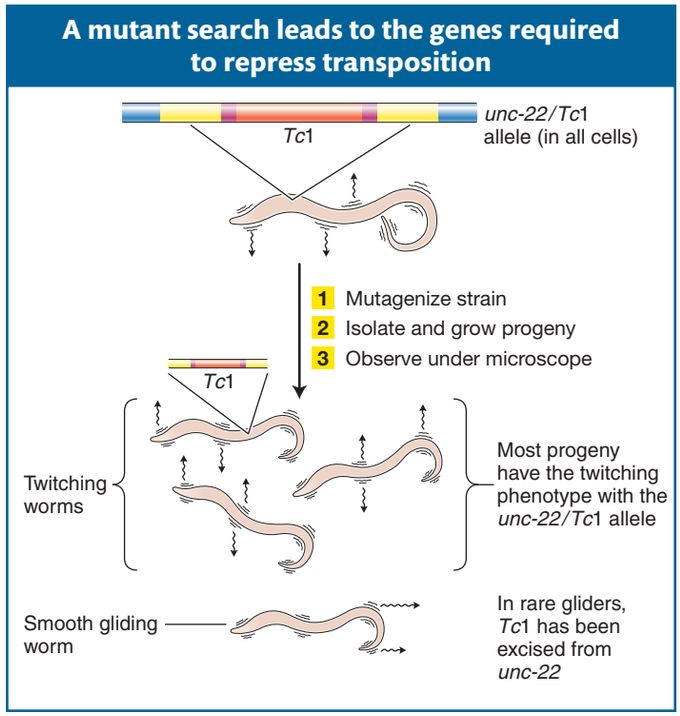


Figure 15-25 Experimental design used to identify genes required to repress transposition. Investigators look for mutants that have regained normal movement, because mutations in these individuals would have disabled the repression mechanism that prevents the transposition of the *Tc*1 element from the *unc-22* gene.

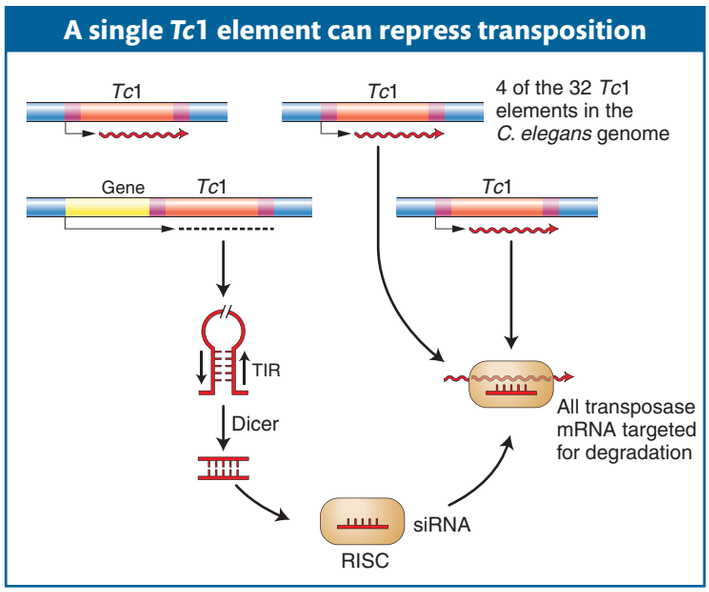


Figure 15-26 The production of dsRNA from only a single *Tc*1 element is sufficient to silence all of the *Tc*1 transposase genes and thereby repress transposition in the germ line. The siRNA derived from *Tc*1 dsRNA is bound to RISC and targets all complementary RNA for degradation.

gene and continues to disrupt its function. Thus, the strain with the mutant *unc-22/Tc*1 gene should express a twitching phenotype from generation to generation. However, Plasterk and co-workers reasoned that mutations that inactivated *C. elegans* genes required for repression would allow *Tc*1 to excise from the *unc-22/Tc*1 allele in the germ line and revert the twitching phenotype to wild type (*unc-22*). To this end, they exposed the mutant *unc-22/Tc*1 strain to a chemical that greatly increased the frequency of mutation (called a mutagen; see Chapter 16) and examined their progeny under a microscope, searching for rare worms that no longer twitched.

This and subsequent genetic screens identified over 25 *C. elegans* genes that, when mutated, allowed the host to excise *Tc*1 in the germ line. Significantly, many of the products of these genes are integral components of the RNAi silencing pathway, including proteins found in Dicer and RISC (see Chapters 8 and 12). Recall from Chapter 8 that Dicer binds to long dsRNAs and cleaves them into small dsRNA fragments. These fragments are then unwound so that one strand, the siRNA, can target RISC to chop up complementary mRNAs (see Figure 8-24).

Beginning with this elegant genetic screen, many years of experimentation have led to the following model for the repression of transposable elements in the germ line of *C. elegans*. With 32 *Tc*1 elements scattered throughout the *C. elegans* genome, a few elements near genes are transcribed along with the nearby gene by "read-through" transcription (see Chapter 8). Because the ends of *Tc*1 are 54-bp terminal inverted repeats, the *Tc*1 RNA spontaneously forms dsRNA (Figure 15-26). Like all dsRNAs produced in most eukaryotes, this RNA is recognized by Dicer and ultimately siRNA is produced, which directs RISC to chop up complementary *Tc*1 transcripts. Because all *Tc*1 RNA is efficiently chopped up in the germ line, the element-encoded transposase gene is silenced. Without transposase, the element cannot excise. It has been hypothesized that *Tc*1 can transpose in somatic cells because RNAi is not as efficient and some transposase can be produced.

Over the past decade, numerous laboratories working with both plants and animals have discovered that mutations that disrupt the RNAi pathway often lead to the activation of transposable elements in their respective genomes. Because of the abundance of transposable elements in eukaryotic genomes, it has been suggested that the natural function of the RNAi pathway is to maintain genome stability by repressing the movement of transposable elements.

Message Eukaryotic hosts use RNAi to repress the expression of active transposable elements in their genomes. In this way, a single element that inserts near a gene can be transcribed to produce dsRNA that will trigger the silencing of all copies of the element in the genome.

What Geneticists Are Doing Today

The RNAi silencing pathway is akin to radar in that the host is able to detect new insertions into the genome. The host can then respond by silencing the transposase gene and thus prevent the movement of all family members. However, much like planes that evade radar by flying close to the ground, some transposons have evolved mechanisms that allow them to evade the RNAi silencing pathway. These transposons can attain very high copy numbers. Evidence for these mechanisms can be found in the genomes of all characterized plants and animals containing transposon families (such as *Alu* in humans) that have thousands of members.

How do some transposons avoid detection by the RNAi silencing pathway? The short answer is that in most cases, we do not know. To understand how a transposon avoids detection, it is necessary to study actively transposing elements. To date, scientists have detected very few transposon families with high copy numbers that are still actively transposing. One of the best-characterized elements among this small group is a special type of nonautonomous DNA transposon called a **miniature inverted repeat transposable element** (abbreviated **MITE**). Like other nonautonomous elements, MITEs can form by deletion of the transposase gene from an autonomous element. However, unlike most nonautonomous elements, MITEs can attain very high copy numbers, particularly in the genomes of some grasses (see Figure 15-23). Some MITEs in grasses have been amplified to thousands of copies.

The only actively transposing MITE isolated to date is the *mPing* element of rice, which is formed from the autonomous *Ping* element by deletion of the entire transposase gene (Figure 15-27). This element was discovered in the laboratory of Susan Wessler by Ning Jiang. Another member of the Wessler laboratory, Ken Naito, documented that in individuals of some rice strains there are only 3 to 7 copies of *Ping* and over 1000 copies of *mPing*. Remarkably, the copy number of *mPing* in these strains is increasing by almost 40 new insertions per plant per generation.

Two questions about the rapid increase in *mPing* copy number immediately come to mind. First, how does a rice strain survive a transposable-element burst of this magnitude? To address this question, the Wessler laboratory used next-generation sequencing technology (see Chapter 14) to determine the insertion sites of over 1700 *mPing* elements in the rice genome. Surprisingly, they found that the element avoided inserting into exons, thus minimizing the impact of insertion on rice gene expression. The mechanism underlying this preference is currently being investigated.

The second question is, why does the rice host apparently fail to repress *mPing* transposition? While this question is also an active area of current research, a reasonable hypothesis is that *mPing* can fly under the hosts' RNAi radar because it does not contain any part of the transposase gene that resides on the *Ping* element (Figure 15-27). Thus, read-through transcription into *mPing* elements inserted throughout the rice genome will produce lots of dsRNA and siRNA. However, because siRNAs derived from *mPing* share no sequence with the source of transposase, siRNAs produced from *mPing* will not induce silencing mechanisms aimed at transposase. Instead, the transposase gene will remain active and will continue to catalyze the movement of *mPing*. According to this hypothesis, *mPing* transposition will be repressed only when a much rarer *Ping* insertion generates dsRNA that triggers the silencing of its transposase gene.

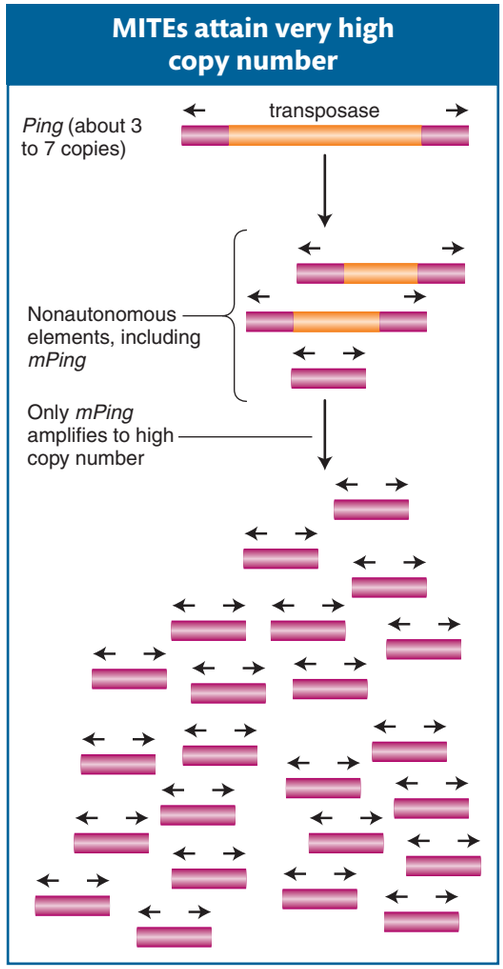


Figure 15-27 MITEs are nonautonomous DNA transposons that can attain very high copy number because they do not encode the transposase necessary for their transposition. The active MITE *mPing* is the only deletion derivative of the autonomous *Ping* element that has attained high copy number in certain strains of rice.

Message MITEs are nonautonomous DNA transposons that can attain high copy numbers. While MITEs can utilize the transposase of autonomous elements, they probably evade host repression because their amplification does not lead to the silencing of the transposase gene.

Summary

Transposable elements were discovered in maize by Barbara McClintock as the cause of several unstable mutations. An example of a nonautonomous element is *Ds*, the transposition of which requires the presence of the autonomous *Ac* element in the genome.

Bacterial insertion-sequence elements were the first transposable elements isolated molecularly. There are many different types of IS elements in *E. coli* strains, and they are usually present in at least several copies. Composite transposons contain IS elements flanking one or more

genes, such as genes conferring resistance to antibiotics. Transposons with resistance genes can insert into plasmids and are then transferred by conjugation to nonresistant bacteria.

There are two major groups of transposable elements in eukaryotes: class 1 elements (retrotransposons) and class 2 elements (DNA transposons). The *P* element was the first class 2 DNA transposon to be isolated molecularly. It was isolated from unstable mutations in *Drosophila* that were induced by hybrid dysgenesis. *P* elements have been developed into vectors for the introduction of foreign DNA into *Drosophila* germ cells.

Ac, Ds, and *P* are examples of DNA transposons, so named because the transposition intermediate is the DNA element itself. Autonomous elements such as *Ac* encode a transposase that binds to the ends of autonomous and nonautonomous elements and catalyzes excision of the element from the donor site and reinsertion into a new target site elsewhere in the genome.

Retrotransposons were first molecularly isolated from yeast mutants, and their resemblance to retroviruses was immediately apparent. Retrotransposons are class 1 elements, as are all transposable elements that use RNA as their transposition intermediate.

The active transposable elements isolated from such model organisms as yeast, *Drosophila, E. coli,* and maize constitute a very small fraction of all the transposable elements in the genome. DNA sequencing of whole genomes, including the human genome, has led to the remarkable finding that almost half of the human genome is derived from transposable elements. Despite having so many transposable elements, eukaryotic genomes are extremely stable as transposition is relatively rare because of two factors. First, most of the transposable elements in eukaryotic genomes cannot move because inactivating mutations prevent the production of normal transposase and reverse transcriptase. Second, expression of the vast majority of the remaining elements is silenced by the host RNAi pathway. Some high-copy-number elements, such as MITEs, may evade silencing because they do not trigger the silencing of the transposase that catalyzes their transposition.

KEY TERMS

Activator (*Ac*) (p. 524)
Alu (p. 543)
autonomous element (p. 527)
class 1 element (retrotransposon) (p. 536)
class 2 element (DNA transposon) (p. 537)
cointegrate (p. 532)
composite transposon (p. 530)
conservative transposition (p. 531)
copia-like element (p. 536)
"copy and paste" (p. 536)
"cut and paste" (p. 531)
C-value (p. 542)
C-value paradox (p. 542)
Dissociation (*Ds*) (p. 524)
DNA transposon (p. 537)
excise (p. 527)
gene therapy (p. 523)
hybrid dysgenesis (p. 537)
insertion-sequence (IS) element (p. 529)
inverted repeat (p. 530)
long interspersed element (LINE) (p. 543)
long terminal repeat (LTR) (p. 534)
LTR-retrotransposon (p. 536)
M cytotype (p. 538)
miniature inverted repeat transposable element (MITE) (p. 549)
negative selection (p. 544)
nonautonomous element (p. 527)
P cytotype (p. 538)
P element (p. 537)
provirus (p. 534)
replicative transposition (p. 531)
retrotransposon (p. 536)
retrovirus (p. 530, 534)
reverse transcriptase (p. 534)
R factor (p. 530)
safe haven (p. 545)
short interspersed element (SINE) (p. 543)
simple transposon (p. 530)
solo LTR (p. 537)
synteny (p. 545)
targeting (p. 546)
target-site duplication (p. 533)
transposase (p. 529)
transpose (p. 527)
transposition (p. 530)
transposon (Tn) (p. 530)
transposon tagging (p. 540)
Ty element (p. 533)
unstable phenotype (p. 527)

SOLVED PROBLEMS

SOLVED PROBLEM 1. Transposable elements have been referred to as "jumping genes" because they appear to jump from one position to another, leaving the old locus and appearing at a new locus. In light of what we now know concerning the mechanism of transposition, how appropriate is the term "jumping genes" for bacterial transposable elements?

Solution

In bacteria, transposition takes place by two different modes. The conservative mode results in true jumping genes, because, in this case, the transposable element excises from its original position and inserts at a new position. The other mode is the replicative mode. In this pathway, a transposable element moves to a new loca-

tion by replicating into the target DNA, leaving behind a copy of the transposable element at the original site. When operating by the replicative mode, transposable elements are not really jumping genes, because a copy does remain at the original site.

SOLVED PROBLEM 2. Following from the question above, in light of what we now know concerning the mechanism of transposition, how appropriate is the term "jumping genes" for the vast majority of transposable elements in the human genome and in the genomes of most other mammals?

Solution

The vast majority of transposable elements in the characterized mammalian genomes are retrotransposons. In humans, two retrotransposons (the LINE called *L1* and the SINE called *Alu*) account for a whopping one-third of our entire genome. Like bacterial elements, retrotransposons do not excise from the original site, so they are not really jumping genes. Instead, the element serves as a template for the transcription of RNAs that can be reverse-transcribed by the enzyme reverse transcriptase into double-stranded cDNA. Each cDNA can potentially insert into target sites throughout the genome. Note that while both bacterial elements and retrotransposons do not leave the original site, their respective mechanisms of transposition are dramatically different. Finally, while LTR retrotransposons do not excise, they can become much shorter insertions due to the production of solo LTRs by recombination.

PROBLEMS

Most of the problems are also available for review/grading through the GENETICSPORTAL www.yourgeneticsportal.com.

WORKING WITH THE FIGURES

1. In the chapter-opening photograph of kernels on an ear of corn, what is the genetic basis of the following (**Hint:** Refer to Figure 15-4 for some clues):
   a. the fully pigmented kernel?
   b. the unpigmented kernels? Note that they can arise in two different ways.
2. In Figure 15-3a, what would the kernel phenotype be if the strain was homozygous for all dominant markers on chromosome 9?
3. For Figure 15-7, draw out a series of steps that could explain the origin of this large plasmid containing many transposable elements.
4. For Figure 15-8, draw a figure for the third mode of transposition, retrotransposition.
5. In Figure 15-10, show where the transposase would have to cut to generate a 6-bp target-site duplication. Also show the location of the cut to generate a 4-bp target-site duplication.
6. If the transposable element in Figure 15-14 were a DNA transposon that had an intron in its transposase gene, would the intron be removed following transposition? Justify your answer.
7. For Figure 15-22, draw the pre-mRNA that is transcribed from this gene and then draw its mRNA.

BASIC PROBLEMS

8. Describe the generation of multiple-drug-resistant plasmids.
9. Briefly describe the experiment that demonstrates that the transposition of the *Ty1* element in yeast takes place through an RNA intermediate.
10. Explain how the properties of *P* elements in *Drosophila* make gene-transfer experiments possible in this organism.
11. Although class 2 elements are abundant in the genomes of multicellular eukaryotes, class 1 elements usually make up the largest fraction of very large genomes such as those from humans (~2500 Mb), maize (~2500 Mb), and barley (~5000 Mb). Given what you know about class 1 and class 2 elements, what is it about their distinct mechanisms of transposition that would account for this consistent difference in abundance?
12. As you saw in Figure 15-22, the genes of multicellular eukaryotes often contain many transposable elements. Why do most of these elements not affect the expression of the gene?
13. What are safe havens? Are there any places in the much more compact bacterial genomes that might be a safe haven for insertion elements?
14. Nobel prizes are usually awarded many years after the actual discovery. For example, James Watson, Francis Crick, and Maurice Wilkens were awarded the Nobel Prize in Medicine or Physiology in 1962, almost a decade after their discovery of the double-helical structure of DNA. However, Barbara McClintock was awarded the Nobel Prize in 1983, almost four decades after her discovery of transposable elements in maize. Why do you think it took this long?

CHALLENGING PROBLEMS

15. The insertion of transposable elements into genes can alter the normal pattern of expression. In the following situations, describe the possible consequences on gene expression.

a. A LINE inserts into an enhancer of a human gene.

b. A transposable element contains a binding site for a transcriptional repressor and inserts adjacent to a promoter.

c. An *Alu* element inserts into the 3′ splice (AG) site of an intron.

d. A *Ds* element that was inserted into the exon of a gene excises imperfectly and leaves 3 base pairs behind in the exon.

e. Another excision by that same *Ds* element leaves 2 base pairs behind in the exon.

f. A *Ds* element that was inserted into the middle of an intron excises imperfectly and leaves 5 base pairs behind in the intron.

16. Before the integration of a transposon, its transposase makes a staggered cut in the host target DNA. If the staggered cut is at the sites of the arrows below, draw what the sequence of the host DNA will be after the transposon has been inserted. Represent the transposon as a rectangle.

```
        ↓
AATTTGGCCTAGTACTAATTGGTTGG
TTAAACCGGATCATGATTAACCAACC
                  ↑
```

17. In *Drosophila,* M. Green found a *singed* allele (*sn*) with some unusual characteristics. Females homozygous for this X-linked allele have singed bristles, but they have numerous patches of sn^+(wild-type) bristles on their heads, thoraxes, and abdomens. When these flies are mated with *sn* males, some females give only singed progeny, but others give both singed and wild-type progeny in variable proportions. Explain these results.

18. Consider two maize plants:

a. Genotype C/c^m ; Ac/Ac^+, where c^m is an unstable allele caused by *Ds* insertion

b. Genotype C/c^m, where c^m is an unstable allele caused by *A*c insertion

What phenotypes would be produced and in what proportions when (1) each plant is crossed with a base-pair-substitution mutant *c*/*c* and (2) the plant in part *a* is crossed with the plant in part *b*? Assume that *Ac* and *c* are unlinked, that the chromosome-breakage frequency is negligible, and that mutant *c*/*C* is Ac^+.

19. You meet your friend, a scientist, at the gym and she begins telling you about a mouse gene that she is studying in the lab. The product of this gene is an enzyme required to make the fur brown. The gene is called *FB* and the enzyme is called FB enzyme. When *FB* is mutant and cannot produce the FB enzyme, the fur is white. The scientist tells you that she has isolated the gene from two mice with brown fur and that, surprisingly, she found that the two genes differ by the presence of a 250-bp SINE (like the human *Alu* element) in the *FB* gene of one mouse but not in the gene of the other. She does not understand how this difference is possible, especially given that she determined that both mice make the FB enzyme. Can you help her formulate a hypothesis that explains why the mouse can still produce FB enzyme with a transposable element in its *FB* gene?

20. The yeast genome has class 1 elements (*Ty1, Ty2,* and so forth) but no class 2 elements. Can you think of a possible reason why DNA elements have not been successful in the yeast genome?

21. In addition to *Tc*1, the *C. elegans* genome contains other families of DNA transposons such as *Tc*2, *Tc*3, *Tc*4, and *Tc*5. Like *Tc*1, their transposition is repressed in the germ line but not in somatic cells. Predict the behavior of these elements in the mutant strains where *Tc*1 is no longer repressed due to mutations in the RNAi pathway. Justify your answer.

22. Based on the mechanism of gene silencing, what features of transposable elements does the RNAi pathway exploit to ensure that the host's own genes are not also silenced?

23. What are the similarities and differences between retroviruses and retrotransposons? It has been hypothesized that retroviruses evolved from retrotransposons. Do you agree with this model? Justify your answer.

24. You have isolated a transposable element from the human genome and have determined its DNA sequence. How would you use this sequence to determine the copy number of the element in the human genome if you just had a computer with an Internet connection? (**Hint:** see Chapter 14.)

25. Following up on the previous question, how would you determine whether other primates had a similar element in their genomes?

26. Of all the genes in the human genome, the ones with the most characterized *Alu* insertions are those that cause hemophilia, including several insertions in the factor VIII and factor IX genes. Based on this fact, your colleague hypothesizes that the *Alu* element prefers to insert into these genes. Do you agree? What other reason can you provide that also explains these data?

27. If all members of a transposable element family can be silenced by dsRNA synthesized from a single family member, how is it possible for one element family (like *Tc*1) to have 32 copies in the *C. elegans* genome while another family (*Tc*2) has fewer than 5 copies?