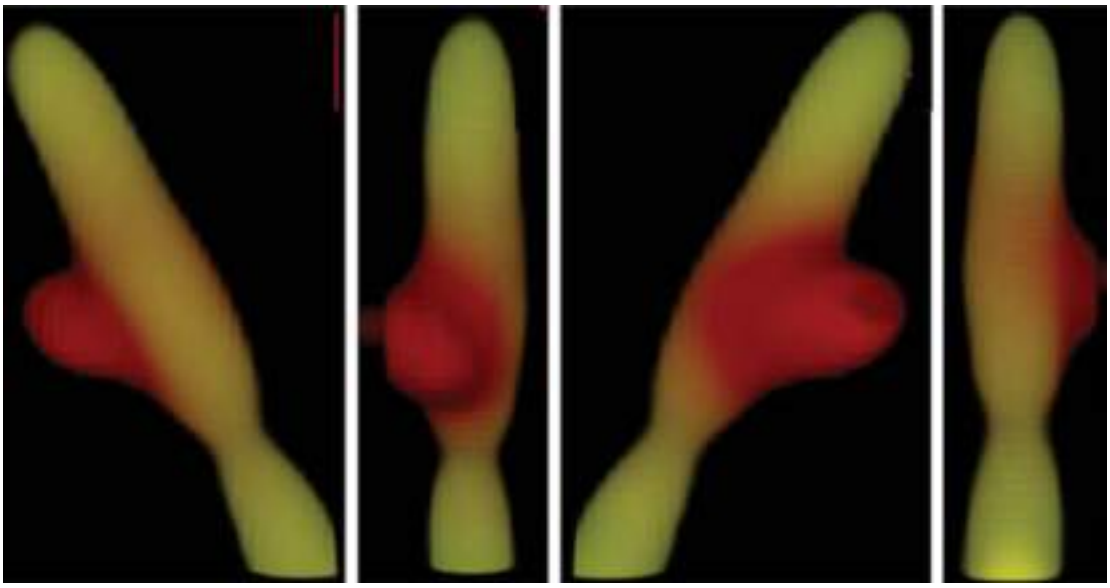
CHAPTER 9

Genetic Change and Genome Evolution



***Bacillus subtilis* in the process of importing naked DNA from the environment.** Epifluorescence microscopy was used to visualize yellow fluorescent protein–expressing *Bacillus subtilis* cells (yellow) coincubated with DNA (red) labeled with the fluorophore rhodamine. The series of panels shows successive 90° rotations of the field of view for this interaction. This method of incorporation of DNA into the genome of the recipient is referred to as transformation and is one of several means of horizontal gene transfer that microbes employ during evolution.

Chapter Sections		Learning Objectives
9.1 Mutations	»	Discuss the different classes of mutations and how they can arise as spontaneous mutations or by exposure to mutagens.
	»	Explain how mutations may affect phenotype or genotype and how mutagenicity is assessed.
9.2 DNA Repair	»	List the important mechanisms used by cells to repair damaged DNA.
9.3 Gene Transfer Mechanisms	»	Compare and contrast the various means by which microbes exchange genetic material.
9.4 Mobile Genetic Elements	»	Explain how genetic elements can catalyze their own movement between DNA molecules and how

Chapter Sections		Learning Objectives
		this behavior can be exploited by scientists wishing to explore gene function through knockout mutations.
9.5 Genome Evolution	»	Describe how microbial genomes evolve through the processes of gene gain and loss.
Special Topic 9 Gut Microbiome Caught in the Act of Gene Transfer		
eResearch Activity 9 How Do Mobile Genomic Islands Exploit Incoming Plasmids for Their Ticket to New Hosts?		

D NA is a dynamic molecule. Genome sequences change over generations through mutations that include single-base-pair changes, DNA rearrangements, and gene transfers between species. Bacterial and archaeal genomes can shuttle large clusters of genes between members of different taxonomic domains. The process of genetic change—through gene mutation, gene gain, and gene loss—is the essential first step in evolution.

In Chapter 9 we explore the mechanisms of genetic change that facilitate evolution in microbes. First we look at how individual genes mutate, and how the cell repairs mutations. We then look at

the various ways cells acquire foreign DNA via horizontal gene transfer. Next we describe the role of transposable elements in both gene mutation and horizontal gene transfer. Finally, we describe how mechanisms of genetic change expand or shrink the genomes of microbes. This chapter explores the mechanisms of mutation and gene flow, while the consequences of these processes during evolution will be discussed in Chapter 17 .

9.1 Mutations

Any permanent, heritable alteration in the nucleotide sequence of the genome, whether harmful, beneficial, or neutral, is called a mutation . Mutations are the foundational events that facilitate evolutionary change. In addition, mutations provide the researcher with the ability to investigate the function of genes through manipulation of their DNA sequence. In this section we discuss how various types of mutation occur at one DNA position. Sections 9.3 and 9.4 discuss genetic changes that occur by transfer of DNA from one site to another, including transfer to another microbe.

Mutations Can Change Genes in Many Ways

Changes in the DNA sequence of genes through mutation can have a variety of effects on those genes. Mutations can change the structure and function of the gene's products, be they protein or RNA, and can also change the regulation of the gene's expression. Gene mutations can have important consequences for the microbe, such as gaining resistance to antibiotics that normally target the gene's product. And as we will see in Section 9.5 , mutation is an important mechanism that adds or subtracts genes from genomes.

Mutations that alter gene sequence fall into several different physical and structural classes:

- A point mutation is a change in a single nucleotide (**Fig. 9.1A**). Replacing a purine with a different purine or a pyrimidine with a different pyrimidine is called a transition . Swapping a purine for a pyrimidine (or vice versa) is a transversion .
- Insertions and deletions involve, respectively, the addition or subtraction of one or more nucleotides (**Fig. 9.1B** and **C**), making the sequence either longer or shorter than it was originally.

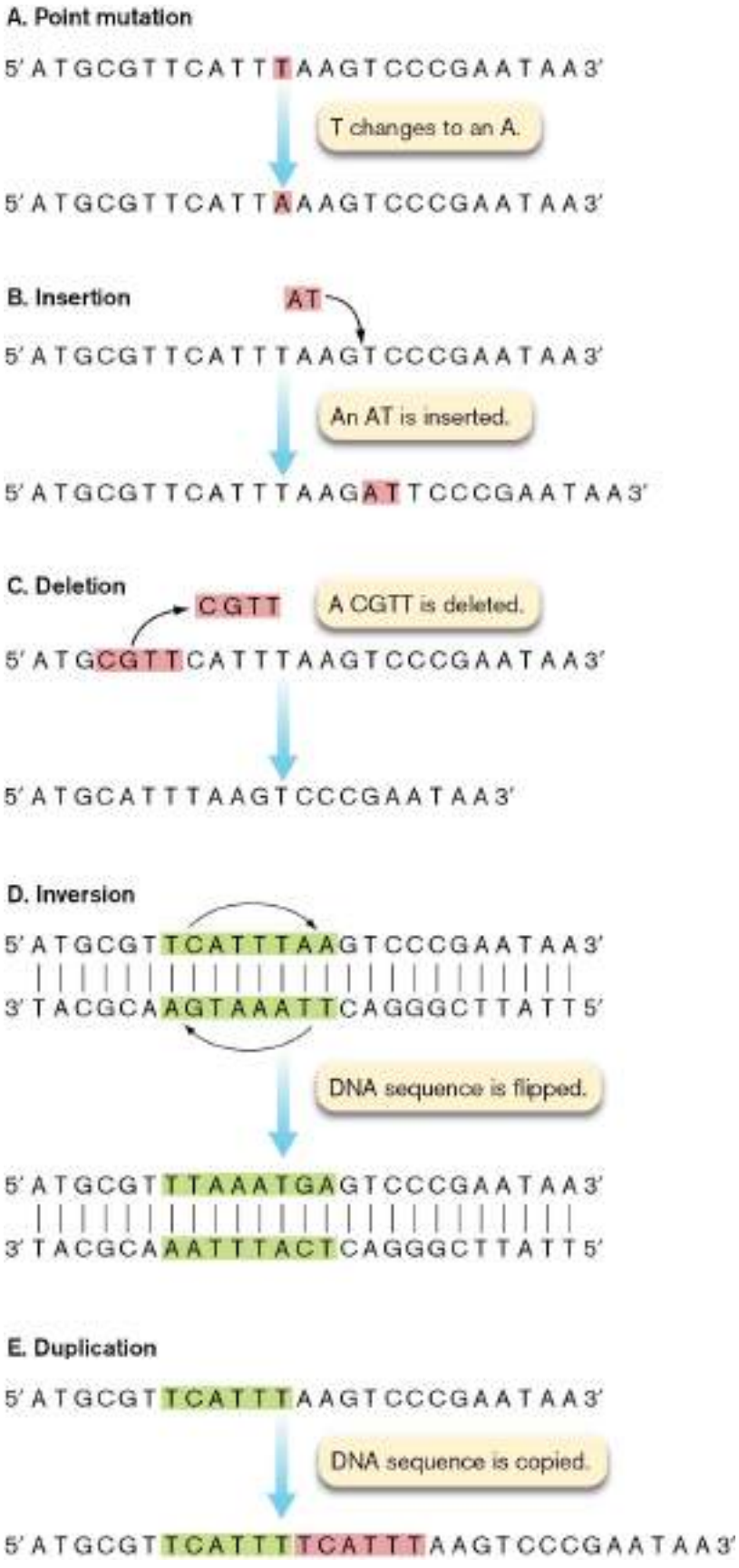

A. Point mutation
5′ A T G C G T T C A T T T A A G T C C C G A A T A A 3′
T changes to an A.
5′ A T G C G T T C A T T A A A G T C C C G A A T A A 3′
B. Insertion
AT
5′ A T G C G T T C A T T T A A G T C C C G A A T A A 3′
An AT is inserted.
5′ A T G C G T T C A T T T A A G A T T C C C G A A T A A 3′
C. Deletion
CGTT
A CGTT is deleted.
5′ A T G C G T T C A T T T A A G T C C C G A A T A A 3′
5′ A T G C A T T T A A G T C C C G A A T A A 3′
D. Inversion
5′ A T G C G T T C A T T T A A G T C C C G A A T A A 3′
3′ T A C G C A A G T A A A T T C A G G G C T T A T T 5′
DNA sequence is flipped.
5′ A T G C G T T T A A A T G A G T C C C G A A T A A 3′
3′ T A C G C A A A T T T A C T C A G G G C T T A T T 5′
E. Duplication
5′ A T G C G T T C A T T T A A G T C C C G A A T A A 3′
DNA sequence is copied.
5′ A T G C G T T C A T T T T C A T T T A A G T C C C G A A T A A 3′

FIGURE 9.1 ■ Changes in a DNA sequence that result in different classes of mutations. Reversion mutations reverse the direction of the arrows shown.

- An inversion results when a fragment of DNA is flipped in orientation relative to the flanking DNA on either side (**Fig. 9.1D**).
- A duplication produces a second copy of a sequence fragment on the DNA molecule, usually adjacent to the original copy (**Fig. 9.1E**).
- A transposition is the movement of a sequence fragment from one location to another. These movements are usually catalyzed by special enzymes, and they can involve insertions, deletions, and duplications, depending on the mechanism of transposition. Transpositions are discussed in Section 9.4 .
- A reversion restores a mutated sequence to its original sequence (**Fig. 9.1**).

Mutations can be further categorized into informational classes on the basis of how they affect the gene product . To illustrate these effects, let us reexamine the DNA sequence of **Figure 9.1** when placed within the context of an open reading frame (ORF), such that each triplet of bases constitutes a codon (refer to **Figure 8.12** for the genetic code). Mutations that do not change the amino acid sequence of a translated ORF are called silent mutations . For example, a point mutation changing TTT to TTC in the sense DNA strand (corresponding to a UUU-to-UUC codon change in messenger RNA) still codes for phenylalanine (**Fig. 9.2A**). Thus, even though the DNA sequence has changed, the protein sequence remains the same (hence, it was a synonymous base substitution). However, if the UUU codon were changed to UUA (a U-to-A transversion), then the protein would have a leucine where a phenylalanine was (**Fig. 9.2B**). This type of mutation is a missense mutation because it changes the amino acid sequence of the protein (hence, it was a nonsynonymous base substitution).

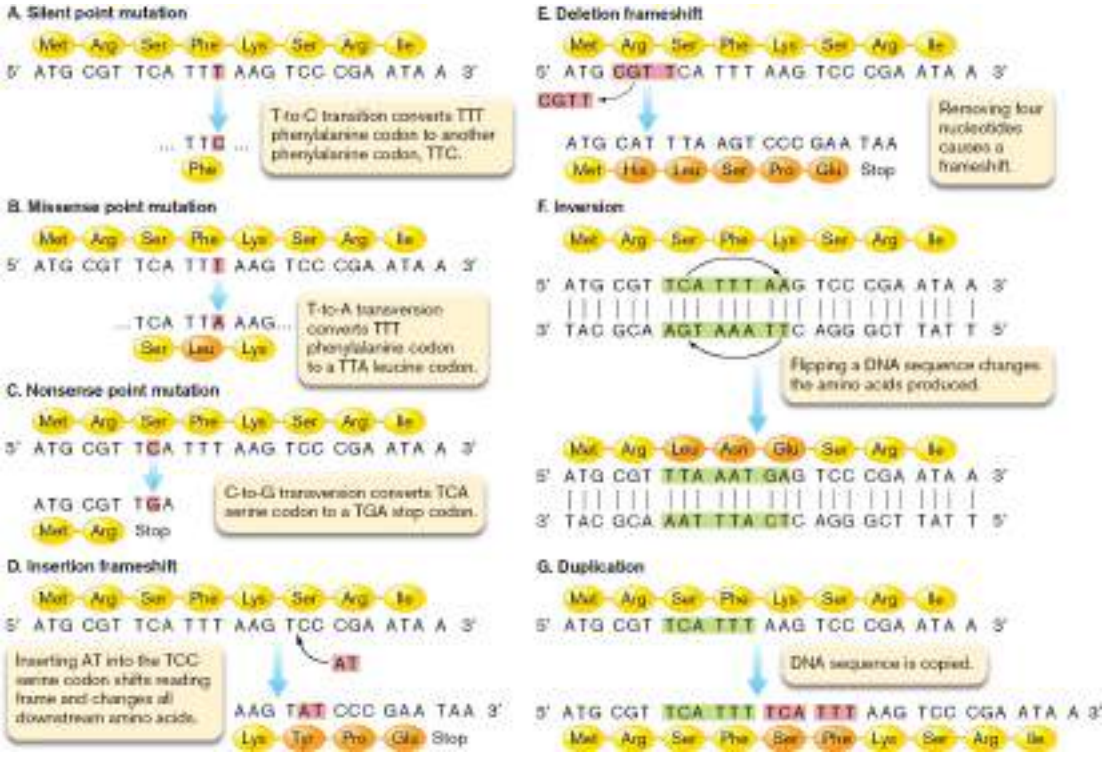


FIGURE 9.2 ■ Changes in amino acid sequence that result from different classes of mutation.

The amino acid substitution resulting from a missense mutation may or may not alter protein function. The outcome depends on the structural importance of the original amino acid and how close in structure and chemical properties the replacement amino acid is to the original. Missense mutations result in either conservative amino acid replacements, in which the new amino acid is structurally similar to the original (for example, leucine is substituted for isoleucine), or nonconservative replacements, in which a very different amino acid is substituted (for example, tyrosine for alanine). A missense change may decrease or eliminate the activity of the protein (a loss-of-function mutation) or it may make the protein more active or even make the protein gain a new activity, such as an expanded substrate specificity or a completely different substrate specificity (these are called gain-of-function mutations).

A mutation that eliminates function is known as a knockout mutation . Knockout mutations can include multiple-base insertions

and deletions, as well as nonsense mutations. A nonsense mutation is a point mutation that changes an amino acid codon into a translation termination codon; for example, TCA (serine) to TGA (**Fig. 9.2C**). The result is a truncated protein. Truncations can completely knock out the function of proteins, especially in cases where the mutation occurred early in the open reading frame. Typically, these defective, truncated proteins are degraded by cellular proteases (see Section 8.4). Nonsense mutations can turn genes into pseudogenes, which subsequently can be lost from the genome during evolution (see Section 9.5).

Insertions and deletions can alter the reading frame of the DNA sequence (**Fig. 9.2D** and **E**). Remarkable as it is, the ribosome simply reads RNA sequences one codon at a time, stringing amino acids together in the process. It translates each codon "word" but cannot understand the overall protein "sentence." It does not recognize when bases have been added or removed through mutation; instead, it keeps reading the sequence in triplets. If the number of bases inserted or deleted is not a multiple of three, the reading frame of translation changes. The result is a frameshift mutation , in which the ribosome produces a garbled protein product. Frameshift mutations often cause the ribosome to encounter a premature stop codon, originally in a different reading frame. Thus, like nonsense mutations, frameshifts can also turn genes into pseudogenes. However, if the insertion or deletion involves multiples of three bases, the reading frame is not changed, but one or more amino acids are added or removed.

Note: Frameshift, missense, and nonsense mutations do not exist for genes whose product is untranslated RNA, such as ribosomal RNA (rRNA) and transfer RNA (tRNA). However, insertions, deletions, base substitutions, and other changes can all still affect the function of the RNA product.

Thought Question

9.1 How could frameshift mutations be used to confirm that codons consist of three bases, versus two or four? *Hint:* Think of how a series of "like" frameshifts (for example, single-basepair additions) along a gene would affect the reading frame.

An inversion mutation flips a DNA sequence (**Fig. 9.2F**). Imagine the sequence highlighted in the figure rotating 180° while the adjacent sequences remained right where they were. The rotation would retain the 5′-to-3′ polarity in the new molecule. But if the inversion occurred within a gene, it would likely change the codons in the area and alter the resulting protein. **Figure 9.2F** shows a small inversion.

However, inversions often involve large tracts of DNA encompassing several genes. If an entire gene with its promoter inverts, the gene will likely remain functional, and its encoded protein may very well still be made. Inversions occur within a genome as a result of recombination events between similar DNA sequences or as a consequence of mobile genetic elements jumping between different areas of a genome (discussed in Section 9.4).

Finally, duplications within an open reading frame can create repeats of codons (**Fig. 9.2G**), thus increasing the length of the protein product. If the length of the duplicated segment is not divisible by three, a frameshift will also occur.

Mutations can affect both the genotype and the phenotype of an organism. The genotype of an organism reflects its genome sequences. Regardless of whether a mutation causes a change in a trait (phenotype), every mutation causes a change in the genotype. In contrast to genotype, phenotype comprises only observable characteristics, such as biochemical, morphological, or growth traits.

It is also important to realize that the size of a mutation does not always correlate with the extent of a phenotypic change. As

illustration, consider that a single point mutation in the gene encoding HPr will render a bacterium incapable of growing on many sugars. HPr is an enzyme that delivers phosphate to a number of phosphotransferase sugar transport systems (these systems are discussed in Section 4.2). But inserting 5 kb of DNA just past the *hpr* stop codon will have no effect on cell growth. For mutations, it's often not the size that counts; it's the location.

Mutations Arise by Diverse Mechanisms

What causes DNA mutation in microbes? Some mutations can arise spontaneously. Despite the high accuracy of the replication apparatus (see Section 7.3), mistakes do occur, albeit at a very low rate. For example, spontaneous mutations in *Escherichia coli* have a frequency of occurrence ranging from 10^{-8} to 10^{-6} per cell division in a given gene.

Spontaneous mutations in a genome arise for many reasons; for example, tautomeric shifts in the chemical structure of the bases (**Fig. 9.3**). Tautomeric shifts involve a change in the bonding properties of amino ($-NH_2$) and keto (CO) groups. Normally, the amino and keto forms predominate (over 85%), but when an amino group shifts to an imino (=NH) group, for example, base pairing changes. A thymine that normally base-pairs with adenine will, in its rare enol form, base-pair with guanine.

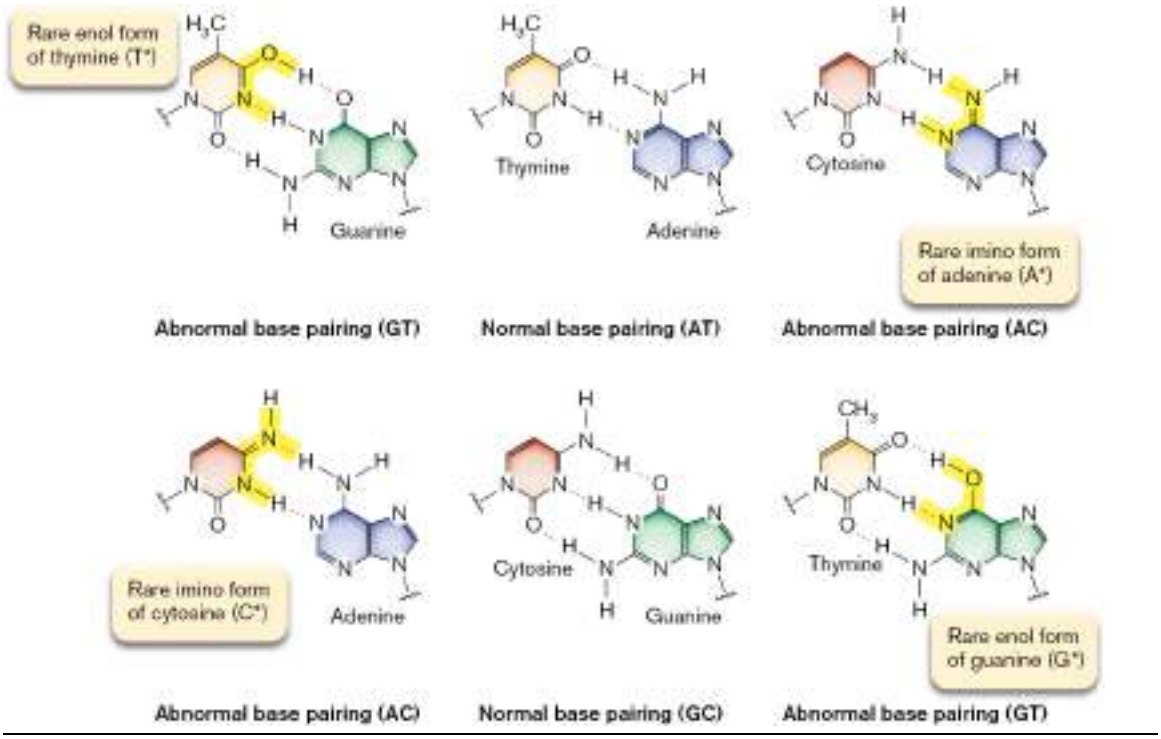


FIGURE 9.3 ■ Rare tautomeric forms of bases have altered base-pairing properties. Tautomeric transitions can lead to permanent mutations.

Tautomeric shifts that occur during DNA replication will increase the number of mutation events. **Figure 9.4** shows an example of how a tautomeric shift in thymine during replication results in an AT-to-GC transition mutation. Note that after the second round of replication, the mutation is "fixed" on both DNA strands in the mutant, and the mutation will therefore be passed on to all of the mutant's progeny.

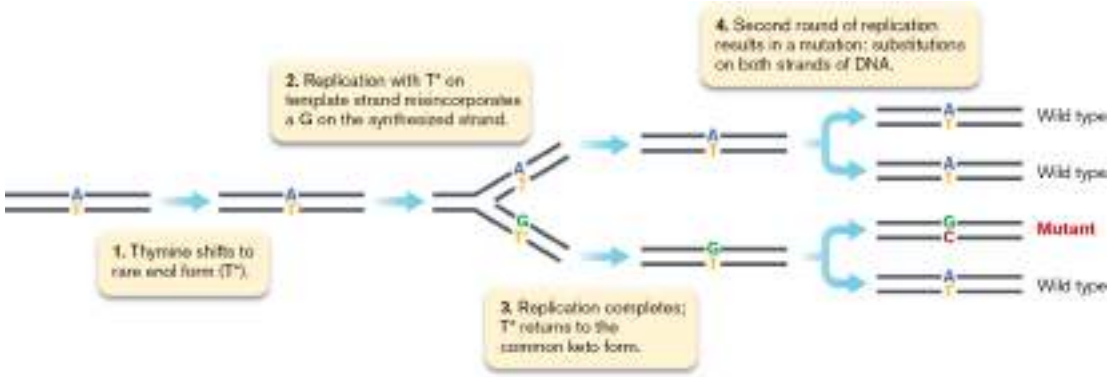


FIGURE 9.4 ■ Mutation arising from the tautomeric shift of thymine prior to DNA replication. This AT-to-GC transition mutation in one of the progeny occurs after the second round of replication.

Spontaneous mutations in DNA can also be caused by chemical reactions with water (hydrolysis). For example, cytosine spontaneously deaminates to yield uracil, which base-pairs with adenine instead of guanine (**Fig. 9.5**).

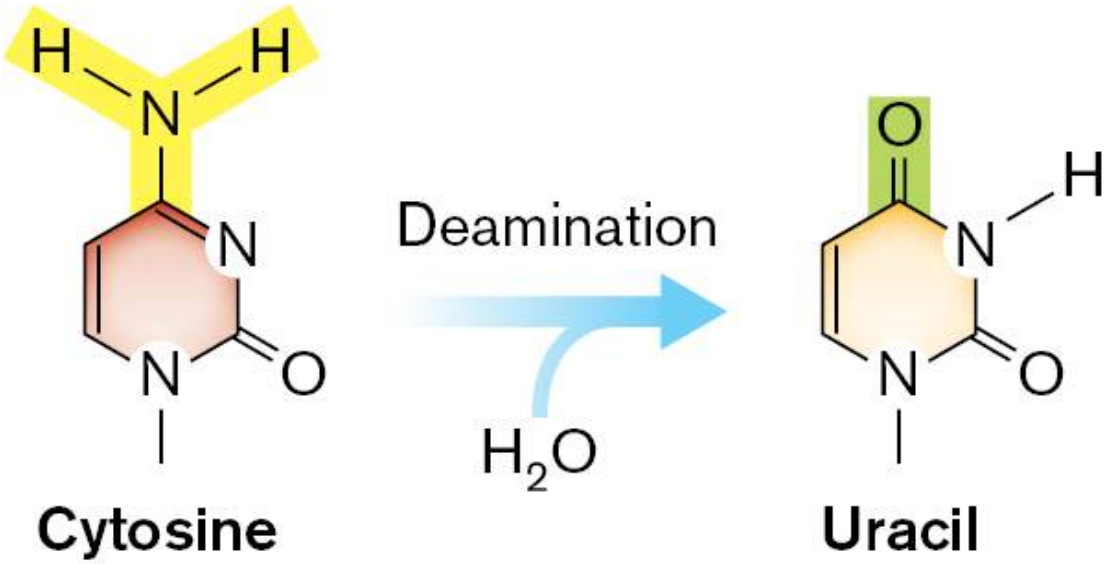


FIGURE 9.5 ■ Spontaneous deamination of cytosine. Oxidative deamination changes cytosine to uracil.

Thought Question

9.2 Does deamination of cytosine to uracil lead to a transition mutation or a transversion mutation? Work this out in a drawing, and use **Figure 9.4** as a guide.

In addition, purines are particularly susceptible to spontaneous ejection from DNA via breakage of the glycosidic bond connecting the base to the sugar backbone (**Fig. 9.6**). The result of this loss is the formation of an apurinic site (one missing a purine base) in the DNA. Lack of a purine would obviously hinder transcription and replication.

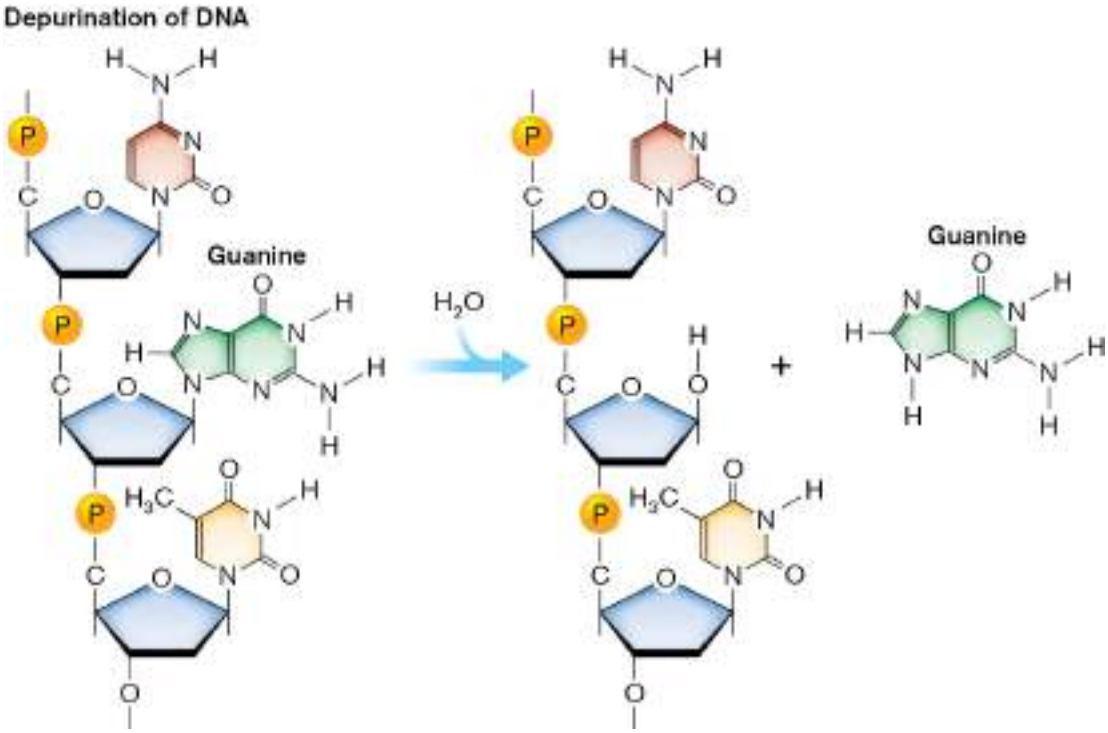


FIGURE 9.6 ■ Spontaneous formation of an apurinic site.

Naturally occurring intracellular methylation agents (for example, *S* -adenosylmethionine) can spontaneously methylate DNA to produce a variety of altered bases. The spontaneous methylation of the N-7 position of guanine, for example, weakens the glycosidic bond and spontaneously releases the base (forming an apurinic site) or opens the imidazole ring (forming a methylformamide pyrimidine). In addition to mispairing, some of these spontaneous events can lead to major chromosomal rearrangements, such as duplications, inversions, and deletions.

Spontaneous mutations such as those just described are just one class of mutations that occur in microbes. Microbial DNA is also susceptible to damage inflicted by a variety of physical and chemical agents. Mutagens are chemical agents or forms of electromagnetic radiation that can damage DNA, often at rates that far exceed spontaneous mutation (**Table 9.1**). For example, irradiation by X-rays can cause a massive number of double-stranded DNA breaks. When this happens, the integrity of the chromosome is lost, and if the DNA is left unrepaired, the damage will lead to severe problems with chromosome stability and replication and, ultimately, cause cell death. Other agents can directly modify the bases in DNA while leaving its overall structure intact. In this case, the modified bases have altered hydrogen bond base-pairing properties that result in the incorporation of an inappropriate base during replication. When that happens, a mutation results. Mutagens can also increase the mutation rate by inducing repair pathways that themselves introduce mutations (error-prone DNA polymerases such as Pol IV and Pol V are discussed in Section 9.2).

TABLE 9.1	**Mutagenic Agents and Their Effects**
Mutagenic agent	**Effects**
Chemical agent	
Base analog *Examples:* caffeine, 5-bromouracil	Substitutes "look-alike" molecule for normal nitrogenous base during DNA replication: point mutation
Alkylating agent *Example:* nitrosoguanidine	Adds alkyl group, such as methyl group ($-CH_3$), to nitrogenous

TABLE 9.1	Mutagenic Agents and Their Effects
	base, resulting in incorrect pairing: point mutation
Deaminating agent *Examples:* nitrous acid, nitrates, nitrites	Removes amino group ($-NH_2$) from nitrogenous base: point mutation
Acridine derivative *Examples:* acridine dyes, quinacrine	Inserts (intercalates) into DNA ladder between backbones to form a new rung, distorting the helix: can cause frameshift mutations
Electromagnetic radiation	
Ultraviolet rays	Link adjacent pyrimidines to each other, as in thymine dimer formation, thereby impairing replication; lethal if not repaired
X-rays and gamma rays	Ionize and break molecules in cells to form free radicals, which in turn break DNA; lethal if not repaired

Ultraviolet (UV) light will produce striking structural alterations in DNA molecules. Pyrimidines (more than purines) are highly susceptible to UV radiation. The energy absorbed by a pyrimidine hit with UV light boosts the energy of its electrons to the point where the molecule is unstable. If two pyrimidines are neighbors on a

single DNA strand, their energized electrons can react to form a four-membered cyclobutane ring. The result is a pyrimidine dimer that will block replication and transcription (**Fig. 9.7**).

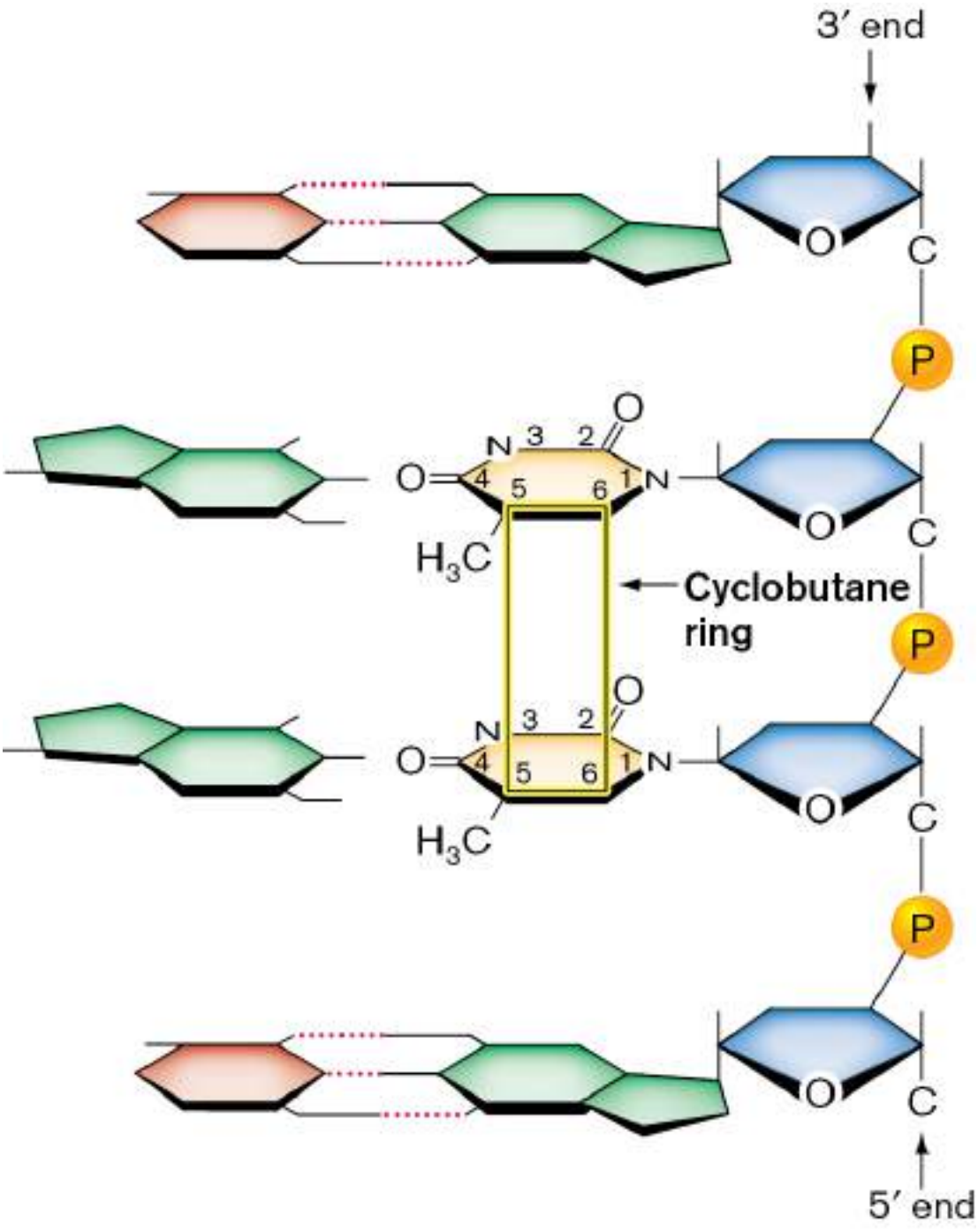


FIGURE 9.7 ■ Production of a pyrimidine dimer. The energy from UV irradiation can be absorbed by pyrimidine

molecules. The excited electrons of carbons 5 and 6 on adjacent pyrimidines can then be shared to form a four-membered cyclobutane ring between adjacent pyrimidines. The pyrimidine dimer blocks replication and transcription.

DNA can also be damaged by metabolic activities of the cell that produce reactive oxygen species, such as hydrogen peroxide (H_2O_2), superoxide radicals ($^{\bullet}O_2^-$), and hydroxyl radicals ($^{\bullet}OH$). Even though bacteria have biochemical mechanisms to detoxify reactive oxygen species, the systems can be overwhelmed. Oxidative damage causes the production of thymidine glycol or 8-oxo-7-hydrodeoxyguanosine in DNA (**Fig. 9.8**).

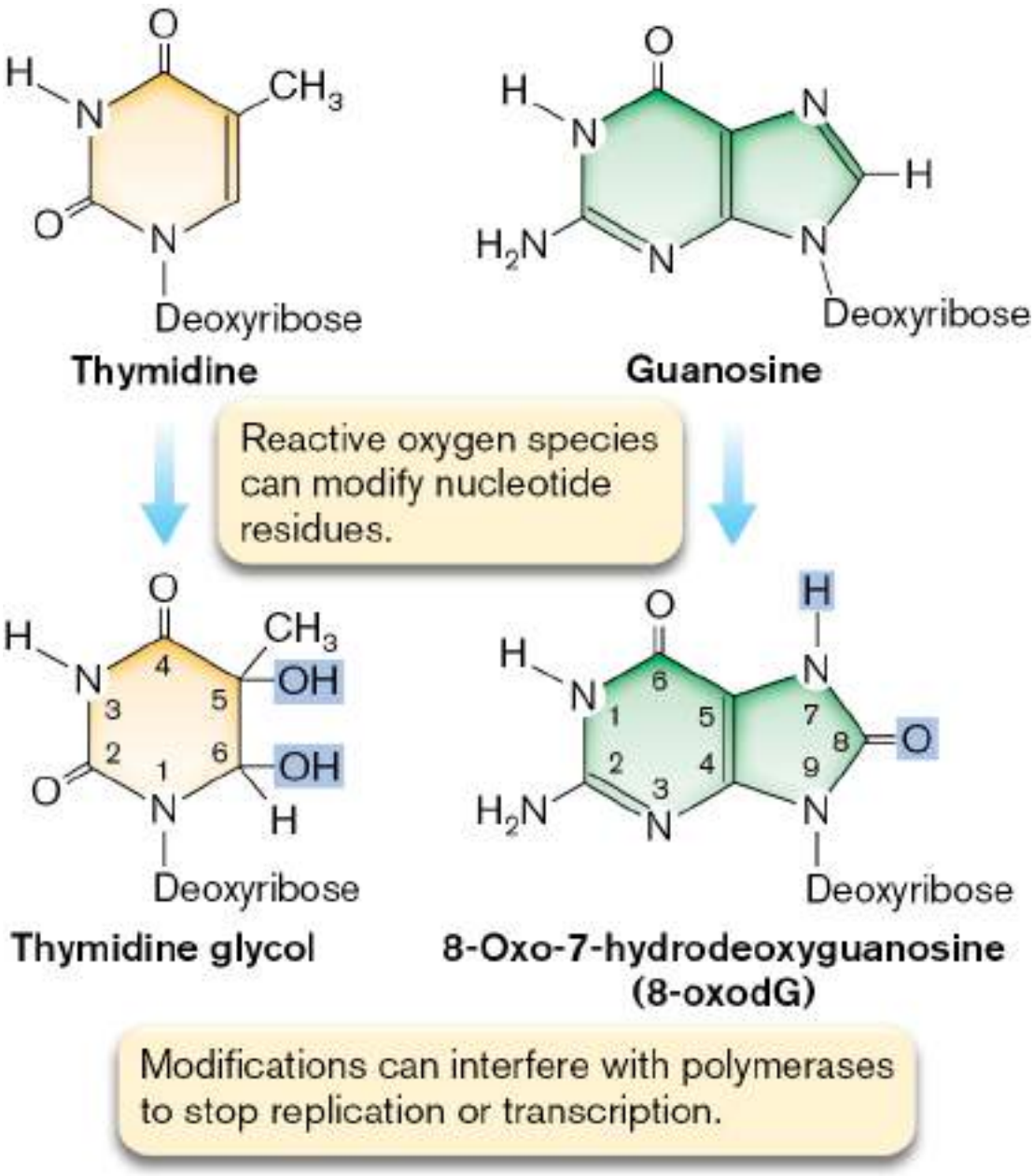


FIGURE 9.8 ■ Examples of damage caused by reactive oxygen species. The modifications can interfere with polymerase function and stop replication or interfere with the transcription of affected genes. The blue highlighting identifies modifications to thymidine and guanosine residues.

Although DNA can be damaged in numerous ways, the cell can repair that damage before it becomes fixed as a mutation. But the

repair mechanisms are not perfect. Repair errors contribute heavily to the formation of heritable mutations and, thus, to evolution.

Identifying Mutagens Using Bacterial "Guinea Pigs"

In a world where we are continually exposed to new chemicals, it is important to determine which ones are potential mutagens and therefore potential carcinogens in humans. Bruce Ames and his colleagues invented a simple assay that uses bacteria as a rapid initial screen, for which Ames received the National Medal of Science in 1998. This method accelerates the process of drug discovery by providing an inexpensive preliminary screen for weeding out mutagenic chemicals before more expensive animal testing is undertaken. The method relies on a mutant of *Salmonella enterica* that is defective in the *hisG* gene, whose product is involved in histidine biosynthesis. The *hisG* mutant is auxotrophic for histidine and thus cannot grow on a defined minimal medium lacking histidine. However, if a reversion mutation occurs in the *hisG* gene and restores the gene to its original functional state, the new mutant cell will form a colony even in the absence of histidine. Generally, this method is called a reversion test, and the specific test that Ames developed is called the Ames test.

Today we use a modified form of the Ames test that takes into account that some chemicals become mutagenic (and may cause cancer) only after they are processed by the liver. The liver is the chief organ for detoxifying the body—a task that liver enzymes accomplish by chemically modifying foreign substances. The potential mutagen, *hisG* mutant bacteria, and liver homogenate are combined and mixed with agar (**Fig. 9.9**). The combination is poured onto a Petri plate. If the liver extract enzymes act on the test compound and the metabolites produced are mutagenic, then increasing numbers of His^+ revertants will be observed with increasing doses of mutagen. If the compound is not mutagenic, the number of colonies will not exceed those found in a control group that was not exposed to the mutagen.

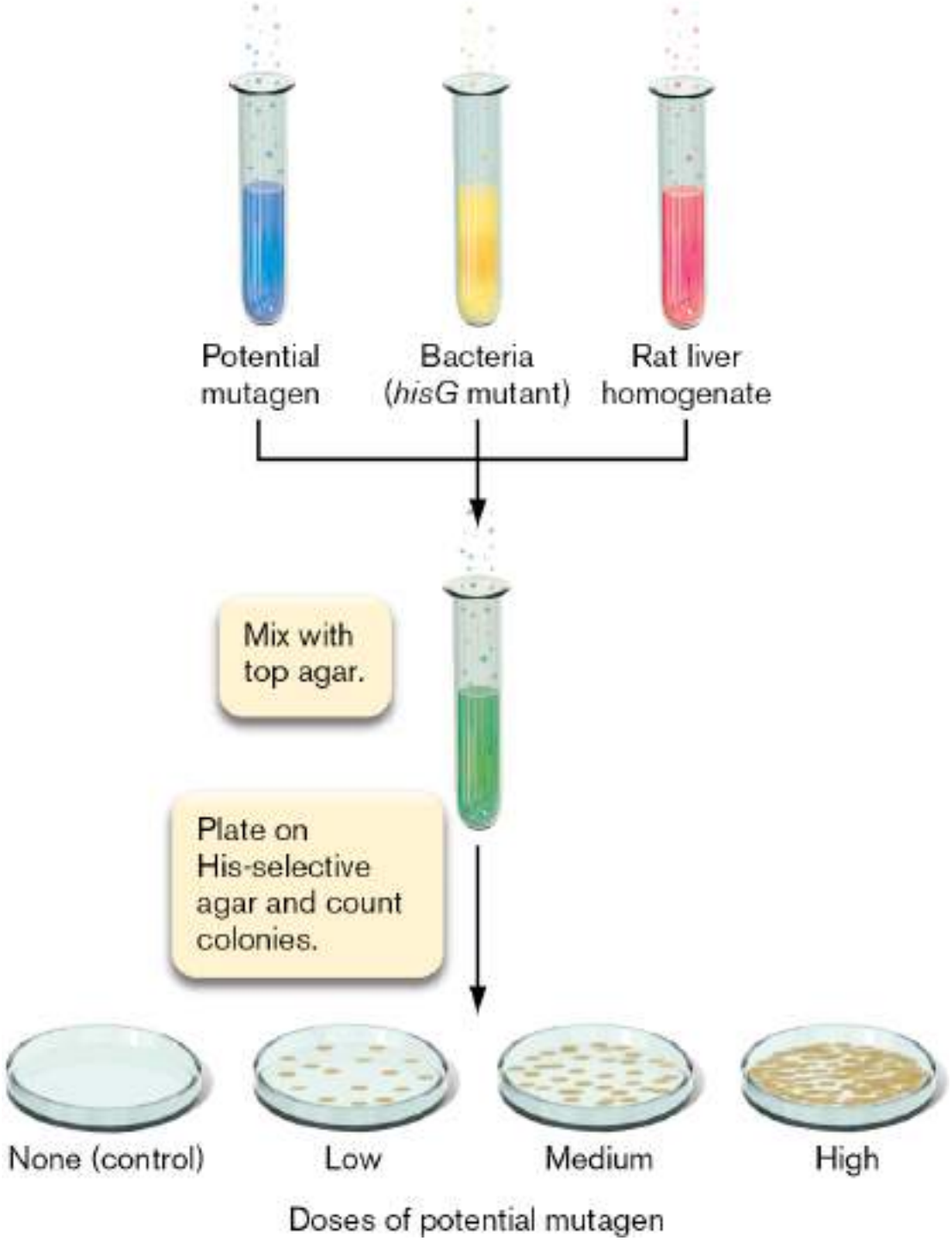


FIGURE 9.9 ■ The modified Ames test to assay for the mutagenic properties of chemicals processed through the liver.

To Summarize

- **A mutation** is any heritable change in DNA sequence, regardless of whether a change in gene function results.
- **Classes of mutations** include point mutations, insertions, deletions, frameshift mutations, and transpositions.
- **Mutations can be further categorized by their effect on gene products** and include silent, missense, and nonsense mutations.
- **Genotype** reflects the genetic makeup of an organism, whereas **phenotype** reflects its physical traits.
- **Spontaneous mutations** reflect tautomeric shifts in DNA nucleotides during replication, accidental incorporation of noncomplementary nucleotides during replication, or "natural" levels of chemical or physical (irradiation) mutagens in the environment.
- **Chemical mutagens** can alter purine and pyrimidine structure and change base-pairing properties.
- **The mutagenicity** of a chemical can be assessed by its effect on bacterial cultures.

Glossary

mutation
: A heritable change in a DNA sequence.

point mutation
: A change in a single nucleotide within a nucleic acid sequence.

transition
: A point mutation in which a purine is replaced by a different purine or a pyrimidine is replaced by a different pyrimidine.

transversion
: A point mutation in which a purine is replaced by a pyrimidine or vice versa.

insertion
: The addition of nucleotides to the middle of a DNA sequence.

deletion
: The loss of nucleotides from a DNA sequence.

inversion
: A mutation in which a DNA fragment is flipped within a chromosome. It may allow or repress the transcription of a particular gene.

duplication
: The production of a second copy of a sequence fragment on a DNA molecule, usually adjacent to the original copy.

transposition
: The process of moving a transposable element from one DNA region to another.

reversion
: A mutation that changes a previous mutation back to its original state.

silent mutation
: A mutation that does not change the amino acid sequence encoded by an open reading frame. The changed codon encodes the same amino acid as the original codon.

missense mutation
: A point mutation that alters the sequence of a single codon, leading to a single amino acid substitution in a protein.

loss-of-function mutation
: A mutation that eliminates or decreases the function of the gene product.

gain-of-function mutation
: A mutation that enhances the activity or allows new activity of a gene product.

knockout mutation
: A mutation that completely eliminates the activity of a gene product.

nonsense mutation
: A mutation that changes an amino acid codon into a premature stop codon.

frameshift mutation
: A gene mutation involving the insertion or deletion of nucleotides that cause a shift in the codon reading frame.

genotype
: The genome sequence of an organism.

phenotype
: The observable characteristics of an organism.

apurinic site
: A DNA site missing a purine base because the bond linking the base to the sugar has been hydrolyzed. Also called **AP site** .

mutagen
: A chemical that damages DNA and increases the rate of mutations.

auxotrophic
: Describing a mutant state in which the cell has lost the ability to synthesize a substance required for growth. An auxotroph has a nutritional requirement not shared by the parent.

Figure 8.12

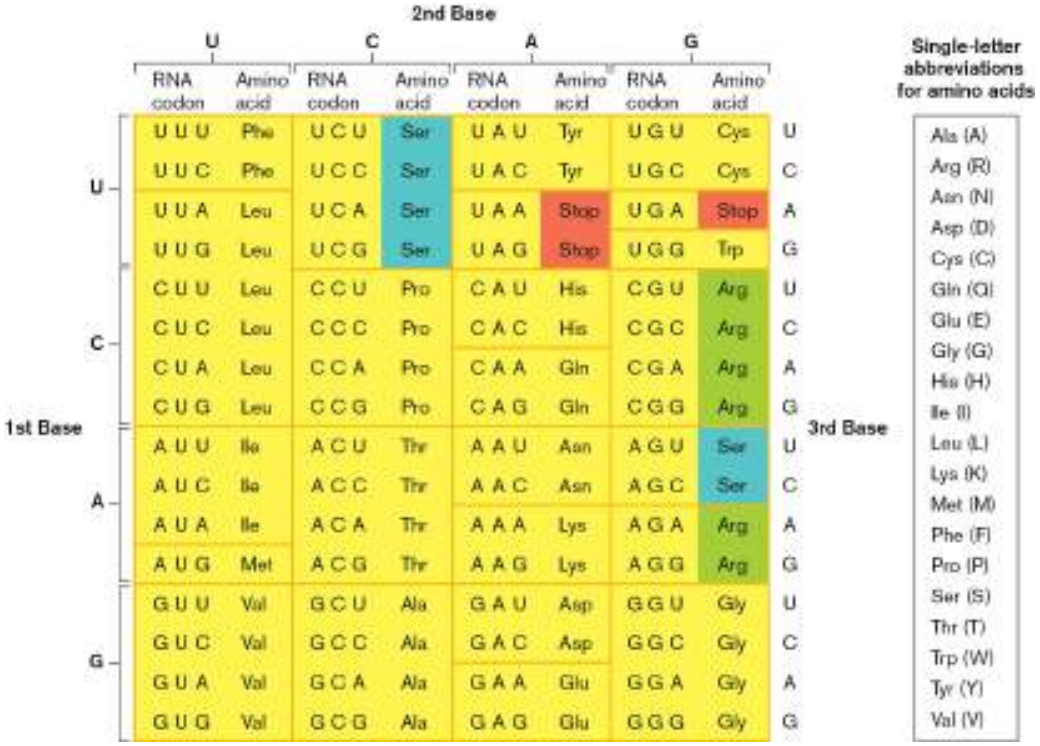


FIGURE 8.12 ■ The standard genetic code. Codons within a single box encode the same amino acid. Blue- and green-highlighted amino acids are encoded by codons in two boxes. Stop codons are highlighted red. Often, single-letter abbreviations for amino acids are used to convey protein sequences (see legend).

9.2 DNA Repair

Microorganisms are equipped with a variety of molecular tools that repair DNA damage before the damage becomes a heritable mutation (**Table 9.2**). The type of repair mechanism used (and when it is used) depends on two things: the type of mutation needing repair and the extent of damage. Some repair mechanisms are error proof and do not introduce mutations; others are error prone and require "emergency" DNA polymerases expressed under dire circumstances. These polymerases sacrifice replication accuracy to rescue the damaged genome. Whether damage is introduced by mutagens or by inaccurate DNA synthesis, microbial survival depends on the ability to repair DNA.

TABLE 9.2 Types of DNA Repair

System	Proteins	Mutations recognized	Repair mechanism	Accuracy
Photoreactivation	PhrB	Pyrimidine dimers	Cyclobutane ring cleaved	Error proof
Nucleotide excision	UvrABCD	Helical destabilization (e.g., pyrimidine dimers)	Patch of nucleotides excised	Error proof
Base excision	Fpg, Ung, Tag, MutY, Nfo	Various modified bases	Glycosylases remove base from phosphodiester backbone; apurinic (AP) sites formed	Error proof

TABLE 9.2	Types of DNA Repair			
Methyl mismatch	MutHSL, Dam	Transitions, transversions	Nick on nonmethylated strand; excision of nucleotides	Error proof
Recombination	RecA	Single-strand gaps and double-strand breaks	Recombination	Error proof
Translesion bypass replication	UmuDC	Gaps	Part of SOS system	Error prone (generates mutations)
Nonhomologous end joining	Ku, LigD	Double-strand breaks	Ligation of strands after processing with exonuclease and polymerase	Error prone

We will first discuss the error-proof repair pathways that prevent mutations. These include photoreactivation, nucleotide excision repair, base excision repair, methyl mismatch repair, and recombinational repair. Then we will turn our attention to error-prone repair pathways. These pathways risk introducing mutations and operate only when damage is so severe that the cell has no other choice but to die.

Error-Proof Repair Pathways

Repair of DNA replication errors. What happens if DNA polymerase simply makes a mistake and incorporates a normal but incorrect base? The inherent error rate of DNA polymerase III is approximately one mistake per 10^8 bases synthesized, after proofreading. However, the mutation rate in a live cell is actually only one mistake every 10^{10} replicated bases. This difference indicates that the cell has a way to recognize and reverse mutations even after DNA is synthesized and (insufficiently) proofread.

One way to reveal the incorrect base is methyl-directed mismatch repair , in which repair enzymes recognize the methylation pattern in DNA bases. Many bacteria methylate their DNA at specific sites. In *Escherichia coli* , for example, DNA adenine methyltransferase (Dam) methylates the palindromic sequence GATC to produce GA ^{ME} TC. Timing is critical to Dam's role in repair. Before replication, both strands of DNA are fully methylated at GATC sites. Immediately after replication, newly synthesized strands are unmethylated. Soon after synthesis, Dam methylates the new strand, but in the window of time between replication and methylation, the DNA mismatch repair system can distinguish parental from new strands by the difference in methylation pattern.

Misincorporation of a base during replication produces a mismatch between the incorrect base in the newly synthesized, unmethylated strand and the correct base residing in the parental, methylated strand (**Fig. 9.10**). Methyl-directed mismatch repair enzymes (MutS, MutL, and MutH) bind to the mismatch. MutS identifies the mismatch as a distortion in the usual base stacking in the DNA helix and recruits MutL and MutH (**Fig. 9.10** , steps 1 and 2). MutL recognizes the methylated strand (GA ^{ME} TC) and brings it in a loop to meet MutS and MutH (steps 3 and 4). Then, MutH cleaves the unmethylated strand containing the mutation, near the GATC sequence (step 5). A DNA helicase called UvrD then unwinds the cleaved strand, exposing it to a variety of exonucleases (step 6). The result is a gap that is filled in by DNA polymerase I (Pol I) and sealed by DNA ligase.

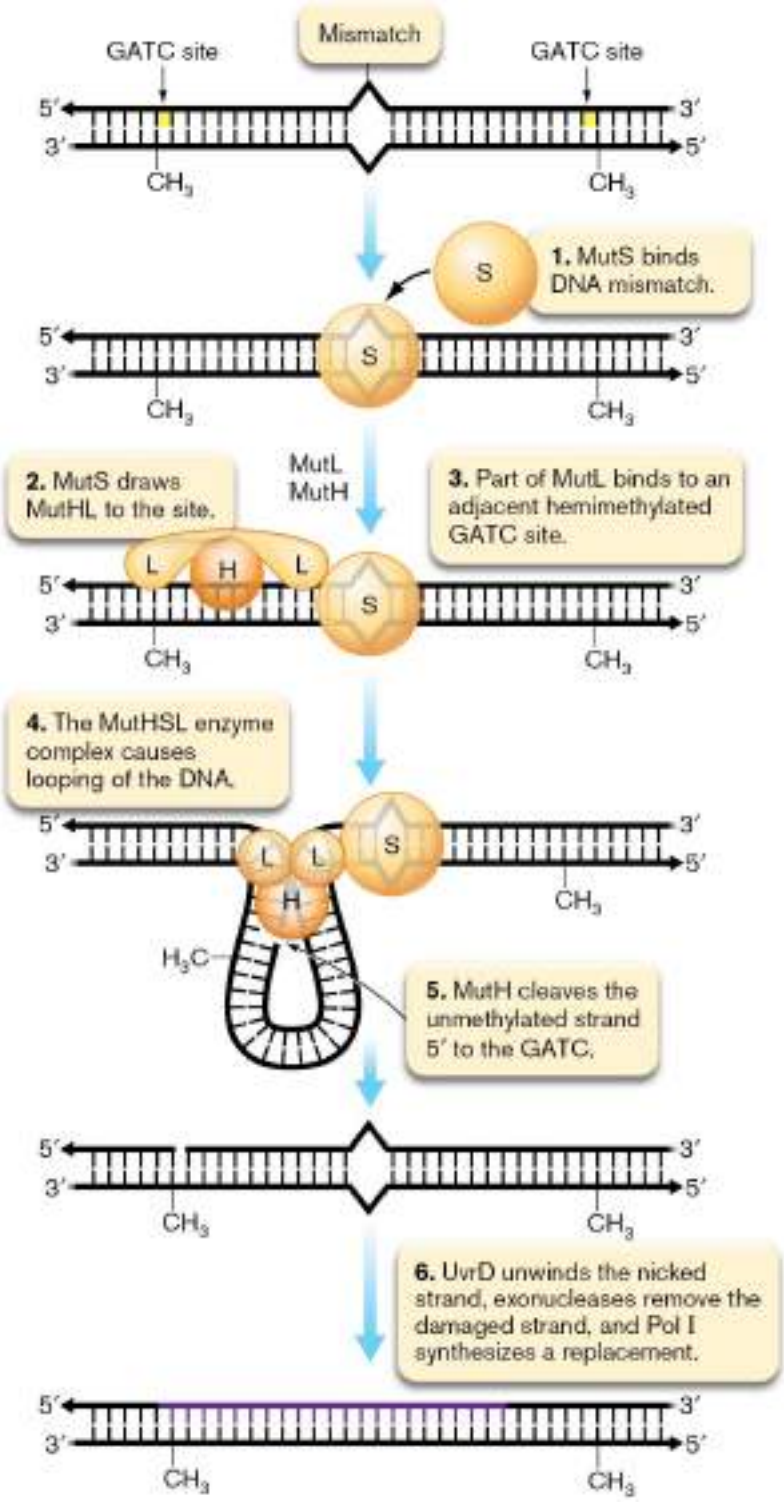


FIGURE 9.10 ■ Methyl-directed mismatch repair. The bacterial cell, in this case *E. coli* , can use specific methylations on DNA to recognize parental DNA strands for preferential DNA repair. Newly replicated strands are not immediately methylated. So when a mismatch is found, the

mismatch repair system views the newly synthesized strand as suspect and replaces the section of unmethylated DNA encompassing the mismatch.

The methyl-directed mismatch repair proteins (and genes) are called Mut (and *mut*) because a high mutation rate results when strains are defective in one of these proteins. A bacterial strain with a high spontaneous mutation rate is called a mutator strain . Mutator strains of *E. coli* can be identified by the presence of blue papillae, or microcolonies, that form on an unpigmented colony (**Fig. 9.11**). *E. coli* engineered with a loss-of-function mutation in *lacZ* cannot grow on lactose or enzymatically convert a nonpigmented lactose analog, X-Glu (5-bromo-4-chloro-3-indolyl-beta-d-glucoside), to a blue pigment. On plates containing glucose and lactose, these mutants grow as colonies on the glucose but leave the lactose behind. Mutator variants of this mutant, including those defective in mismatch repair, revert the *lacZ* mutation to wild type at a high rate, and revertant cells within the colony can exploit the remaining lactose, allowing them to grow as papillae and at the same time reveal themselves by turning blue.

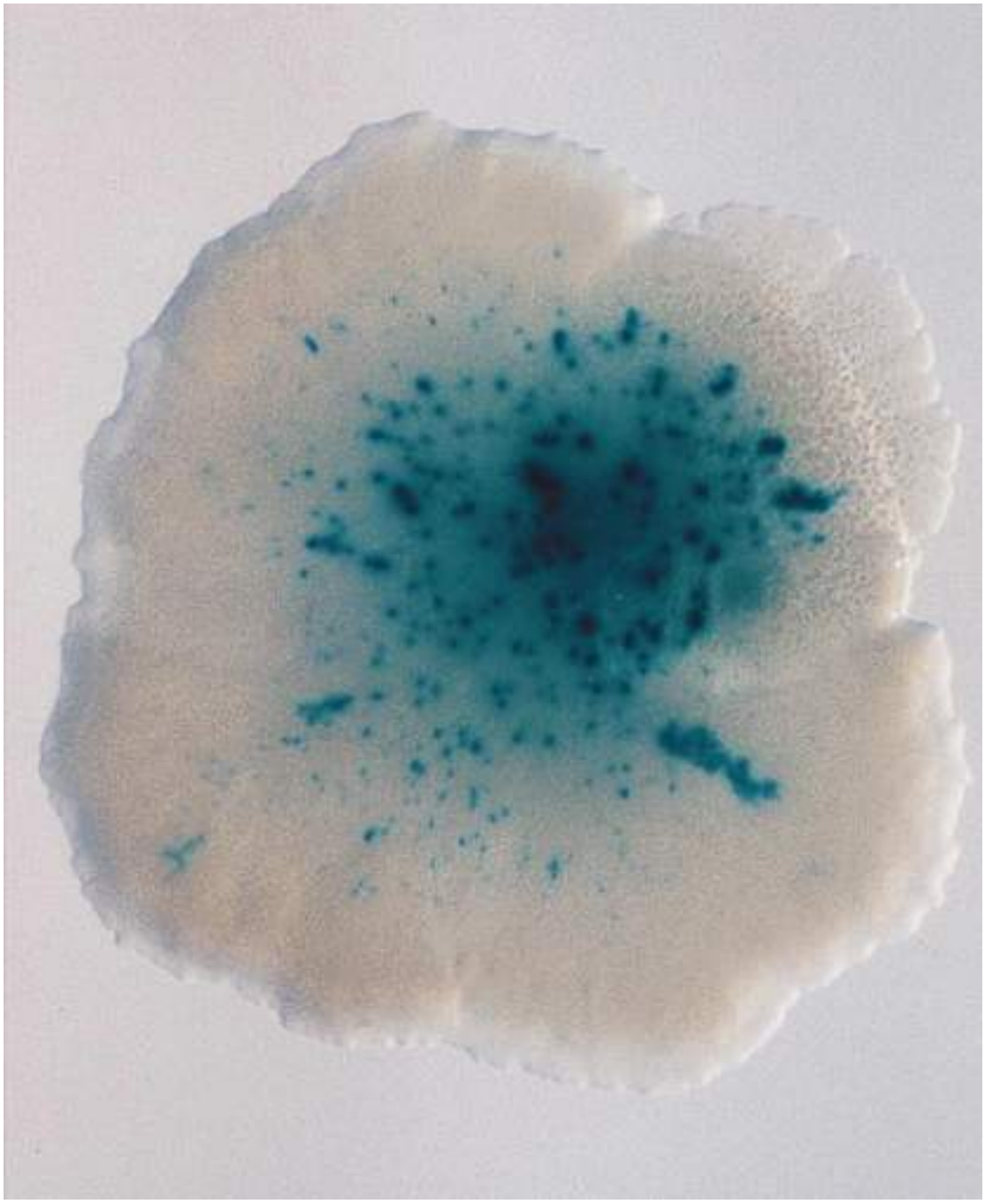


FIGURE 9.11 ■ Methyl-directed mismatch repair mutant of *Escherichia coli* identified by blue papillae. In mutator strains with defects in mismatch repair, a high frequency of cells within a *lacZ*$^-$ colony revert to *lacZ*$^+$. These revertants use lactose and grow as papillae and also convert the lactose analog X-Glu to a blue pigment.

J. H. MILLER ET AL. 1999. *J BACTERIOL* . **181** :1576–1584

Note: Not all DNA methylations signal methyl-directed mismatch repair. DNA methylation by restriction-modification systems, for example, prevents cleavage by DNA restriction endonucleases but is not used to designate parental DNA after replication.

Thought Questions

9.3 Would mutants that lack Dam have a mutator phenotype? Explain why or why not. What about mutants that overexpress Dam?

9.4 It has been reported that hypermutable bacterial strains are overrepresented in clinical isolates. Out of 500 isolates of *Haemophilus influenzae* , for example, 2%–3% were mutator strains having mutation rates 100–1,000 times higher than those of lab reference strains. Why might the mutator phenotype be beneficial to pathogens?

Repair of UV damage. Many microbes commonly encounter ultraviolet light in their natural habitat. The pyrimidine dimers that form as a result of ultraviolet irradiation can be repaired by several different mechanisms in bacteria. In nucleotide excision repair , the UvrABC nuclease recognizes the dimer, cleaves and removes a short section of DNA that includes the dimer, and leaves behind a gap that can be filled by DNA Pol I and ligase. UV dimers can't be read by DNA polymerase III (Pol III) during chromosome replication. Consequently, Pol III skips this damaged region, and the region is subsequently repaired through recombinational repair involving homologous recombination (discussed later) between daughter strands, followed by nucleotide excision repair to remove the dimer.

Photoreactivation is a third mechanism that repairs DNA damage from UV exposure. Notably, the photolyase enzymes involved in photoreactivation are activated by a lower-energy wavelength of light (in the visible range) and actually use this light to repair the damage from UV. In photoreactivation, photolyase binds to the dimer and uses light energy to cleave the cyclobutane ring linking the two adjacent, damaged nucleotides. The damage is repaired without any bases being excised.

DNA photolyases have been implicated in the evolution and range expansion of *Prochlorococcus* , the most abundant photosynthetic organism of the ocean. *Prochlorococcus* is found throughout the euphotic zone, which is the upper, sun-exposed layer of the ocean where photosynthesis can occur. The *Prochlorococcus* lineage is thought to have its origins in the deeper part of the euphotic zone, where UV light does not reach. Present-day lineages that live exclusively in the lower euphotic zone are most closely related to the ancestor of *Prochlorococcus* , and they lack DNA photolyases. In contrast, lineages found in the upper euphotic zone, where UV is present, have one or more photolyases. These lineages evolved more recently, and it is thought that the acquisition of photolyases by horizontal gene transfer (see Section 9.5) helped *Prochlorococcus* invade and colonize the upper euphotic zone.

Replacement of damaged bases. Another error-proof process, known as base excision repair (BER) , detects and replaces damaged bases. Uracil can be found in DNA either as a spontaneous deamination product of cytosine (see Fig. 9.5) or as a result of inappropriate incorporation during replication. Spontaneous deamination of adenine residues produces hypoxanthine. Uracil and hypoxanthine have base-pairing properties very different from those of the original bases, so their formation can lead to mutations.

Another kind of DNA damage is alkylation, the addition of an alkyl group such as methyl or ethyl to a ring nitrogen. Mutagens such as methyl methanesulfonate can alkylate adenine to produce 3-methyladenine. A 3-methyladenine residue will totally block DNA replication, making this a lethal form of damage. Consequently, repairing these mutations is vital for survival.

At the beginning of base excision repair, enzymes called glycosylases cleave the bond connecting the damaged base to the deoxyribose moiety in the phosphodiester backbone (**Fig. 9.12** , step 1). The result is an intact phosphodiester backbone missing a base. The site is called an AP site because it is either ap urinic (missing a purine) or ap yrimidinic (missing a pyrimidine). In step 2 of BER, AP endonucleases specifically cleave the phosphodiester backbone at AP sites. The 5′-to-3′ exonuclease activity of the gap-filling DNA polymerase I will degrade the cleaved strand downstream of the AP site and at the same time synthesize in its stead a replacement strand containing the proper base (step 3). DNA ligase seals the remaining nick, and the repair process is complete (step 4).

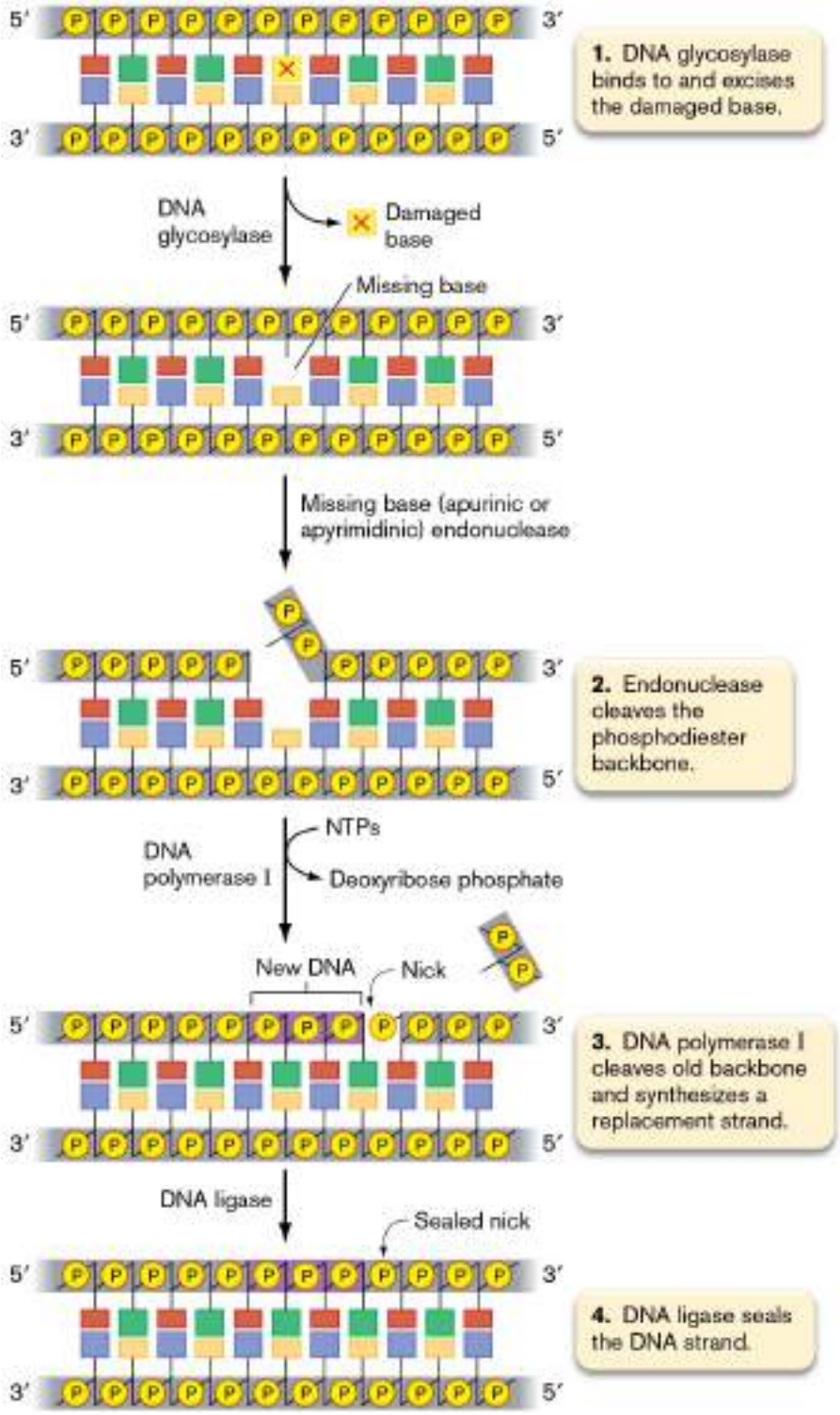


FIGURE 9.12 ■ Base excision repair. Replacement of a damaged base requires a succession of four enzymatic activities: glycosylase, endonuclease, polymerase, and ligase. NTPs = nucleoside triphosphates.

All of the repair processes described thus far (mismatch repair, nucleotide excision repair, photoreactivation, and base excision repair) are error-proof pathways that rely on the presence of an intact template strand opposite the damaged strand. Replication of the template strand is used to replace the damaged bases accurately. But what happens when both strands are damaged?

Repair of double-strand breaks by homologous recombination. Double-strand breaks that generate "blunt ends" of the chromosome are challenging to repair. Blunt ends do not have overlapping "sticky ends" that allow the DNA strands to reanneal and provide a stable substrate for DNA ligase to perform the repair. While possible, ligation of blunt ends is highly inefficient, because it relies on rare and short-lived conditions in which the two DNA ends have contacted each other through random collision. However, if the bacterial cell contains a second, intact chromosome copy in the cell, enzymes can use this undamaged copy as a guide to repair the broken chromosome. This repair mechanism involves homologous recombination , which is a form of strand exchange that occurs at shared regions of homology between DNA molecules. Notably, humans have homologs of these repair enzymes that protect human cells from DNA damage.

Figure 9.13 illustrates the mechanism of double- strand-break repair by homologous recombination. First, the two ends of the broken chromosome are trimmed back by nucleases to generate 3′ single-stranded ends (overhangs; **Fig. 9.13** , step 1). In *E. coli* , this trimming is performed by the RecBCD protein complex. The 3′ overhangs are recognized by the protein RecA, which binds in multiple copies to form a filament (step 2). The RecA filament mediates strand invasion of one 3′ overhang into the double helix of the second chromosome. This strand swapping occurs at the region of the second chromosome that is complementary to the sequence of the overhang. Using the undamaged complementary DNA as a template, the 3′ end is extended via a repair DNA polymerase (step 3). This extended DNA can now dissociate from the intact chromosome and anneal to the other 3′ overhang of the damaged chromosome by complementary base pairing (step 4). Repair polymerases can now complete the replacement of missing nucleotides on both strands of the damaged chromosome (step 5). Finally, DNA ligase seals the two nicks and restores the chromosome (step 6).

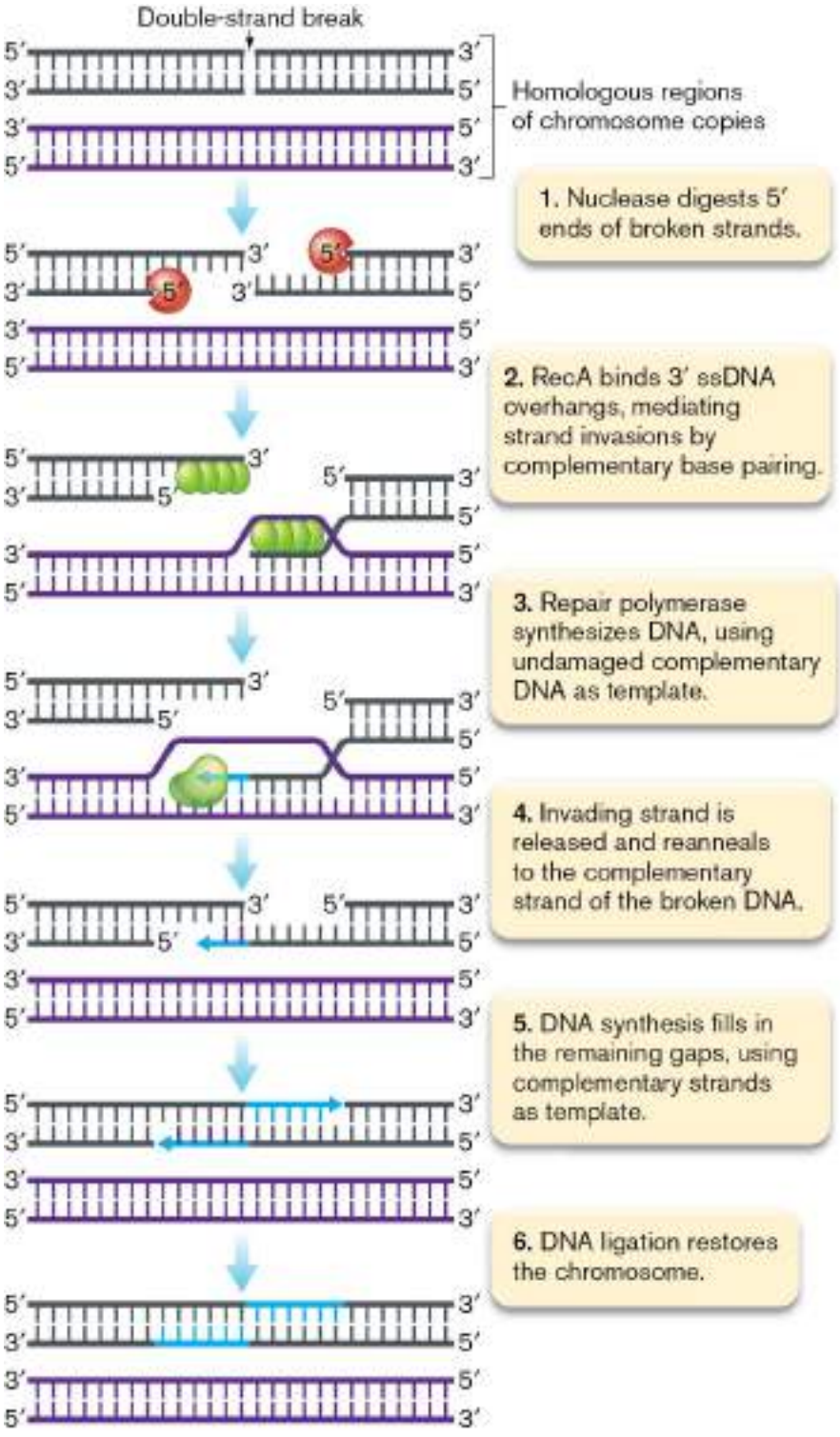


FIGURE 9.13 ■ Repair of a double-strand break by homologous recombination involving a second, undamaged chromosome.
ssDNA = single-stranded DNA.

Error-Prone DNA Repair

SOS ("Save Our Ship") repair. When DNA damage is extensive, repair strategies that excise pieces of DNA or that require recombination will destroy the integrity of the chromosome and kill the cell. To save the chromosome, and itself, the cell must take more drastic measures and induce the SOS response , a system that introduces mutations into severely damaged DNA. The cell relaxes replication fidelity to maintain an intact chromosome even if incorrect bases are introduced. In the SOS system, the RecA protein, normally located at the cell poles, senses the extent of DNA damage by monitoring the level of single-stranded DNA (ssDNA) produced. For example, excessive ultraviolet irradiation produces numerous ssDNA gaps because DNA polymerase cannot replicate through pyrimidine dimers. RecA interacts with the ssDNA and disengages from the poles.

Formation of RecA filaments (see **Fig. 9.13** , step 2) on ssDNA activates a second function of RecA, called coprotease activity. Coprotease activity stimulates autodigestion of LexA, a repressor protein that binds to the promoters of DNA repair genes and prevents their transcription (**Fig. 9.14A** and **B**). Autodigestion of LexA repressor unleashes the production of DNA repair enzymes. Among these enzymes are two "sloppy" DNA polymerases that lack proofreading activity: DinB (also called DNA polymerase IV, or Pol IV) and UmuDC (also called DNA polymerase V, or Pol V; **Fig. 9.14C**). The UmuDC enzyme is perfect for replicating through damaged bases (a process called translesion bypass replication) because it sacrifices accuracy for continuity. When the enzyme encounters an undecipherable damaged base, it will insert whatever nucleotide is available. The result, of course, will be numerous permanent mutations, but the benefit is that the cell has a chance to live if it can tolerate the mutations. The choice really is "mutate or die," because housekeeping DNA polymerases like Pol III cannot move through damaged DNA. (Note that the term "housekeeping" is applied to proteins or enzymes that keep the cell running at all times.)

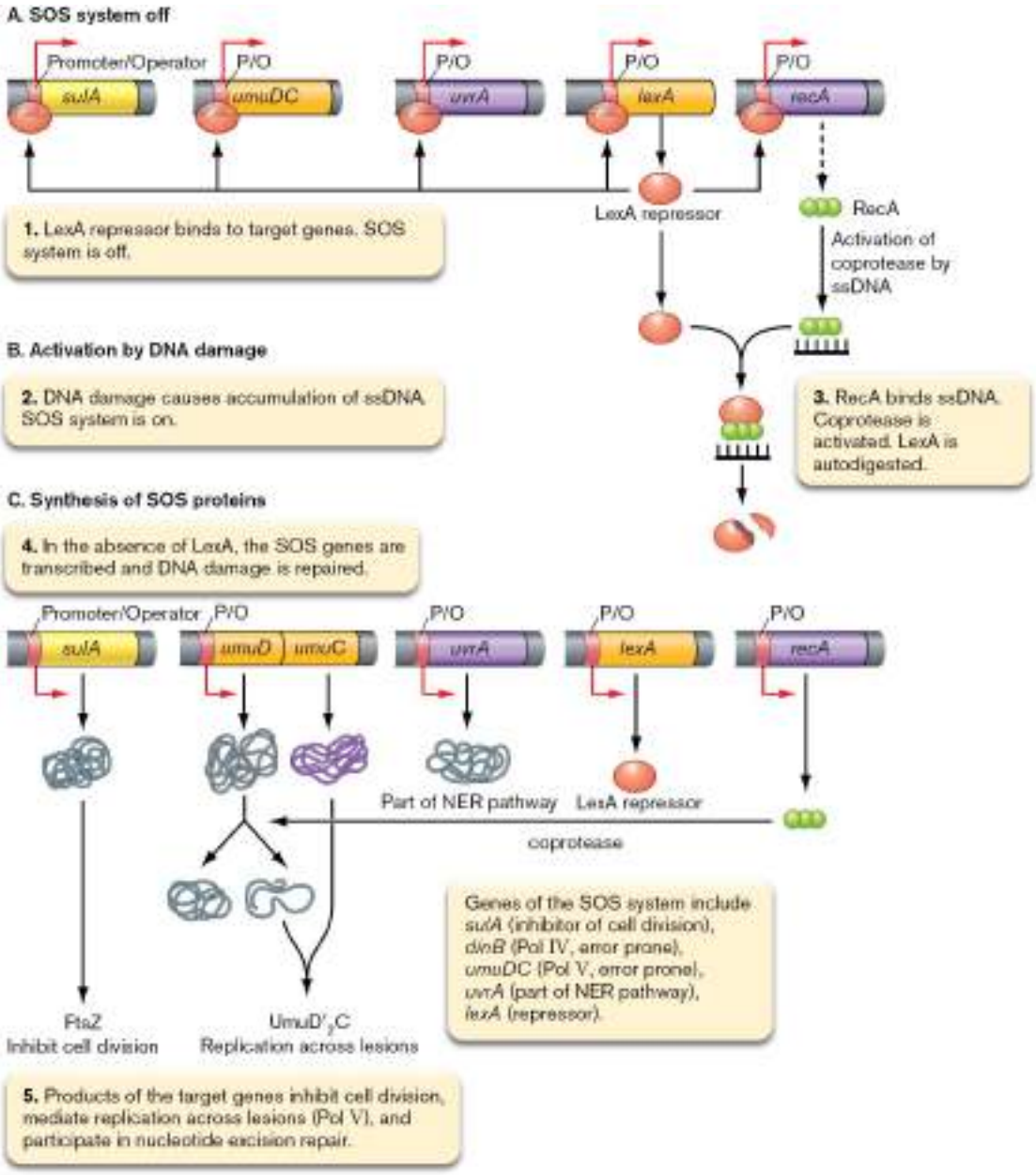


FIGURE 9.14 ■ Regulation of the SOS response system. The emergency DNA repair system known as the SOS response is induced when there is extensive DNA damage. The system is not a single repair mechanism but a set of different mechanisms that collaborate to rescue the cell. NER = nucleotide excision repair.

When the cell is severely compromised by mutation, it temporarily halts cell division. Like an emergency pit stop during an auto race, this pause in cell division allows time for repair enzymes to fix the damage. The product of another SOS-regulated gene, called *sulA* , causes this pause by binding to the

FtsZ cell division protein, keeping it from initiating cell division (see Sections 3.1 and 3.4). Once the damage has been repaired, RecA coprotease is inactivated and LexA repressor accumulates, turning off all the SOS genes, including *sulA* . The lingering SulA protein is then degraded by a protease called Lon. The protein is called Lon because when it is missing, SulA inappropriately accumulates, inhibits cell division, and produces long filamentous cells.

The coprotease activity of RecA can also activate prophages by stimulating autocleavage of proteins that prevent phage replication. This can be considered a form of eavesdropping on the host cell by the prophage: In environmental conditions that trigger the SOS response, the virus may have a better chance of survival and replication as a viral particle rather than as a lysogen.

The SOS induction of viruses can be exploited by some organisms to displace competitors in an environmental niche. For instance, *Streptococcus pneumoniae* , a cause of pneumonia, displaces *Staphylococcus aureus* in the human nasopharynx by producing hydrogen peroxide. Hydrogen peroxide damages *Staphylococcus aureus* DNA, which activates the SOS response and triggers replication of resident bacteriophages that lyse the cell. The loss of *Staphylococcus aureus* means more room for *Streptococcus pneumoniae* .



Nonhomologous end joining. Double-strand breaks in DNA are particularly dangerous to a cell because the loss of chromosome integrity is immediate. In rapidly growing bacteria such as *E. coli* , double-strand breaks can be repaired by homologous recombination as long as another copy of the chromosome is present to guide the repair. But in slow-growing bacteria such as *Mycobacterium tuberculosis* , a second chromosome copy is usually not available. These bacteria can use an intriguing repair mechanism called n on h omologous end joining (NHEJ) , which is also used by mammals (**Fig. 9.15**).

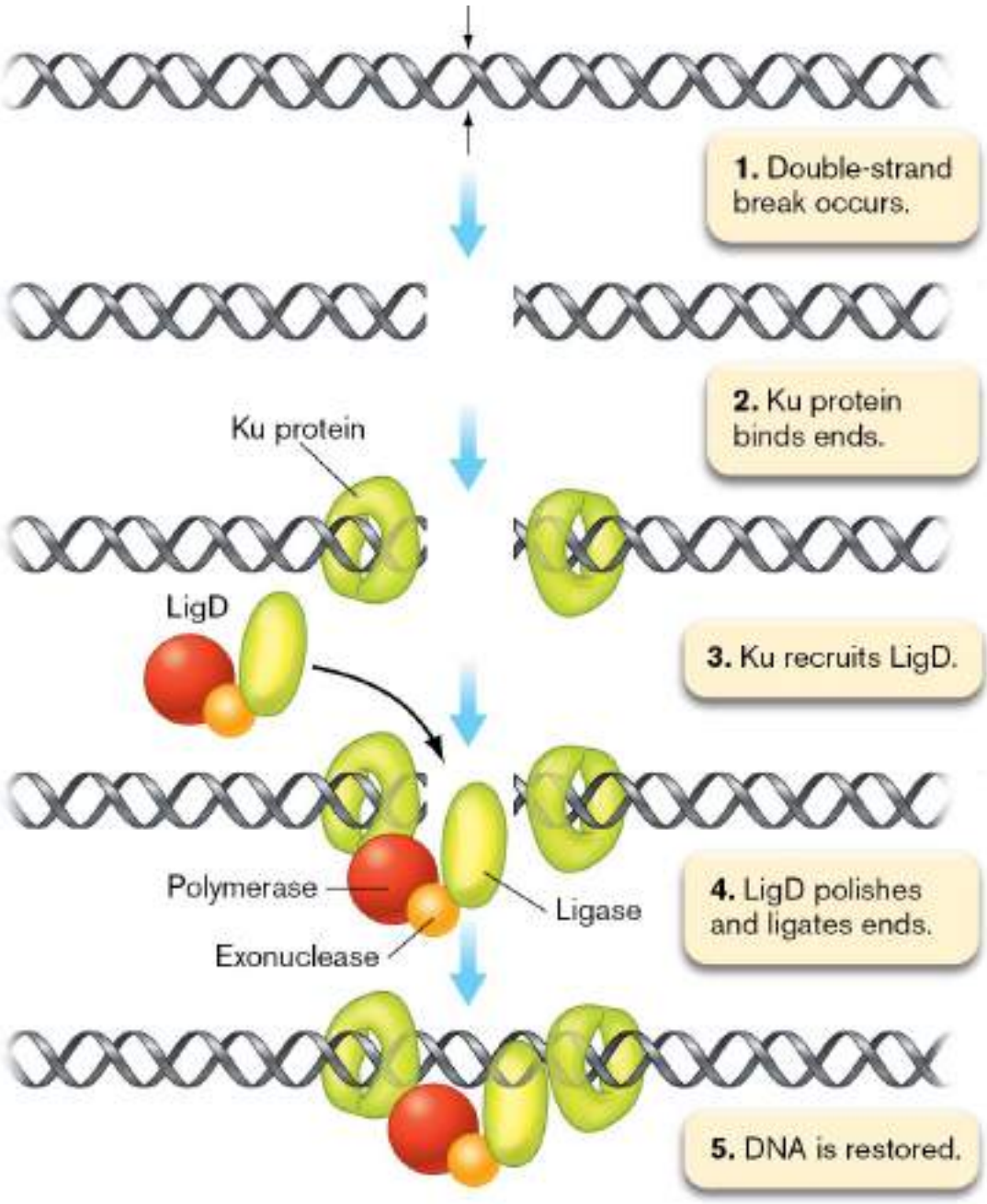


FIGURE 9.15 ■ Nonhomologous end joining. NHEJ, a common repair mechanism in eukaryotes, has been documented in some bacteria, such as *Mycobacterium tuberculosis* . These pathogens can survive within macrophages that produce antibacterial compounds (nitric oxide and hydrogen peroxide) that generate double-strand breaks in DNA. NHEJ mechanisms can repair those breaks in double-stranded DNA (dsDNA).
NIH/COURTESY OF THE UNIVERSITY OF WISCONSIN-MADISON ARCHIVES (ID S04387)

Two bacterial proteins, Ku and LigD, carry out NHEJ repair. Ku protein binds to the ends of a double-strand break and recruits LigD, a protein that has polymerase and 3′ exonuclease activities that fill in or remove single-strand overhangs and a ligase activity that joins two double-strand breaks. Because the system does not require homology, it can be error prone, causing the loss or addition of a few nucleotides at the break site or even the joining of two previously unlinked DNA molecules.

To Summarize

- **DNA repair pathways** in microorganisms include error-proof and error-prone mechanisms.
- Methyl-directed mismatch repair uses methylation of the parental DNA strand to distinguish it from newly replicated DNA. The premise is that the parental strand will contain the proper DNA sequence.
- Photoreactivation cleaves the cyclobutane rings of pyrimidine dimers.
- Base excision repair (BER) excises structurally altered bases without cleaving the phosphodiester backbone. The resulting AP (apurinic or apyrimidinic) site is targeted by AP nucleases.
- Recombinational repair uses an intact copy of the DNA to repair a double-strand break in the damaged copy.
- **Extensive DNA damage leads to induction of the SOS response** , producing increased levels of the error-proof repair systems, as well as error-prone translesion bypass DNA polymerases that introduce mutations.
- Nonhomologous end joining (NHEJ) mechanisms repair double-strand DNA breaks in some, but not all, bacteria.

Glossary

error-proof repair
: DNA repair mechanisms that minimize the occurrence of mutations.

error-prone repair
: Low-accuracy DNA repair mechanisms that allow mutations.

methyl-directed mismatch repair

A DNA repair system that fixes misincorporation of a nucleotide after DNA synthesis. The unmethylated daughter strand is corrected to complement the methylated parental strand.

mutator strain

A strain of cells with a high mutation rate, usually due to a mutation in a DNA repair enzyme.

nucleotide excision repair (NER)

A DNA repair mechanism that cuts out damaged DNA. New, correctly base-paired DNA is synthesized by DNA polymerase I.

photoreactivation

A light-induced, photolyase-catalyzed repair of pyrimidine dimers.

base excision repair (BER)

A DNA repair mechanism that cleaves damaged bases off the sugar-phosphate backbone. After endonuclease activity at the AP site, a new, correct DNA strand is synthesized complementary to the undamaged strand.

AP site

See **apurinic site** .

homologous recombination

The process by which two DNA molecules exchange arms by cutting and splicing their helix backbones. Exchange occurs between sequences that are identical or nearly identical, as the machinery requires complementary base pairing to exchange the DNA molecules.

SOS response

A coordinated cellular response to extensive DNA damage. It includes error-prone repair.

nonhomologous end joining (NHEJ)

A pathway that repairs double-strand breaks in DNA by direct ligation without the need for large regions of homology.

methyl-directed mismatch repair

A DNA repair system that fixes misincorporation of a nucleotide after DNA synthesis. The unmethylated daughter strand is corrected to complement the methylated parental strand.

photoreactivation

A light-induced, photolyase-catalyzed repair of pyrimidine dimers.

base excision repair (BER)

A DNA repair mechanism that cleaves damaged bases off the sugar-phosphate backbone. After endonuclease activity at the AP site, a new, correct DNA strand is synthesized complementary to the undamaged strand.

recombinational repair

A DNA repair mechanism that relies on recombination between an undamaged chromosome and a gap that occurred during replication of damaged DNA.

nonhomologous end joining (NHEJ)

A pathway that repairs double-strand breaks in DNA by direct ligation without the need for large regions of homology.

Fig. 9.5

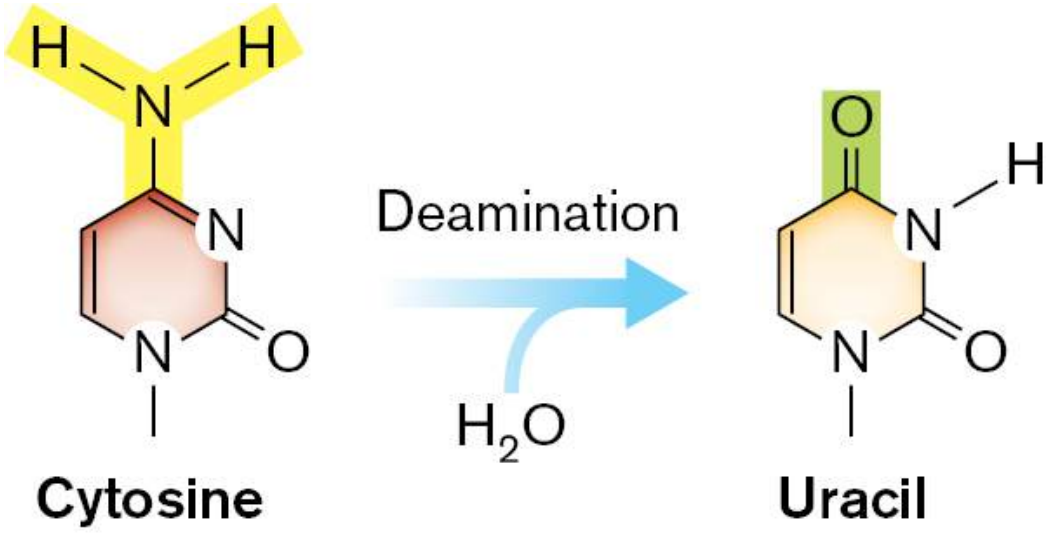


FIGURE 9.5 ■ Spontaneous deamination of cytosine. Oxidative deamination changes cytosine to uracil.

9.3 Gene Transfer Mechanisms

Microbial genomes demonstrate a remarkable ability to quickly undergo extensive gene exchange. One striking example involves the pathogen *Streptococcus pneumoniae* , which causes meningitis and serious lung infections in any age group and also recurrent ear infections in infants. A research team headed by Fen Hu and Garth Ehrlich (Drexel University) found evidence of evolution in a strain of *S. pneumoniae* isolated from an 8-month-old child stricken with recurrent ear infections over 7 months. Genome sequences of strains isolated during separate bouts revealed that 16 distinct gene transfer events had occurred from the earliest isolate to the latest. Over those 7 months the child had been colonized by multiple strains of *S. pneumoniae* that readily received DNA from other microbes. The resultant gene transfers enabled the species to quickly evolve and cause recurrent disease.

In another example of gene transfer, John Sullivan and Clive Ronson from New Zealand witnessed microbial evolution taking place in soil within a mere decade. These researchers were studying the microbe *Mesorhizobium loti* , a symbiotic bacterium that forms nitrogen-fixing nodules on plant roots. Genes encoding symbiosis are located as a group on the mesorhizobial chromosome. In a remarkable experiment, Sullivan and Ronson inoculated a single strain of *M. loti* into an area of land devoid of natural nodulating rhizobia. Seven years later, they discovered that the area contained many genetically diverse symbiotic mesorhizobia now able to nodulate the flowering plant *Lotus corniculatus* . These microbes did not exist before the experiment. The 500-kb genome segment encoding symbiosis had somehow made its way from *M. loti* into these other bacteria—essentially generating new species.

We can also study the microbial gene transfer that has occurred over much longer timescales. One remarkable feature of gene exchange in microbes is that the donors are often distantly related to the recipient. Archaea, for example, arose from a common

eukaryotic-archaeal phylogenetic branch and, as a result, possess many traits in common with eukaryotes, such as the structure and function of their DNA and RNA polymerases. However, many archaeal genes whose products are involved with intermediary metabolism are probably of bacterial origin. In fact, 37% of the proteins found in the archaeon *Methanocaldococcus jannaschii* are found in all three domains—Archaea, Eukarya, and Bacteria. Another 26% are otherwise found only among bacteria and archaea, while a mere 5% are confined to archaea and eukaryotes. Archaeal genomes are not alone in this mosaicism, as recent bioinformatic studies of the well-known *Escherichia coli* bacterium have discovered numerous genes with foreign origins in its genome. All three domains, then, enjoy a mixed heritage. How does all this genomic blending happen during microbial evolution?

Before we describe the different mechanisms of gene transfer, it is worth stepping back and asking why it might be advantageous for bacteria and archaea to take up foreign DNA. Several advantages have been proposed:

- First, species that indiscriminately import DNA may use the DNA as food ; it is an energy-rich molecule and a good source of carbon, nitrogen, and phosphorus.
- Second, imported DNA can be used to repair damaged chromosomes. This capability is especially relevant in instances when DNA from genetically similar microbes is taken up, because it can undergo homologous recombination to replace damaged segments of DNA.
- Finally, gene transfer can play a role in evolution by introducing new genes into the genome. Once expressed, these new genes can provide novel functions to the cell, which may enable them to compete at higher fitness in their current environment and also allow them to invade new environments. Gene transfer may have enabled pathogens such as *Neisseria gonorrhoeae* , the cause of gonorrhea, to acquire genes whose protein products now help the organism evade the host immune system.

Gene Transfer by Conjugation

In 1946, a 20-year-old student at Yale University named Joshua Lederberg (1925–2008; **Fig. 9.16A**) reported the first instance of recombination in prokaryotes. His adviser, Ed Tatum (of "one gene–one enzyme" fame), provided Lederberg with *E. coli* strains that each contained a different set of auxotrophic mutations (**Fig. 9.16B**). These mutants could not grow on agar plates unless the media were supplemented with the metabolites that their auxotrophic mutations prevented them from synthesizing. Lederberg hypothesized that if *E. coli* were capable of gene transfer and recombination, then when two different auxotrophs were mixed together, the alleles (auxotrophic and wild type) of the genes in question would sort out in various combinations in the progeny. Among these recombinant progeny would be a special class: those that received the wild-type alleles of all four genes. Such "prototrophic" cells could be found easily because they would be the only ones that could grow on agar plates without metabolite supplementation. This is exactly what Lederberg observed.

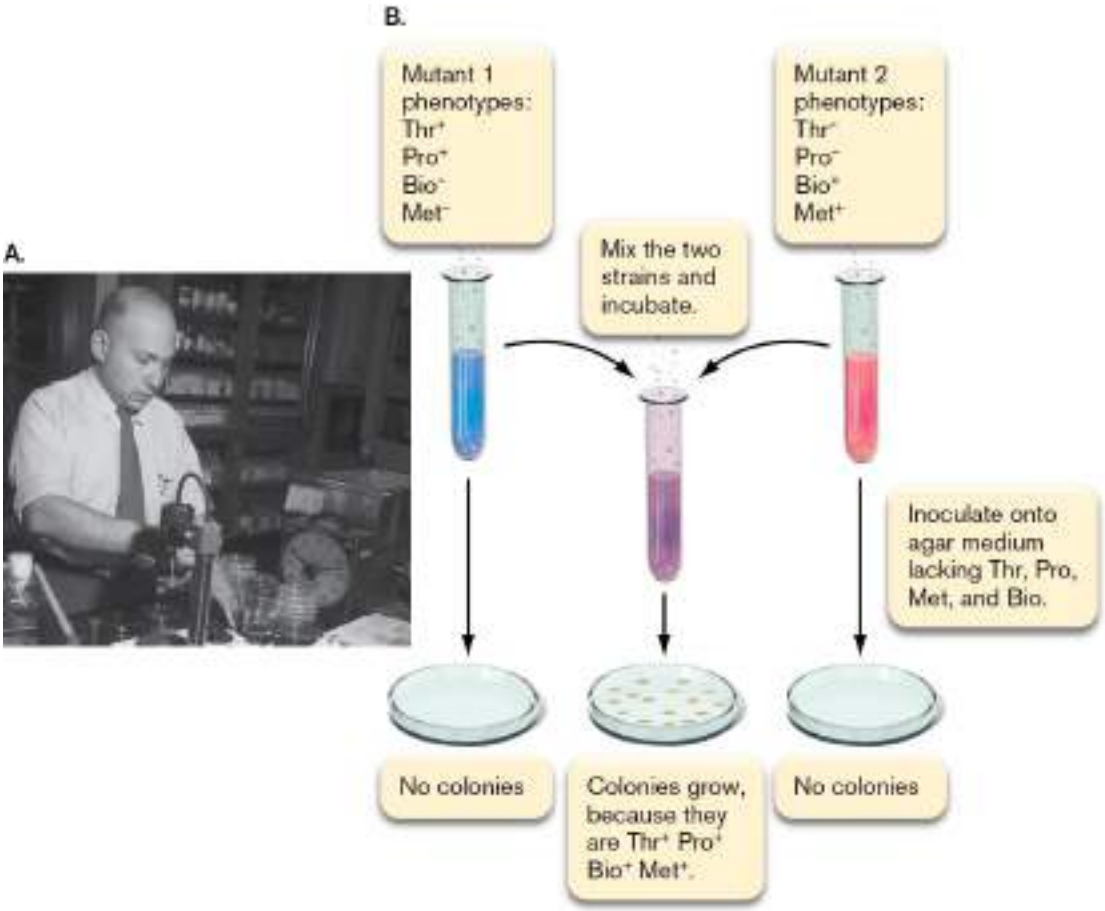


FIGURE 9.16 ■ The discovery of sexual recombination in *Escherichia coli* . A. Joshua Lederberg in the lab (circa 1958). **B.** Mutants with two sets of auxotrophies, which prevent the cells from making either their own biotin and methionine (mutant 1) or threonine and proline (mutant 2), cannot grow on agar medium lacking these molecules. However, after these two strains are mixed and incubated, they generate recombinant offspring, some of which are prototrophic and are able to grow as colonies on plates where the parental auxotrophs could not.

NIH/COURTESY OF THE UNIVERSITY OF WISCONSIN-MADISON ARCHIVES (ID S04387)

Lederberg's discovery that alleles can be recombined opened the gateway for the exploration of gene function and regulation in *E. coli* , with the major consequence that the best-studied bacterium on Earth is a member of the human gut microbiome. For his discovery

of recombination in bacteria, Lederberg won the 1958 Nobel Prize in Physiology or Medicine. Subsequent studies revealed that the mechanism of gene transfer required cell-cell contact and was thus fundamentally distinct from the transformation method of *Streptococcus pneumoniae* (see Section 7.1). Lederberg had discovered bacterial conjugation.



Conjugation is a form of unidirectional gene transfer that requires contact between donor cell and recipient cell. In Gram-negative bacteria, cell-cell contact is typically mediated by a special retractable pilus (**Fig. 9.17**) protruding from a donor cell. The pilus is a component of the type IV secretion system (T4SS; see Section 25.4) that also delivers the DNA to the recipient cell. Conjugation occurs in many species of bacteria and archaea, even in hyperthermophiles such as *Sulfolobus* species. Conjugation can transfer DNA to cells of the same species or different species, even different domains.

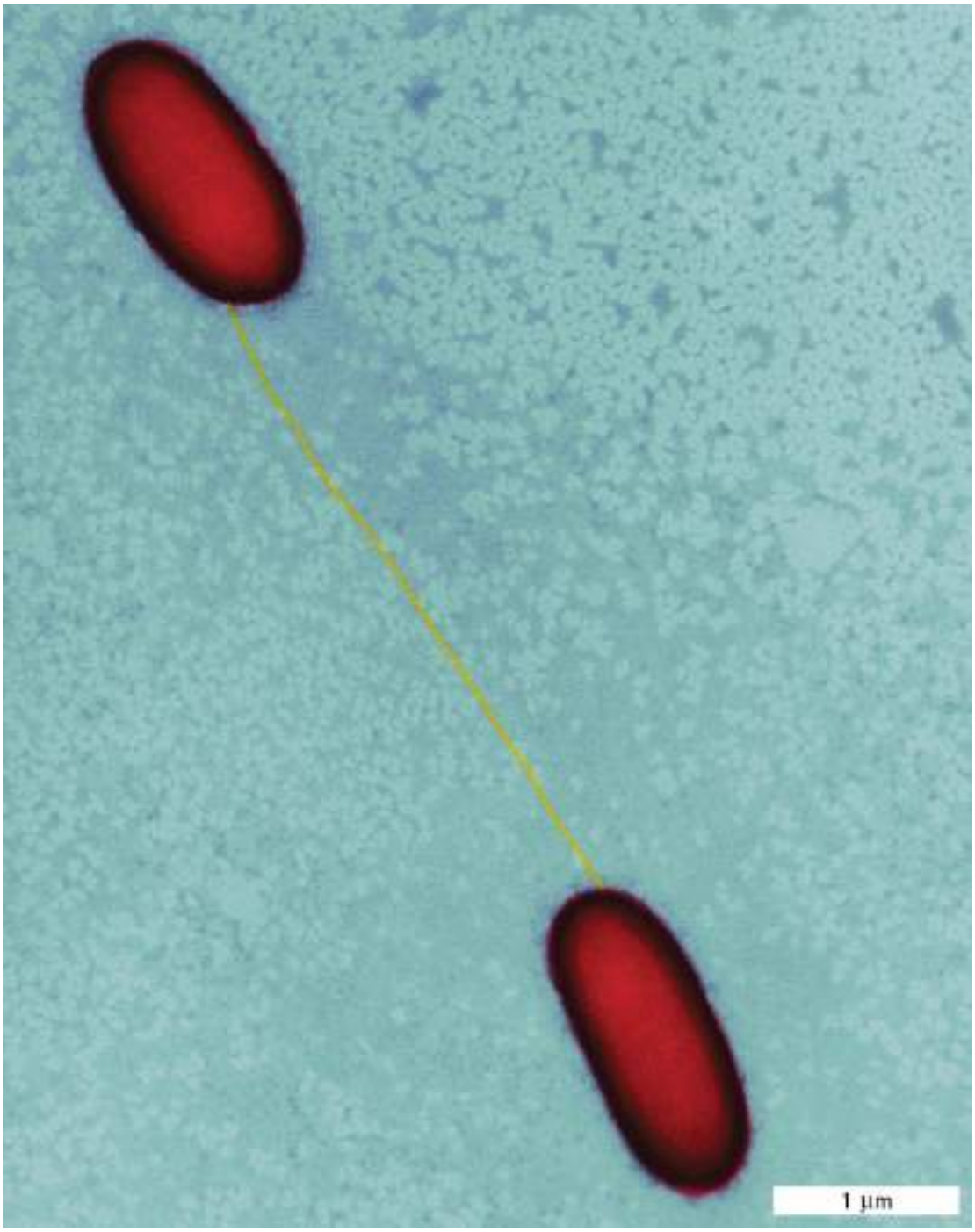


FIGURE 9.17 ■ Sex pilus connecting two *Escherichia coli* cells. Pseudocolor added.

DENNIS KUNKEL MICROSCOPY, INC./SCIENCE SOURCE

One striking example is *Agrobacterium tumefaciens* , which causes crown gall disease in plants (**Fig. 9.18A**). Using its type IV secretion machinery, *A. tumefaciens* transfers its tumor-inducing (Ti) plasmid via conjugation into plant cells. The bacteria detect and swim toward phenolic compounds released from the wound of a damaged plant. At the wound site, the bacteria attach to the plant cell (**Fig. 9.18B**). Subsequent transfer, integration, and expression of the Ti plasmid in the plant cell genome triggers the release of plant hormones that stimulate tumorous growth of the plant. Plant cells within the tumor release amino acid derivatives, called "opines," which the microbe can then use as a source of carbon and nitrogen. The unique mode of action of *A. tumefaciens* has made this bacterium an indispensable tool for plant breeding and enables entirely new (nonplant) genes to be engineered into crops (discussed in Chapter 18).

A.



B.

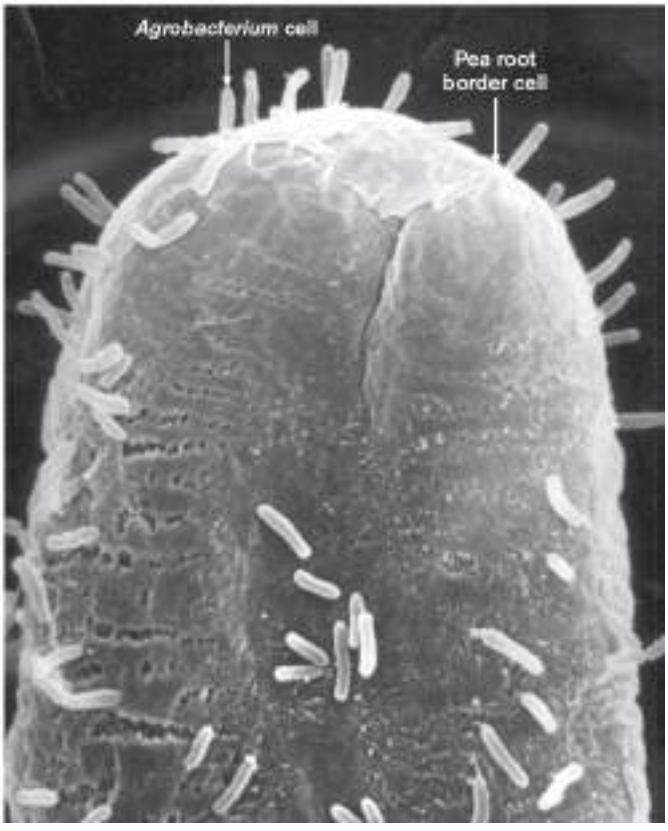


FIGURE 9.18 ■ An example of gene transfer between bacteria and plants. A. Crown gall disease tumor caused by the bacterium *Agrobacterium tumefaciens* . **B.** Electron micrograph of *A. tumefaciens* attached to a pea root border cell.

DR. ROBERT CALENTINE/VISUALS UNLIMITED, INC.
MARTHA HAWES, UNIVERSITY OF ARIZONA

Note: Conjugation in bacteria and archaea is mechanistically different from conjugation among eukaryotic microbes. Some eukaryotic microbes exchange nuclei through a structure, called a conjugation bridge, that is very different from the T4SS of bacteria (see Chapter 20).

Bacterial conjugation typically involves special transferable plasmids that contain all the genes needed for pilus formation and DNA export. In some cases, plasmids can also contain genes encoding antibiotic resistance. Lederberg's discovery of recombination involved a well-studied, transferable plasmid in *E. coli* called fertility factor , or F factor (**Fig. 9.19**). To carry out plasmid replication and DNA transfer, F factor uses two replication origins, *oriV* and *oriT* , located at different positions on the plasmid. The origin *oriV* is used to replicate and maintain the plasmid in nonconjugating cells, whereas *oriT* is used only to replicate DNA during DNA transfer. F factor also contains several *tra* genes, whose protein products carry out the DNA transfer process.

Conjugation of the F factor begins with cell-cell contact between a donor F^+ cell that carries the plasmid and a recipient F^- cell (**Fig. 9.19** , step 1). Contact causes pilus retraction and triggers activation of the plasmid-encoded relaxosome complex. The relaxosome is physically associated with the pilus, and before activation the relaxase subunit of the relaxosome binds the plasmid at *oriT* (see step 1). Upon activation, the relaxosome produces a

single-stranded DNA bubble in *oriT* , which recruits a second relaxase molecule (step 2). This relaxase unwinds the DNA to separate the strand that will be transferred from the strand that will be retained. At this point, the original relaxase nicks the phosphodiester backbone of one strand at the *nic* site in *oriT* (step 3). The relaxase remains covalently bound to the strand of DNA, and this complex is delivered to the T4SS for transfer. The DNA-relaxase complex is then transferred to the recipient cell through the interior channel of the T4SS. The intact circular strand remains in the donor. DNA polymerase III (Pol III) is then recruited to *oriT* in the donor, where replication begins (step 4). In contrast to bidirectional replication, which is used to duplicate the chromosome, replication for conjugative transfer occurs unidirectionally, by rolling-circle replication (step 5; see Section 7.4).

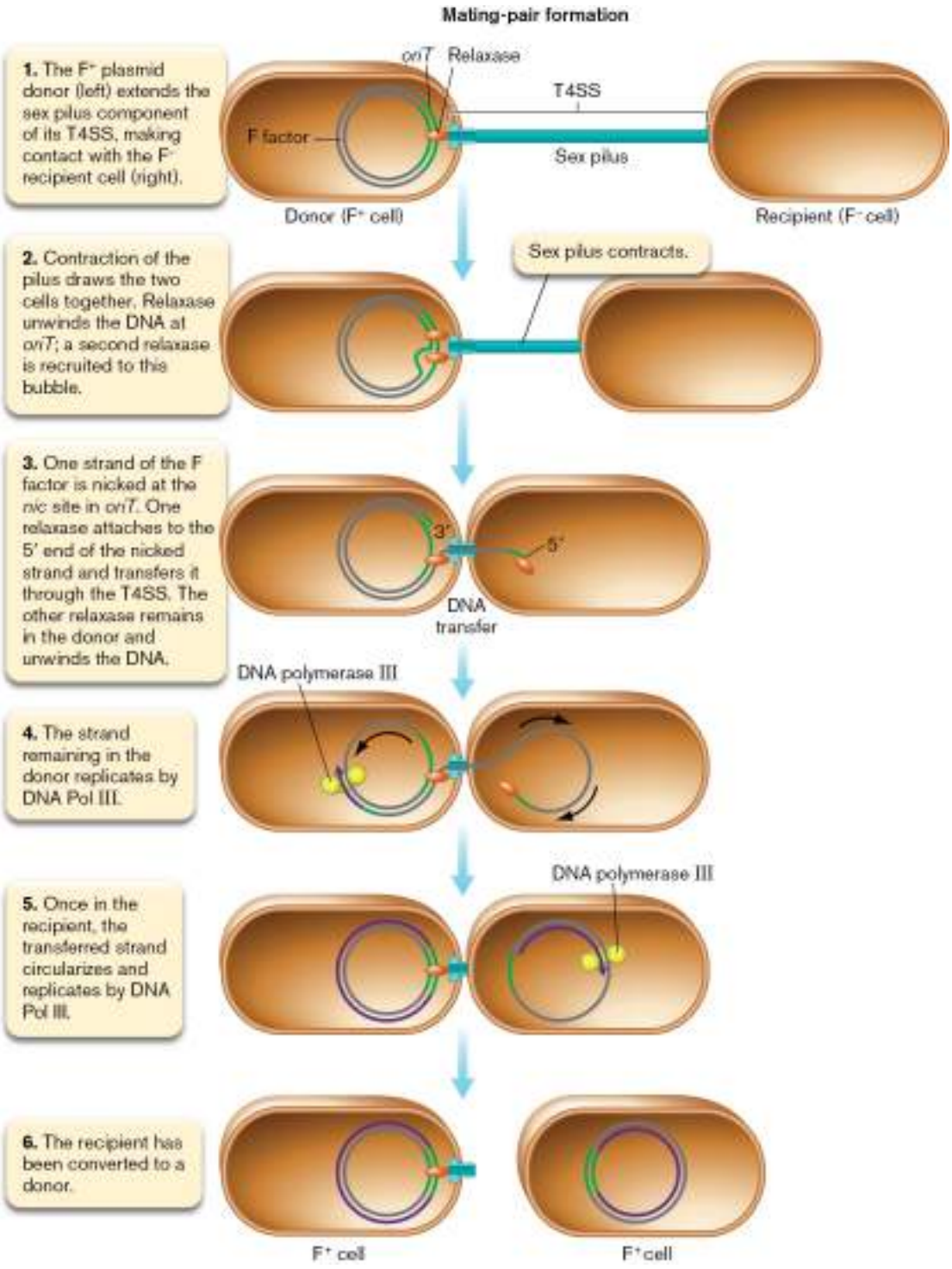


FIGURE 9.19 ■ The conjugation process. Some plasmids, such as F factor in *E. coli* , can mediate cell-to-cell transfer of DNA. Purple arrowheads mark the 3′ end of replicating DNA.

Once the transfer is complete, the relaxase re-ligates the 5′ end it was holding to the 3′ tail of the transferred strand in the recipient. Thus, the last portion of F factor moved to the recipient is *oriT* . Once circularized, the plasmid is replicated by DNA polymerase III, and the F^+ recipient cell becomes a new F^+ donor cell (**Fig. 9.19** , step 6). Ultimately, the conjugation complex spontaneously comes apart, and the membranes seal. This transfer process is very quick, taking less than 5 minutes to transfer the entire 110-kb F factor.

Note: The sex pilus is hollow, and some evidence suggests it can serve as the channel for DNA transfer between cells. However, this role is not universally accepted in the field. Most evidence suggests DNA transfer occurs when there is close contact between cells, and may involve a T4SS whose pilus has been fully retracted (depolymerized into pilin subunits). As this is currently an unresolved question, for clarity we show DNA transfer steps with a T4SS that lacks the pilus component.

Is it possible for two donors to exchange DNA with each other? For bacteria, the answer is generally no (however, some eukaryotic microbes can reciprocally exchange genomes by a very different mechanism, as seen in Chapter 20). In bacteria, donors have mechanisms in place to prevent reciprocal DNA exchange. For example, a membrane protein encoded by the F factor will inhibit formation of a conjugation complex with another donor that possesses the same protein. This mechanism prevents the pointless transfer of a plasmid to a cell that already has that plasmid.

Many different types of plasmids can be found in the microbial world. Most are not transferable by themselves. However, one group of plasmids that cannot transfer themselves can be mobilized if a transferable plasmid such as F factor is also present in the same cell. Mobilizable plasmids usually contain an *oriT* -like DNA replication origin recognized by the conjugation apparatus of the

transferable plasmid. As a result, when the transferable plasmid begins conjugating, so does the mobilizable plasmid. This is one way in which antibiotic resistance genes on plasmids called R factors can be spread throughout a microbial population. Note that not all mobilizable plasmids contain antibiotic resistance genes.

The F-factor plasmid can transfer chromosomal genes. Thus far we have observed how the F factor of *E. coli* can transfer itself as a plasmid from cell to cell, but how was Lederberg able to use this system to recombine genes on the chromosome? This process required the F-factor plasmid to integrate into the chromosome (**Fig. 9.20** , step 1). Integration of the plasmid into the genome involves recombination between regions of homology between the plasmid and chromosome, such as the IS *3* element. A strain with the F factor integrated into the chromosome is called an Hfr (h igh-f requency r ecombination) **strain** . The strain is called "high-frequency" because every cell is capable of transferring chromosomal DNA to an F^- cell. Hfr strains are rare in nature, and it was quite fortuitous that one such strain (mutant 1 in **Fig. 9.16B**) was in the hands of Lederberg.

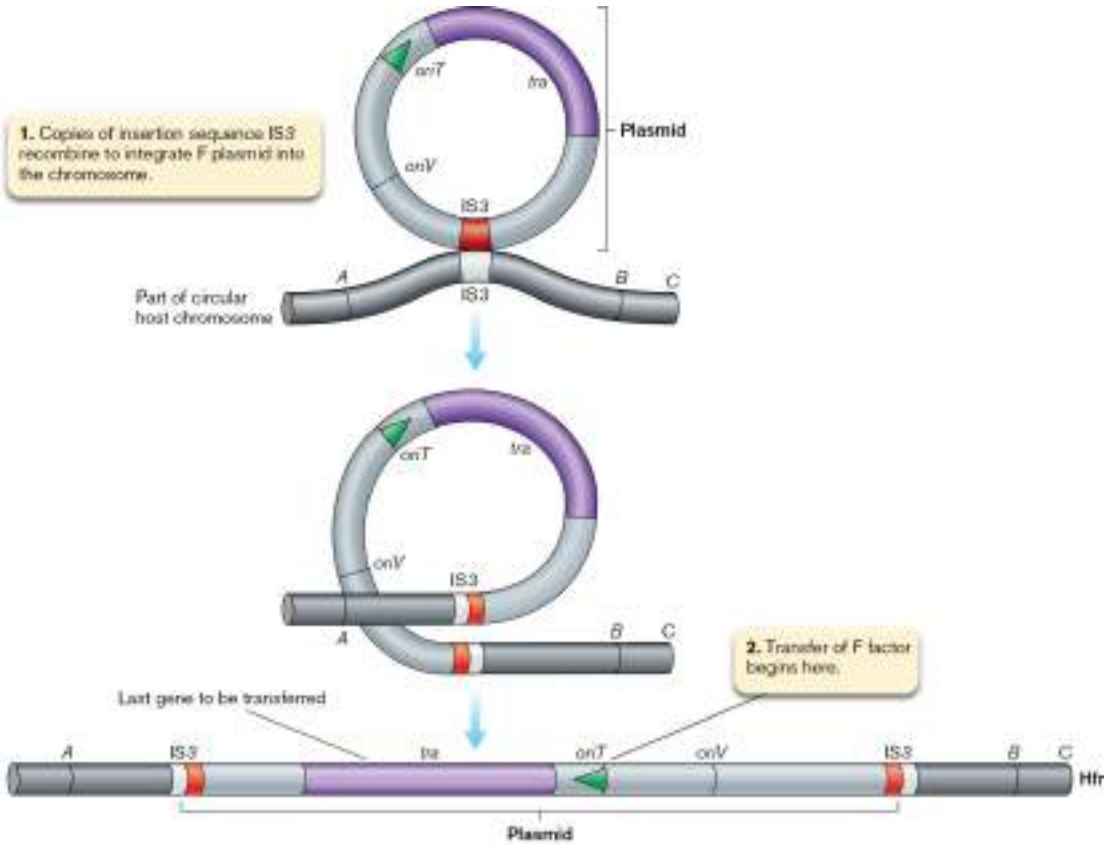


FIGURE 9.20 ■ Hfr formation. Insertion sequence 3 (IS *3* ; see Section 9.4) serves as a region of sequence homology between an F factor and the host chromosome. Chromosome and F-factor copies of IS *3* are recognized by host cell recombination proteins that then integrate F factor into the host chromosome. The transfer genes (*tra*) and replication origins (*oriV* = vegetative; *oriT* = transfer) are shown.

The integrated F factor of the Hfr strain can transfer one strand of DNA through the T4SS as it would in plasmid form, except now it transfers chromosomal genes (**Fig. 9.21**). Essentially, the entire circular chromosome becomes the F factor. Chromosomal genes adjacent to and behind *oriT* are transferred first (**Fig. 9.20** , step 2), and the remainder of the F factor, including the *tra* gene, is transferred last.

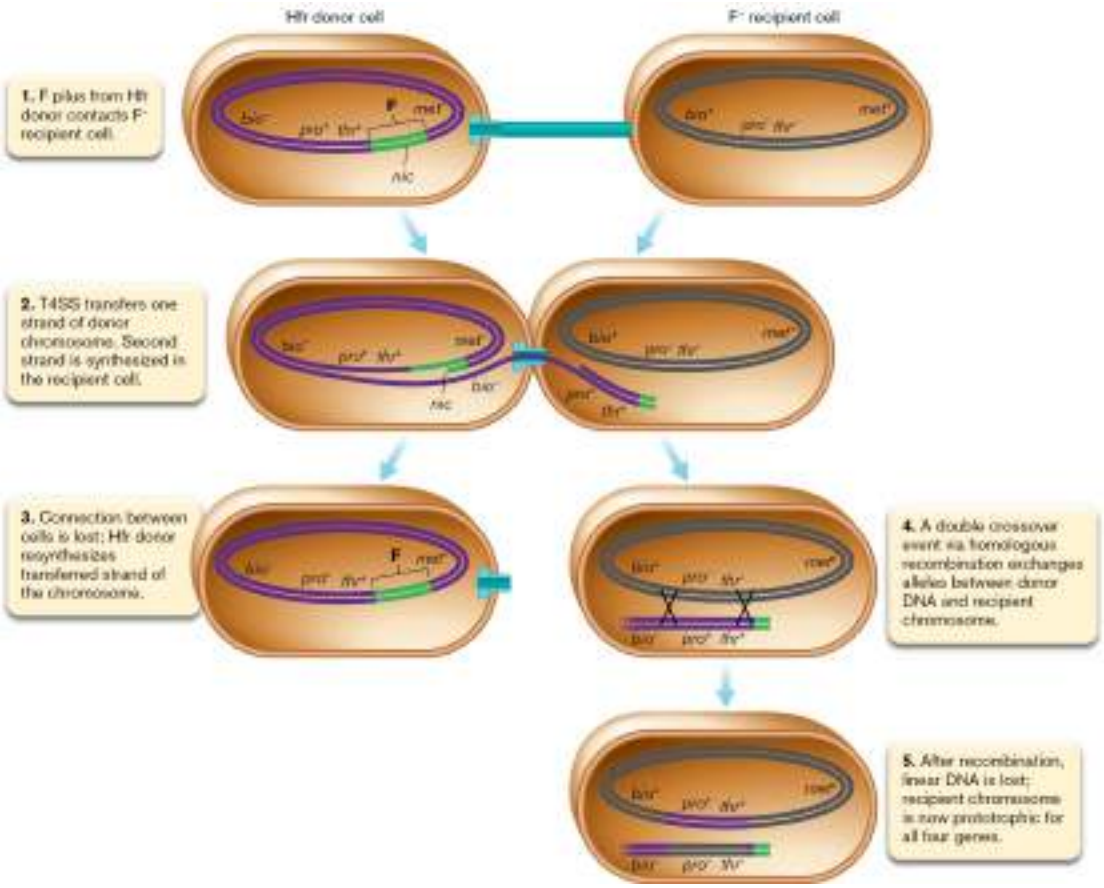


FIGURE 9.21 ■ Transfer of chromosomal genes of an Hfr strain.

Figure 9.21 shows a reconstruction of the Hfr mating experiment that gave Lederberg the results of **Fig. 9.16B** , here using current knowledge of the relative position of the genes and the F factor he used. The F factor transferred the *thr* $^{+}$, *pro* $^{+}$, and *bio* $^{-}$ alleles into the *thr* $^{-}$ *pro* $^{-}$ *bio* $^{+}$ *met* $^{+}$ recipient (**Fig. 9.21** , step 4). The Hfr strain was unlikely to have transferred the *met* $^{-}$ allele, because it would have been one of the last genes transferred, and the pilus connection is usually broken long before this can occur. Once the donor DNA strand entered the cell, it was used as a template to synthesize the second strand. This double-stranded DNA could then exchange alleles with the recipient chromosome via homologous recombination (step 4). For matings involving recombination (crossover) events that flanked the *thr* and *pro* loci, the auxotrophic

alleles in the recipient (*thr* $^{-}$ and *pro* $^{-}$) were replaced with the wild-type alleles of the Hfr donor, and a *thr* $^{+}$ *pro* $^{+}$ *bio* $^{+}$ *met* $^{+}$ prototrophic strain was generated (step 5).

Hfr conjugation enabled researchers to construct the genetic map of *E. coli* . Today, DNA sequencing is used as a rapid and inexpensive technique to map genes in other organisms. Nevertheless, conjugation remains important as a tool for scientists to move genes between species, and in nature it is an important means of microbial evolution via horizontal gene transfer (see Section 9.5).

As a final note, the F factor in an Hfr strain can excise from the chromosome and replicate once again as an autonomous plasmid. In rare instances, excision is imperfect and the plasmid takes with it part of the chromosome. These plasmids are called F-primes (F′). If transferred to an F $^{-}$ wild-type *E. coli* , the F-prime plasmid provides a second copy of the chromosomal genes it carries. The ability to generate partial diploids made F-primes highly valuable in the early days of *E. coli* genetics. Today we use recombinant DNA technology to clone specific DNA sequences on plasmids, as described in eAppendix 3 .

Thought Question

9.5 Transfer of an F factor from an F $^{+}$ cell to an F $^{-}$ cell converts the recipient to F $^{+}$. Why doesn't transfer of an Hfr do the same?

Gene Transfer by Phage Transduction

Bacteriophages harbor their own genomes, distinct from those of their host cells, but could phages also pluck host genes from one cell and move them to another? Norton Zinder (1928–2012) was the first

to suspect that bacteriophages could taxi chromosomal DNA between cells. To test this idea, he grew *Salmonella* in two tubes separated by a fine filter through which viruses could pass, but not bacteria. The experiments, reported in 1966, showed that a filterable agent, or virus, could carry genetic material between bacterial strains; direct contact between bacteria was unnecessary. The process in which bacteriophages carry payloads of host DNA from one cell to another is known as transduction . There are two basic types of transduction: generalized and specialized. Generalized transduction can take any gene from a donor cell and transfer it to a recipient cell, whereas specialized transduction can transfer only genes located near the chromosomal integration site of the bacteriophage.

How does transduction happen? Bacteriophages capable of generalized transduction, such as the *Salmonella* P22 bacteriophage investigated by Zinder, have trouble distinguishing their own DNA from that of the host when attempting to package DNA into their capsids, so pieces of bacterial host DNA accidentally become packaged in the phage capsid instead of phage DNA. As a result, the packaging system sometimes mistakenly packages host DNA instead of phage DNA.

The result of this mistake is that 1% of the phage particles in any population of P22 contain no phage DNA but carry host DNA plucked from around the chromosome, up to about 44 kb in length, or almost 1% of the genome (**Fig. 9.22**). The phages that carry host DNA are called transducing particles. Any single transducing particle will contain only one segment of host DNA, but different particles in the phage population will contain different segments of host DNA.

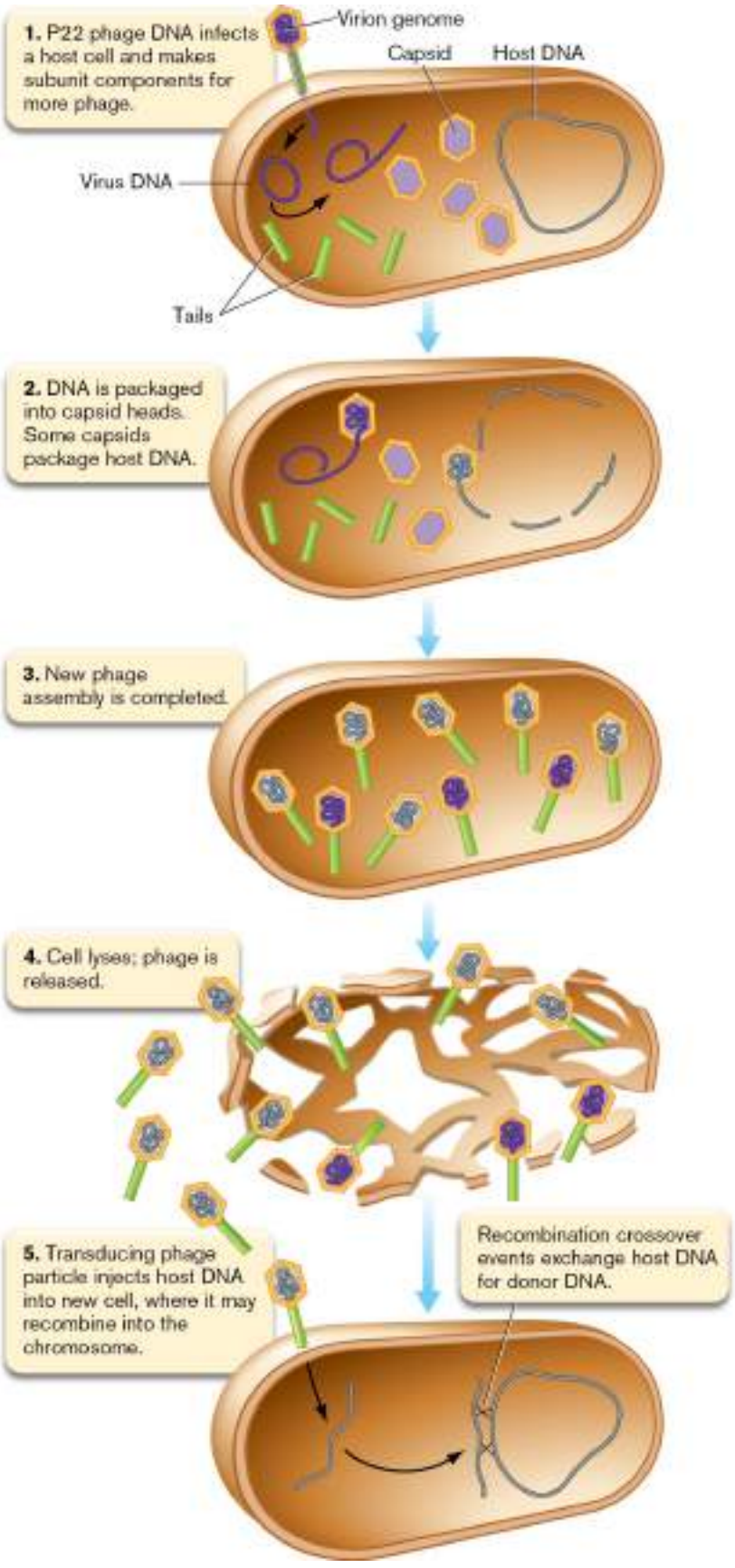

1. P22 phage DNA infects a host cell and makes subunit components for more phage.
Virion genome
Capsid
Host DNA
Virus DNA
Tails
2. DNA is packaged into capsid heads. Some capsids package host DNA.
3. New phage assembly is completed.
4. Cell lyses; phage is released.
5. Transducing phage particle injects host DNA into new cell, where it may recombine into the chromosome.
Recombination crossover events exchange host DNA for donor DNA.

FIGURE 9.22 ■ Generalized transduction. Generalized transduction by phage vectors can move any segment of donor chromosome to a recipient cell. The number of genes transferred in any one phage capsid is limited, however, to what can fit in the phage head.

When a transducing particle injects its DNA into a cell, no new phages are made, but the hijacked host DNA can recombine, or exchange, with sequences in the host chromosome of the newly infected cell, thus changing the genetic makeup of the recipient. In *E. coli* the phage P1 is able to perform generalized transduction in a manner similar to that of P22 in *Salmonella* . The ability of these phages to transfer large sections of the genome (1%–2%) has been exploited extensively by microbiologists to construct strains of varying genotypes. And, in the days before high-throughput genome sequencing, they were also used to map the relative position of genes at very fine scales.

Thought Question

9.6 In a transductional cross between an $A^+ B^+ C^+$ genotype donor and an $A^- B^- C^-$ genotype recipient, 100 A^+ recombinants were selected. Of those 100, 15 were also B^+ , while 75 were C^+ . Is gene B or gene C closer to gene *A?*

Specialized (restricted) transduction is a phage-mediated gene transfer mechanism that requires alternating lysogenic and lytic events (see Section 6.4) to move genes from a donor cell to a recipient cell. In contrast to generalized transduction, specialized transduction can move only a limited number of donor genes. *E. coli* phage lambda (discussed in Chapter 11) provides the classic

example of specialized transduction. Lambda phage DNA is linear when it first enters the cell; then it circularizes at cohesive ends called *cos* sites. Next, a 15-bp DNA sequence in lambda called *attP* can recombine with a similar host DNA sequence (called *attB*) located between the *gal* (galactose catabolism) and *bio* (biotin synthesis) genes on the *E. coli* chromosome, forming a lysogen (**Fig. 9.23** , step 1). This site-specific recombination event is carried out by the phage integrase protein.

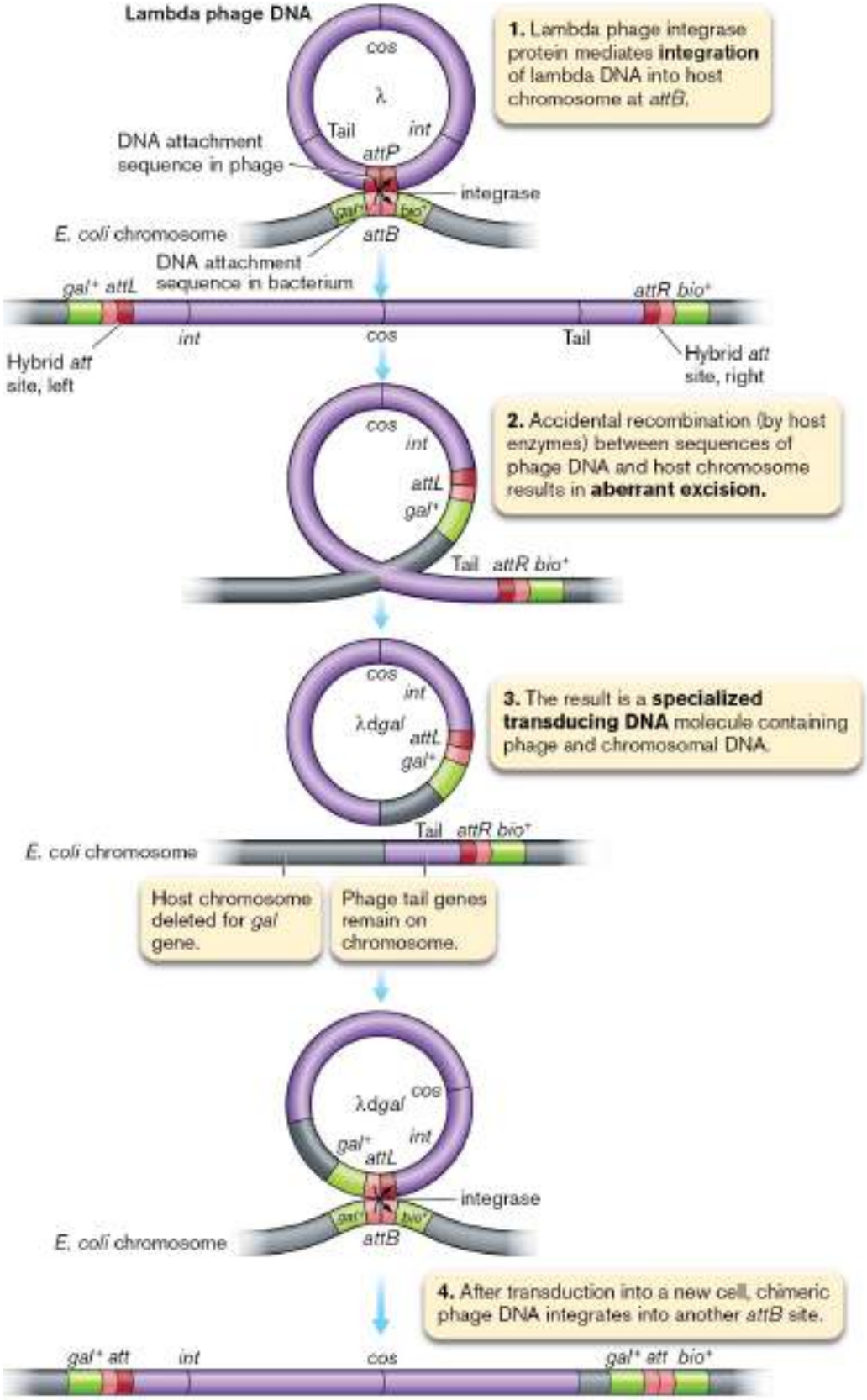

Lambda phage DNA
cos
λ
Tail
int
attP
DNA attachment sequence in phage
integrase
gal+
bio+
E. coli chromosome
attB
DNA attachment sequence in bacterium
1. Lambda phage integrase protein mediates integration of lambda DNA into host chromosome at attB.
gal+ attL
attR bio+
int
cos
Tail
Hybrid att site, left
Hybrid att site, right
2. Accidental recombination (by host enzymes) between sequences of phage DNA and host chromosome results in aberrant excision.
attL
gal+
Tail attR bio+
λdgal
3. The result is a specialized transducing DNA molecule containing phage and chromosomal DNA.
Host chromosome deleted for gal gene.
Phage tail genes remain on chromosome.
4. After transduction into a new cell, chimeric phage DNA integrates into another attB site.
gal+ att
gal+ att bio+

FIGURE 9.23 ■ Specialized transduction. Specialized transduction is restricted to moving host genes flanking the phage attachment site. The number of genes transferred is limited by the size of the phage head. The resulting recipient, or transductant, chromosome becomes a partial diploid, in this case, for the *gal* gene.

The machinery used at this site is distinct from the RecA-dependent machinery used in generalized homologous recombination. Site-specific recombination produces a chromosome with an integrated phage genome, referred to as a prophage, flanked by chimeric *att* sites called *attL* and *attR* . These sites are called "chimeric" because each one is made half from the bacterial and half from the phage *att* sites. Prophage DNA remains latent until something happens to the cell to activate it.

Specialized transduction begins with the reactivation of this prophage DNA, which is often induced by DNA damage. Usually, the phage enzymes that excise the lambda DNA do so precisely. The chromosome and viral DNAs are restored to their native states. The viral DNA will then replicate and make more phage particles containing normal phage DNA. On rare occasions, however, improper excision (mediated by host recombination enzymes) can take place between host DNA sequences that lie adjacent to the phage insertion site (*attL* and *attR* in **Fig. 9.23** , step 2) and similar DNA sequences within the prophage. Improper excision yields a virus that will lack a few viral genes (tail genes remain on the chromosome in step 3) but will include host genes lying adjacent to the phage attachment site (the galactose utilization gene *gal* in step 3). With some phages, the specialized transducing particles can replicate unaided. However, in **Figure 9.23** the result is a defective, specialized transducing phage DNA (lambda d *gal* , or λd *gal*). Specialized transducing phage particles that are defective cannot replicate by themselves but require the presence of a second, helper phage that supplies missing gene products.

Once formed, specialized transducing phages can deliver the hybrid DNA molecule to a new recipient cell. This is the transduction process. Once in that cell, the phage DNA can integrate into the host *attB* site, carrying the donor host gene(s) with it and forming a lysogen once more (**Fig. 9.23** , step 4). The result will be a partial diploid situation in which the new recipient contains two copies of a host gene: one originally present on its chromosome, and one brought in by the transducing DNA.

Transformation of Naked DNA

In 1928, a perceptive English medical officer, Frederick Griffith (1879–1941), found that he could kill mice by injecting them with dead cells of a virulent pneumococcus (*Streptococcus pneumoniae* , a cause of pneumonia), together with live cells of a nonvirulent mutant. Even more extraordinary was that he recovered live, virulent bacteria from the dead mice. Were the dead bacteria brought back to life? Unfortunately, Griffith was killed by a German bomb during an air raid on London in 1941 and never learned the answer.

In a landmark series of experiments published in 1944, Oswald Avery (1877–1955), Colin MacLeod (1909–1972), and Maclyn McCarty (1911–2005) proved that Griffith's experiment was not a case of resurrecting dead cells. Instead, DNA released from dead cells of the virulent strain entered into the harmless living strain of *S. pneumoniae* —an event that transformed the live strain into a killer. The process of importing free DNA into bacterial or archaeal cells is now known as transformation .

Organisms in which transformation is a natural part of the growth cycle include Gram-positive bacteria such as *Streptococcus* and *Bacillus* , as well as some Gram-negative genera including the human pathogens *Haemophilus* and *Neisseria* and the nonpathogenic photosynthetic cyanobacterium *Synechocystis* . At least 82 species of bacteria have been shown to be capable of transformation. Transformation also happens in some archaea, and

although the mechanisms of DNA uptake are still under investigation, they appear to be distinct from those used by bacteria.

Taking up free DNA comes with several challenges. First, the cell needs a way to contact and recognize the DNA as a substrate for uptake. Second, the cell needs to pass the DNA through the cell envelope, which in Gram-negative organisms includes both the inner and outer membranes. The marine bacterium *Vibrio cholerae* , the causative agent of cholera, meets these challenges with a series of protein complexes at the cell surface, beginning with a retractable pilus.

Rather than rely on passive cell contact with the extracellular DNA, *V. cholerae* extends a type IV pilus beyond the outer membrane, which can grab double-stranded DNA(dsDNA) by its tip (**Fig. 9.24**). While some transformable species can take up only DNA that has species-specific sequence tags, the type IV pilus of *V. cholerae* will bind and take up any free dsDNA molecule. Ankur Dalia and colleagues at Indiana University discovered that once the pilus binds DNA, it can pull the DNA into the periplasm by retraction.

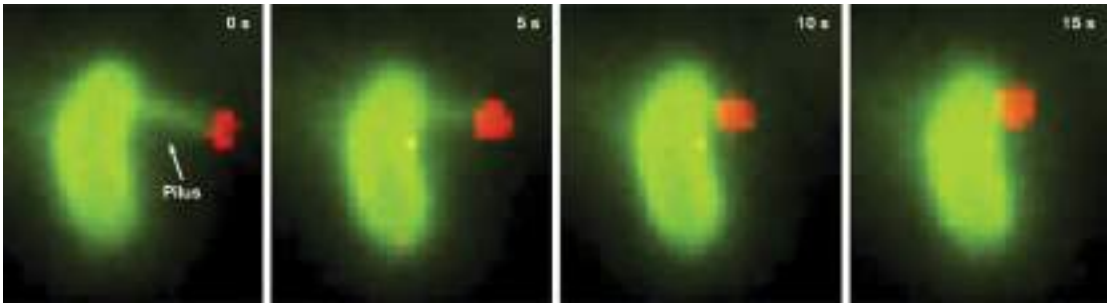


FIGURE 9.24 ■ Time-lapse microscopy of dsDNA binding and retraction of the type IV pilus of *Vibrio cholerae* . Fluorescence microscopy of the same cell over time in seconds (s), showing the cell and its pilus (stained green) and the bound DNA (stained red). See video.

C. ELLISON ET AL. 2018. *NAT MICROBIOL* . **3** :773–780

Retraction occurs through the disassembly of the pilus filament at the base of the pilus. The PilA monomers that are removed from the filament are stored in the membrane for the next time the filament is assembled. Pilus retraction pulls the dsDNA molecule through the open channel that the pilus filament occupied (**Fig. 9.25**). Once in the periplasm, the dsDNA is bound by the ComEA protein. ComEA is too large to pass through the pilus channel, so each addition of ComEA protein to the incoming dsDNA serves to ratchet the dsDNA into the periplasm. ComEA-bound dsDNA is then delivered to the ComEC protein channel that spans the inner membrane.

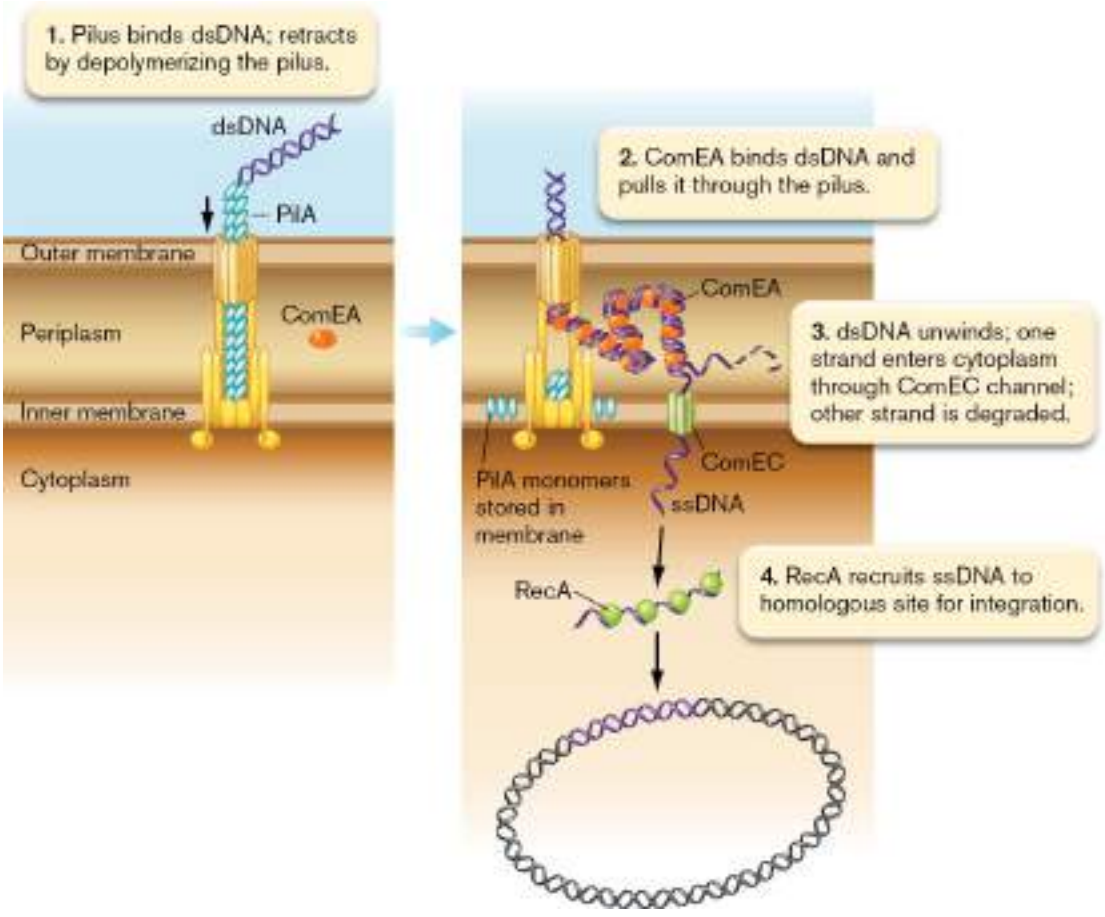


FIGURE 9.25 ■ Model for the role of the type IV pilus in transformation for *Vibrio cholerae* . The pilus spans the inner and outer membranes to access the environment. Details

of the homologous recombination event are not well understood and not shown.

The subsequent steps of transformation are not fully resolved for *V. cholerae* , but they are summarized in the model illustrated in **Figure 9.25** , on the basis of evidence from other transformable species. At the ComEC channel the dsDNA is unwound into two single strands (ssDNA). One strand is degraded in the periplasm, while the other is transported through the ComEC channel into the cytoplasm. Once in the cytoplasm, the ssDNA is bound by protective single-stranded DNA-binding proteins (SSBs). RecA proteins eventually replace the SSB proteins and mediate homologous recombination between the ssDNA and regions of homology in the chromosome.

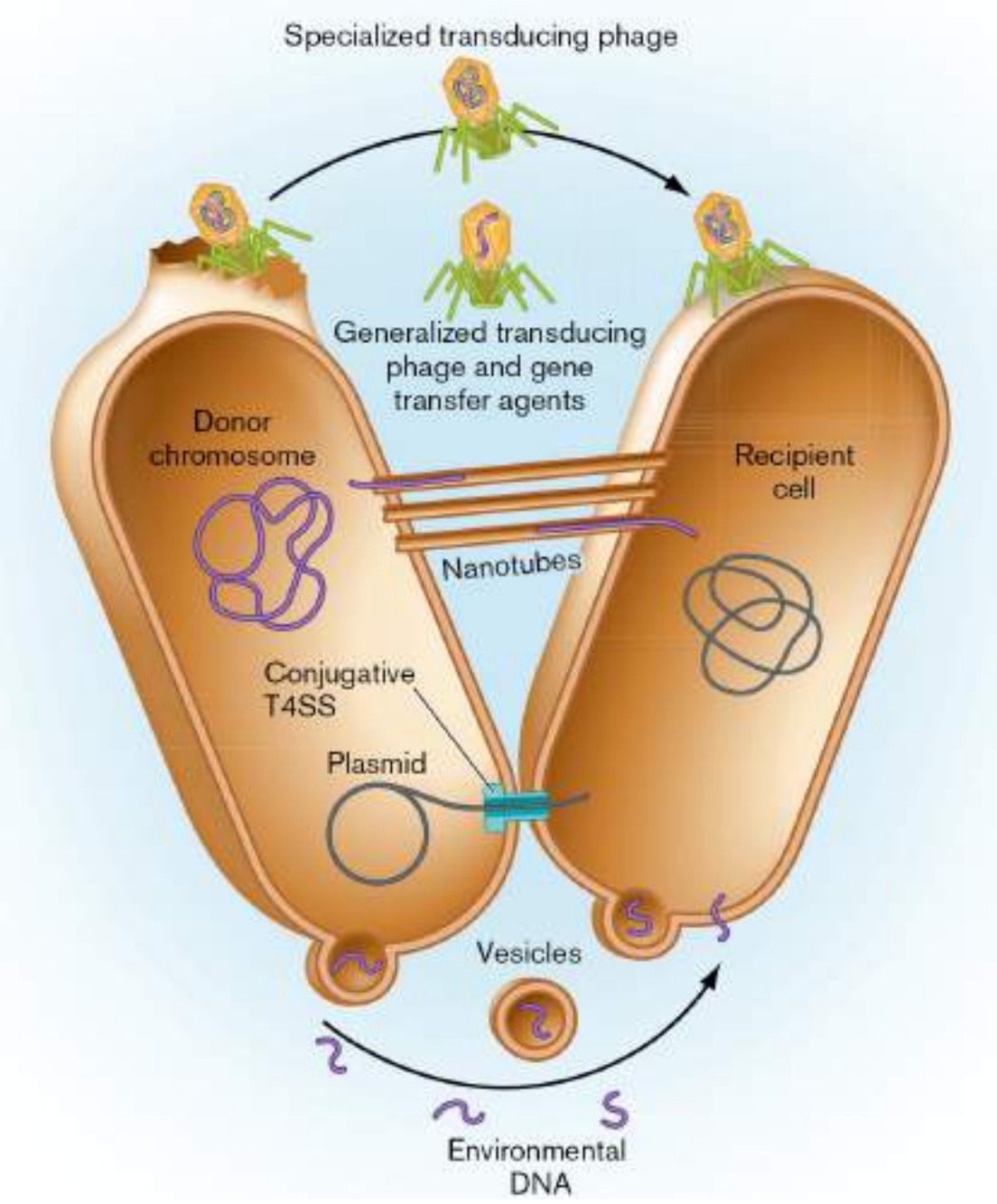


FIGURE 9.26 ■ Summary of mechanisms of gene transfer in bacteria and archaea.

Additional Modes of Gene Transfer between Bacteria

While conjugation, transduction, and transformation are probably the most prevalent means of transferring genes from one cell to another, some bacteria and archaea have developed alternative mechanisms (**Fig. 9.26**). Two of these mechanisms, membrane vesicles and nanotubes, are described in Chapter 3 . Both membrane vesicles and nanotubes can transfer DNA between different species, which can play an important role in genome evolution (discussed in Section 9.5).

A third alternative DNA delivery mechanism—gene transfer agents—shares properties with generalized transduction and transformation. Like P22 and P1 transducing viruses, gene transfer agents encode a phage-like capsid structure that packages random sections of chromosomal DNA. These phage-like particles are released when the host cell dies, and they deliver DNA to the recipient. Unlike phages, however, gene transfer agents do not package their own genes into the phage-like particle, and thus they cannot use the viral lifestyle to move from host to host. Another important difference from transducing phages is that gene transfer agents rely on the recipient's transformation machinery to release the DNA into the recipient's cytoplasm, rather than an independent mechanism of injection.

Note: Current studies suggest that even more mechanisms of gene transfer may exist in archaea. Importantly, some of these may involve bidirectional DNA transfer, meaning that each cell involved in the mating sends as well as receives DNA. This is unlike the unidirectional transfers of the mechanisms described in this chapter. Bidirectional DNA transfer may happen in some archaea by cell fusion (cytoplasmic exchange) and in others by transfers at discrete intercellular bridge structures.

To Summarize

- Conjugation is a DNA transfer process mediated by a transferable plasmid that requires cell-cell contact and formation of a protein complex between mating cells.
- **Some bacteria can transfer DNA across phylogenetic domains.** For example, *Agrobacterium tumefaciens* conjugates with plant cells.
- **Integration of the F factor** into the chromosome creates an Hfr strain, which can transfer chromosomal genes via conjugation.
- Transduction is the process whereby bacteriophages transfer fragments of bacterial DNA from one bacterium to another. In generalized transduction , a phage can move any gene in a bacterial genome to another bacterium. In specialized transduction , a phage can move only a limited number of bacterial genes.
- Transformation is the uptake by living cells of free-floating DNA from dead, lysed cells.
- **Membrane vesicles, nanotubes, and gene transfer agents** contribute to gene transfer by mechanisms distinct from conjugation, transduction, or transformation.

Glossary

conjugation

: Horizontal gene transfer involving cell-to-cell contact. In bacteria, pili draw together the donor and recipient cell envelopes, and a protein complex transmits DNA across. In ciliated eukaryotes, a conjugation bridge forms between two cells connecting their cytoplasm, through which micronuclei are exchanged.

fertility factor (F factor)
: A specific plasmid (transferred by an F^+ donor cell) that contains the genes needed for pilus formation and DNA export.

F factor
: See **fertility factor** .

F^+ cell
: The DNA donor cell that transmits the fertility factor F^+ to an F^- cell during bacterial conjugation.

F^- cell
: The DNA recipient cell in bacterial conjugation.

Hfr strain
: A high-frequency recombination bacterial strain, caused by the presence of a chromosomally integrated F factor.

transduction
: The transfer of host genes between bacterial cells via a phage head coat.

generalized transduction
: A phage-mediated gene transfer process in which any donor gene can be transferred to a recipient cell.

specialized transduction
: Also called *restricted transduction* . Transduction in which the phage can transfer only a short DNA sequence adjacent to a prophage (integrated phage genome) from a donor cell to the recipient cell.

transformation
: The internalization of free DNA from the environment into bacterial cells.

conjugation
: Horizontal gene transfer involving cell-to-cell contact. In bacteria, pili draw together the donor and recipient cell envelopes, and a protein complex transmits DNA across. In ciliated eukaryotes, a conjugation bridge forms between two cells connecting their cytoplasm, through which micronuclei are exchanged.

transduction

The transfer of host genes between bacterial cells via a phage head coat.

generalized transduction

A phage-mediated gene transfer process in which any donor gene can be transferred to a recipient cell.

specialized transduction

Also called *restricted transduction* . Transduction in which the phage can transfer only a short DNA sequence adjacent to a prophage (integrated phage genome) from a donor cell to the recipient cell.

transformation

The internalization of free DNA from the environment into bacterial cells.

9.4 Mobile Genetic Elements

In 1948, Barbara McClintock (1902–1992) noticed that certain traits of maize defied the laws of Mendelian inheritance. The genes encoding these traits seemed to hop from one chromosome to another. Although McClintock's theories of these so-called jumping genes were provocative at the time, we now know that these types of genes, referred to as transposable elements , exist in virtually all life forms and can move both within and between chromosomes. These mobile fragments of DNA have contributed greatly to genome rearrangements during the evolution of all species.

Transposons and Transposition

Transposable elements exist primarily as hitchhikers integrated into some other DNA molecule (**Fig. 9.27**). All transposable elements include a gene that encodes a transposase , which is an enzyme that catalyzes the transfer or copying of the element from one DNA molecule into another. Bacteria have two types of transposable elements: insertion sequences and transposons. An insertion sequence (IS) is a transposable element (typically 700–1,500 bp) consisting of a transposase gene flanked by short inverted-repeat sequences that are targets of the transposase (**Fig. 9.27**). Transposons are mobile elements that carry other genes in addition to those required for transposition. Genes encoding antibiotic resistance proteins are one important class of "cargo" carried by transposons.

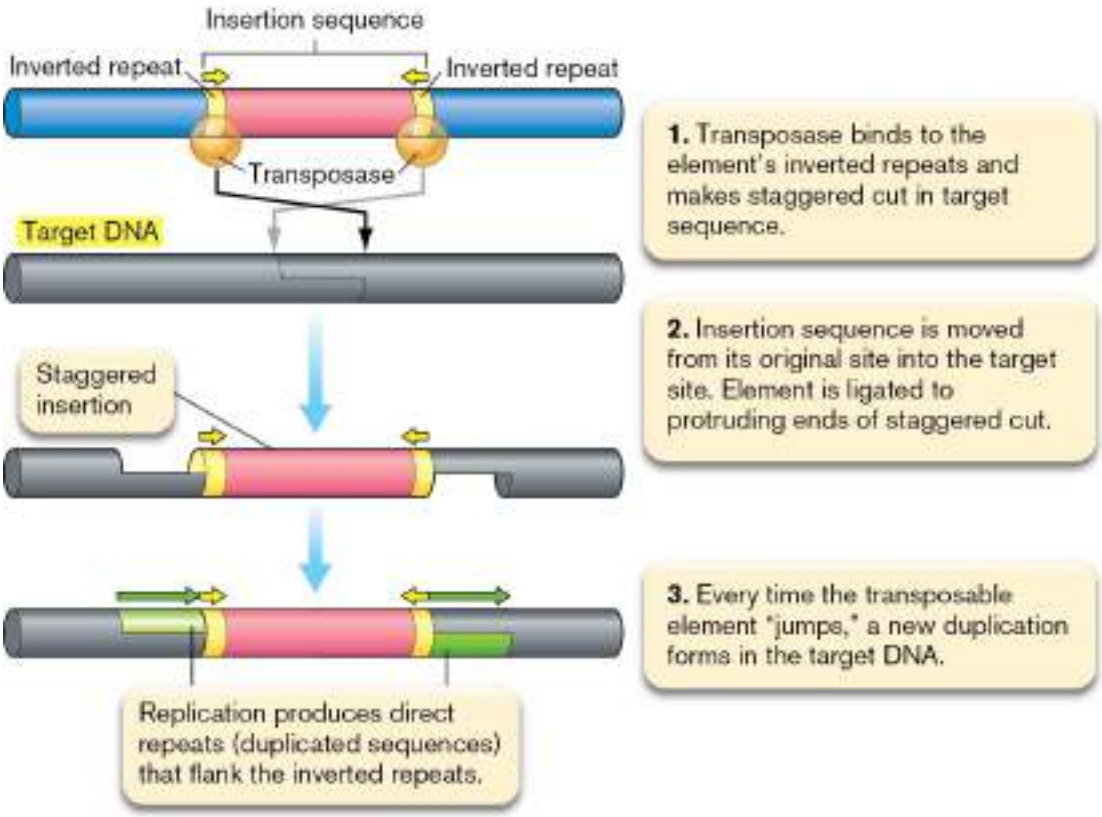


FIGURE 9.27 ■ Basic transposition and the origin of target site duplication. Enzymes that catalyze transposition generate duplications in the target site by ligating the ends of the insertion element to the long ends of a staggered cut at the target DNA site.

JEFFREY H. MILLER

Note: An inverted repeat is a DNA sequence identical (or nearly identical) to another downstream sequence. The repeats have reversed sequences and are separated from each other by intervening sequences:

5′-AATCGAT . . . . . ATCGATT-3′
3′-TTAGCTA . . . . . TAGCTAA-5′

Recall that you must consider both strands of the DNA to see the inversion. Compare each strand in the 5′-to-3′ direction. View the top

strand left to right and bottom strand right to left. When no nucleotides intervene between the inverted sequences, the whole structure is called a palindrome.

Transposition is the process of moving a transposable element within or between DNA molecules. During transposition, a short target DNA sequence on the destination DNA molecule is duplicated so that one copy of the sequence will flank each end of the element, forming direct repeats (**Fig. 9.27**). The transposase randomly selects one of many possible target sequences where it will move the insertion sequence.

A 1978 study by Jeffrey H. Miller (**Fig. 9.28A**) and colleagues provided some of the first evidence that transposable elements produce direct repeats when they insert into new target sequences. DNA sequencing of two spontaneous mutations within the *lacI* gene of *Escherichia coli* revealed that transposable element IS *1* inserted at sites just downstream of the translation start codon (**Fig. 9.28B**). For each mutant they identified IS *1* by its (nearly identical) inverted repeats, and they found that these inverted repeats were flanked by 9-bp direct repeats of the *lacI* gene. They noticed that the direct repeats, while perfect copies of each other, shared no apparent similarity between the two mutants, suggesting that the IS *1* element can "hop" into DNA randomly or nearly randomly.

A.



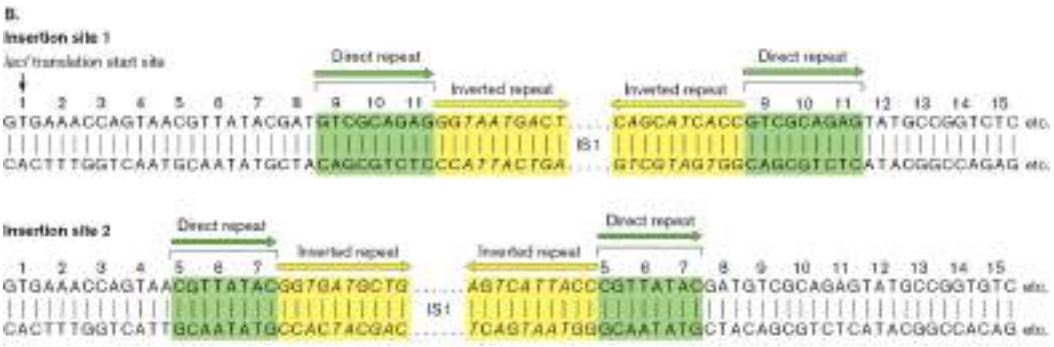


FIGURE 9.28 ■ Transposition of IS *1* into two sites within the *lacI* gene of *Escherichia coli* . Researcher **(A)** Jeffrey H. Miller and colleagues discovered **(B)** that the inverted repeats (IRs) composing the left and right ends of IS *1* (italics, partial IR sequences shown) are flanked by 9-bp direct repeats (DRs) of the *lacI* gene (brackets). Note that the two IR sequences of IS *1* are nearly but not completely identical. This small sequence variation revealed that the IS *1* inserted in opposite orientation at the two insertion sites. Also note that the DR sequences are identical at

each insertion site but share little if any similarity between insertion site 1 and insertion site 2.

JEFFREY H. MILLER

In fact, all transposable elements characterized to date show some degree of target sequence specificity. While IS *1* targets AT-rich regions generally, other elements are more specific, such as IS *91* , which targets 5′-GAAC-3′ or 5′-CAAG-3′ sequences. These biases prevent truly random targeting but still provide many sites within genomes to insert.

Transposition occurs by two different mechanisms: nonreplicative or replicative transposition (**Fig. 9.29**). In nonreplicative transposition, the insertion sequence excises itself from one host DNA and integrates into the destination DNA. In replicative transposition, the sequence copies itself into the new host DNA while a copy remains within the original host.

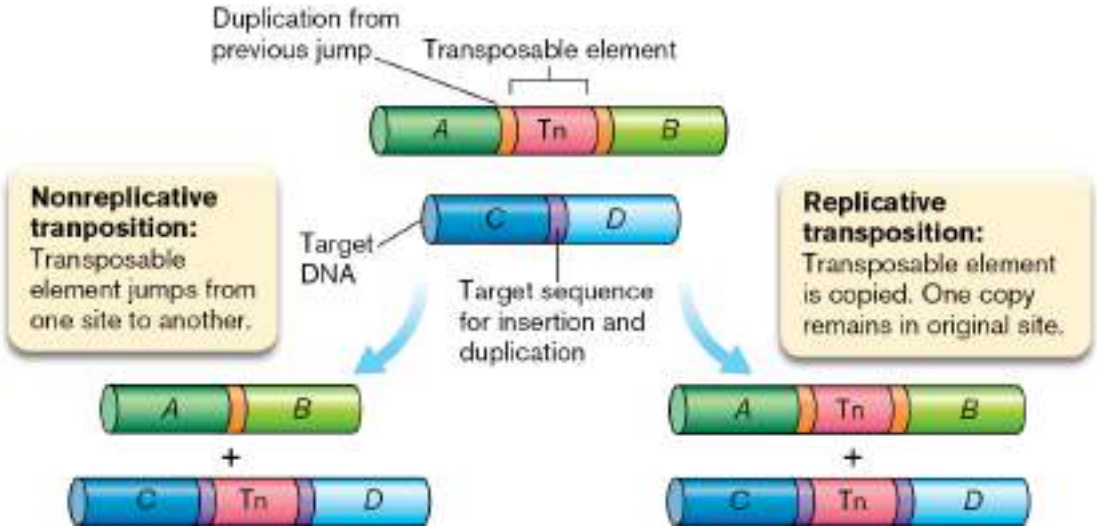


FIGURE 9.29 ■ Products of nonreplicative and replicative transposition. Nonreplicative transposition moves an insertion element from one DNA site to another without leaving a copy of the element at the original site. Replicative transposition leaves the element at the original site and moves a replicated copy to the new site.

See above for *Transposition* animation

Thought Question

9.7 Evolutionarily speaking, why might it be advantageous for a transposon to utilize replicative transposition? Why might nonreplicative transposition be advantageous?

Conjugative Transposons

The transposition events described thus far involve transfer between DNA molecules within the same cell. Some transposons, called conjugative transposons , are able to transfer from one cell to another by conjugation. These mobilizable transposons have been found in both bacteria and archaea. One example of a mobilizable transposon is the transposon SXT of the marine bacterium *Vibrio cholerae* . SXT carries genes that confer resistance to the antibiotics sulfamethoxazole and trimethoprim. Conjugative transposons such as SXT therefore contribute to the spread of antibiotic resistance in the microbial world. Importantly, their mobilization can be triggered by antibiotic exposure.

Like many other conjugative transposons, SXT is induced by DNA-damaging agents, which include some antibiotics. DNA damage triggers the SOS response of the host, which induces expression of the SXT activator proteins (**Fig. 9.30**). In turn, these activator proteins induce expression of the SXT integrase, which excises the SXT element from donor DNA. Excision results in the formation of a circular intermediate. Like the F plasmid of *E. coli* , the SXT transposon encodes its own transfer machinery, including a T4SS used to contact the recipient cell and to provide the transfer channel. Once transferred into a new host, the conjugative transposon expresses its integrase, which recognizes the specific *att*

site in the host chromosome and inserts the SXT transposon at that location.

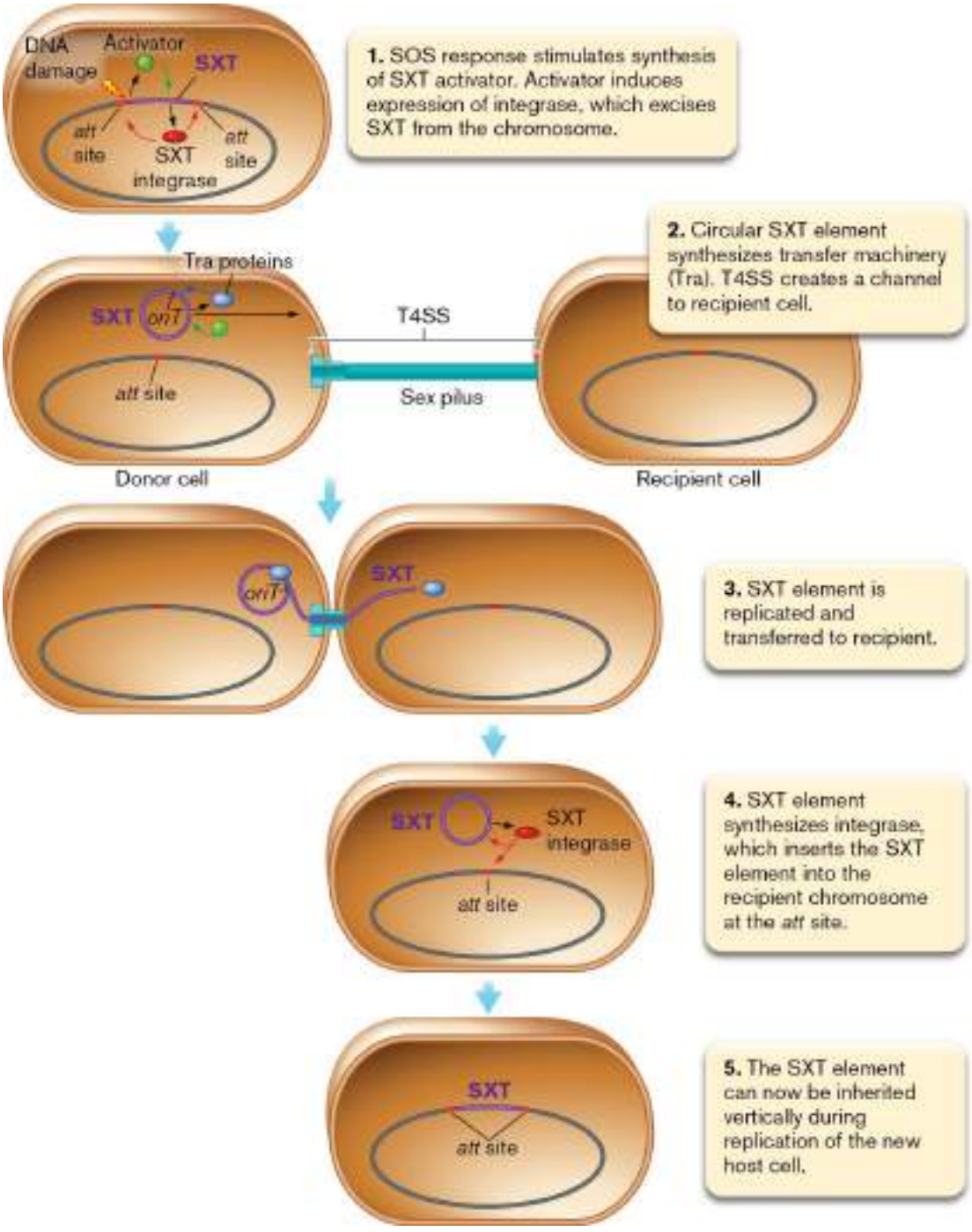


FIGURE 9.30 ■ Activation and transfer of the SXT conjugative transposon to a new cell.

In an unusual but intriguing process, some genetic elements can hijack the conjugation machinery of conjugative transposons such as SXT to move into new hosts. So-called mobilizable genomic islands (MGIs) lack the machinery for conjugation but have an *oriT* that mimics that of the conjugative transposon, and they use this *oriT* to trick the conjugative transposon into transferring the island. For example, the SXT transposon's transcription activator can induce expression of the integrase encoded by the MGI, which then excises and circularizes the MGI, much as the conjugative transposon does (**Fig. 9.31**). The SXT transfer machinery then recognizes the MGI's *oriT* and transfers the island to a recipient cell. Once transferred, the MGI uses its own integrase to recombine into the new host chromosome.

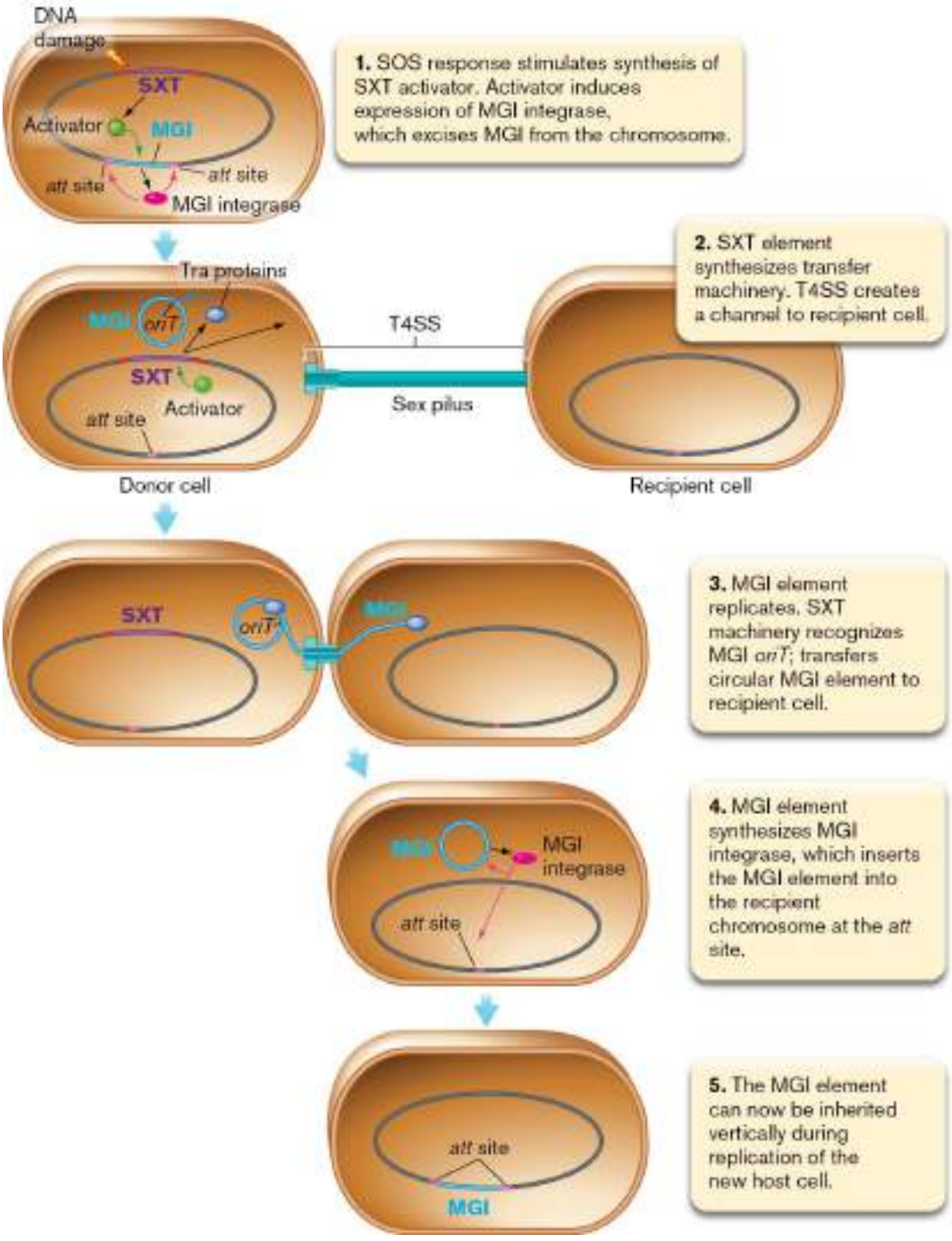


FIGURE 9.31 ■ Activation and transfer of a mobilizable genome island (MGI) by the SXT conjugative transposon.

Like bacteriophage and conjugative plasmids and transposons, MGIs take a more active role in mediating horizontal gene transfer

between microbes (see Section 9.5). MGIs were first discovered in *Vibrio* and have subsequently been found in an array of other marine bacteria. The cargo genes of MGIs can vary, but they often include restriction-modification systems, which could improve the viral defense of their host cells. MGIs can also carry antibiotic resistance genes and exploit the transfer machinery of incoming plasmids to spread antibiotic resistance in pathogens such as *Salmonella* , as described in eResearch Activity 9 .

Transposable Elements in Genetic Analysis

Transposable elements have been of immeasurable service as tools for gene discovery. To identify the genes involved in a microbial function, one genetic approach is to generate mutations in the microbe and assess which mutation(s) affect the function in question. Ideally, only a single gene or operon would be mutated for each mutant cell, and mutants of every gene in the genome would be present in the mutant pool.

Once the pool is generated, an assay is used to identify the mutants that lost the function in question. The assays come in two types: screen and selection. In a screen assay , all mutants can grow, but only the mutants involved in the function of interest show a phenotype different from the wild type. In a selection assay , only the mutants involved in the function of interest can grow, while all other mutants and the wild-type parent die under the assay conditions.

Note: Many genes in the genome are essential for growth, and mutants that lack the activity of these genes do not typically contribute to the mutant pools for genetic screens and selections, unless special measures are taken.

Transposons have characteristics that make them great tools for both genetic screens and selections. First, they usually disrupt only a

single gene or operon per mutant, unlike chemical mutagens and UV light, which generate mutations in multiple locations. Limiting the mutations to one per mutant enables the researcher to assign the phenotypic change of the cell to a single mutation. Second, transposons hop into the chromosome at nearly random locations, such that mutations in essentially every gene in the genome can be generated. Third, transposons that carry antibiotic resistance markers can provide the researcher with a means to exclude all nonmutants from the screen or selection assay.

After mutagenesis, the cells are simply grown in the presence of the antibiotic, and all cells without the transposon are killed. Finally, transposons can generate loss-of-function mutations by inserting within the coding region of the gene. Transposon insertions can truncate the inserted gene's product if the transposon carries transcription and translation terminators. Thus, the region downstream (3′) of the insertion site is usually not transcribed or translated into protein.

For practical considerations, most screen and selection assays use modified transposons called minitransposons. Minitransposons have the transposase enzyme removed from the transposon (**Fig. 9.32**). Typically, the transposase gene is moved to another region of the plasmid or virus vector that delivers the transposon to the recipient. This way, transposase can still be expressed in the recipient cell but will not be included in the minitransposon that hops into the chromosome. The plasmid carrying the transposase cannot replicate in the recipient, so after cell division, transposase activity disappears and the new insertion is stable. Stability means that the minitransposon cannot hop to a second location, where it could further change the genotype and phenotype of the mutant.

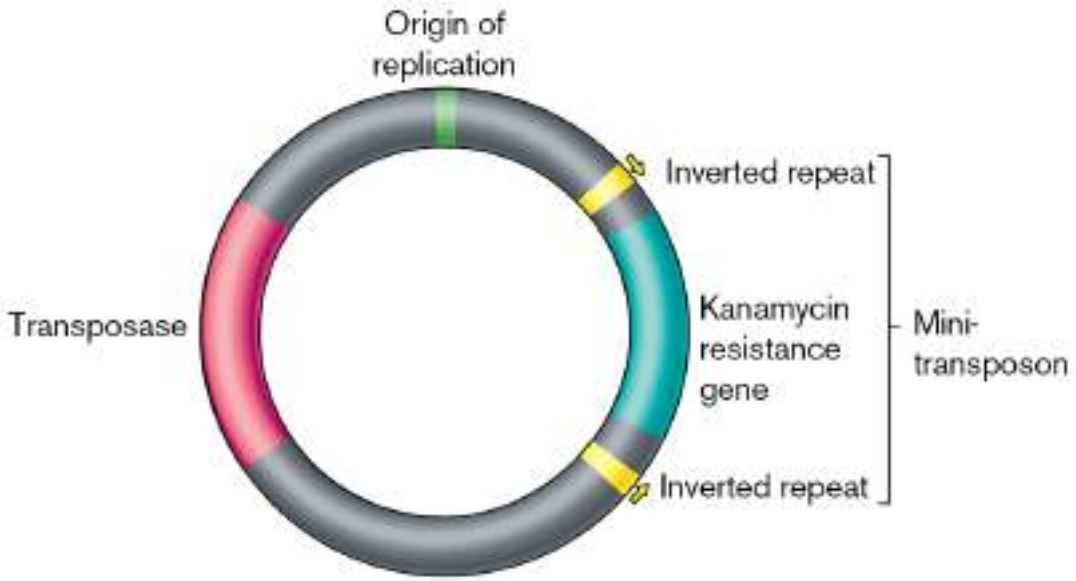


FIGURE 9.32 ■ Simplified model of a typical transposon delivery vector. Note that the origin of replication is usually engineered to function in the donor strain but not the recipient strain.

In an application of transposon mutagenesis, Alison Buchan and colleagues at the University of Tennessee (**Fig. 9.33A**) performed a screen of marine bacteria called roseobacters to identify genes that are involved in the killing of other bacteria. Roseobacters are important decomposers of plant material in coastal salt marshes, habitats where competition for resources is stiff. One way that some microbes, such as roseobacters, can outcompete rival microbes is to eliminate them—an activity that can be observed on agar plates in a killing assay. When the (wild-type) roseobacter *Phaeobacter* strain Y4I is spotted onto a lawn of *Vibrio* (*Aliivibrio*) *fischeri* , a factor released by the roseobacter is able to lyse the *V. fischeri* cells, forming a halo around the patch of roseobacter cells (**Fig. 9.33B**).

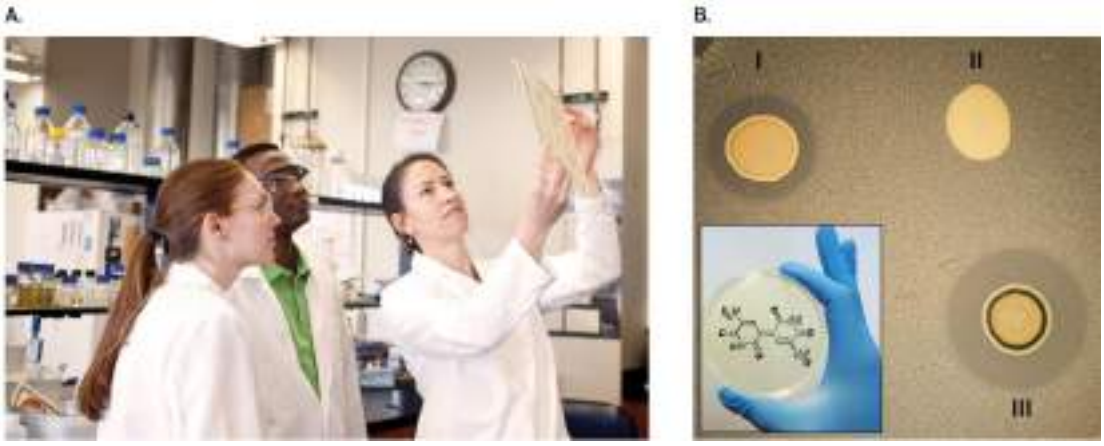


FIGURE 9.33 ■ Transposon mutagenesis of killer roseobacters. A. Alison Buchan (right) and colleagues screen for mutants. **B.** Individual cultures of *Phaeobacter* spp. are spotted onto lawns of *Vibrio* (*Aliivibrio*) *fischeri* . Wild-type cultures formed a zone of *V. fischeri* clearing when they grew (I), whereas one transposon mutant lost the ability to form a halo (II), and another one produced a larger halo than the wild type did (III). Note the band of increased pigment indigoidine in spot III. **Inset:** Indigoidine-producing roseobacters streaked on agar to demonstrate the structure of the pigment.

ALISON BUCHAN, UNIVERSITY OF TENNESSEE, KNOXVILLE, COLLEGE OF ARTS & SCIENCES, DEPARTMENT OF MICROBIOLOGY

ASHLEY FRANK

PHOTO COURTESY OF ALISON BUCHAN

Buchan and colleagues screened a pool of transposon insertion mutants for those that lost the ability to form a halo and found a small subset of mutants with this loss-of-function phenotype. Using genomics to find the location of the transposon insertion, they found that some of these mutants were no longer able to make indigoidine, a molecule that they subsequently demonstrated has the antimicrobial activity. Notably, some other mutants in this screen showed a hyperkilling phenotype, with a larger halo and deeper pigmentation from indigoidine. As this example highlights, screens can also provide unexpected phenotypic outcomes that may spur new lines of investigation.

The development of high-throughput sequencing technology (see Chapters 7 and 12) has enabled researchers to screen for phenotypes among thousands of transposon mutants simultaneously. The TnSeq (also called InSeq) technique is essentially a test of fitness of the pool of mutants in a specific environmental condition. The relative abundances of tens of thousands of insertion mutants are compared before and after the environmental incubation. Those that decrease in abundance have insertions in genes whose activity contributes positively to fitness, and those that increase in abundance involve genes whose activity contributes negatively to fitness.



For example, in a study by Jeffrey Gordon and colleagues at Washington University in St. Louis, transposon mutants of the human gut bacterium *Bacteroides* were inoculated into the guts of mice to look for genes that were important for colonization and persistence. Over the course of several weeks, the population of bacteria in the fecal pellets was monitored by TnSeq (**Fig. 9.34**). Of the mutants that decreased in abundance during incubation in the mice, many were involved in carbohydrate consumption and amino acid biosynthesis. These results indicate that the ability to acquire carbohydrates from the gut environment and convert them to amino acids for protein production contributes significantly to the growth and competitiveness of *Bacteroides* in the gut. The advantages of TnSeq for a study such as this should be apparent: Imagine the time (and number of mice!) required to test each of the thousands of mutants individually for a fitness phenotype in the mouse gut.

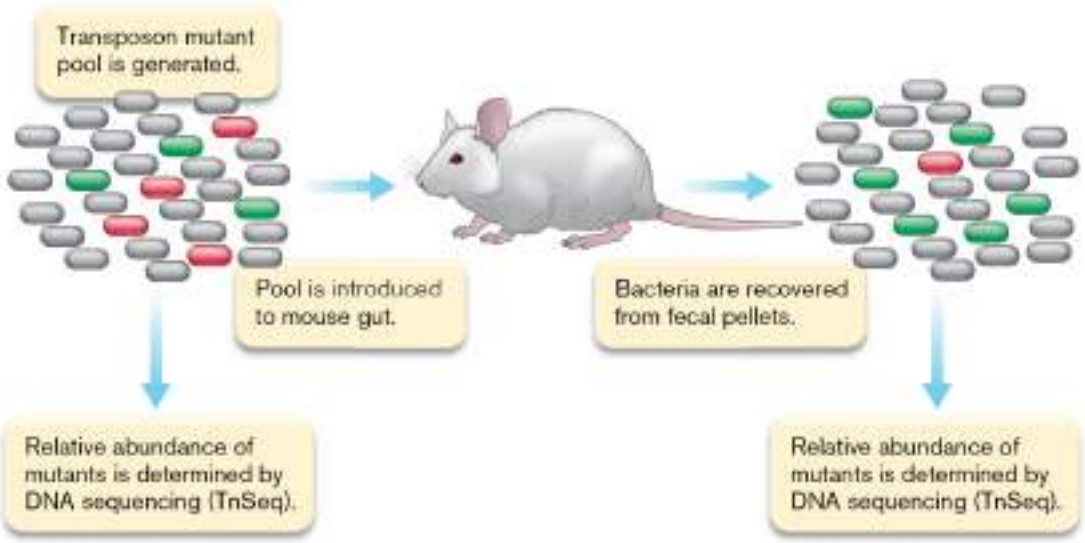


FIGURE 9.34 ■ TnSeq identifies genes involved in fitness in the mouse gut. Green and red cells are mutants that increase or decrease relative abundance, respectively, when introduced into the mouse.

To Summarize

- **Transposable elements** ("jumping genes") move from one DNA molecule to another, usually without replicating separately (that is, they are not plasmids).
- **Transposase** catalyzes the transfer or copying of transposable elements from one DNA molecule into another.
- **Insertion sequences** are simple transposable elements containing a transposase gene flanked by short inverted-repeat sequences.
- **Transposons** are complex transposable elements carrying additional genes (encoding, for example, drug resistance).
- **Transposable elements move** by nonreplicative or replicative mechanisms.
- **Transposons can carry a variety of genes** , including antibiotic resistance genes. Some transposons can transfer themselves or mobilizable genomic islands via conjugation.

- **Transposons can be exploited to identify genes involved in microbial functions** via genetic screen and selection assays.

Glossary

transposable element
: Also called *jumping gene* . A segment of DNA that can move from one DNA region to another.

transposase
: A transposable element–encoded enzyme that catalyzes the transfer of the transposable element from one DNA region to another.

insertion sequence (IS)
: A simple transposable element consisting of a transposase gene flanked by short, inverted-repeat sequences that are the target of transposase.

transposon
: A transposable DNA element that contains genes in addition to those required for transposition. Examples of additional genes include those that encode resistance to antibiotics.

transposition
: The process of moving a transposable element from one DNA region to another.

conjugative transposon
: A transposon that can be transferred from one cell to another via conjugation.

mobilizable genomic island (MGI)
: A genomic island that can be transferred to another cell via the machinery of a conjugative transposon.

screen assay
: A genetic assay, used to identify mutants that have lost a particular function, in which all mutants can grow, but only the

mutants involved in the function of interest show a phenotype different from the wild type.

selection assay

A genetic assay, used to identify mutants that have lost a particular function, in which only the mutants involved in the function of interest can grow, while all other mutants and the wild-type parent die.

9.5 Genome Evolution

Genomes evolve through the gain and loss of genes, with each gene having its own evolutionary history and origin. The genome of the best-studied organism on the planet, *Escherichia coli* strain K-12, consists of 4,489 genes. Barring any losses since its cultivation, this was the collection of genes that was present when the organism was isolated in 1922 from a fecal sample of an unidentified human donor at a hospital in Palo Alto, California. How did this collection of genes assemble into a genome encoding the gut-colonizing bacterium we call *E. coli* K-12? To put it another way, where did all of the *E. coli* genes come from?



In this section we revisit the topics of mutation and gene transfer and discuss them in the context of genomic change: how genes are gained and lost during evolution. The mutation events can be very rapid, such as the mutation from one base to another that gives a gene a new function. The impact that such changes have on microbial populations in longer evolutionary timescales is discussed in Chapter 17 .

Homology, Duplications, and Divergence

Evolution by natural selection favors genotypes (mutants) whose fitness is higher than others in the environment in which selection occurs (see Section 17.3). Often, the selected mutant has acquired —through a random process—a novel function. One way cells can acquire novel functions is through mutation of existing genes. As a mechanism of evolutionary change, a major problem with such a mutation is that the original function of the gene is lost, possibly at great cost to the cell. Cells can circumvent this problem through the mechanism of duplication followed by divergence: the gene is copied such that one copy retains the original function and the other copy is free to evolve a new function.

Gene duplication can be mediated by homologous recombination between direct repeats on the chromosome. Direct repeats are identical stretches of DNA arranged in the same orientation on the chromosome. **Figure 9.35** shows how homologous recombination between direct repeats on sister chromosomes can generate gene duplications. After chromosome replication, the sister chromosomes can associate such that the direct repeat on the left (repeat 1) of one chromosome aligns with the direct repeat on the right (repeat 2) of the other chromosome. Homologous recombination between these repeats results in two mutant chromosomes: one with a tandem duplication of the genes flanked by the repeats, the other with a deletion of those genes. If the chromosomes are circular, a second recombination event at *dif* or elsewhere along the chromosomes would be required to resolve the dimer that formed by the first recombination event (see **Figure 7.20**). Chromosome partitioning followed by cell division generates a daughter cell with the duplication and a daughter cell with the deletion. Gene deletion as an agent of genome reduction is discussed later in this section.

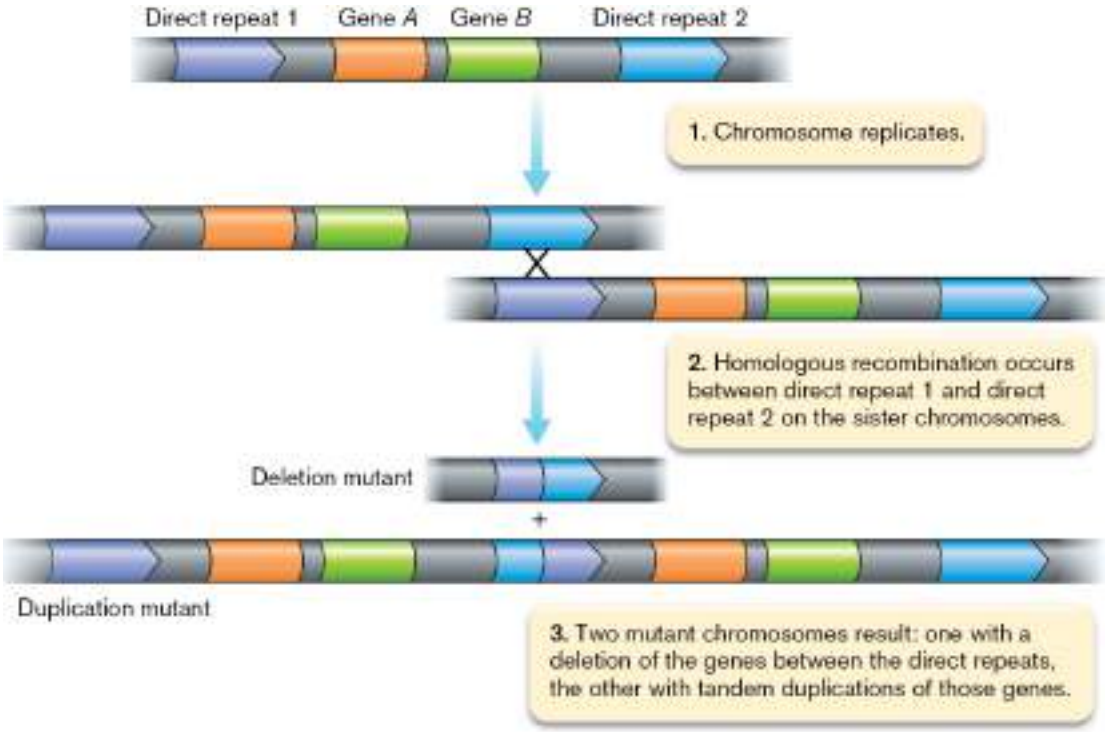


FIGURE 9.35 ■ Gene loss and gene duplication resulting from homologous recombination between two sister chromosomes. Recombination at direct repeats that flank genes *A* and *B* leads to sister chromosomes with a deletion or tandem duplication of those genes. After chromosomes are partitioned and cell division completes, one daughter cell is a deletion mutant, and the other is a duplication mutant.

Once the gene is duplicated, one copy (or both copies) can acquire new function through mutation. Gene duplication followed by divergence is evident when the sequences of genes within and between genomes are compared. Genes with shared ancestry have sequence similarity, or homology, and are referred to as homologs . Homologs can be classified as either orthologs or paralogs. Orthologous genes (orthologs) are found in different species and usually have the same function. Paralogous genes (paralogs) are homologs found within a single genome and typically have related but distinct functions.

Paralogs arise by gene duplication (**Fig. 9.36**). Duplication facilitates divergence because while one copy maintains the preexisting function, the second copy is free to evolve into a paralogous gene. For example, genes encoding certain exporters of virulence factors in pathogenic bacteria evolved from paralogous genes whose products assemble flagella. The two sets of genes share a common ancestor but have since evolved completely different functions. Paralogous genes are maintained in a microbial genome because their distinct functions contribute to the organism's adaptive potential.

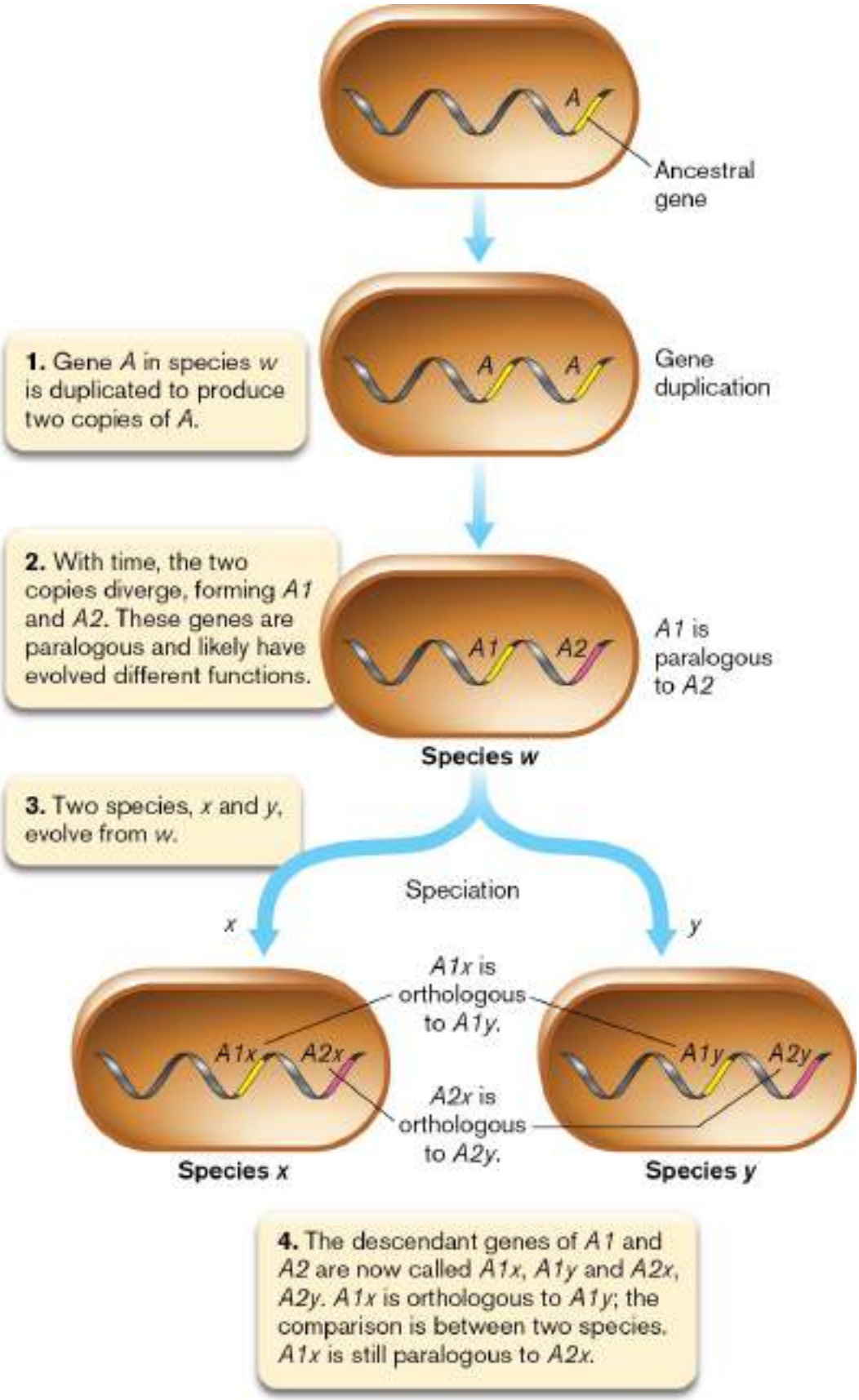

A
Ancestral gene
1. Gene A in species w is duplicated to produce two copies of A.
A
A
Gene duplication
2. With time, the two copies diverge, forming A1 and A2. These genes are paralogous and likely have evolved different functions.
A1
A2
A1 is paralogous to A2
Species w
3. Two species, x and y, evolve from w.
Speciation
x
y
A1x is orthologous to A1y.
A2x is orthologous to A2y.
A1x
A2x
A1y
A2y
Species x
Species y
4. The descendant genes of A1 and A2 are now called A1x, A1y and A2x, A2y. A1x is orthologous to A1y; the comparison is between two species. A1x is still paralogous to A2x.

FIGURE 9.36 ■ Paralogous versus orthologous genes. An ancestral gene can undergo a duplication to evolve an orthologous or paralogous gene.

Annotating DNA Sequences



We know from considerable experimental precedent that orthologous proteins often have the same or similar function, regardless of species. This information allows researchers to predict the functions of newly sequenced genes and even entire genomes. Once the DNA sequence is acquired and assembled into the chromosomes and plasmids that constitute the genome, the next steps are to find the genes and then use the sequence information to infer function. These steps are performed using bioinformatics , an interdisciplinary field that combines computer science, mathematics, and statistics.

To illustrate the value of bioinformatics, **Figure 9.37** shows a circular display map of chromosome I of the Gram-negative pathogen *Vibrio vulnificus* . *V. vulnificus* is a marine bacterium that causes lethal blood or wound infections associated with eating or catching raw shellfish. One powerful approach to understanding how *V. vulnificus* causes disease is to probe its genome for genes associated with disease. Once such genes are located, they can be analyzed by bioinformatics to assign predicted functions on the basis of their sequence. The genome-scale level of predicted microbial activity can be used to generate hypotheses about how this organism causes disease, and these hypotheses can be tested using other techniques.

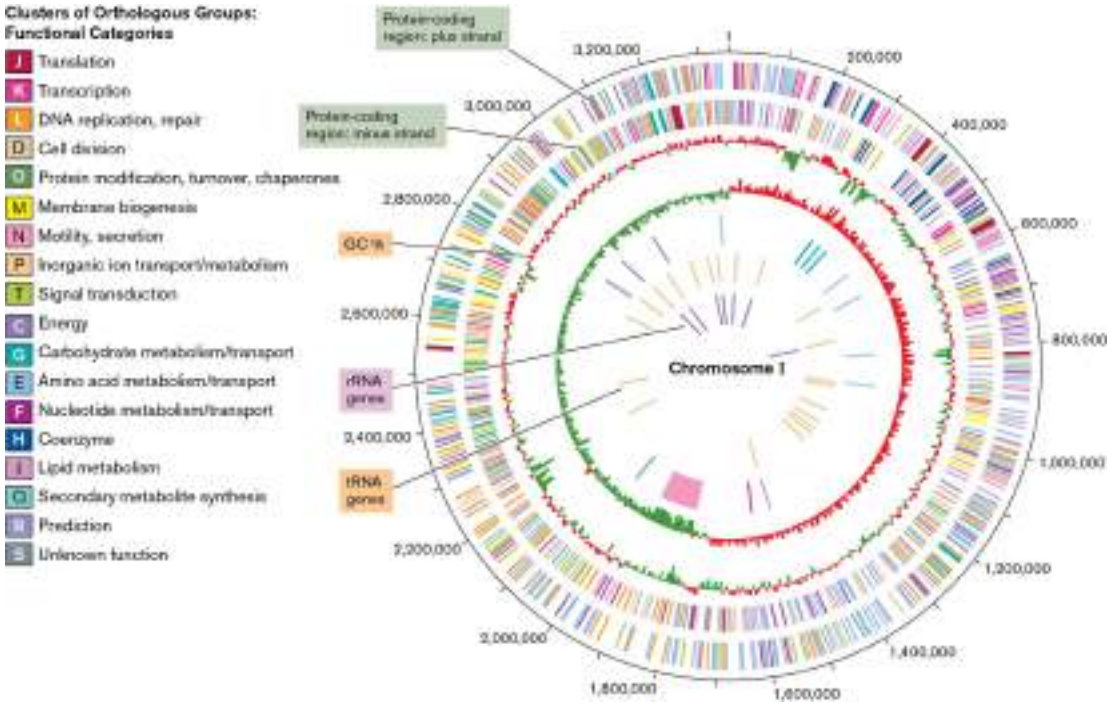


FIGURE 9.37 ■ Circular display map of the *Vibrio vulnificus* genome (chromosome I; 3,377 kb). The color code indicates gene clustering by function. The two outer rings represent genes transcribed from different DNA strands. The innermost circle uses peaks to show GC content levels above (red) or below (green) the genome average.

Bioinformatics has forever changed how the science of microbiology is conducted. Since 1998, over 140,000 complete microbial genomes have been sequenced, and a vast number of partial genomes from uncultured microbes is likewise available. Recall that fewer than 1% of microorganisms can be cultured. Bioinformatics, then, is critical to understanding the complexity of the uncultured microbial world.

In Chapter 7 we introduced the genome and the metagenome, the genes in a single microbe or a microbial community, respectively. But how do we learn which genes are encoded in a newly sequenced genome or metagenome? The first step is annotation . Annotation is analogous to identifying separate sentences and words in an unknown language. Computers annotate DNA sequences by using

the known rules of transcription and translation to identify the start and stop sites of potential genes. DNA sequences are then translated *in silico* to determine amino acid sequences. When a sequence appears to encode a protein, the DNA sequence is called an open reading frame (ORF). Genes that encode transfer RNAs and ribosomal RNAs do not encode proteins but can be identified because of the conservation of these sequences across vast phylogenetic distances.

Seeking ORFs. Computer programs such as ORF finders use the universal genetic code to deduce protein sequences potentially formed in all reading frames on RNA molecules transcribed from either direction on the chromosome (**Fig. 9.38**). An ORF, the equivalent of a sentence in our analogy, is defined as a DNA sequence that can potentially encode a string of amino acids of minimum length—say, 50 residues. The 50 residues of the ORF could encode a 5,500-Da (5.5-kDa) protein, as the average weight of an amino acid is 110 Da. Each ORF begins with a translation start codon (usually ATG or, more rarely, GTG or TTG). A translation start codon marks where ribosomes start to read a messenger RNA (mRNA) molecule. An ORF ends with a translation termination codon (in DNA: TAA, TAG, or TGA). In addition, the computer can identify an ORF by looking for potential ribosome-binding sites upstream of the start codon, but these ribosome-binding sites may differ between species, so finding one is not essential for declaring the presence of an ORF.

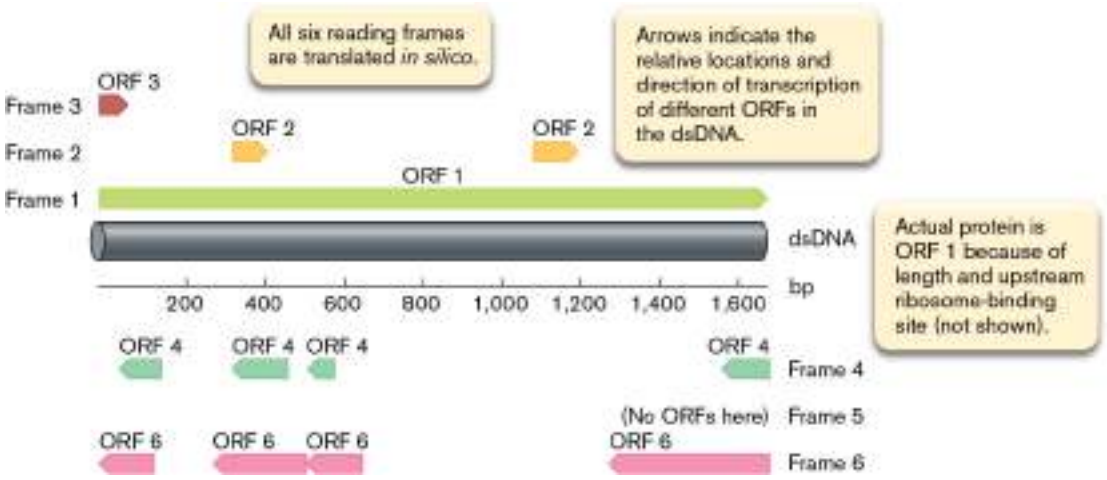


FIGURE 9.38 ■ Predicting open reading frames (ORFs) in a bacterial DNA sequence. Each predicted ORF in this 1,600-bp sequence begins with AUG or GUG and ends with a translation terminator codon.

Note: It is important to recognize that ORF predictions need to be verified experimentally. Additionally, it is important to be aware that arbitrary cutoffs for minimum ORF length, such as 50 amino acids, can exclude from the search smaller proteins that may play important roles in the cell.

Once the ORFs are identified, the next step is to assign predicted function on the basis of homology. During this step in the annotation process, ORFs are compared to hundreds of thousands of known proteins to look for orthologs with experimentally validated function. ORFs with sufficient homology to known proteins can then be assigned putative function. Powerful computer programs are available to carry out these analyses quickly (for examples, visit the Joint Genome Institute website).

If a query protein sequence does not show high levels of similarity with known proteins, "annotation by function" can be assigned when a protein domain exhibits characteristics or motifs

similar to those possessed by proteins of known function. For example, periodic hydrophobic regions may indicate a membrane protein, or an amino acid sequence (or motif) matching those of known ATPases may suggest a similar function for the query protein. Annotation by function provides an extremely valuable starting point toward unraveling a protein's function, but biological validation is required for confirmation.

Horizontal Gene Transfer

The natural movement of genes between organisms is called horizontal gene transfer , or lateral gene transfer. Horizontal gene transfer between cells differs from vertical gene transfer , which is the generational passing of genes from parent to offspring (as in cell division). In contrast, horizontal gene transfer involves the import of genetic material from a foreign source into the cell and incorporation of that material into the genome. Sections 9.3 and 9.4 describe the major ways that genes are transferred between microbes. In this discussion we place these transfer events into the context of evolutionary gene acquisition.

Evolution can occur during both vertical and horizontal gene transfer, but the mechanisms and outcomes can be quite different. During vertical gene transfer, genomes can evolve through mutation of genes already present in the genome, via mechanisms such as gene divergence after duplication. Horizontal gene transfer allows genomes to evolve by introducing new genes from foreign sources. These genes can be entirely unrelated to anything in the original genome and can provide brand-new properties to the organism once expressed.

Horizontal gene transfer is a prevalent mode of genetic innovation in the microbial world. For example, it has been estimated that nearly 20% of the *E. coli* genome may have originated in other microbes. *E. coli* strain O157:H7, the culprit in several fatal outbreaks of food-borne disease throughout the world, contains 1,387 genes that are not in strain K-12. These additional

genes represent about 25% of the O157:H7 genome and encode virulence factors, metabolic pathways, and prophages (phage genes integrated into host chromosomes), all of which were acquired from other species.

Thought Question

9.8 Gene homologs of *dnaK* encoding the heat-shock chaperone HSP70 exist in all three domains of life. All bacteria contain HSP70, but only some species of archaea encode a *dnaK* homolog. The archaeal homologs are closely related to those of bacteria. Knowing this information, how do you suppose *dnaK* genes arose in archaea?

Genomic Islands Are Acquired by Horizontal Transfer

Many of the new functions that horizontal gene transfer can provide require the products of multiple genes. In such cases, these genes are usually all transferred together, forming a genomic island in the recipient chromosome (**Fig. 9.39**). Genomic islands serve a variety of functions, including virulence (pathogenicity islands), symbiosis (symbiosis islands), metabolism (metabolic islands), and antibiotic resistance (resistance islands). Pathogenicity and symbiosis islands can encode remarkable protein secretion systems (called type III and type IV secretion systems) capable of injecting effector proteins directly from the bacterium straight into a eukaryotic cell. Once inside the target cell, these effector proteins alter the function of eukaryotic cell proteins, making the host organism, in the case of symbiosis, more accepting of the symbiont. Type III and type IV secretion systems are discussed further in Chapter 25 .

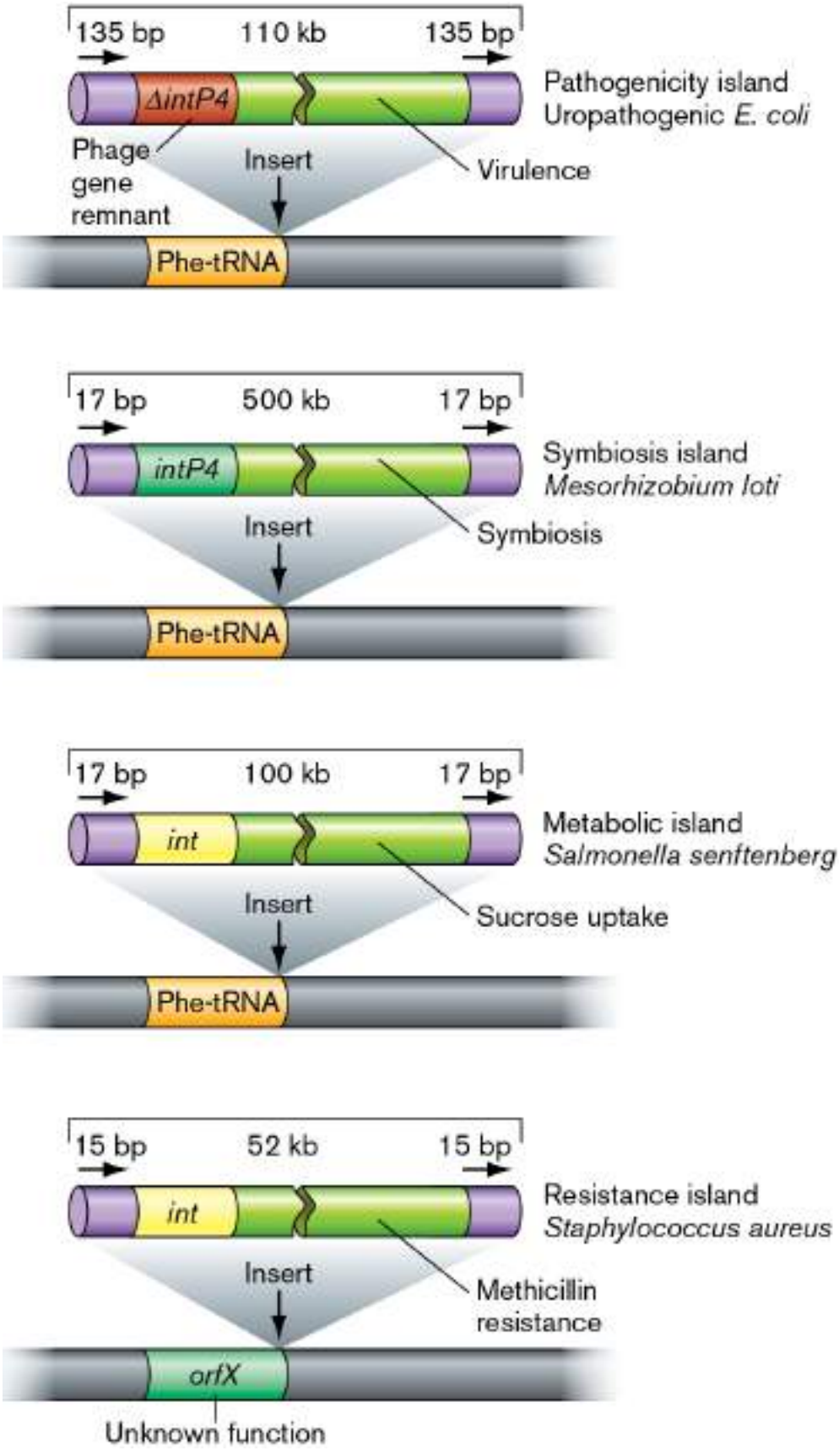

135 bp
110 kb
135 bp
ΔintP4
Pathogenicity island
Uropathogenic E. coli
Phage
gene
remnant
Insert
Virulence
Phe-tRNA
17 bp
500 kb
17 bp
intP4
Symbiosis island
Mesorhizobium loti
Insert
Symbiosis
Phe-tRNA
17 bp
100 kb
17 bp
int
Metabolic island
Salmonella senftenberg
Insert
Sucrose uptake
Phe-tRNA
15 bp
52 kb
15 bp
int
Resistance island
Staphylococcus aureus
Insert
Methicillin
resistance
orfX
Unknown function

FIGURE 9.39 ■ Genomic islands from different microorganisms. Genomic islands are often found inserted adjacent to tRNA genes and have direct repeats at their ends (arrows). Functions for each island are identified.

Genomic islands are often flanked by boundary regions such as direct repeats or insertion elements, usually found near tRNA genes (**Fig. 9.39**). These flanking elements are hot spots for genetic exchange (recombination) and are the usual targets for insertion of genomic islands into the chromosome.



Genomic islands recently acquired by horizontal gene transfer can be identified through analysis of the frequency of GC base pairs (versus AT base pairs) along the chromosome (also known as GC content). A sudden change in the GC content in a discrete section of the chromosome is the equivalent of an evolutionary "footprint" marking a horizontal gene transfer. As an example, genomic analysis indicated that genes involved in synthesizing the cell surface of the marine photosynthetic bacterium *Prochlorococcus* are of foreign origin and arrived via horizontal gene transfer as a genomic island. Whereas the average GC content of the *Prochlorococcus* strain MIT9313 was 51%, the 33-gene cluster encoding the surface polysaccharide genes was, on average, much lower, at 42% (**Fig. 9.40**). Note that there is some variation in GC content of the cluster, such that several cluster regions are actually higher in GC than the 51% average for the genome. As another indicator of its foreign origin, this gene cluster of strain MIT9313 was absent in the related strains *Prochlorococcus* MED4 and *Synechococcus* WH8102 (**Fig. 9.40**).

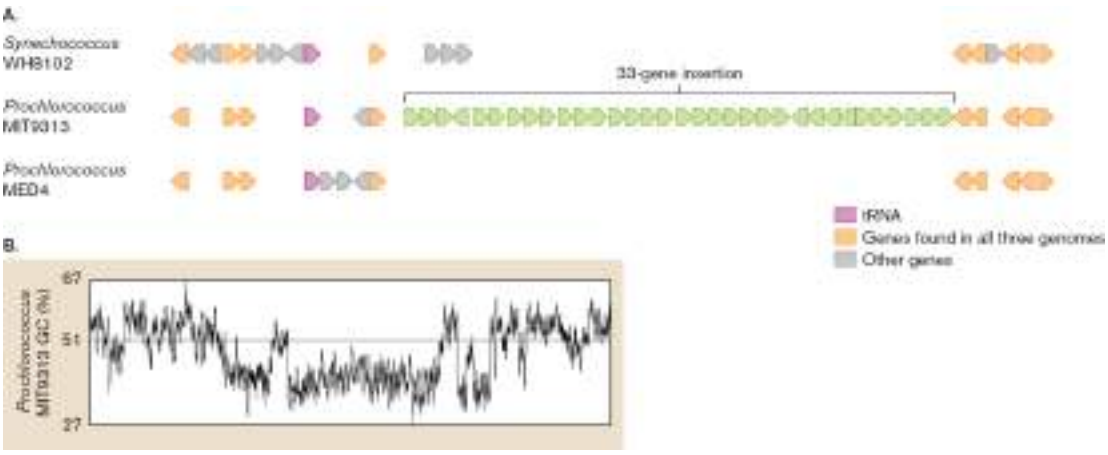


FIGURE 9.40 ■ Genomic island in *Prochlorococcus* strain MIT9313. A. The genomic region of the island, comparing the same genomic region in related strains *Synechococcus* WH8102 and *Prochlorococcus* MED4. Gaps between genes do not represent DNA; they are in place to align the genes shared between strains. Genes in purple are tRNAs, genes in orange are found in all three genomes, genes in gray are found in one or two strains, and genes in green are the genomic island of MIT9313. **B.** GC content for the MIT9313 genome in this region.

Source: Modified from G. Rocap et al. 2003. *Nature* **424** :1042–1047, fig. 3b.

G. ROCAP ET AL. 2003. *NATURE* 424:1042–47, FIG. 3B

Thought Question

9.9 Every strain of *Prochlorococcus* examined to date has a unique cluster of genes that modifies the chemical properties of the cell surface. What sort of selective pressure(s) might account for the high degree of variation in cell-surface properties for this marine bacterium?

Importantly, over long periods of evolutionary time, horizontally transferred genes tend to lose their distinctive GC signature and adopt the "average" GC composition of the recipient cell. There are several reasons for this shift, one being that even though multiple codons can code for the same amino acid during translation (see Chapter 8), cells are usually biased in the codons they use for this purpose. Consequently, there is evolutionary pressure for acquired genes to adopt the recipient's predominant codon usage in their own DNA sequence, which will ultimately change the GC content of the genes. As such, many of the more ancient horizontal gene transfer events cannot be identified by examination of the GC content, and other methods must be used to discover them.

Horizontal gene transfers between Bacteria and Eukarya. In eukaryotes, the sexual exchange of genes is usually only within a single species. Bacteria and archaea, on the other hand, are more promiscuous. Matings between microbial genera are common and perhaps even desirable from an evolutionary viewpoint. By sharing genes, different species can sample genes from each other and keep genes (through natural selection) that increase fitness. But what about transfers between Bacteria and Eukarya? Do they happen? If so, are they useful? Work by Gos Micklem and colleagues at the University of Cambridge suggests that interdomain transfer of DNA over the course of evolution may be more common than was previously realized. These researchers report that DNA sequences of many eukaryotes, including humans, appear to include a variety of genes horizontally transferred from bacteria (**Fig. 9.41**). How the genes jumped domains is unclear.

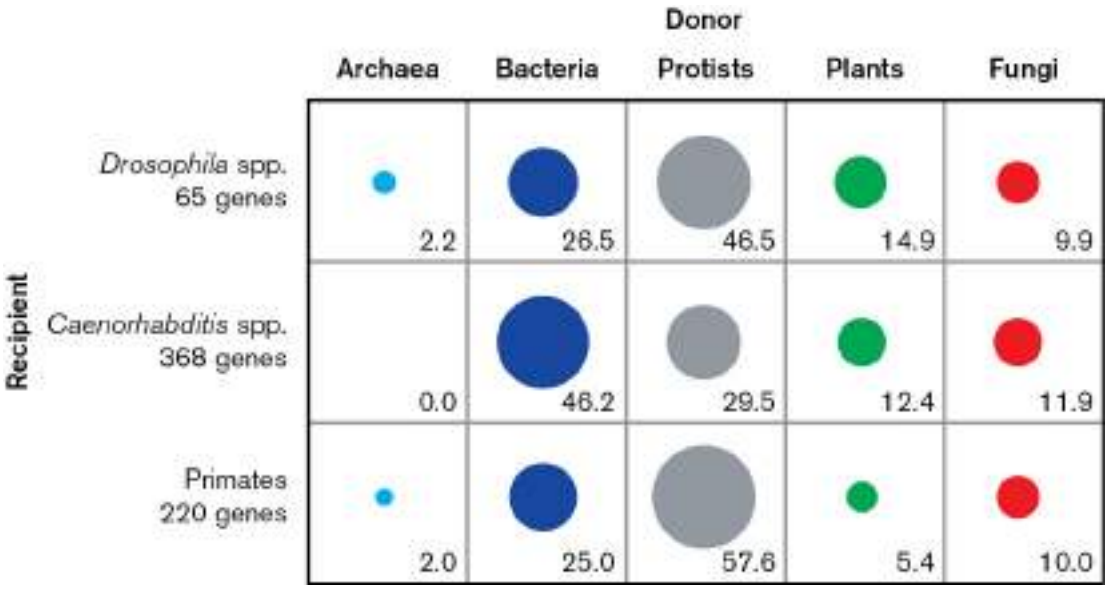


FIGURE 9.41 ■ Genes in *Drosophila, Caenorhabditis* , and primate genomes that are predicted to have been transferred horizontally from bacteria, archaea, and other eukaryotes. Numbers within cells show percent contribution of each donor to the total of all genes acquired horizontally by the recipient.

In a bizarre example, Joseph Mougous (University of Washington) and colleagues described a eukaryotic gene called *dae* (d omesticated a midase e ffector) that was transferred from bacteria to eukaryotic ticks and mites several times over millions of years. The bacterial version of the protein, called Tae, is injected into other bacteria and disassembles the recipients' cell wall. Importantly, eukaryotes lack bacterial cell walls and so are unaffected by the toxin. Among the eukaryotic recipients of the *dae* gene is the deer tick *Ixodes scapularis* , which serves as the reservoir for the Lyme disease spirochete *Borrelia burgdorferi* (Lyme disease is described in Chapter 26). After their horizontal transfer, the *dae* genes became expressed and eventually acquired eukaryotic protein secretion signals.

Figure 9.42 illustrates that the *I. scapularis dae* gene helps the insect control the levels of *B. burgdorferi* colonizing the tick's midgut.

The bacterial gene, therefore, became part of the *Ixodes* innate immune system against *B. burgdorferi* and probably other bacteria. Innate immunity is described in Chapter 23 .

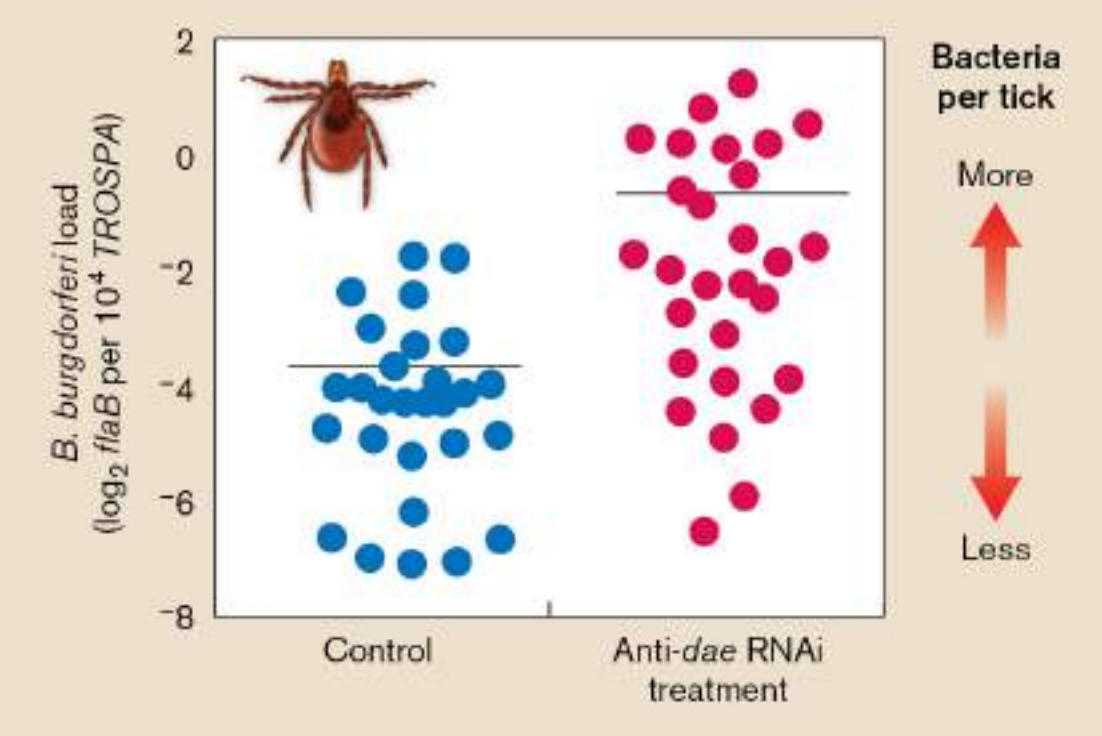


FIGURE 9.42 ■ The *dae* product helps the deer tick control levels of the spirochete *Borrelia burgdorferi* . Deer tick nymphs (inset) fed on spirochete-infected mice, and 2 weeks later *B. burgdorferi* levels were quantified to identify the number of *B. burgdorferi flaB* genes. To estimate relative numbers of bacteria per tick (*B. burgdorferi* load), the ratio of *B. burgdorferi* –specific *flaB* gene copies relative to copies of the tick-specific gene *TROSPA* were determined. Control nymphs received no treatment. Treated nymphs were administered an interfering RNA (RNAi) that binds to *dae* mRNA to reduce *dae* expression. Horizontal lines indicate mean values.

The Intestine: Cauldron of Horizontal Gene Transfer



The human gastrointestinal tract is ideal for exchanging genetic information between widely diverse species (see Section 23.1). The gut contains a high bacterial cell density and mixed-species biofilms that enable close and frequent contact between organisms, both living (a requisite for conjugation) and dead (a source of transforming DNA). The intestine is also rife with phages that can carry DNA by transduction (see Chapter 6). Transduction is one of the key ways that antibiotic resistance genes can be transferred within the gut microbiome. Special Topic 9 describes how scientists can use high-throughput DNA sequencing to catch phage in the act of specialized and generalized transduction in the gut.

SPECIAL TOPIC 9 Gut Microbiome Caught in the Act of Gene Transfer

The gut ecosystem is a hotbed of genetic exchange, which has important consequences for antibiotic resistance among other effects on the microbiome and host. Both specialized and generalized transduction have been implicated as important contributors to horizontal gene transfer in the gut. But important questions remain: Among the highly diverse microbiome, which hosts are transferring their DNA, and how important is transduction for the overall process of horizontal gene transfer in the gut? These questions are quite daunting to answer. Many hosts and phages within the gut may not be involved in transduction. Additionally, there are multiple mechanisms of horizontal gene transfer other than transduction, so how can genes transferred by phage be identified?

A team led by Manuel Kleiner and Breck Duerkop from North Carolina State University and the University of Colorado School

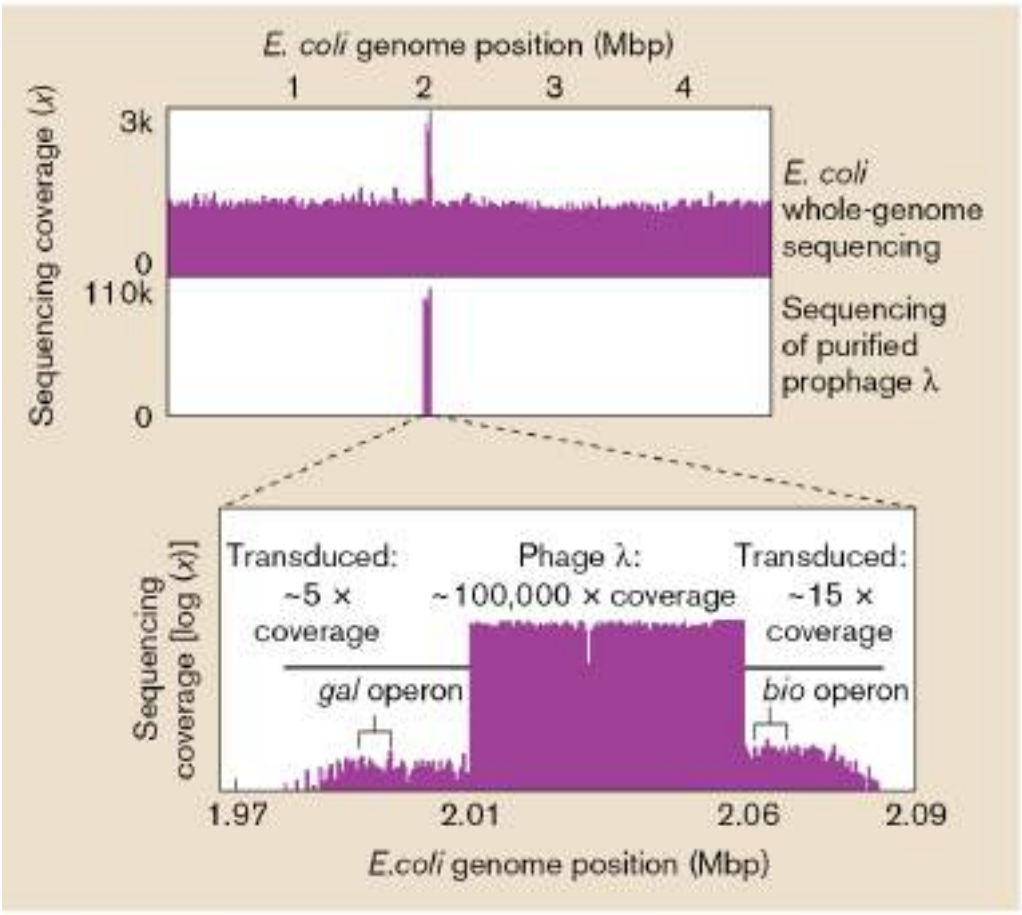
of Medicine, respectively, addressed these questions with a high-throughput sequencing approach designed to catch transducing phage in the act of host DNA transfer. The approach was to compare the DNA carried by released phage particles with the DNA of the host: The sum total of all host genes found within the phage particles constituted the "transductome." Importantly, while not every gene within the transductome would likely be successfully transferred to a new host, this analysis would provide a snapshot of the total genetic potential for horizontal gene transfer by transduction.

To compare the transductome with the metagenome of the whole community, the transducing phage needed to be purified from the mixed community sample. After cesium chloride (CsCl) density gradient centrifugation, purified phages were collected as a discrete band in the centrifugation tube. DNA from both purified phages and whole-community samples was then extracted and sequenced. Short reads from Illumina sequencing of the whole community and of the purified phage particles were assembled into contigs (see Section 7.6) and compared by alignment of homologous regions.

What sequencing results would indicate that a gut microbiome phage was participating in either specialized or generalized transduction? The researchers used the *Escherichia coli* –infecting phages lambda (λ) and P1 as test cases for specialized and generalized transduction, respectively (see Section 9.3). **Figure ST 9.1A** shows the sequencing results of a prophage induction experiment for an *E. coli* strain that is lysogenic for lambda. DNA from the lambda phage particles aligned exactly to the region of the *E. coli* chromosome where lambda integrates as a prophage, confirming that the host is a lambda lysogen. Most of the reads from the DNA isolated from the purified lambda particles mapped to the lambda phage genome as expected, but a small fraction of the particles also carried flanking regions that belonged to the host. One region included the *gal* operon while the other included the *bio*

operon, both of which are known to be transferred by lambda via specialized transduction.

A.



B.

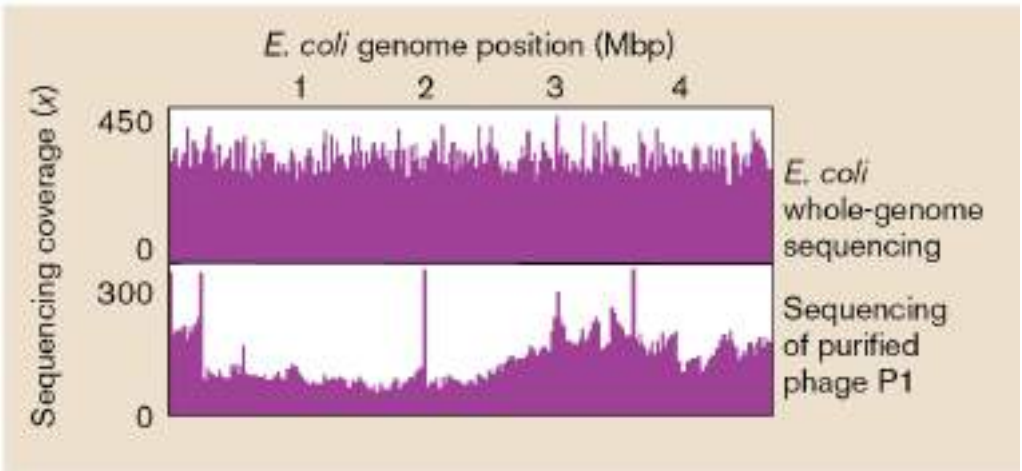


FIGURE ST 9.1 ■ Witnessing specialized and generalized transduction by *Escherichia coli*–infecting phages. A. Comparison of the assembled DNA contigs of the lysogenized *E. coli* host (top) and purified phage lambda (bottom). Sequences of the purified lambda align to the integrated lambda phage in the lysogenized host. **Inset:** A close-up of the purified lambda DNA, revealing that a subset of the DNA copies (that is, less total sequencing coverage) also contains flanking DNA from the host chromosome. **B.** Comparison of the assembled DNA contigs of the *E. coli* host and the fraction of purified P1 phage that contains DNA from the host instead of phage DNA. Note that the alignments of the transduced host DNA to the host chromosome indicate some bias in terms of host DNA transduced, but overall, every section of the host chromosome was being incorporated into the P1 phage capsid for transduction. Mbp = megabase pairs.

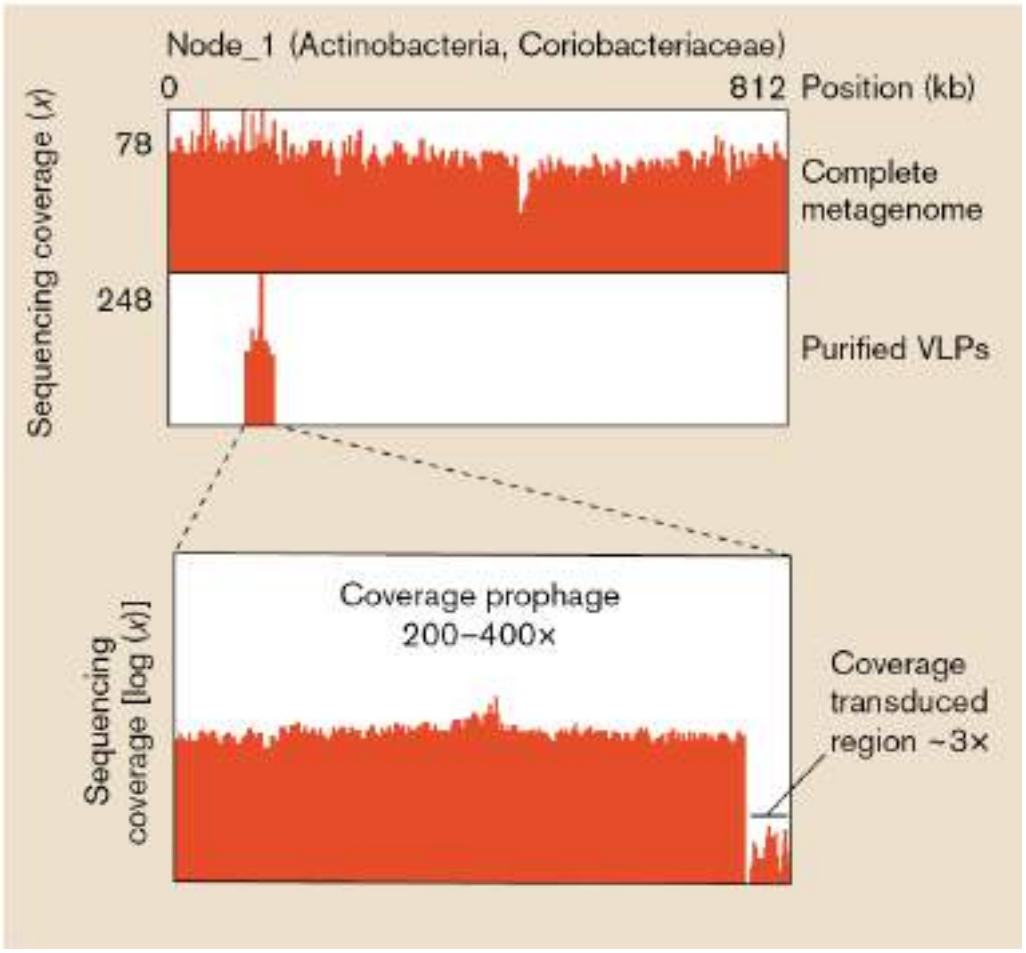
Kleiner, Manuel, Brian Bushnell, Kenneth E. Sanderson, Lora V. Hooper, and Breck A. Duerkop. 2020. Transductomics: Sequencing-based detection and analysis of transduced DNA in pure cultures and microbial communities. *Microbiome* 8:158.

Results with the generalized transducing phage P1 were likewise successful. Most (95%) DNA sequences isolated from P1 phage particles mapped to the P1 genome (data not shown). However, 5% of the reads mapped to the *E. coli* chromosome, and the locations of these reads were broadly distributed (**Fig. ST 9.1B**), as one would expect for generalized transduction.

With the patterns of DNA sequencing results for specialized and generalized transduction demonstrated by these test cases, the researchers then drew their attention to the microbiome of the mouse gut, using fecal samples as indicators of gut community composition. They sequenced 390 million reads of the bacterial host metagenome and 360 million reads

from purified phage particles (VLPs). From the host metagenome, they were able to assemble 2,143 contigs of 40 kilobase pairs (kb) or larger. Of these 2,143 contigs, 186 had sequences that could also be mapped in the phage particle contigs. Twelve of these 186 contigs, such as Node_1 from a member of the Actinobacteria (**Fig. ST 9.2A**), showed evidence of specialized transduction: They contained prophage DNA flanked by DNA belonging to the host. Forty-six of the 186 contigs showed evidence of generalized transduction, as seen in **Figure ST 9.2B** for the contig named Node_251, from an unidentified bacterium. Note that for these putative generalized transducing phages, there is no phage DNA flanking the host DNA, as only the host DNA has been packaged into the capsids. As such, the identities of this class of transducing phages remain a mystery.

A.


Node_1 (Actinobacteria, Coriobacteriaceae)
0
812 Position (kb)
Sequencing coverage (x)
78
Complete metagenome
248
Purified VLPs
Coverage prophage 200–400×
Sequencing coverage [log (x)]
Coverage transduced region ~3×

B.

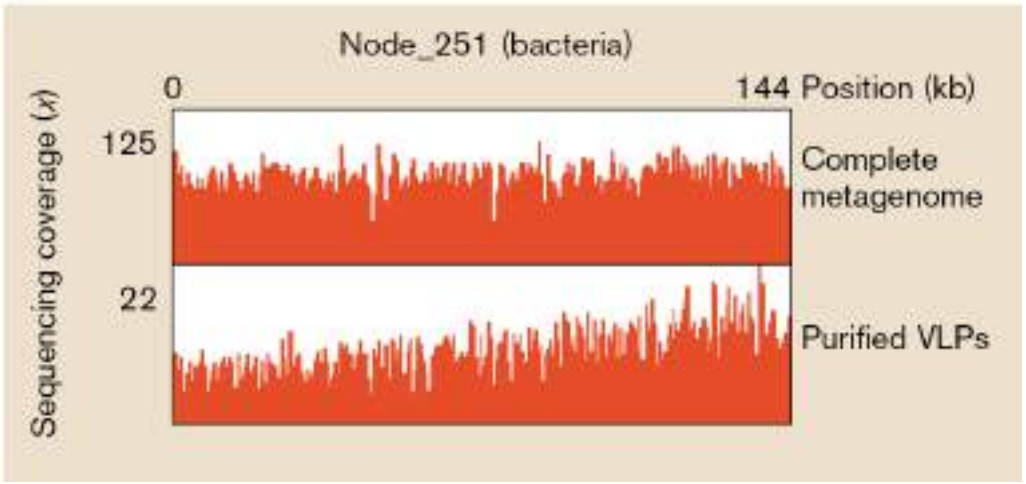

Node_251 (bacteria)
0
144 Position (kb)
Sequencing coverage (x)
125
Complete metagenome
22
Purified VLPs

FIGURE ST 9.2 ■ Specialized and general transducing phages identified in the mouse gut . **A.** An assembled contig called Node_1, identified as belonging to the Actinobacteria, contained a prophage (top). Among the purified phages of the mouse gut, contigs aligning to this prophage were found (bottom), and a subset of these sequenced contigs also carried flanking DNA of the host bacterium (inset). **B.** An assembled contig called Node_251, from an unidentified bacterium (top), aligned to several contigs of the purified phage (bottom). Note that these latter contigs contained no phage DNA, indicating they were in the process of generalized transduction of host DNA.

In this pioneering investigation, the act of DNA transduction by phage was captured in a complex gut microbial community. This study demonstrated that high-throughput sequencing could identify the host bacteria serving as transduction donors, as well as the genes they donated. The study also showed that phage responsible for specialized transduction could likewise be identified, and that both specialized and generalized transduction rates could be estimated in the complex gut community.

RESEARCH QUESTION

How might a similar approach be used to explore the DNA transferred via natural transformation or by membrane vesicles in complex ecosystems such as the animal gut?

Kleiner, Manuel, Brian Bushnell, Kenneth E. Sanderson, Lora V. Hooper, and Breck A. Duerkop. 2020. Transductomics: Sequencing-based detection and analysis of transduced DNA in pure cultures and microbial communities. *Microbiome* **8** :158.

Perhaps the most intriguing and worrisome evidence of horizontal gene transfer in the gut involves antibiotic resistance. The gut microbiome can be considered a repository of many antibiotic resistance genes, usually within anaerobic members of the microbiome. Studies that examined the gut metagenomes of hundreds of individuals have identified upwards of 1,000 antibiotic resistance genes "lurking" among the microbial population. Their presence is not always evident until someone is treated with an antibiotic. For example, a plasmid encoding high-level resistance to carbapenem antibiotics was horizontally transferred from *Klebsiella pneumoniae* to *Escherichia coli* in the intestine of a 91-year-old patient. The resistance enzyme, carbapenemase, confers resistance to all cephalosporins, monobactams, and carbapenems used as first-line drugs for hospitalized patients (see Chapter 27). The man was being treated for sepsis (a severe blood infection) with ertapenem. When the man's stool was tested after treatment, it contained carbapenem-resistant *K. pneumoniae* , but not *E. coli* . One month later, carbapenem-resistant *E. coli* containing the same plasmid was recovered from his stool.

Antibiotic exposure may even promote horizontal gene transfer among microbiome members. Human volunteers whose stools contained erythromycin-susceptible *Bacteroides* strains were given erythromycin. Within 7 days, erythromycin-resistant *Bacteroides* strains were found. These results implicate conjugation of erythromycin resistance plasmids from unknown members of the gut microbiome. **Figure 9.43** illustrates the mechanisms by which antibiotic resistance genes can be transferred from reservoir gut microbes to opportunistic pathogens that can escape the intestine, either by defecation or through intestinal lesion, to infect elsewhere.

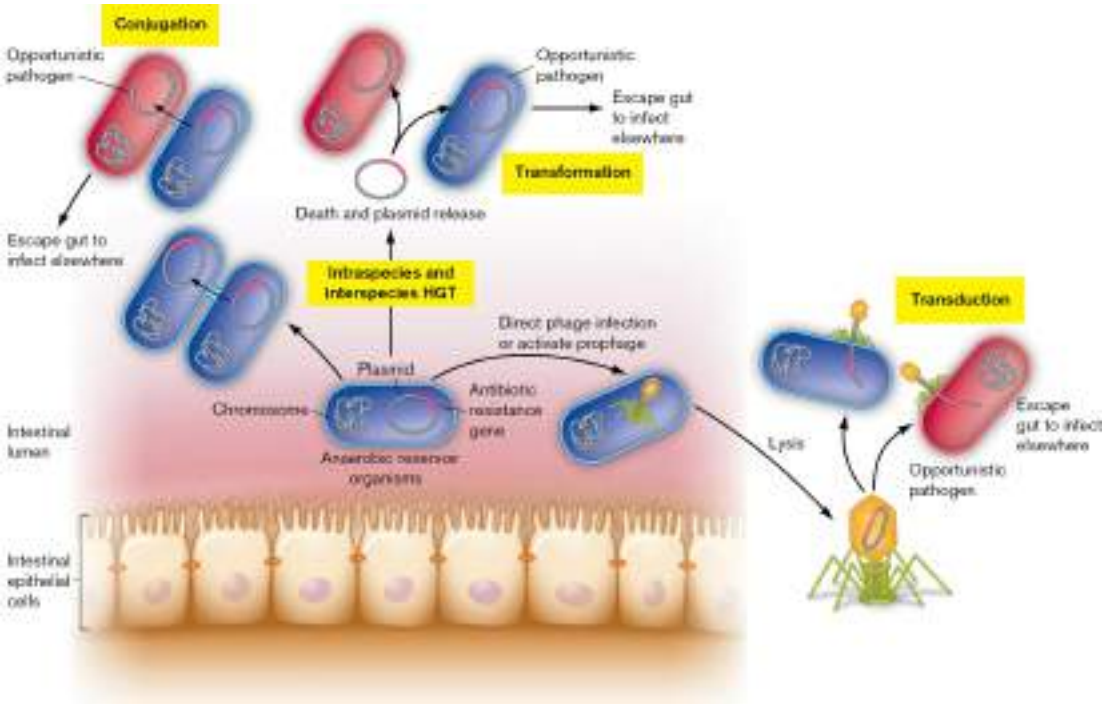


FIGURE 9.43 ■ Model for horizontal gene transfer between gut microbiota. High cell density and biofilm formation in the intestine make the horizontal transfer of antibiotic resistance genes (or other genes) easier. Many anaerobic members of the human microbiome (such as *Bacteroides*) serve as reservoirs for antibiotic resistance genes. Various gene transfer mechanisms will randomly transfer the genes, often on plasmids, to other members of the microbiota, some of which are opportunistic pathogens (*E. coli* , for instance). Opportunistic pathogens can escape the gut to cause disease in other body locations. HGT = horizontal gene transfer.

Source: Modified from W. van Shaick. 2015. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370** :20140087.

Genome Reduction

Evolution involves gene loss as well as gene acquisition. **Figure 9.35** illustrates how genes can be lost if two chromosomes

recombine at direct repeats. Gene deletions can also occur at direct repeats of the same chromosome (**Fig. 9.44**). In this scenario, the DNA loops back to align the direct repeats. Homologous recombination between the repeats leads to a chromosome with the genes between the repeats lost as an extrachromosomal circular fragment. This fragment cannot replicate (it lacks the origin of replication *oriC*) and is eventually lost by degradation. If the deletion is not lethal, the chromosome with the deletion mutation is able to replicate and be passed on to the mutant's progeny.

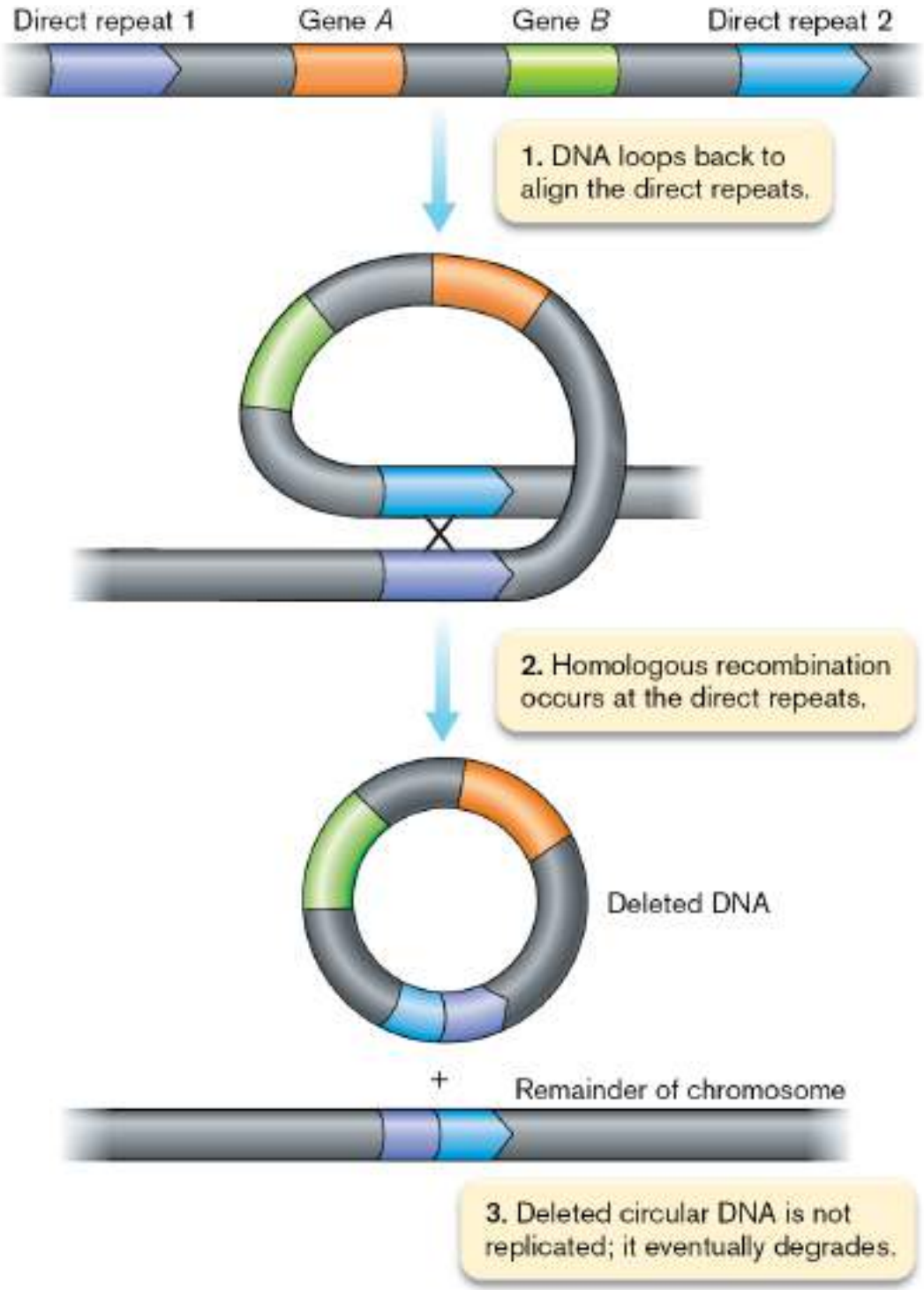


FIGURE 9.44 ■ Gene loss by homologous recombination at direct repeats of a single chromosome.

Gene loss can be "neutral," neither hindering nor helping the microbe, but in some cases the loss of genes can actually benefit the microbe. For example, pathogenic shigellae (the cause of bacillary dysentery) exhibit chromosomal "black holes," regions lacking genes that occur in the closely related *E. coli* . The absence of these genes is required for a fully virulent phenotype. One black hole of shigellae removed *cadA* , which encodes lysine decarboxylase. Lysine decarboxylase produces cadaverine, which inhibits the activity of enterotoxin, an important virulence factor for shigellae.

The large-scale loss of genes through evolution is known as genome reduction . Genome reduction is fairly common in pathogens and symbionts associated with plants and animals, but it can also be found in free-living microbes. The evolutionary pressures leading to genome reduction are covered in more detail in Chapter 17 ; here we provide several interesting examples of genome reduction.



The most abundant organism in the ocean, *Pelagibacter* (also referred to as SAR11), has a highly streamlined genome (1,357–1,576 genes, compared to the 4,489 genes of *E. coli* K-12). Streamlining is thought to be a consequence of selection for efficient growth under the nutrient-poor conditions of the oceanic habitat in which *Pelagibacter* has evolved. Notably, many of the genes found in the genomes of different *Pelagibacter* strains have been found in every *Pelagibacter* genome examined to date. These universally shared (core) genes are found within each *Pelagibacter* chromosome at roughly the same position (**Fig. 9.45**). Therefore, much of the chromosome's gene content seems to diverge only minimally in the different strains.

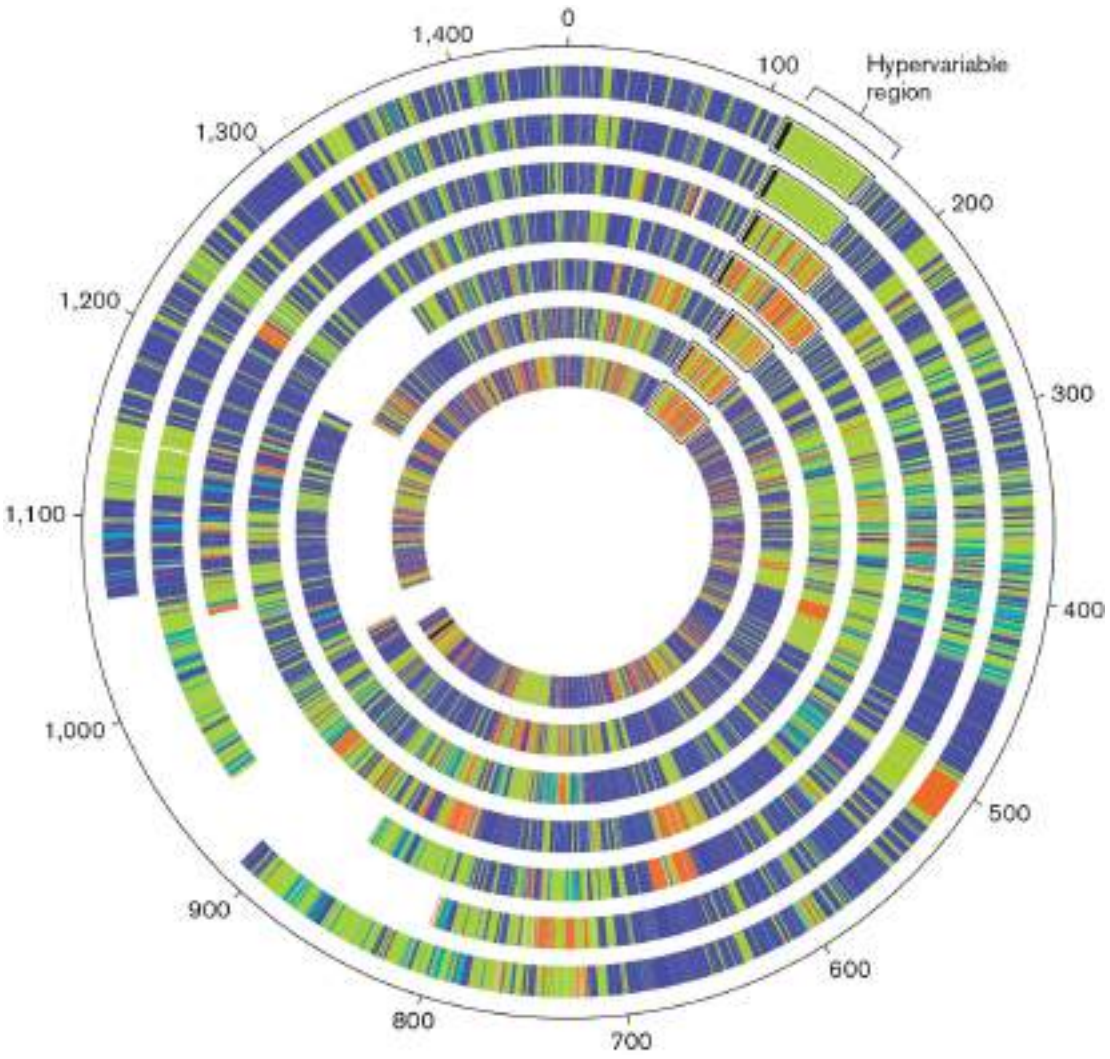


FIGURE 9.45 ■ Circular alignment of *Pelagibacter* genomes reveals a hypervariable region: a horizontal gene transfer hot spot. Each ring represents a *Pelagibacter* genome. Colors identify genes by category: blue = core *Pelagibacter* genes; bright green = additional core genes for *Pelagibacter* subclade Ia; orange = shared non-core genes; red = unique genes; black = rRNA genes. The numerical scale is measured in increments of 10,000 base pairs. Because of the disparity in genome sizes, large gaps were necessary to display the genomes in this circular manner.

Source: Modified from J. Grote et al. 2012. *mBio* **3** :e00252-12.

However, all *Pelagibacter* strains have a large, conspicuous region at roughly the same position in their chromosomes where they show vast differences in gene content between strains (**Fig. 9.45**). This hypervariable region is a spot where rampant horizontal gene transfer has provided strains with unique complements of genes. For some strains, genes in this region are used to scavenge and metabolize sugars and sulfur molecules. It's a mystery why extensive gene transfer is restricted to a hot spot in *Pelagibacter* , but the variety within this hot spot likely helps these bacteria occupy vast regions of the ocean. Chapter 17 further explores the role of genetic variation, including the concept of the "pangenome," in the evolution and ecology of bacteria.

Sometimes we see evidence of genome reduction in the making. Many genomes contain pseudogenes , genes that by homology appear to encode an enzyme but are nonfunctional because a portion is missing as a result of mutation. A pseudogene is the remnant of a gene whose function became superfluous to the organism, and consequently the selective pressure to maintain the gene's functionality was lost. Cells with such pseudogenes are not eliminated by natural selection, because the mutated gene did not contribute to fitness. Pseudogenes can eventually be removed completely from the genome via deletion, and in this way they contribute to the process of genome reduction.

Genes become superfluous and subject to mutation into pseudogenes for multiple reasons. One copy of a duplicated gene can become a pseudogene if the second copy is "covering" the gene's function for the cell. Genes can also become superfluous if the organism changes its lifestyle or invades a new environment. One pathogen that has been "caught in the act" of losing genes because of its lifestyle is *Mycobacterium leprae* , the cause of Hansen's disease (leprosy). Over half of the *M. leprae* genome is made of pseudogenes, presumably formed from genes it no longer needs after becoming an obligate intracellular pathogen (see **Fig. 18.22**).

To Summarize

- **Homologs** are genes with shared ancestry and have similar sequences. **Orthologs** are homologs found in different species; **paralogs** are homologs found within a single genome.
- **The field of bioinformatics** uses computer algorithms, mathematics, and statistics to annotate genes.
- **Annotation** requires computers that look for patterns in DNA sequences. Annotation predicts regulators, ORFs, and genes encoding untranslated RNAs such as tRNAs and rRNAs. Similarities in protein sequence (deduced from the DNA sequence) are used to predict protein structure and function.
- **Horizontal gene transfers** occur between cells of the same species and also between different species, even members of different domains.
- **Genomic islands** result from horizontal gene transfers that expand the competitiveness of a recipient by contributing to its pathogenicity, symbiosis, or fitness in a hostile environment.
- A DNA sequence with a **GC content** different from that of flanking chromosomal DNA is one sign of horizontal gene transfer.
- **Gene loss is an important part of microbial evolution** , and some pathogens, symbionts, and free-living microbes have undergone extensive **genome reduction** .

Glossary

homolog *or* homologous gene
: A gene derived from a common ancestral gene. Homologs may be orthologs or paralogs.

ortholog *or* orthologous gene

A gene present in more than one species that derived from a common ancestral gene and encodes the same function.

paralog *or* paralogous gene
: A gene that arises by gene duplication within a species and evolves to carry out a different function from that of the original gene.

bioinformatics
: A discipline at the intersection of biology and computing that analyzes gene and protein sequence data.

annotation
: The deciphering of genome sequences, including identification of genes and prediction of gene function.

horizontal gene transfer
: Also called *lateral gene transfer* . The natural movement of genes from one genome into another, nonprogeny genome.

vertical gene transfer
: The generational movement of genes from parent to offspring through reproduction.

genomic island
: A region of DNA sequence whose properties indicate that it has been transferred from another genome. Genomic islands usually comprise a set of genes with shared function, such as pathogenicity or symbiosis support.

pathogenicity island
: A type of genomic island in which the stretch of DNA contains virulence factors and may have been transferred from another genome.

symbiosis island
: A type of genomic island in which the stretch of DNA expresses proteins that enable a symbiotic relationship with another organism.

metabolic island
: A genomic island that contributes genes involved in metabolism.

resistance island
: A genomic island that contributes genes involved in antibiotic resistance.

genome reduction
: The large-scale loss of genes through evolution.

pseudogene
: A nonfunctional gene-like sequence that evolved by degenerative evolution.

Figure 7.20

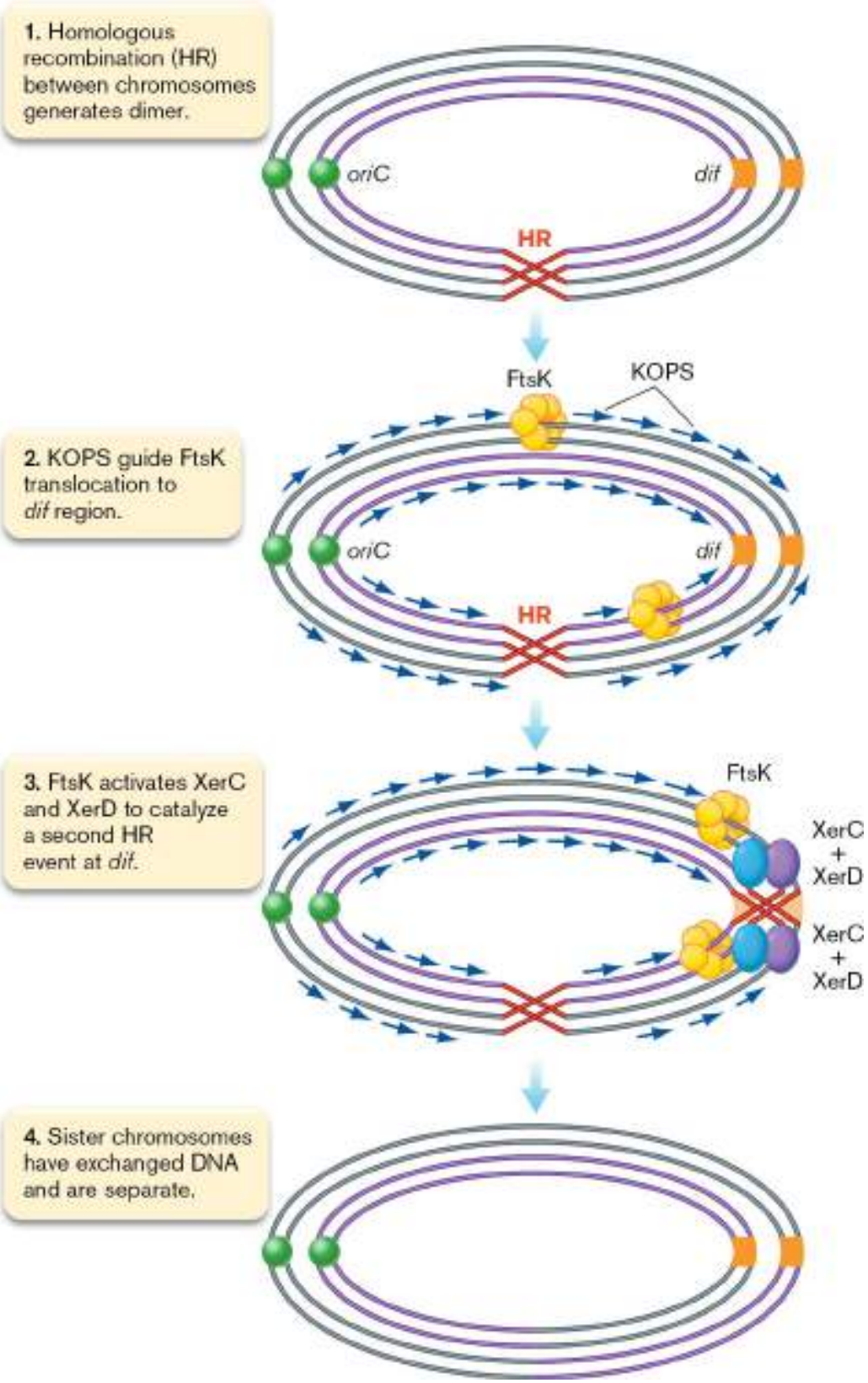

1. Homologous recombination (HR) between chromosomes generates dimer.
oriC
dif
HR
FtsK
KOPS
2. KOPS guide FtsK translocation to dif region.
oriC
dif
HR
3. FtsK activates XerC and XerD to catalyze a second HR event at dif.
FtsK
XerC + XerD
XerC + XerD
4. Sister chromosomes have exchanged DNA and are separate.

FIGURE 7.20 ■ Resolution of chromosome dimers by XerC and XerD at the *dif* locus. The initial homologous recombination (HR) event results in the dimer, and resolution at the *dif* site results in separated chromosomes that have exchanged a segment of their DNA.

Fig. 18.22

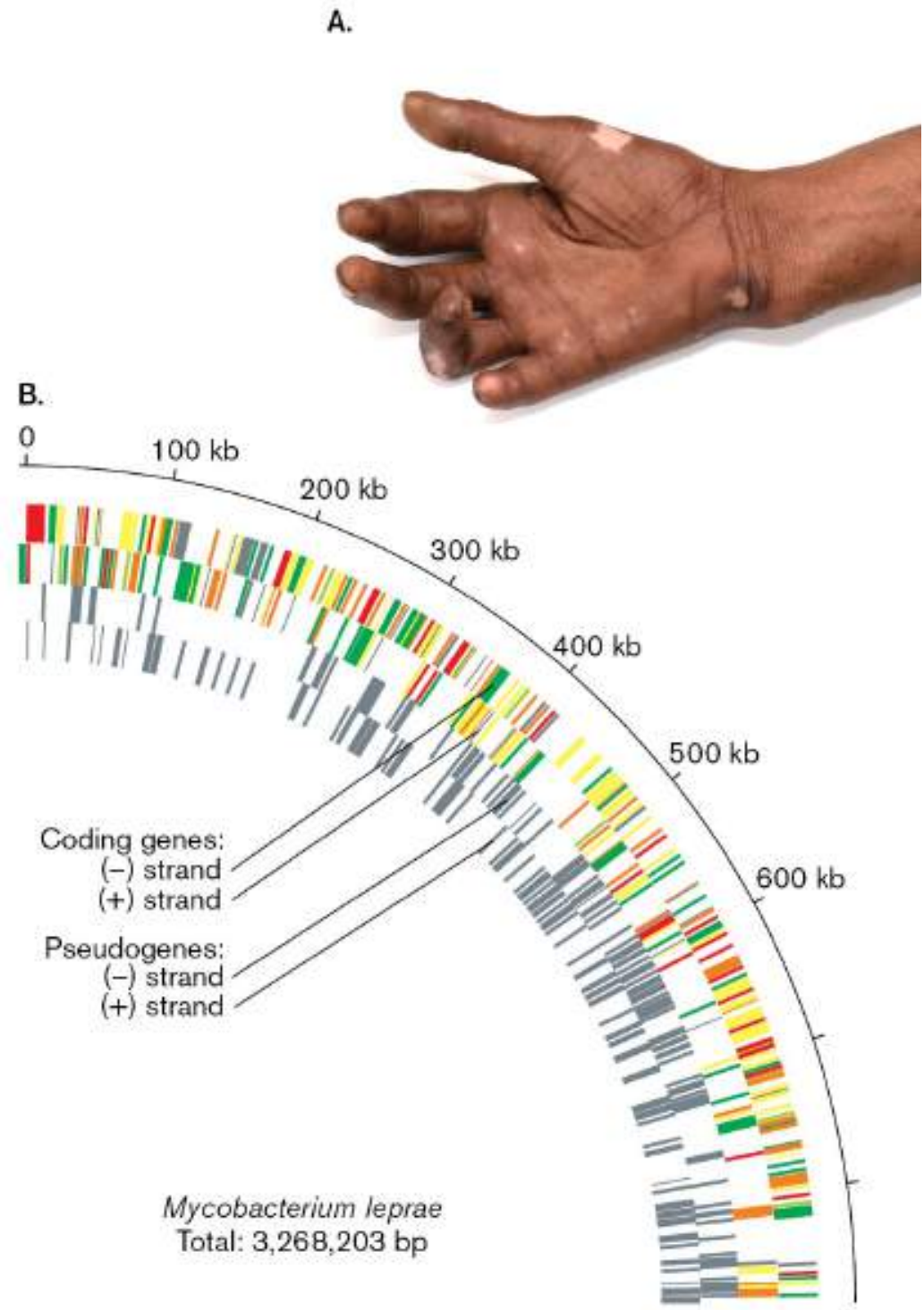


FIGURE 18.22 ■ *Mycobacterium leprae* causes leprosy. A. Hand disfigured by leprosy. **B.** The genome of *M. leprae* shows a high content of decaying pseudogenes

eResearch Activity 9

How Do Mobile Genomic Islands Exploit Incoming Plasmids for Their Ticket to New Hosts?

The genomic island 1 (SGI1) of the human pathogen *Salmonella* confers resistance to multiple antibiotics including beta-lactams, chloramphenicol, and sulfamethoxazole (refer to Sections 3.3 , 7.2 , 8.2 , 8.3 , and 27.2 in the printed book for examples of antibiotic cellular targets). SGI1 has been shown to excise itself from the chromosome and integrate into the chromosome of another *Salmonella* cell, providing this new cell with antibiotic resistance. Just how this genomic island moves between cells, and how these transfer events are triggered, are important topics for research in the effort to study and limit the spread of antibiotic resistance in human pathogens.

Prior studies provided valuable clues to the mechanism of SGI1 transfer between *Salmonella* cells. SGI1 carries genes whose products catalyze the excision from the host chromosome and integration into the new chromosome. SGI1 lacks its own conjugal pilus, but it hijacks the type IV secretion system (T4SS) produced by incoming IncC conjugative plasmids, in much the same way the mobilizable genomic island (MGI) of *Vibrio cholerae* hijacks the T4SS of the conjugative SXT element (see **Figure 9.31**). In fact, SGI1 produces proteins that bind to the T4SS of these plasmids to bias their transfer of SGI1 over their transfer of the plasmid.

Vincent Burrus (**Fig. ERA 9.1**) and colleagues at Université de Sherbrooke in Quebec, Canada, addressed several outstanding questions about SGI1 transfer. They asked whether SGI1 replicates once excised from the chromosome, how this replication affects the fate of the conjugative plasmid, and how the SGI1 plasmid senses the presence of an incoming conjugative plasmid to trigger its excision.

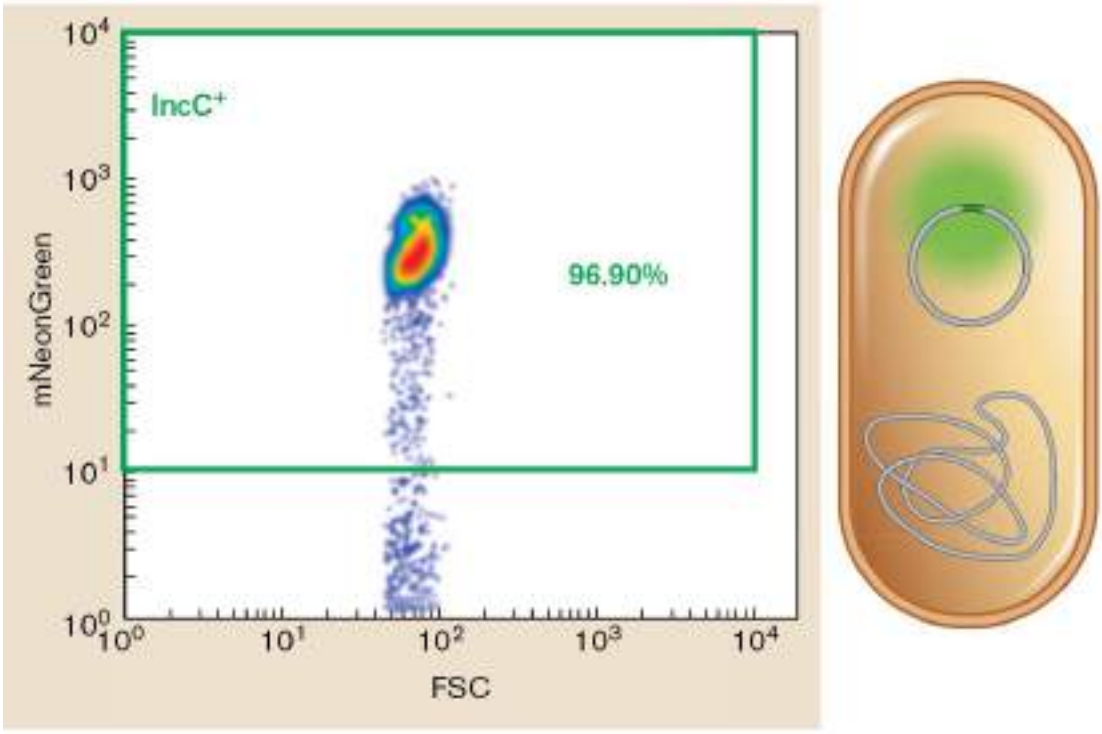


FIGURE ERA 9.1 ■ Vincent Burrus at the Université de Sherbrooke.

VINCENT BURRUS

Central to the researchers' investigation was the ability to track separately the SGI1 element and the plasmid. They engineered the SGI1 to express the red fluorescent protein mCherry and the IncC plasmid, called pVCR94, to express the green fluorescent protein mNeonGreen. Cells containing SGI1 or pVCR94 could be detected by flow cytometry (see Section 4.3 in the printed book), with forward scatter (FSC) used to detect the cell particles and red or green emission filters to detect the mCherry or mNeonGreen proteins within the cell. Almost every cell (>96%) containing one of these constructs had enough fluorescent protein to be detected in this manner (**Fig. ERA 9.2**). Importantly, cells could now be identified as having one of four states: containing only SGI1, containing only pVCR94, containing both elements, or containing neither element.

A.


IncC+
96.90%
mNeonGreen
FSC
10^0
10^1
10^2
10^3
10^4

B.

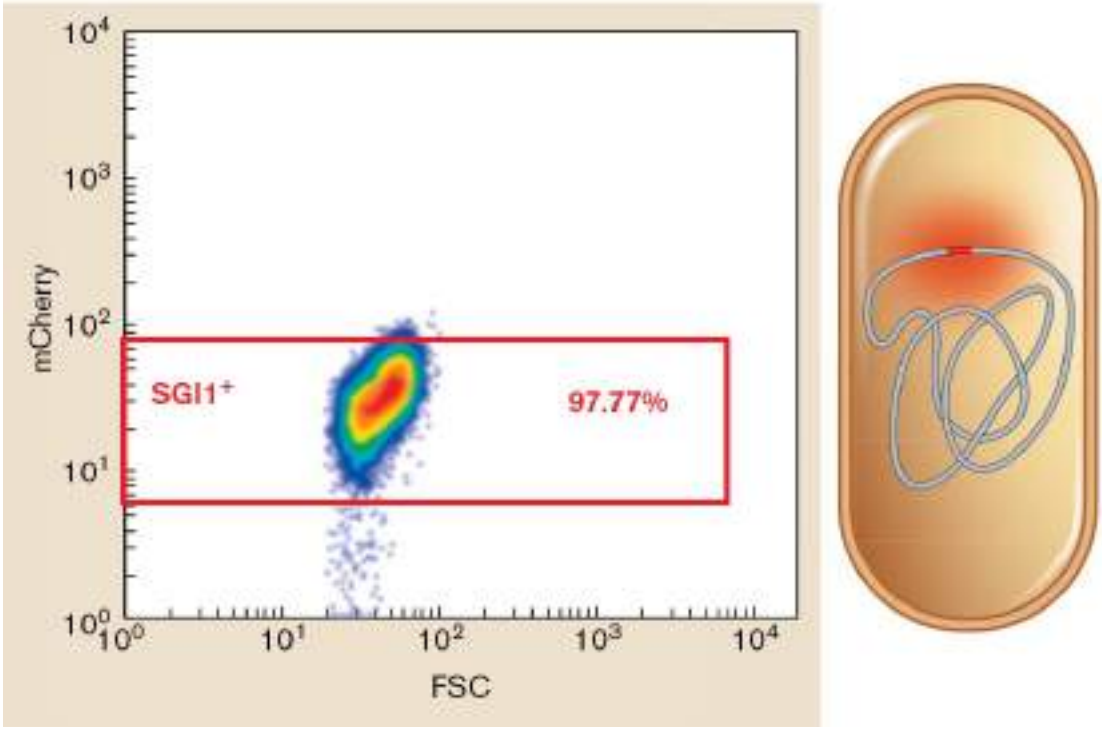

SGI1+
97.77%
mCherry
FSC
10^0
10^1
10^2
10^3
10^4

FIGURE ERA 9.2 ■ Detection of the IncC and SGI1 elements by flow cytometry. The gene encoding mNeonGreen was placed in the IncC plasmid (**A** , left), and the gene encoding mCherry was placed in the SGI1 element on the chromosome of *Salmonella* (**B** , left). Flow cytometry (**A** and **B** , right) detected the *Salmonella* cells via their forward scatter (FSC; *x* -axis), and those cells containing IncC or SGI1 could be identified by the fluorescence emission from mNeonGreen or mCherry (*y* -axis). Colored rectangles indicate regions where particles are assessed as positive for the fluorescent marker (IncC $^{+}$ or SGI1 $^{+}$).

Cells initially containing both elements were placed in medium that contained antibiotics that required the resistance markers from both SGI1 and pVC94 to be present. This established a culture that had each element in every cell. Then the antibiotics were removed, and the fates of the two elements were tracked over multiple generations of growth in culture. Initially, all cells had both elements, but by the 36th generation, almost every cell either had just one element or lacked both entirely (**Fig. ERA 9.3**). From these data, the researchers concluded that once the IncC plasmid enters an SGI1-positive cell, the coexistence of the two elements is not stable—and that cells will keep only one element, with a strong bias toward keeping SGI1 (**Fig. ERA 9.3**).

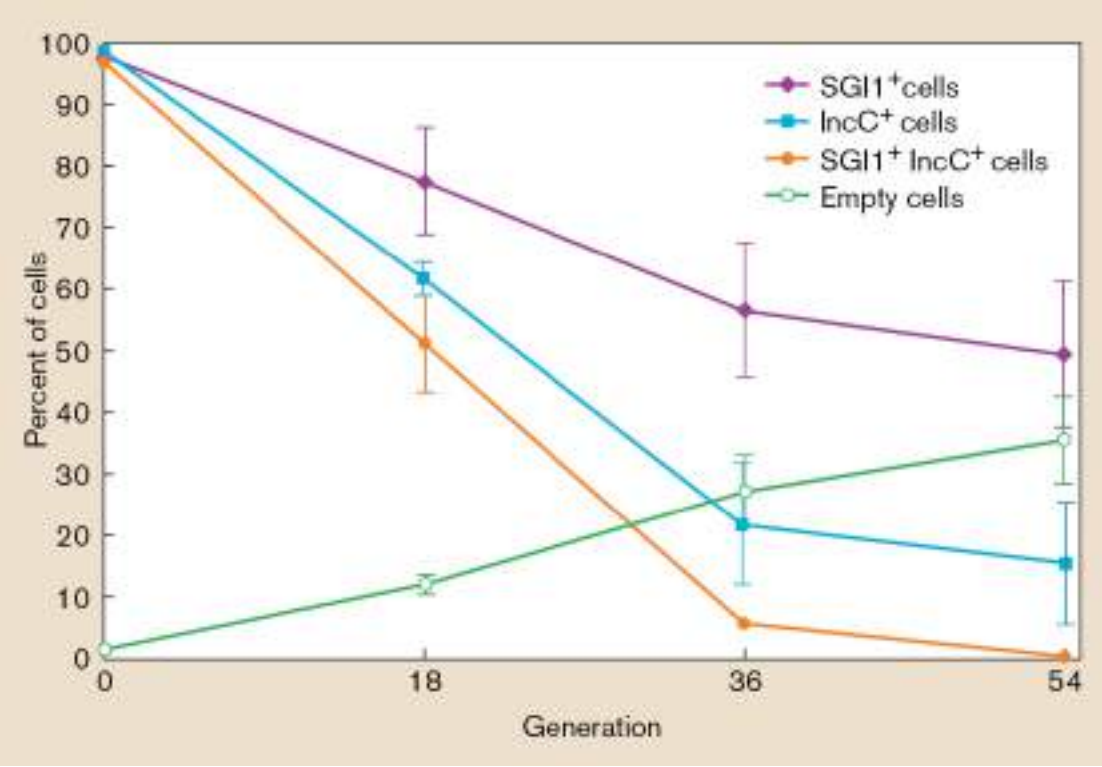


FIGURE ERA 9.3 ■ The IncC plasmid and SGI1 element are incompatible. Flow cytometry monitored the maintenance of IncC and SGI1 within cells of *Salmonella* over 54 generations in the absence of antibiotics. Note that cells categorized as IncC $^{+}$ or SGI $^{+}$ were also categorized as SGI1 $^{+}$ IncC $^{+}$ if the cells contained both elements. By generation 36, most cells lost IncC or SGI1 (<10% of the cells were still in the SGI $^{+}$ IncC $^{+}$ class), and some of these cells lost both elements ("Empty cells").

Could this instability be due to SGI1 element interference with the IncC plasmid, after SGI1 excision from the chromosome? To answer this question, the researchers first examined a mutant of SGI1 defective in its ability to excise. This mutant lacks the *int* sequence of SGI1 recognized by the integrase responsible for excision from the chromosome. As seen in **Figure ERA 9.4** , cells of the Δ *int* mutant retained both the SGI1 element and the IncC plasmid, demonstrating that excision of SGI1 was essential for the loss of IncC.

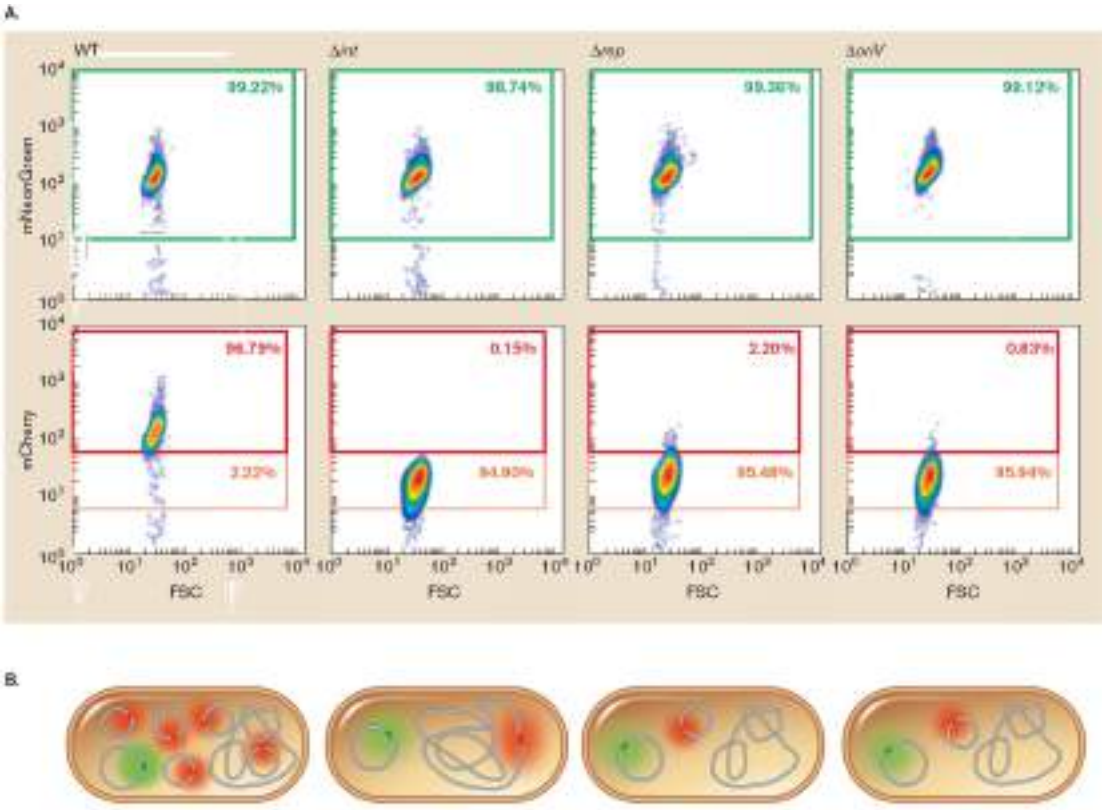


FIGURE ERA 9.4 ■ SGI1 replicates after excision from the chromosome. A. Flow cytometry plots of the IncC plasmid (top) and SGI1 element (bottom). **B.** Interpretations of the plot data. SGI1 elements in the chromosome were wild type (WT) or they were deleted for the excision/integration sequence (Δ *int*), the gene encoding the Rep protein for replication initiation (Δ *rep*), or the origin of replication for the excised SGI1 (Δ *oriV*). While the IncC plasmids were unchanged, each of the tested deletions resulted in a drop in mCherry fluorescence, indicating fewer copies per cell for the SGI1 element.

Genomic analysis of the SGI1 element revealed a potential system for replication post-excision: Homologs to the Rep proteins of plasmids and an origin of replication (*oriV*) were present within SGI1. To assess their role in replicating SGI1, mutants lacking *rep* or *oriV* were observed by flow cytometry. Compared to wild type, cells lacking *rep* or *oriV* had much lower red fluorescence—in fact, as low

as cells that lack *int* (**Fig. ERA 9.4**). Because fluorescence intensity is proportional to copy number of the mCherry-expressing SGI1, this result provided strong evidence that the *rep* and *oriV* genes of SGI1 were driving the replication of SGI1 after excision.

Finally, the researchers asked whether replication after excision was essential for the loss of the IncC plasmid. They repeated the experiment of incubating cells that contained both IncC and SGI1, but this time with SGI1 mutants lacking *int* , *rep* , or *oriV* . The retention of IncC in cells lacking *int* , *rep* , or *oriV* showed that, indeed, excision and subsequent replication of SGI1 was critical to the IncC loss (**Fig. ERA 9.5**).

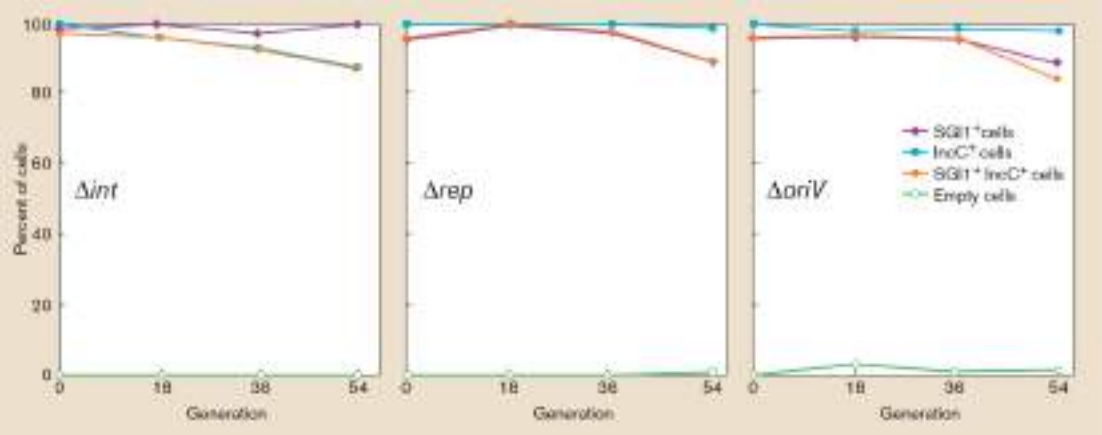


FIGURE ERA 9.5 ■ Excision and replication of SGI1 are necessary for exclusion of IncC. The same experiment of FIGURE ERA 9.3 was performed with SGI1 mutants lacking *int* , *rep* , or *oriV* . Note that for each of these mutants, both IncC and SGI were retained by almost every cell over the 54-generation incubation.

Altogether, the results from this and prior studies demonstrate just how crafty the SGI1 element is in exploiting incoming conjugal plasmids for their own transfer to new cells. Replication of SGI1 interferes with IncC plasmid persistence, but before IncC is lost, it produces the T4SS machinery for conjugal transport. Meanwhile, SGI1 makes proteins that bias the transfer of itself instead of the

IncC plasmid, a bias that is increased by the replication of SGI1 into multiple copies.

Further Exploration

Propose a research experiment to explore the nature of IncC instability in the presence of SGI1, considering active processes such as plasmid replication inhibition or plasmid destruction, as well as competition for replication or partitioning machinery.

Huguet, Kévin T., Nicolas Rivard, Daniel Garneau, Jason Palanee, and Vincent Burrus. 2020. Replication of the *Salmonella* genomic island 1 (SGI1) triggered by helper IncC conjugative plasmids promotes incompatibility and plasmid loss. *PLoS Genetics* **16** :e1008965.

Glossary

Figure 9.31

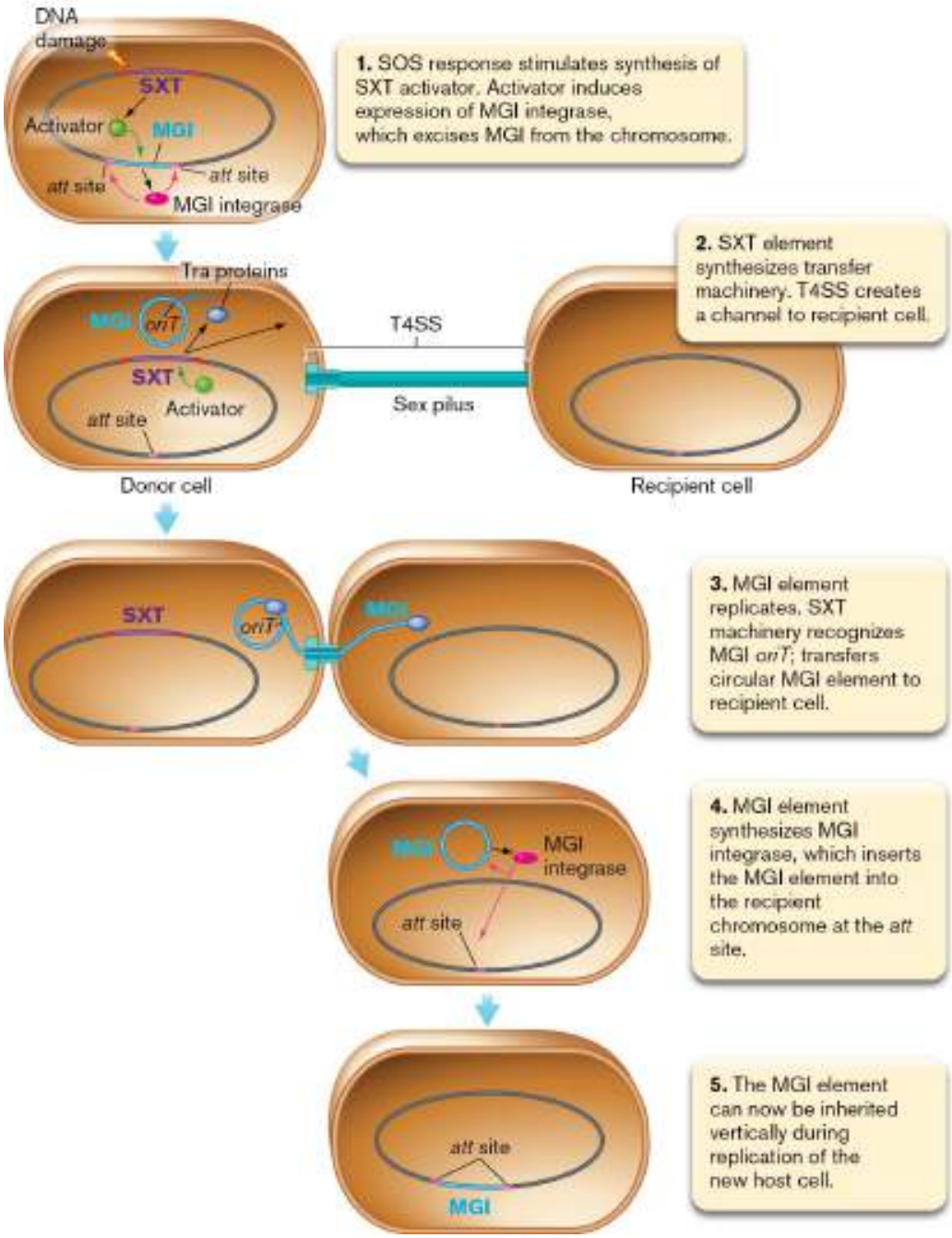


FIGURE 9.31 ■ Activation and transfer of a mobilizable genome island (MGI) by the SXT conjugative transposon.

CHAPTER REVIEW

Review Questions

1. List and explain the different types of mutations.
2. How are genotype and phenotype different? How are they related?
3. Describe several different DNA repair mechanisms. Which ones contribute to mutations?
4. What is the value of recombination to a species?
5. What are the basic ways microorganisms exchange DNA?
6. What is an F factor, and how does it (and other factors like it) contribute to gene exchange?
7. What does microbial gene exchange have to do with the plant disease called crown gall disease?
8. Compare specialized versus generalized transduction.
9. Describe transformation and how it occurs.
10. Explain the basic process of transposition. Why are insertion sequences always flanked by direct repeats of host DNA? How are transposons different from plasmids?
11. Explain the distinction between an ortholog and a paralog. How do they differ in their evolutionary history?
12. Once the DNA sequence of a gene is known, what specific methods can be used to gain clues as to the possible function of the gene product?
13. How can bioinformatics predict a metabolic pathway for an organism that cannot be grown in the laboratory?

14. Discuss horizontal versus vertical gene transfer in the context of genome evolution.
15. What are pathogenicity, metabolic, and resistance islands? What are their characteristics?
16. What is genome reduction, and why might this evolutionary strategy be tolerated or even advantageous for intracellular or free-living microbes?

Thought Questions

1. How would you use transposon mutagenesis to identify genes involved in the repair of UV-damaged DNA? Would this be a genetic screen or a selection?
2. What genetic features would allow you to determine whether a chromosome region contains a conjugative transposon versus a mobilizable genomic island?
3. *Agrobacterium tumefaciens* Ti plasmid has been used as a tool to genetically modify plants. Why would a plant biologist use a tumor-causing plasmid to breed new plants? Wouldn't the genetically altered plant develop a tumor?
4. Though we can identify sets of genes that have been horizontally transferred from one species of microbe to another, rarely can we identify the source species. What might account for this failure?
5. You have just isolated a new temperate bacteriophage for *Salmonella enterica* . How can you determine whether this phage mediates generalized or specialized transduction?
6. Is it possible for a microbe to be too good at repairing mutations? What might the trade-off be?

Key Terms

- annotation (357)
- AP site (336)
- apurinic site (332)
- auxotrophic (333)
- base excision repair (BER) (336)
- bioinformatics (357)
- conjugation (341)
- conjugative transposon (352)
- deletion (328)
- duplication (329)
- error-prone repair (334)
- error-proof repair (334)
-
-
- F $^{-}$ cell (343)
-
-
- F $^{+}$ cell (343)
- fertility (F) factor (342)
- frameshift mutation (330)
- gain-of-function mutation (330)
- generalized transduction (346)
- genome reduction (364)
- genomic island (359)
- genotype (330)
- Hfr strain (345)
- homolog (356)
- homologous recombination (337)
- horizontal gene transfer (359)
- insertion (328)

- insertion sequence (IS) (350)
- inversion (329)
- knockout mutation (330)
- loss-of-function mutation (330)
- metabolic island (359)
- methyl-directed mismatch repair (335)
- missense mutation (329)
- mobilizable genomic island (MGI) (352)
- mutagen (332)
- mutation (328)
- mutator strain (335)
- nonhomologous end joining (NHEJ) (338)
- nonsense mutation (330)
- nucleotide excision repair (336)
- ortholog (356)
- paralog (356)
- pathogenicity island (359)
- phenotype (330)
- photoreactivation (336)
- point mutation (328)
- pseudogene (365)
- resistance island (359)
- reversion (329)
- screen assay (353)
- selection assay (353)
- silent mutation (329)
- SOS response (337)
- specialized transduction (346)
- symbiosis island (359)
- transduction (346)
- transformation (348)
- transition (328)
- transposable element (350)

- transposase (350)
- transposition (329 , 351)
- transposon (350)
- transversion (328)
- vertical gene transfer (359)

Recommended Reading

Ambur, O. H., J. Engelstadter, P. J. Johnsen, E. L. Miller, and D. E. Rozen. 2016. Steady at the wheel: Conservative sex and the benefits of bacterial transformation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **371** :20150528.

Arnold, B. J., I-T. Huang, and W. P. Hanage. 2022. Horizontal gene transfer and adaptive evolution in bacteria. *Nature Reviews. Microbiology* **20** :206–218.

Baidya, A. K., S. Bhattacharya, G. P. Dubey, G. Mamou, and S. Ben-Yehuda. 2017. Bacterial nanotubes: A conduit for intercellular molecular trade. *Current Opinion in Microbiology* **42** :1–6.

Bertrand, Claire, Annabelle Thibessard, Claude Bruand, Francois Lecointe, and Pierre Leblond. 2019. Bacterial NHEJ: A never ending story. *Molecular Microbiology* **111** :1139–1151.

Bondy-Denomy, Joseph, and Alan R. Davidson. 2014. To acquire or resist: The complex biological effects of CRISPR–Cas systems. *Trends in Microbiology* **22** :218–225.

Cabezón, Elena, Jorge Ripoll-Rozada, Alejandro Pena, Fernando de la Cruz, and Ignacio Arechaga. 2015. Towards an integrated model of bacterial conjugation. *FEMS Microbiology Reviews* **39** :81–95.

Chou, Seemay, Matthew D. Daugherty, S. Brook Peterson, Jacob Biboy, Youyun Yang, et al. 2015. Transferred

interbacterial antagonism genes augment eukaryotic innate immune function. *Nature* **518** :98–101.

Dini-Andreote, Francisco, Fernando D. Andreote, Welington L. Araujo, Jack T. Trevors, and Jan D. van Elsas. 2012. Bacterial genomes: Habitat specificity and uncharted organisms. *Microbial Ecology* **64** :1–7.

Hu, Bo, Pratick Khara, and Peter J. Christie. 2019. Structural bases for F plasmid conjugation and F pilus biogenesis in *Escherichia coli* . *Proceedings of the National Academy of Sciences USA* **116** :14222–14227.

Ilangovan, A., C. W. M. Kay, S. Roier, H. El Mkami, E. Salvadori, et al. 2017. Cryo-EM structure of a relaxase reveals the molecular basis of DNA unwinding during bacterial conjugation. *Cell* **169** :708–721.

Johnston, Calum, Bernard Martin, Gwennaele Fichant, Patrice Polard, and Jean-Pierre Claverys. 2014. Bacterial transformation: Distribution, shared mechanisms and divergent control. *Nature Reviews. Microbiology* **12** :181–196.

Lang, A. S., A. B. Westbye, and J. T. Beatty. 2017. The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annual Review of Virology* **4** :87–104.

Louwen, Rogier, Raymond H. J. Staals, Hubert P. Endtz, Peter van Baarlen, and John van der Oost. 2014. The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiology and Molecular Biology Reviews* **78** :74–88.

Million-Weaver, Samuel, Ariana Nakta Samadpour, and Houra Merrikha. 2015. Replication restart after replication-transcription conflicts requires RecA in *Bacillus subtilis* . *Journal of Bacteriology* **197** :2374–2382.

Polz, Martin F., Eric J. Alm, and Willian P. Hanage. 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics* **29** :170–175.

van der Veen, Stijn, and Christoph M. Tang. 2015. The BER necessities: The repair of DNA damage in human-adapted bacterial pathogens. *Nature Reviews. Microbiology* **13** :83–94.

Van Houten, Bennett, and Neil Kad. 2014. Investigation of bacterial nucleotide excision repair using single-molecule techniques. *DNA Repair* (Amsterdam) **20** :41–48.

Wagner, A., R. J. Whitaker, D. J. Krause, J. H. Heilers, M. van Wolferen, et al. 2017. Mechanisms of gene flow in archaea. *Nature Reviews. Microbiology* **15** :492–501.

Glossary

annotation
: The deciphering of genome sequences, including identification of genes and prediction of gene function.

AP site
: See **apurinic site** .

apurinic site
: A DNA site missing a purine base because the bond linking the base to the sugar has been hydrolyzed. Also called **AP site** .

auxotrophic
: Describing a mutant state in which the cell has lost the ability to synthesize a substance required for growth. An auxotroph has a nutritional requirement not shared by the parent.

base excision repair (BER)
: A DNA repair mechanism that cleaves damaged bases off the sugar-phosphate backbone. After endonuclease activity at the AP site, a new, correct DNA strand is synthesized complementary to the undamaged strand.

bioinformatics
: A discipline at the intersection of biology and computing that analyzes gene and protein sequence data.

conjugation

Horizontal gene transfer involving cell-to-cell contact. In bacteria, pili draw together the donor and recipient cell envelopes, and a protein complex transmits DNA across. In ciliated eukaryotes, a conjugation bridge forms between two cells connecting their cytoplasm, through which micronuclei are exchanged.

conjugative transposon

A transposon that can be transferred from one cell to another via conjugation.

deletion

The loss of nucleotides from a DNA sequence.

duplication

The production of a second copy of a sequence fragment on a DNA molecule, usually adjacent to the original copy.

error-prone repair

Low-accuracy DNA repair mechanisms that allow mutations.

error-proof repair

DNA repair mechanisms that minimize the occurrence of mutations.

F^- cell

The DNA recipient cell in bacterial conjugation.

F^+ cell

The DNA donor cell that transmits the fertility factor F^+ to an F^- cell during bacterial conjugation.

fertility factor (F factor)

A specific plasmid (transferred by an F^+ donor cell) that contains the genes needed for pilus formation and DNA export.

frameshift mutation

A gene mutation involving the insertion or deletion of nucleotides that cause a shift in the codon reading frame.

gain-of-function mutation

A mutation that enhances the activity or allows new activity of a gene product.

generalized transduction

A phage-mediated gene transfer process in which any donor gene can be transferred to a recipient cell.

genome reduction

The large-scale loss of genes through evolution.

genomic island

A region of DNA sequence whose properties indicate that it has been transferred from another genome. Genomic islands usually comprise a set of genes with shared function, such as pathogenicity or symbiosis support.

genotype

The genome sequence of an organism.

Hfr strain

A high-frequency recombination bacterial strain, caused by the presence of a chromosomally integrated F factor.

homolog *or* homologous gene

A gene derived from a common ancestral gene. Homologs may be orthologs or paralogs.

homologous recombination

The process by which two DNA molecules exchange arms by cutting and splicing their helix backbones. Exchange occurs between sequences that are identical or nearly identical, as the machinery requires complementary base pairing to exchange the DNA molecules.

horizontal gene transfer

Also called *lateral gene transfer* . The natural movement of genes from one genome into another, nonprogeny genome.

insertion

The addition of nucleotides to the middle of a DNA sequence.

insertion sequence (IS)

A simple transposable element consisting of a transposase gene flanked by short, inverted-repeat sequences that are the target of transposase.

inversion

A mutation in which a DNA fragment is flipped within a chromosome. It may allow or repress the transcription of a particular gene.

knockout mutation
: A mutation that completely eliminates the activity of a gene product.

loss-of-function mutation
: A mutation that eliminates or decreases the function of the gene product.

metabolic island
: A genomic island that contributes genes involved in metabolism.

methyl-directed mismatch repair
: A DNA repair system that fixes misincorporation of a nucleotide after DNA synthesis. The unmethylated daughter strand is corrected to complement the methylated parental strand.

missense mutation
: A point mutation that alters the sequence of a single codon, leading to a single amino acid substitution in a protein.

mobilizable genomic island (MGI)
: A genomic island that can be transferred to another cell via the machinery of a conjugative transposon.

mutagen
: A chemical that damages DNA and increases the rate of mutations.

mutation
: A heritable change in a DNA sequence.

mutator strain
: A strain of cells with a high mutation rate, usually due to a mutation in a DNA repair enzyme.

nonhomologous end joining (NHEJ)
: A pathway that repairs double-strand breaks in DNA by direct ligation without the need for large regions of homology.

nonsense mutation
: A mutation that changes an amino acid codon into a premature stop codon.

nucleotide excision repair (NER)
: A DNA repair mechanism that cuts out damaged DNA. New, correctly base-paired DNA is synthesized by DNA polymerase I.

ortholog *or* orthologous gene
: A gene present in more than one species that derived from a common ancestral gene and encodes the same function.

paralog *or* paralogous gene
: A gene that arises by gene duplication within a species and evolves to carry out a different function from that of the original gene.

pathogenicity island
: A type of genomic island in which the stretch of DNA contains virulence factors and may have been transferred from another genome.

phenotype
: The observable characteristics of an organism.

photoreactivation
: A light-induced, photolyase-catalyzed repair of pyrimidine dimers.

point mutation
: A change in a single nucleotide within a nucleic acid sequence.

resistance island
: A genomic island that contributes genes involved in antibiotic resistance.

reversion
: A mutation that changes a previous mutation back to its original state.

screen assay
: A genetic assay, used to identify mutants that have lost a particular function, in which all mutants can grow, but only the mutants involved in the function of interest show a phenotype different from the wild type.

selection assay
: A genetic assay, used to identify mutants that have lost a particular function, in which only the mutants involved in the function of interest can grow, while all other mutants and the wild-type parent die.

silent mutation

A mutation that does not change the amino acid sequence encoded by an open reading frame. The changed codon encodes the same amino acid as the original codon.

SOS response

A coordinated cellular response to extensive DNA damage. It includes error-prone repair.

specialized transduction

Also called *restricted transduction* . Transduction in which the phage can transfer only a short DNA sequence adjacent to a prophage (integrated phage genome) from a donor cell to the recipient cell.

symbiosis island

A type of genomic island in which the stretch of DNA expresses proteins that enable a symbiotic relationship with another organism.

transduction

The transfer of host genes between bacterial cells via a phage head coat.

transformation

The internalization of free DNA from the environment into bacterial cells.

transition

A point mutation in which a purine is replaced by a different purine or a pyrimidine is replaced by a different pyrimidine.

transposable element

Also called *jumping gene* . A segment of DNA that can move from one DNA region to another.

transposase

A transposable element–encoded enzyme that catalyzes the transfer of the transposable element from one DNA region to another.

transposition

The process of moving a transposable element from one DNA region to another.

transposon

A transposable DNA element that contains genes in addition to those required for transposition. Examples of additional genes include those that encode resistance to antibiotics.

transversion

A point mutation in which a purine is replaced by a pyrimidine or vice versa.

vertical gene transfer

The generational movement of genes from parent to offspring through reproduction.

pseudogene

A nonfunctional gene-like sequence that evolved by degenerative evolution.