

# Microbial Genomics and Chromosome Organization

- Prokaryotic Chromosome Organization
- Generation of full genome sequences
- Genomic Structure & Functional Genomics
- Genome size vs. No. of orfs
- Minimal genome concept
- Lessons from full genomes

**Table 7.1 Kinds of genetic elements**

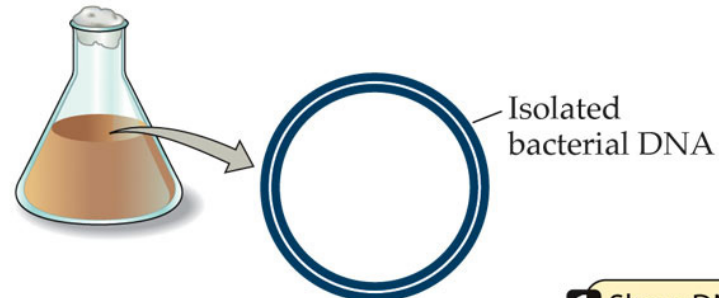
| <b>Organism</b>              | <b>Element</b>        | <b>Description</b>  |
|------------------------------|-----------------------|---|
| Prokaryote                   | Chromosome            | Extremely long, usually circular, double-stranded DNA molecule  |
|                              | Plasmid               | Typically a relatively short, usually circular, double-stranded DNA molecule, which is extrachromosomal |
| Eukaryote                    | Chromosome            | Extremely long, linear, double-stranded DNA molecule  |
|                              | Plasmid <sup>a</sup>  | Typically a relatively short circular or linear double-stranded DNA molecule, which is extrachromosomal |
| All Organisms                | Transposable elements | Double-stranded DNA molecule always found within another DNA molecule                                   |
| Mitochondrion or chloroplast | Chromosome            | Intermediate-length DNA molecules, usually circular   |
| Virus                        | Genome                | Single- or double-stranded DNA or RNA molecule  |

## Chromosomal & nonchromosomal genetic elements

<sup>a</sup>Plasmids are uncommon in eukaryotes.

# Whole-genome shotgun sequencing

## (A) Construction of DNA library

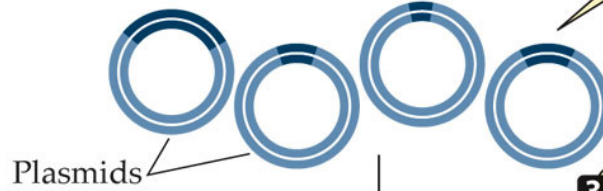


**1** Shear DNA into fragments.



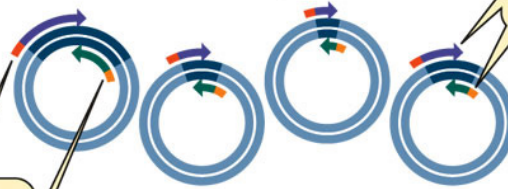
**2** Insert fragments and produce library.

## (B) Random sequencing

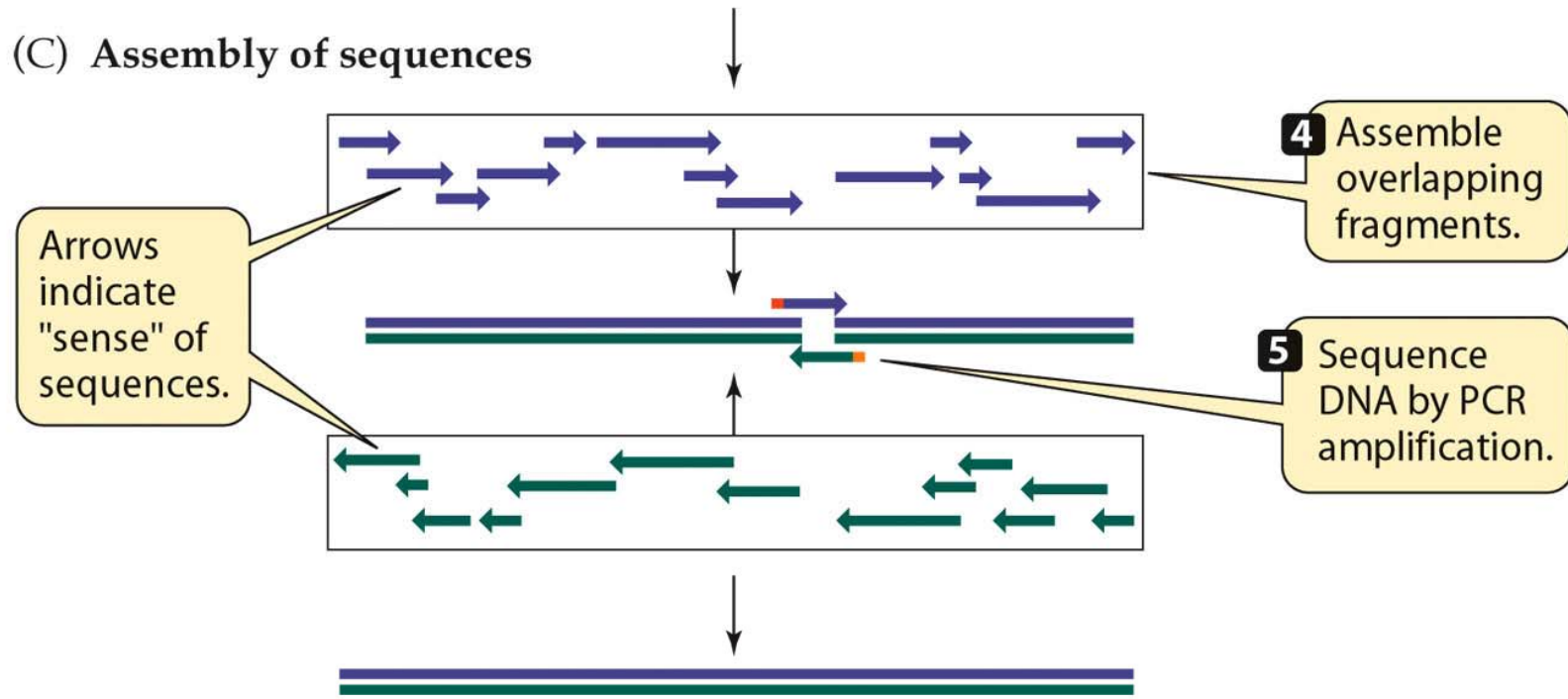


**3** Sequence DNA fragments.

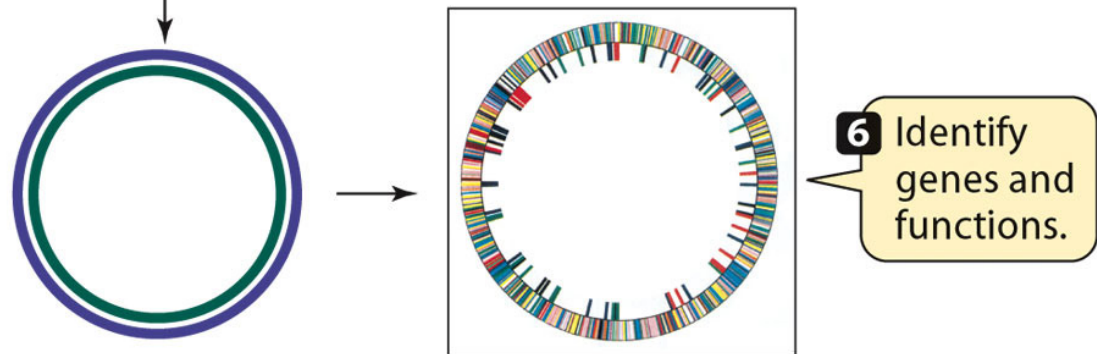
Oligonucleotides act as primers.



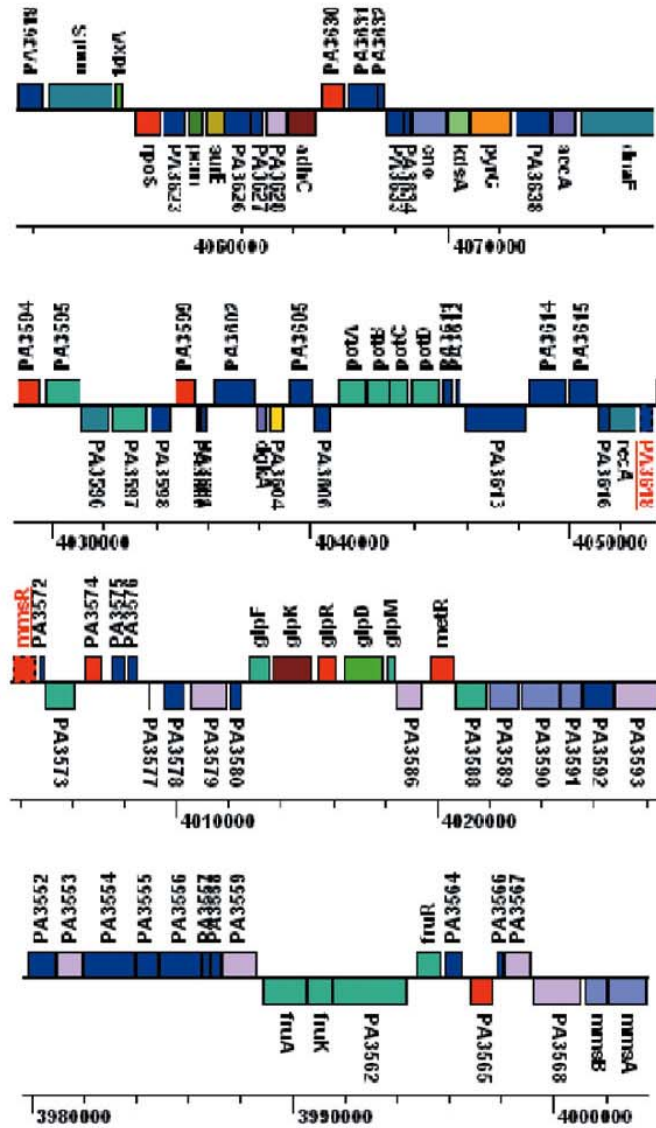
(C) Assembly of sequences

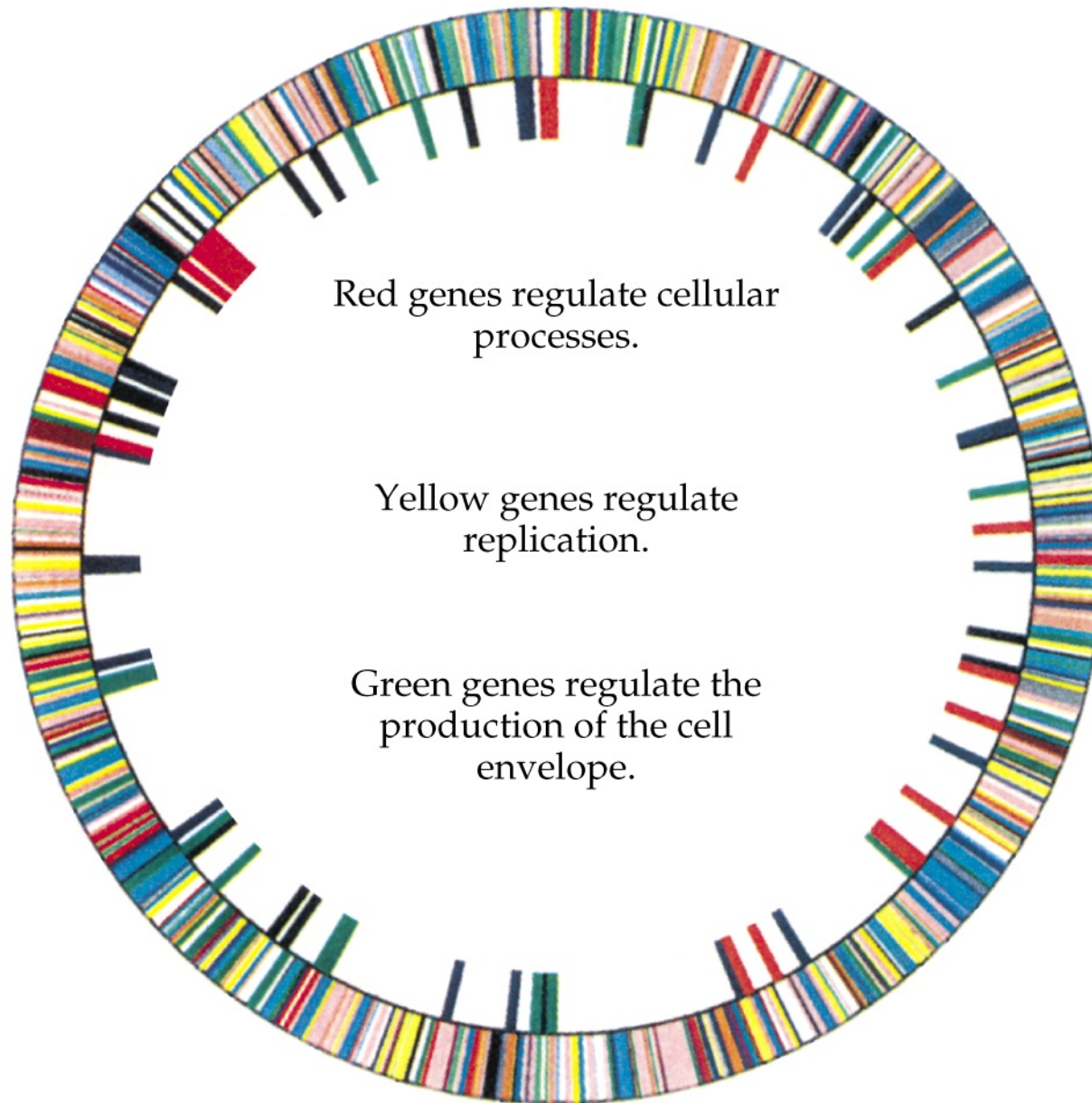


(D) Annotation



# Genes in a portion of bacterial genome



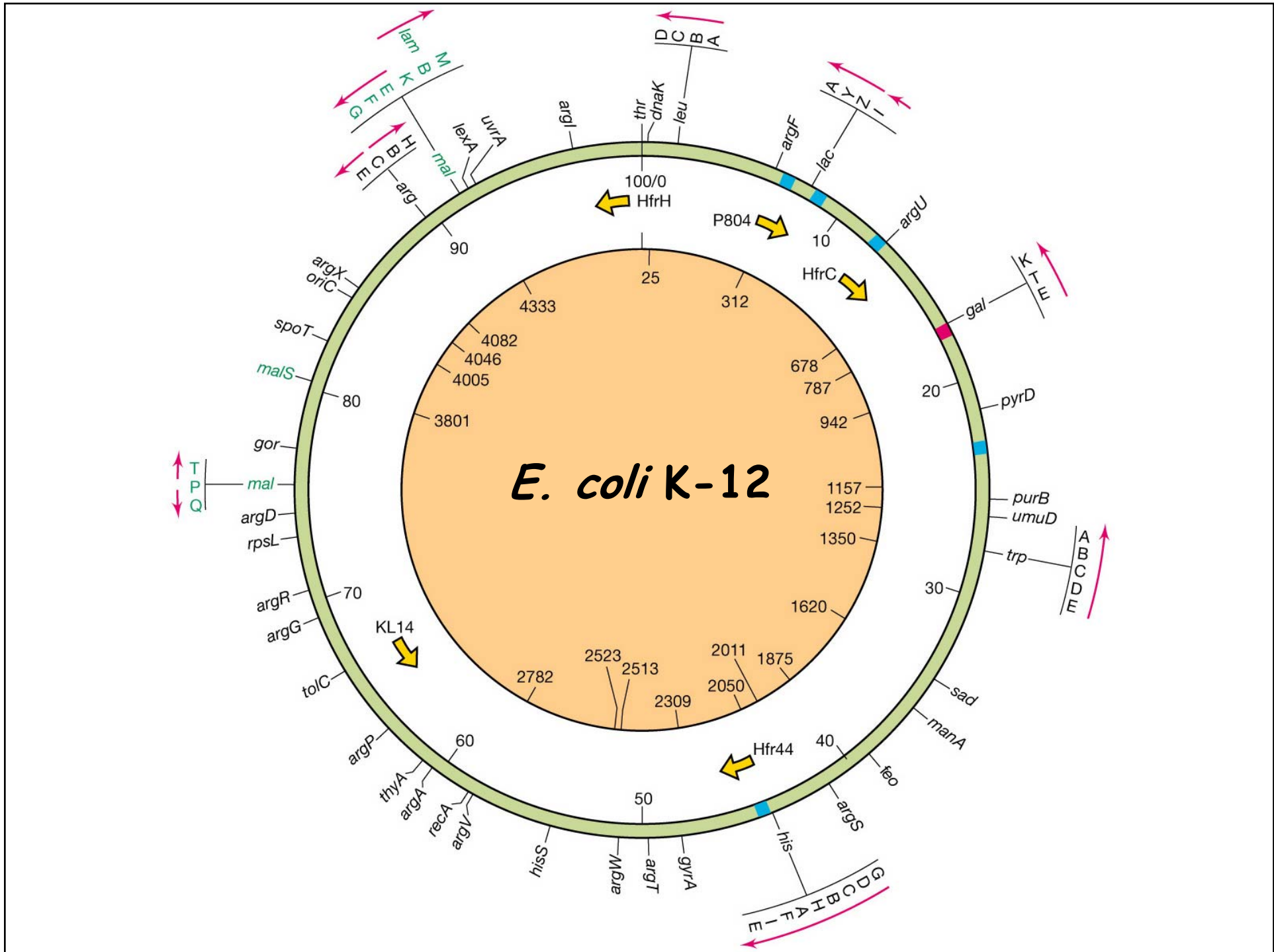


**Functional Organization of *H. influenzae*: 1.8 Megabases**



**Table 16.2****Comparison of regulatory genes in selected bacterial genomes**

| <b>Microorganism</b>              | <b># Genes in the Genome</b> | <b># Regulatory Proteins</b> | <b>% of Total</b> |
|-----------------------------------|------------------------------|------------------------------|-------------------|
| <i>Pseudomonas aeruginosa</i>     | 5570                         | 468                          | 8.4               |
| <i>Escherichia coli</i>           | 4289                         | 250                          | 5.8               |
| <i>Bacillus subtilis</i>          | 4100                         | 217                          | 5.3               |
| <i>Mycobacterium tuberculosis</i> | 3918                         | 117                          | 3.0               |
| <i>Helicobacter pylori</i>        | 1566                         | 18                           | 1.1               |





## *E. coli* K-12 Genome Organization Lessons

Cured of F-plasmid and Lambda phage: 4.6 Mbs

88% of genome comprised in 4200 orfs

1% tRNAs and rRNAs

0.5% noncoding repetitive sequences

10% regulatory sequences, including promoters, operators, origin and terminus of DNA replication

70% single copy genes

Some gene clusters, e.g., lac, trp, his, and NOT arg

~6% polycistronic mRNAs

## *E. coli* K-12 Genome Organization Lessons

Gene orientation can be in both directions, however, highly expressed genes in the same direction as DNA replication forks travel.

For example, all 7 *rrn* operons  
Rem: transcription always 5' to 3'

Also found were several different cryptic, defective prophages and IS elements.  
**18%** of genome from horizontal gene transfer, including large regions (~40Kb) known as **islands of pathogenicity**.

**Table 15.2****Gene function in bacterial genomes****Percentage of genes on  
chromosome in that category**

| <b>Functional categories</b> | <b><i>Escherichia coli</i><br/>(4.64 Mbp)<sup>a</sup></b> | <b><i>Haemophilus influenzae</i><br/>(1.83 Mbp)<sup>a</sup></b> | <b><i>Mycoplasma genitalium</i><br/>(0.58 Mbp)<sup>a</sup></b> |
|------------------------------|---|---|--|
| Metabolism                   | 21.0  | 19.0  | 14.6   |
| Structural                   | 5.5   | 4.7   | 3.6  |
| Transport                    | 10.0  | 7.0   | 7.3  |
| Regulation                   | 8.5   | 6.6   | 6.0  |
| Translation                  | 4.5   | 8.0   | 21.6   |
| Transcription                | 1.3   | 1.5   | 2.6  |
| Replication                  | 2.7   | 4.9   | 6.8  |
| Other, known                 | 8.5   | 5.2   | 5.8  |
| Unknown                      | 38.1  | 43.0  | 32.0   |

<sup>a</sup> Chromosome size. Each organism listed contains only a single circular chromosome.

# Genome size vs. ORFs

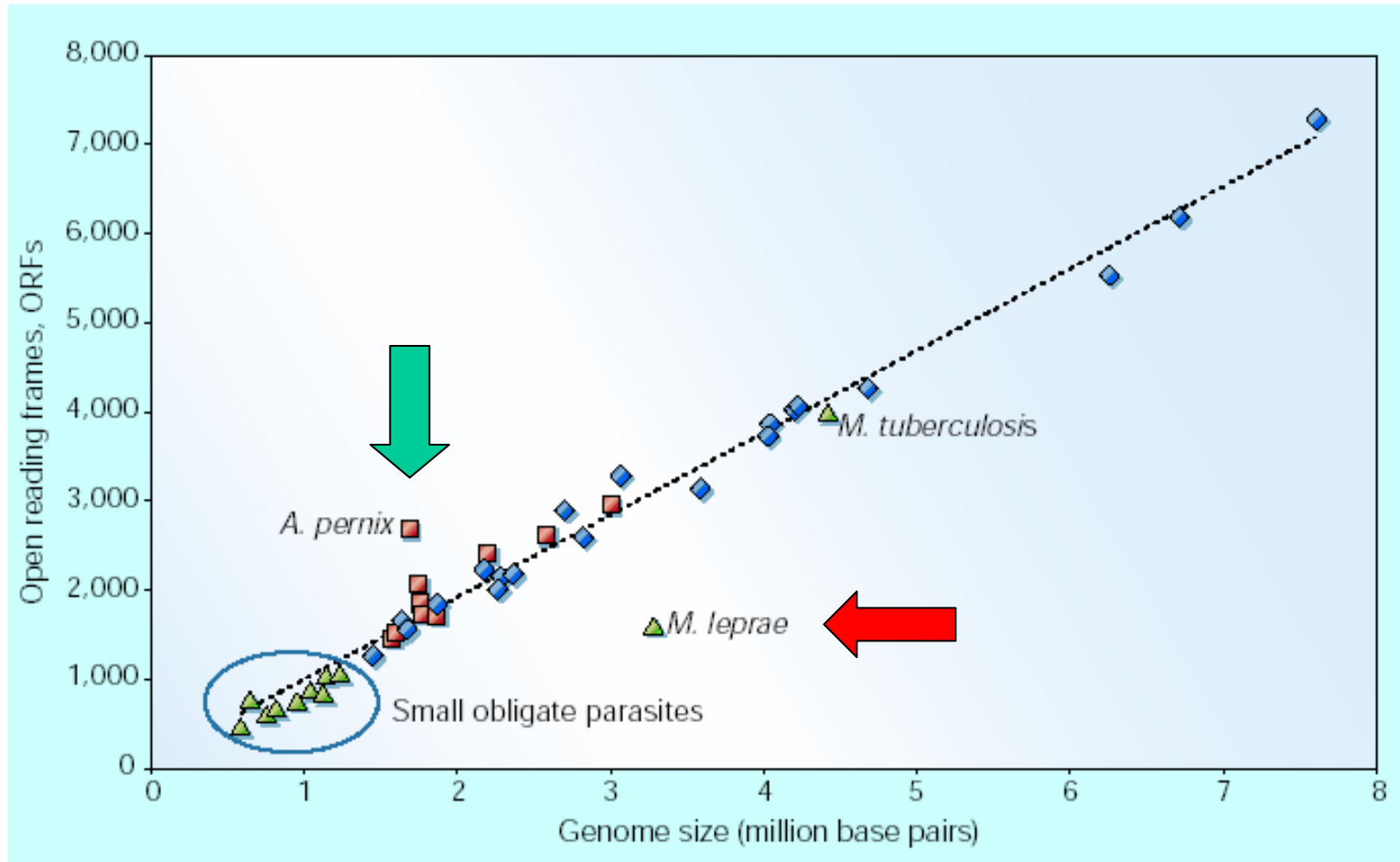


Figure 1 Number of genes (ORFs) plotted against genome size for 44 fully sequenced genomes, including ten Archaea (squares) and 34 Bacteria. Obligate bacterial parasites are denoted by triangles; all other bacteria are shown as diamonds. *Mycobacterium leprae* is a genome 'in decay' that has a large number of pseudogenes. The archaeon *Aeropyrum pernix* is unusual in having an excessive number of duplicated ORFs.

| Organism<br>(number of genes)          | Glycolysis | Tricarboxylic-<br>acid cycle | Amino-acid<br>biosynthesis | Purine<br>biosynthesis | Pyrimidine<br>biosynthesis | Ancestral stock             |
|--|------------|------------------------------|----------------------------|------------------------|----------------------------|-----------------------------|
| <i>Mycoplasma genitalium</i><br>(470)  | +          | -                            | -                          | -                      | -                          | <i>Bacillus-Clostridium</i> |
| <i>Buchnera</i> species<br>(588)       | +          | -                            | +                          | +                      | +                          | Gamma-<br>proteobacteria    |
| <i>Rickettsia prowazekii</i><br>(834)  | -          | +                            | -                          | -                      | -                          | Alpha-<br>proteobacteria    |
| <i>Chlamydia trachomatis</i><br>(894)  | +          | -                            | +                          | -                      | -                          | Main line                   |
| <i>Treponema pallidum</i><br>(1,041)   | +          | -                            | -                          | -                      | -                          | Main line                   |
| <i>Mycobacterium leprae</i><br>(1,604) | Partial    | In decay                     | +                          | +                      | +                          | <i>Bacillus-Clostridium</i> |

Figure 2 Many routes to intracellular adaptation. The differing presence (+) or absence (-) of certain metabolic pathways in the streamlined genomes of parasitic bacteria shows how variable the process may be.



# Global Transposon Mutagenesis and a Minimal Mycoplasma Genome

Clyde A. Hutchison III,<sup>1,2\*</sup> Scott N. Peterson,<sup>1\*†</sup> Steven R. Gill,<sup>1</sup>  
Robin T. Cline,<sup>1</sup> Owen White,<sup>1</sup> Claire M. Fraser,<sup>1</sup>  
Hamilton O. Smith,<sup>1‡</sup> J. Craig Venter<sup>1‡§</sup>

*Mycoplasma genitalium* with 517 genes has the smallest gene complement of any independently replicating cell so far identified. Global transposon mutagenesis was used to identify nonessential genes in an effort to learn whether the naturally occurring gene complement is a true minimal genome under laboratory growth conditions. The positions of 2209 transposon insertions in the completely sequenced genomes of *M. genitalium* and its close relative *M. pneumoniae* were determined by sequencing across the junction of the transposon and the genomic DNA. These junctions defined 1354 distinct sites of insertion that were not lethal. The analysis suggests that 265 to 350 of the 480 protein-coding genes of *M. genitalium* are essential under laboratory growth conditions, including about 100 genes of unknown function.

**265 to 350 genes are the minimum necessary genome**

---

# The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*

Gerard Deckert\*†, Patrick V. Warren\*†, Terry Gaasterland‡, William G. Young\*, Anna L. Lenox\*, David E. Graham§, Ross Overbeek‡, Marjory A. Snead\*, Martin Keller\*, Monette Aujay\*, Robert Huber||, Robert A. Feldman\*, Jay M. Short\*, Gary J. Olsen§ & Ronald V. Swanson\*

\* *Diversa Corporation, 10665 Sorrento Valley Road, San Diego, California 92121, USA*

‡ *Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*

§ *Department of Microbiology, University of Illinois, Urbana, Illinois 61801, USA*


|| *Lehrstuhl für Mikrobiologie, Universität Regensburg W-8400, Regensburg W-8400, Germany*

---

*Aquifex aeolicus* was one of the earliest diverging, and is one of the most thermophilic, bacteria known. It can grow on hydrogen, oxygen, carbon dioxide, and mineral salts. The complex metabolic machinery needed for *A. aeolicus* to function as a chemolithoautotroph (an organism which uses an inorganic carbon source for biosynthesis and an inorganic chemical energy source) is encoded within a genome that is only one-third the size of the *E. coli* genome. Metabolic flexibility seems to be reduced as a result of the limited genome size. The use of oxygen (albeit at very low concentrations) as an electron acceptor is allowed by the presence of a complex respiratory apparatus. Although this organism grows at 95 °C, the extreme thermal limit of the Bacteria, only a few specific indications of thermophily are apparent from the genome. Here we describe the complete genome sequence of 1,551,335 base pairs of this evolutionarily and physiologically interesting organism.



**Table 16.3****Distribution of genes of unknown function among selected bacterial genomes (Part 1)**

| Organism   | Genome Size (Mbp) | No. of ORFs (% coding) |       | Unknown Function |       | Unique ORFs |       |
|--|-------------------|------------------------|-------|------------------|-------|-------------|-------|
| <i>Aeropyrum pernix</i> K1   | 1.67              | 1,885                  | (89%) |                  |       |             |       |
|  <i>A. aeolicus</i> VF5 | 1.50              | 1,749                  | (93%) | 663              | (44%) | 407         | (27%) |
| <i>A. fulgidus</i>   | 2.18              | 2,437                  | (92%) | 1,315            | (54%) | 641         | (26%) |
| <i>B. subtilis</i>   | 4.20              | 4,779                  | (87%) | 1,722            | (42%) | 1,053       | (26%) |
| <i>B. burgdorferi</i>  | 1.44              | 1,738                  | (88%) | 1,132            | (65%) | 682         | (39%) |
| <i>Chlamydia pneumoniae</i> AR39   | 1.23              | 1,134                  | (90%) | 543              | (48%) | 262         | (23%) |
| <i>Chlamydia trachomatis</i> MoP <sub>n</sub>  | 1.07              | 936                    | (91%) | 353              | (38%) | 77          | (8%)  |
| <i>C. trachomatis</i> serovar D  | 1.04              | 928                    | (92%) | 290              | (32%) | 255         | (29%) |
| <i>Deinococcus radiodurans</i>   | 3.28              | 3,187                  | (91%) | 1,715            | (54%) | 1,001       | (31%) |
| <i>E. coli</i> K-12-MG1655   | 4.60              | 5,295                  | (88%) | 1,632            | (38%) | 1,114       | (26%) |
| <i>H. influenzae</i>   | 1.83              | 1,738                  | (88%) | 595              | (35%) | 237         | (14%) |
| <i>H. pylori</i> 26695   | 1.66              | 1,589                  | (91%) | 744              | (45%) | 539         | (33%) |
| <i>Methanobacterium thermautotrophicum</i>   | 1.75              | 2,008                  | (90%) | 1,010            | (54%) | 496         | (27%) |

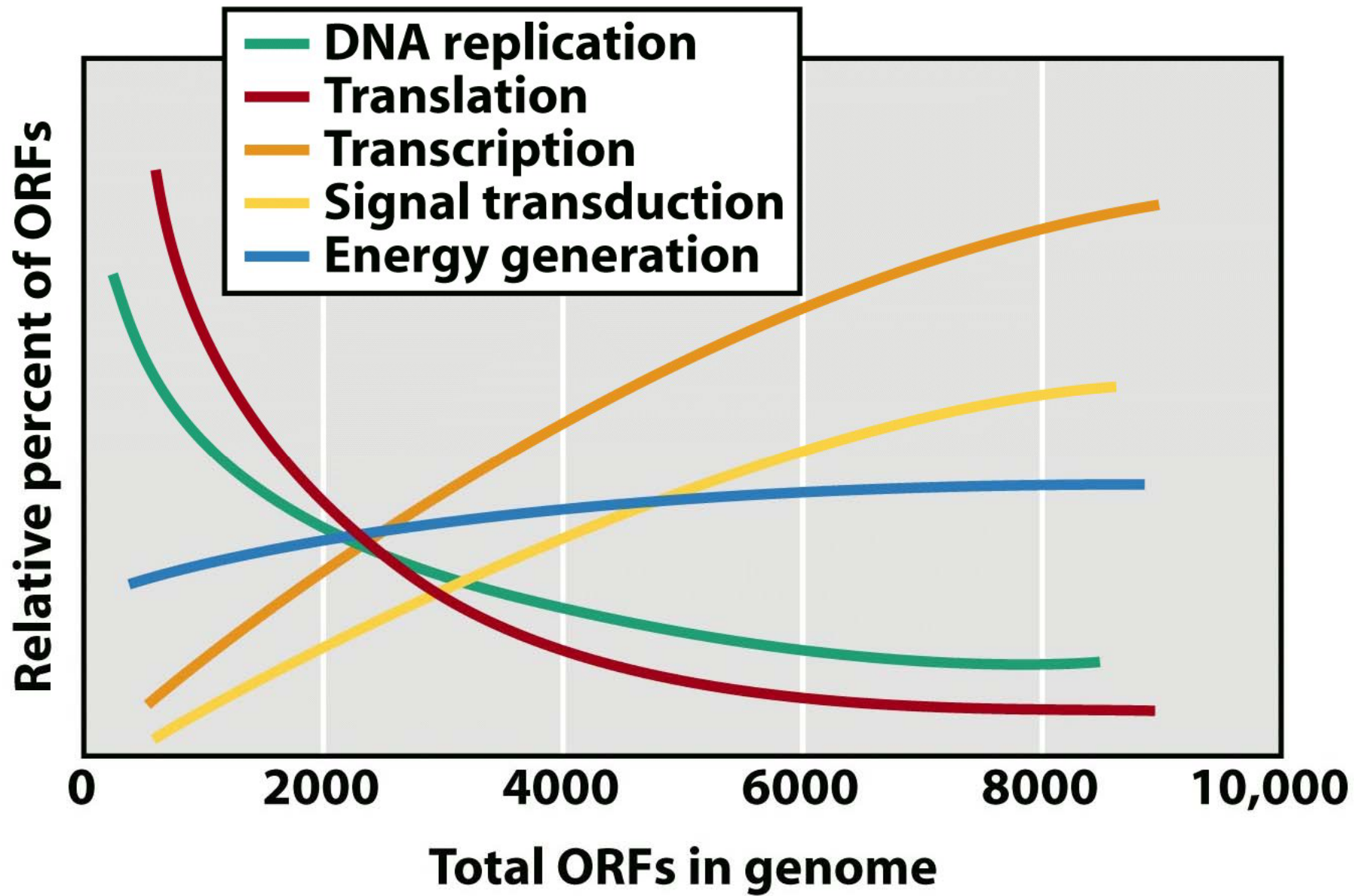


**Table 16.3****Distribution of genes of unknown function among selected bacterial genomes (Part 2)**

| Organism                              | Genome Size (Mbp) | No. of ORFs (% coding) |       | Unknown Function |       | Unique ORFs |       |
|---------------------------------------|-------------------|------------------------|-------|------------------|-------|-------------|-------|
| <i>Methanococcus jannaschii</i>       | 1.66              | 1,783                  | (87%) | 1,076            | (62%) | 525         | (30%) |
| <i>M. tuberculosis</i> CSU#93         | 4.41              | 4,275                  | (92%) | 1,521            | (39%) | 606         | (15%) |
| <i>M. genitalium</i>                  | 0.58              | 483                    | (91%) | 173              | (37%) | 7           | (2%)  |
| <i>M. pneumoniae</i>                  | 0.81              | 680                    | (89%) | 248              | (37%) | 67          | (10%) |
| <i>N. meningitidis</i> MC58           | 2.24              | 2,155                  | (83%) | 856              | (40%) | 517         | (24%) |
| <i>Pyrococcus horikoshii</i> OT3      | 1.74              | 1,994                  | (91%) | 589              | (42%) | 453         | (22%) |
| <i>Rickettsia prowazekii</i> Madrid E | 1.11              | 878                    | (75%) | 311              | (37%) | 209         | (25%) |
| <i>Synechocystis</i> sp.              | 3.57              | 4,003                  | (87%) | 2,384            | (75%) | 1,426       | (45%) |
| <i>T. maritima</i> MSB8               | 1.86              | 1,879                  | (95%) | 863              | (46%) | 373         | (26%) |
| <i>T. pallidum</i>                    | 1.14              | 1,039                  | (93%) | 461              | (44%) | 280         | (27%) |
| <i>Vibrio cholerae</i> El Tor N1696   | 4.03              | 3,890                  | (88%) | 1,806            | (46%) | 934         | (24%) |
| <b>Totals:</b>                        | 50.60             | 52,462                 | (89%) | 22,358           | (43%) | 12,161      | (23%) |

From Fraser et al., *Nature* 2000, vol. 406. p. 800.





## Lessons from full genomes

- Size range 600Kb to 12Mb
- Vast number of putative genes with no known function
- Pathogenicity conferred by "Pathogenicity Islands"  
44.5Kb in *Bacillus anthracis*
- Symbiotic Island of >600Kb in *Sinorhizobium loti*  
including genes for nodulation and N-fixation
- Adaptive gene losses in parasitic bacteria  
*Rickettsia* and *Chlamidia* are ATP thieves using  
the same "alien" ADP/ATP translocase
- Relative proportions of functional genes

# Microbial Genetic Exchange and Plasmids

- Microbial Genetic Exchange is unidirectional!

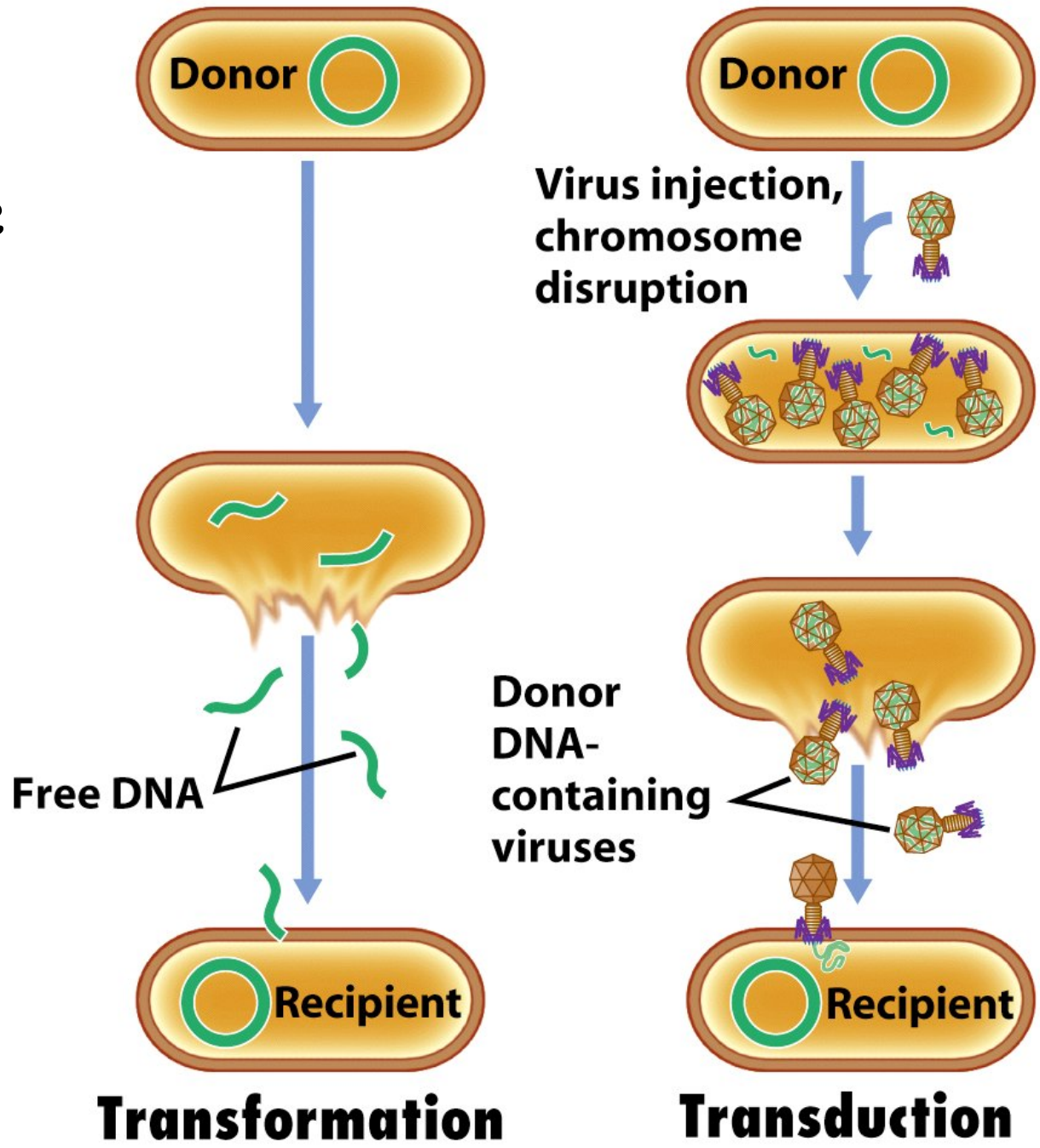
Transformation

Transduction

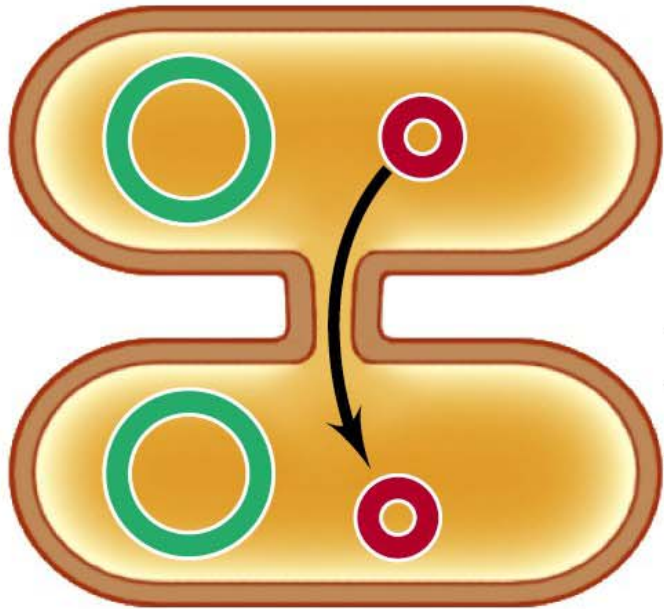
Conjugation

- Each requires Homologous Recombination
- Types of plasmids

# Microbial Genetic Exchange

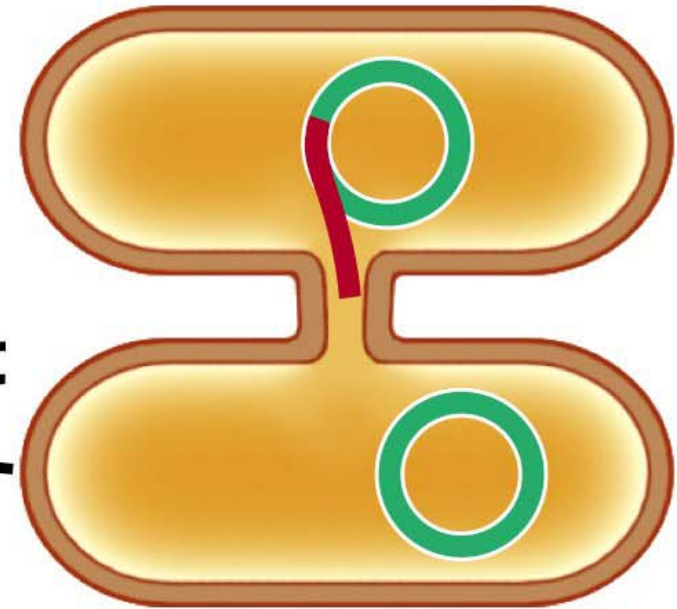


**Plasmid-containing donor cell**



**Conjugation:  
Plasmid  
transfer**

**Donor cell with integrated plasmid**

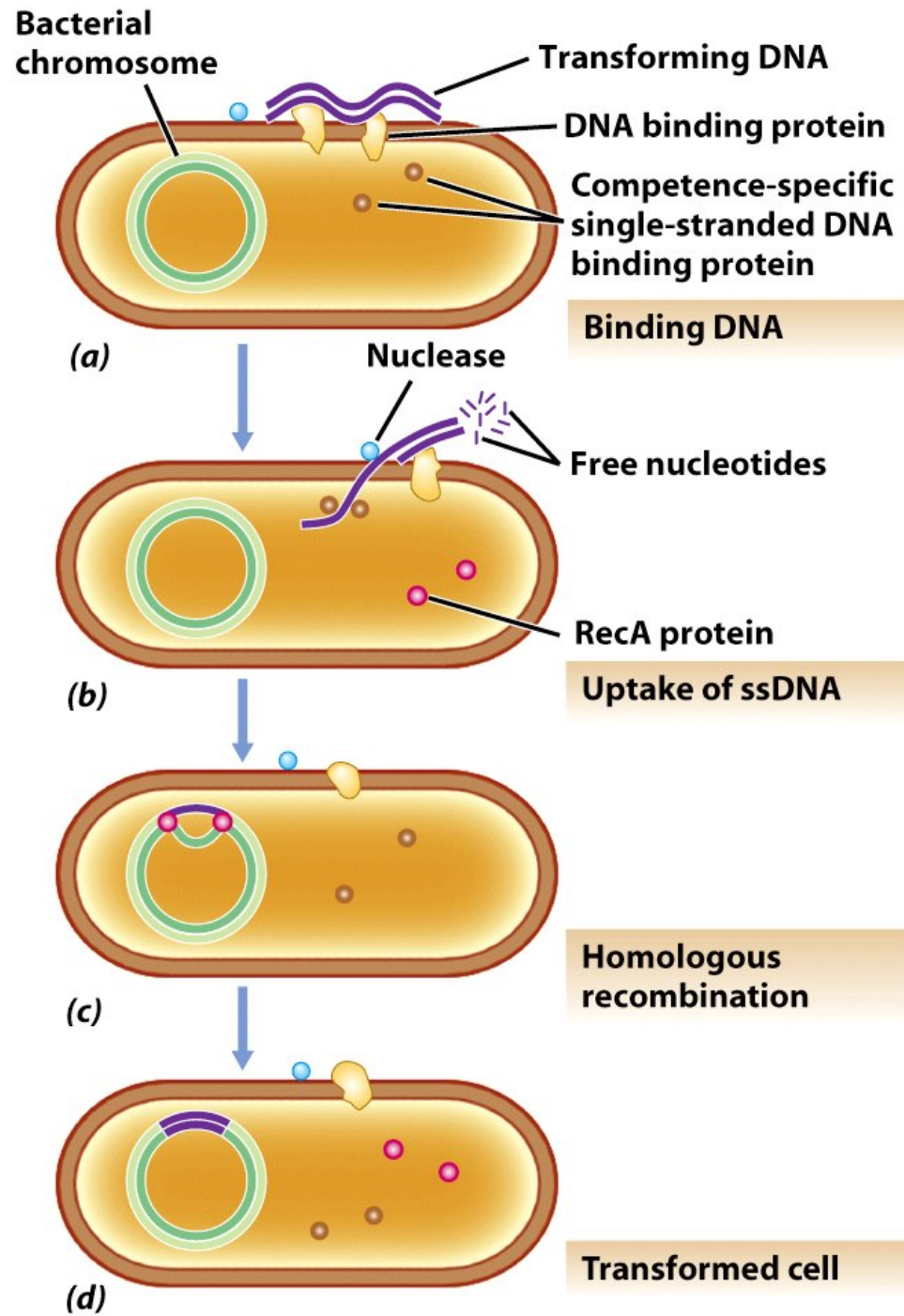


**Recipient  
cells**

**Conjugation:  
Chromosome  
transfer**



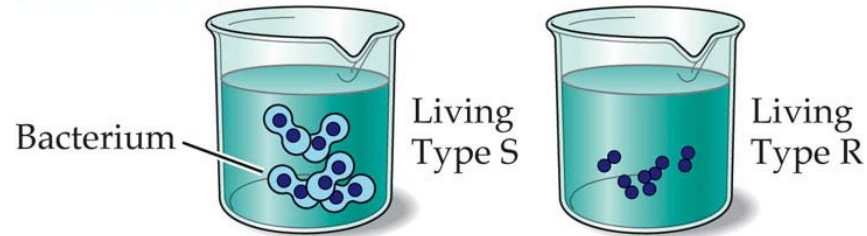
# Transformation by a Gram + competent cell



# Demonstration of transformation

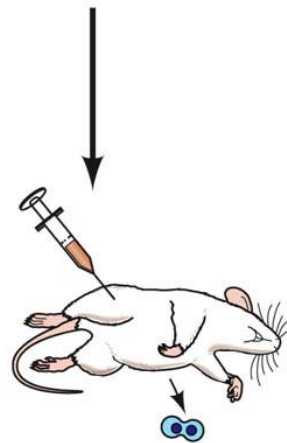
(A)

Treatments

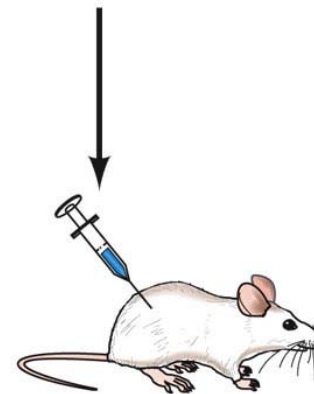


*Streptococcus pneumoniae*

Results



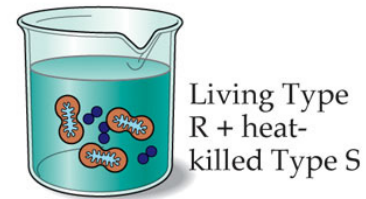
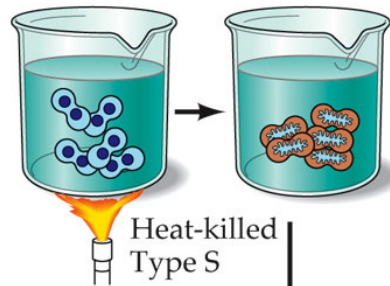
Living Type S  
bacteria recovered



No live bacteria  
recovered

# Demonstration of transformation

## Treatments



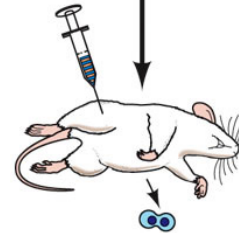
Living Type R + heat-killed Type S

## Transformation

## Results



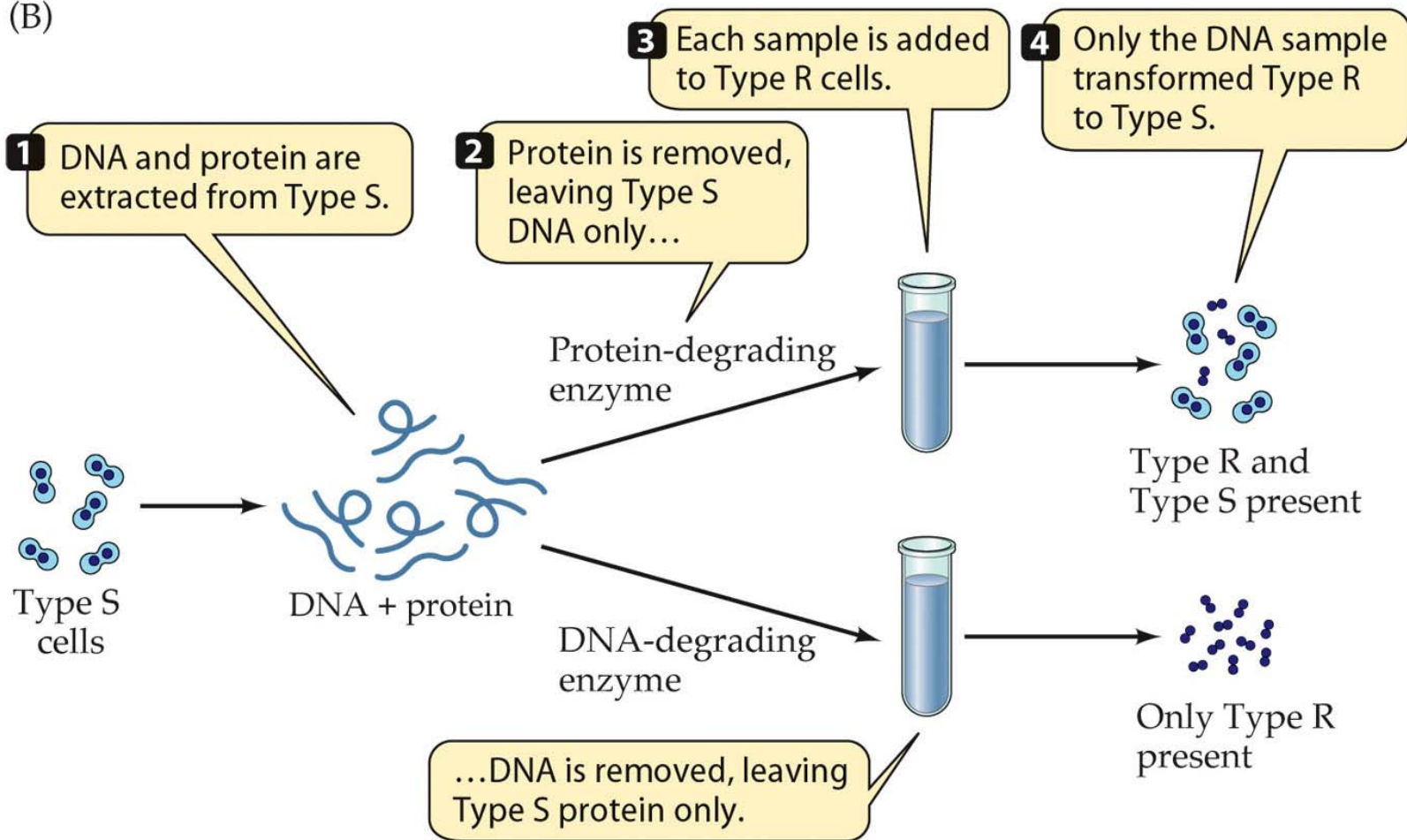
No live bacteria recovered



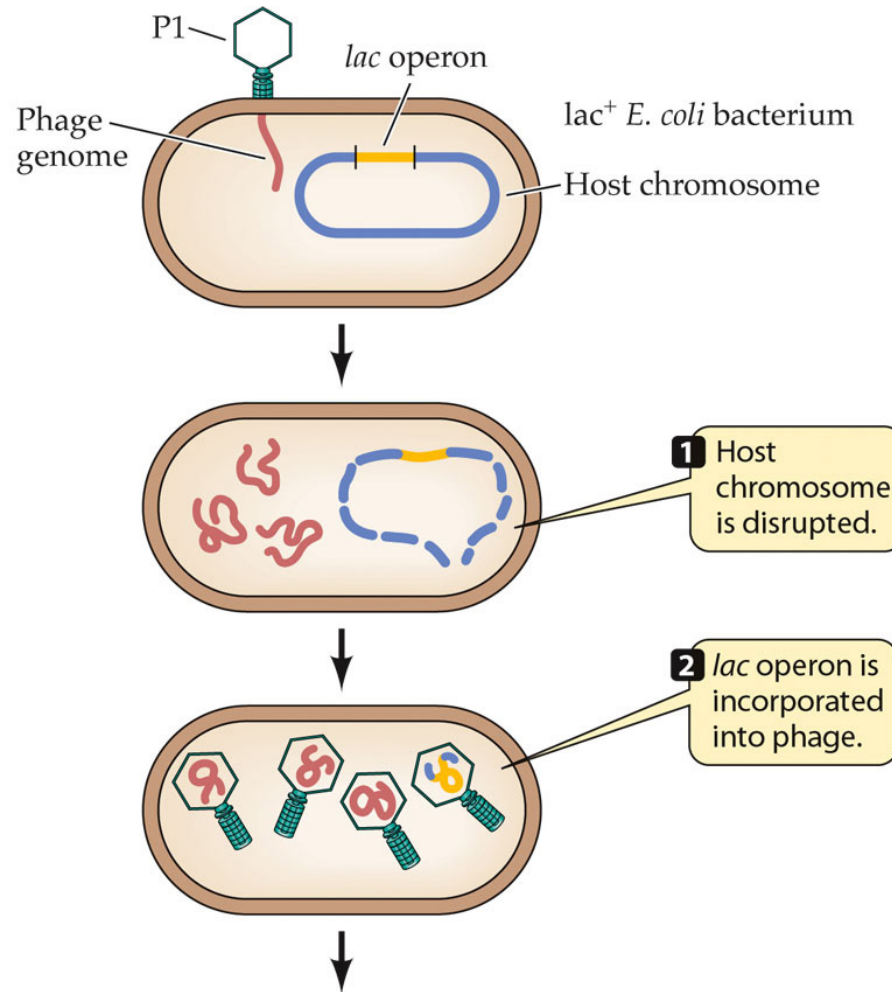
Living Type S bacteria recovered

Type R has been transformed to Type S.

(B)

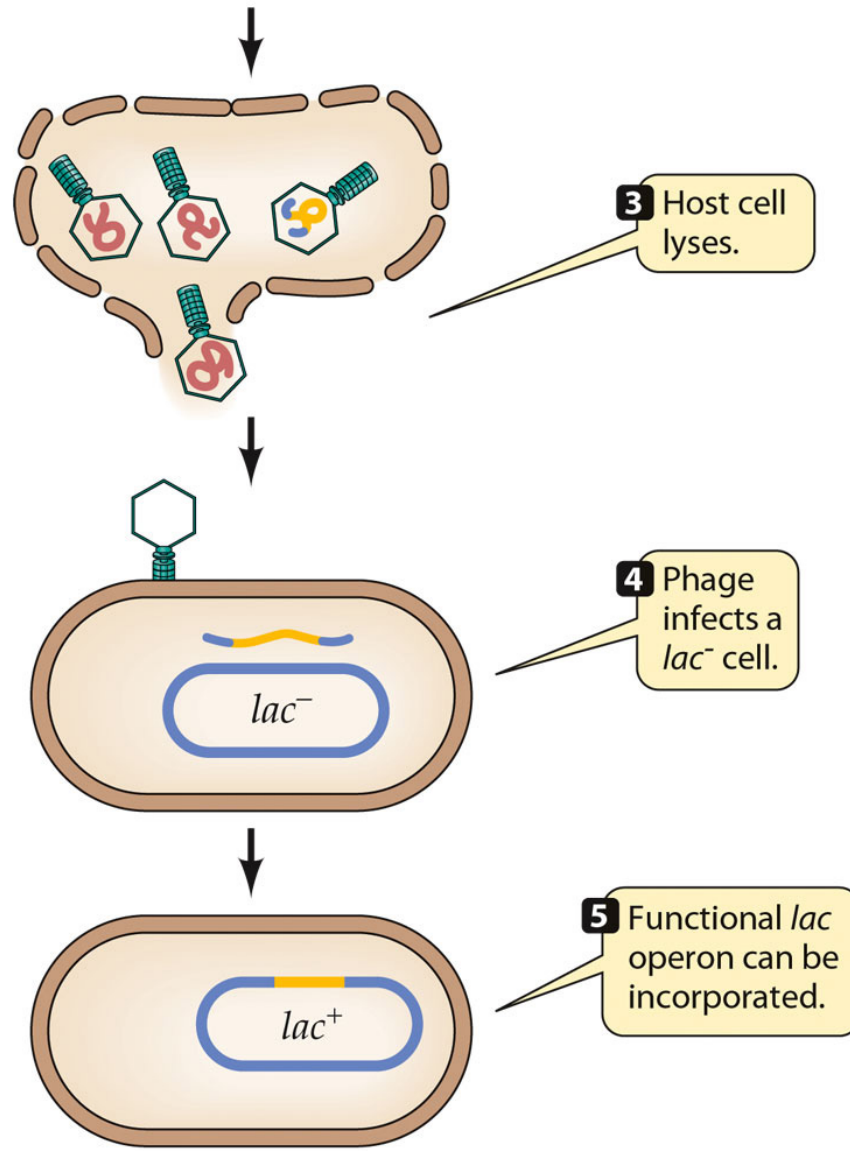


# Generalized transduction

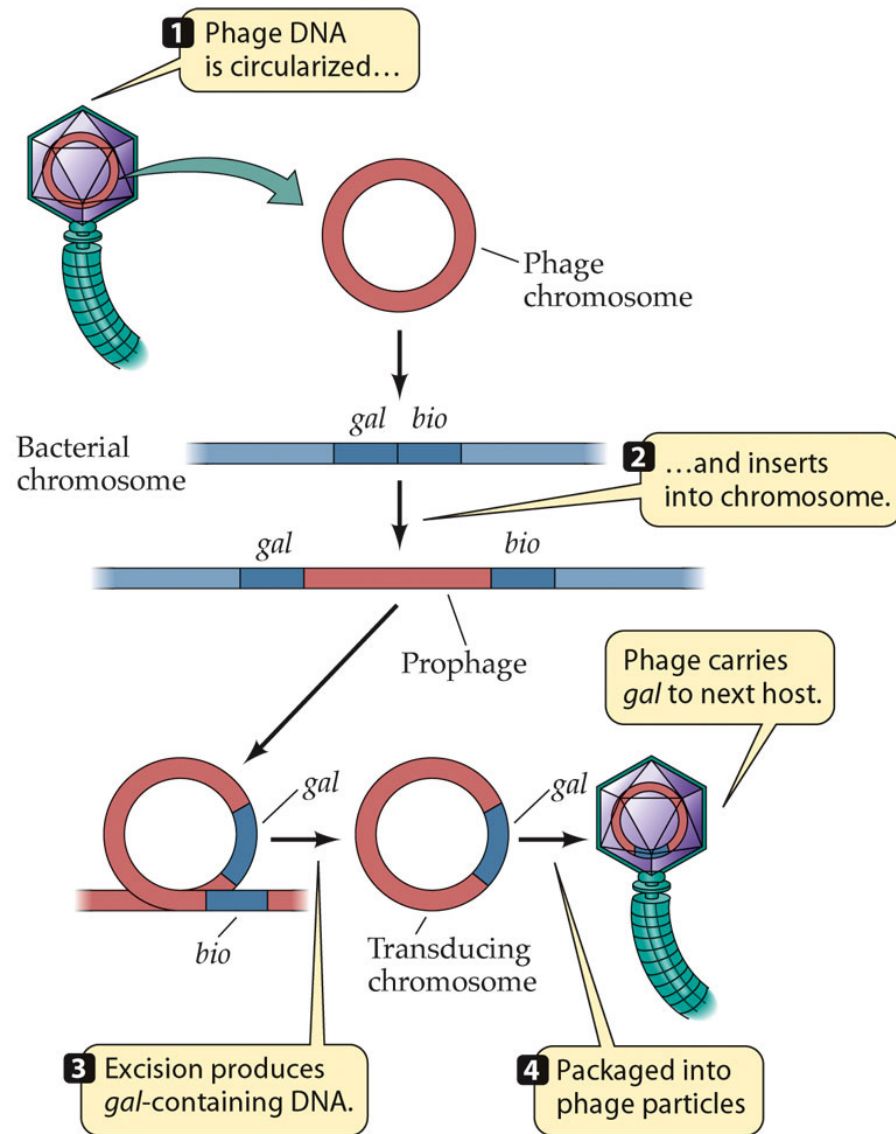


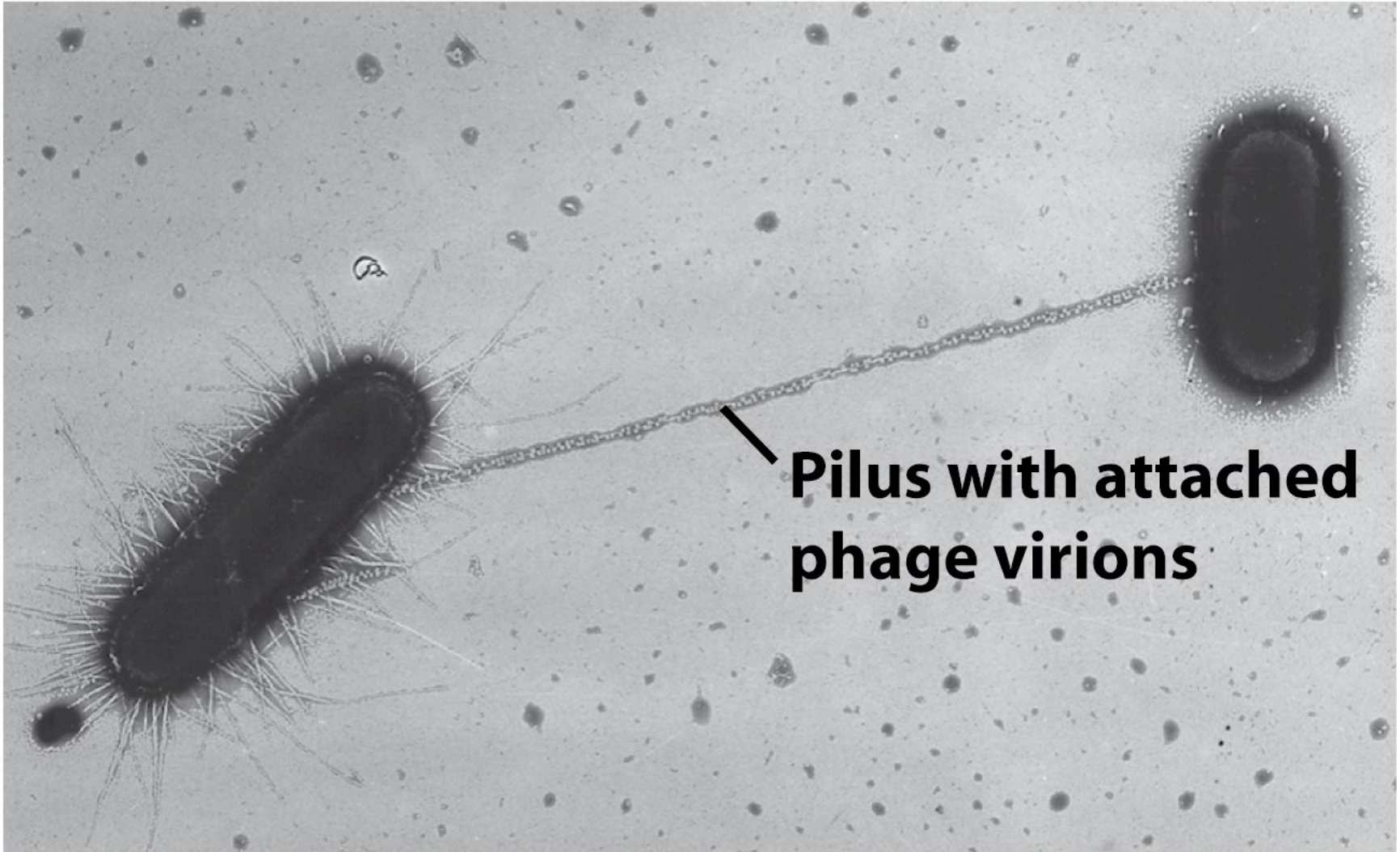


# Generalized transduction (cont.)



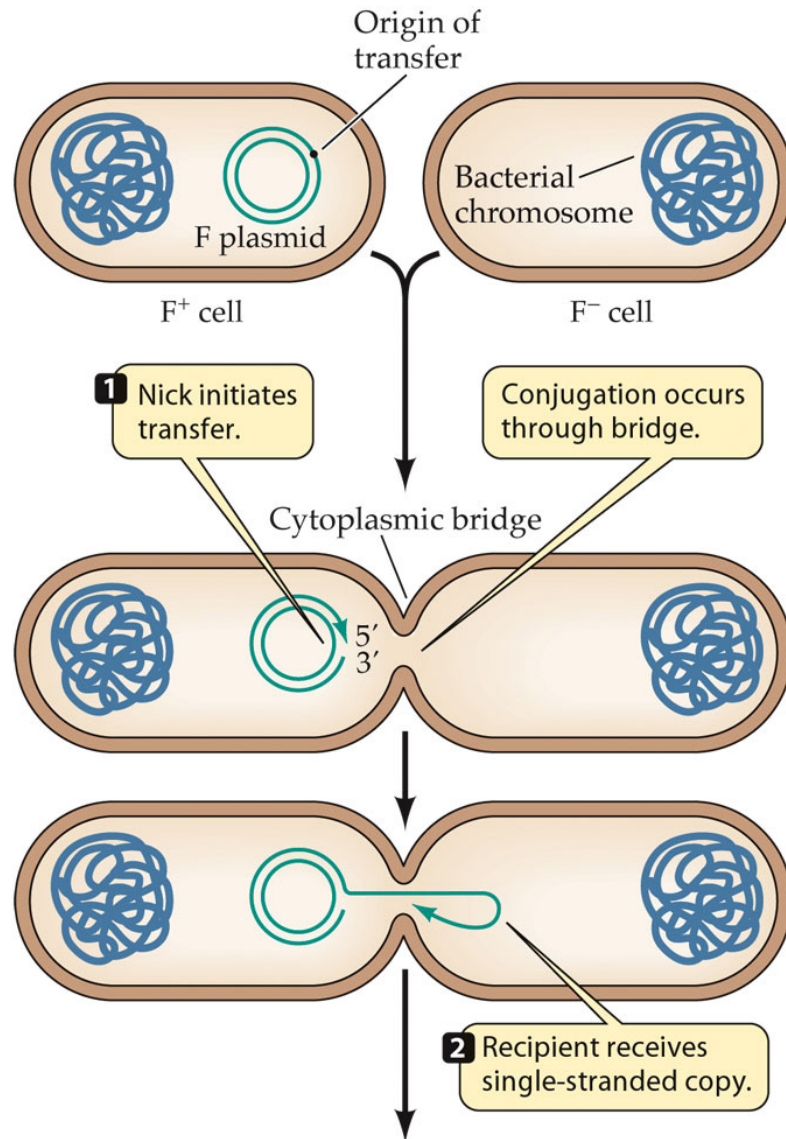
# Specialized transduction

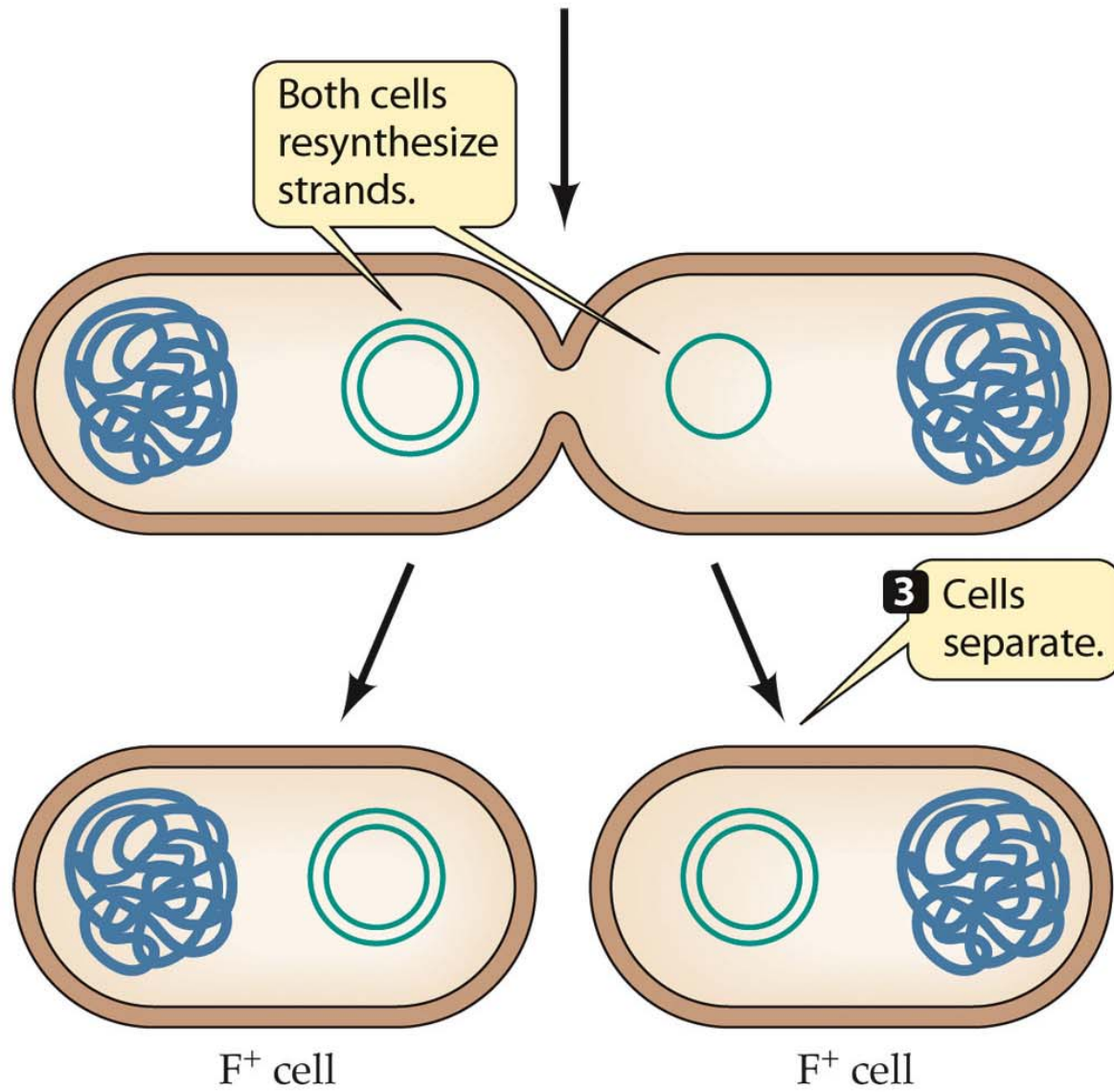




**Pilus with attached  
phage virions**

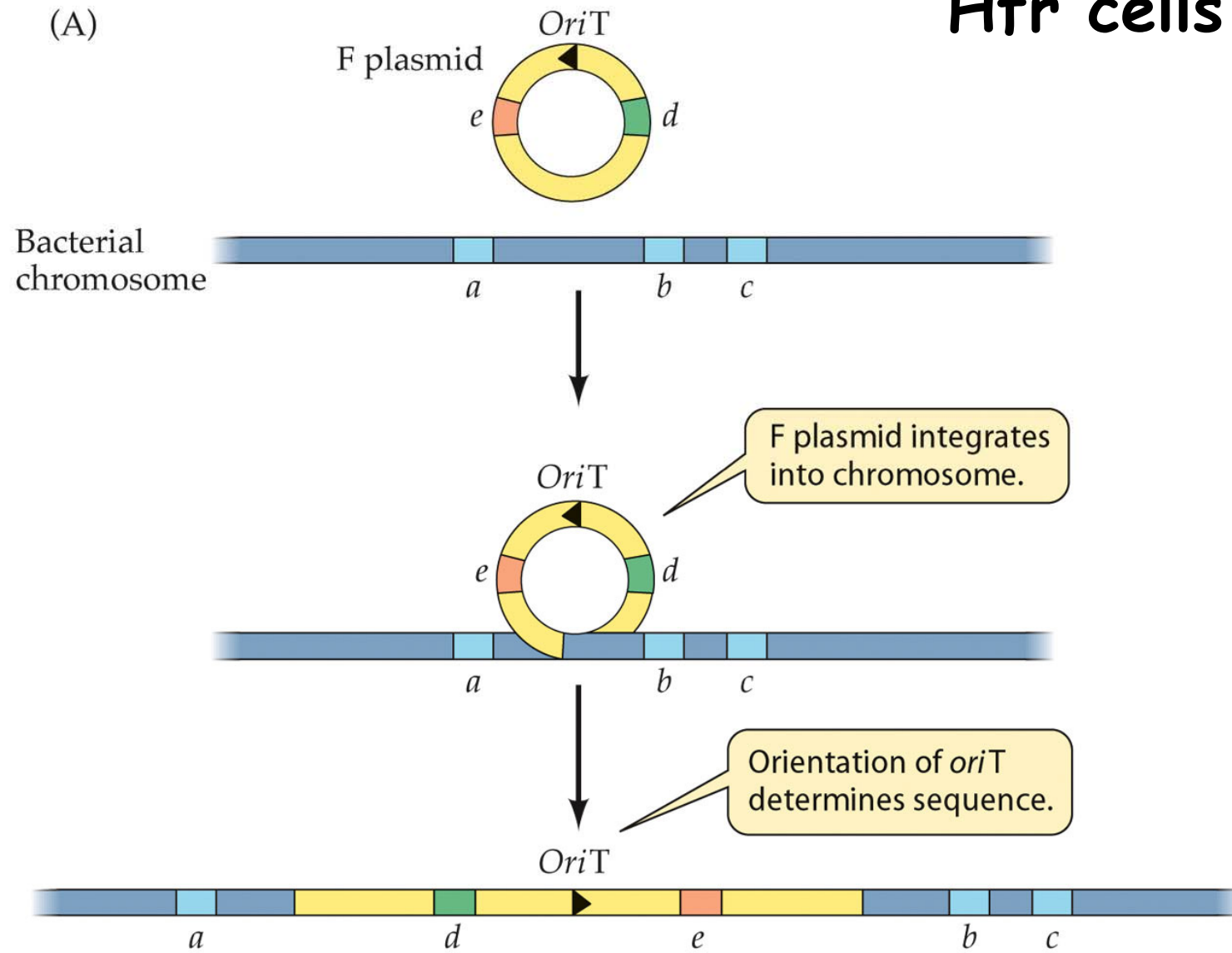
# Bacterial conjugation





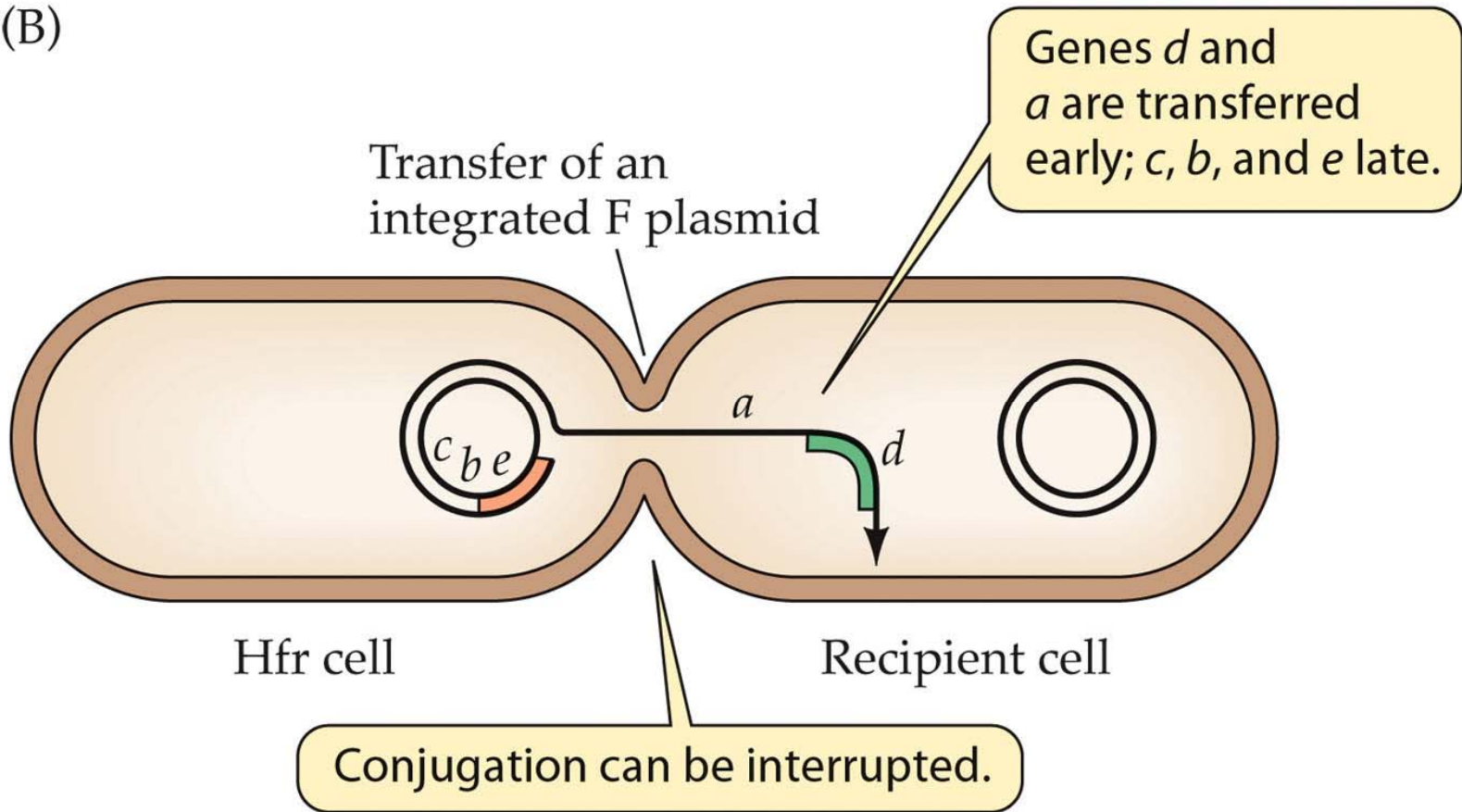


# Hfr cells

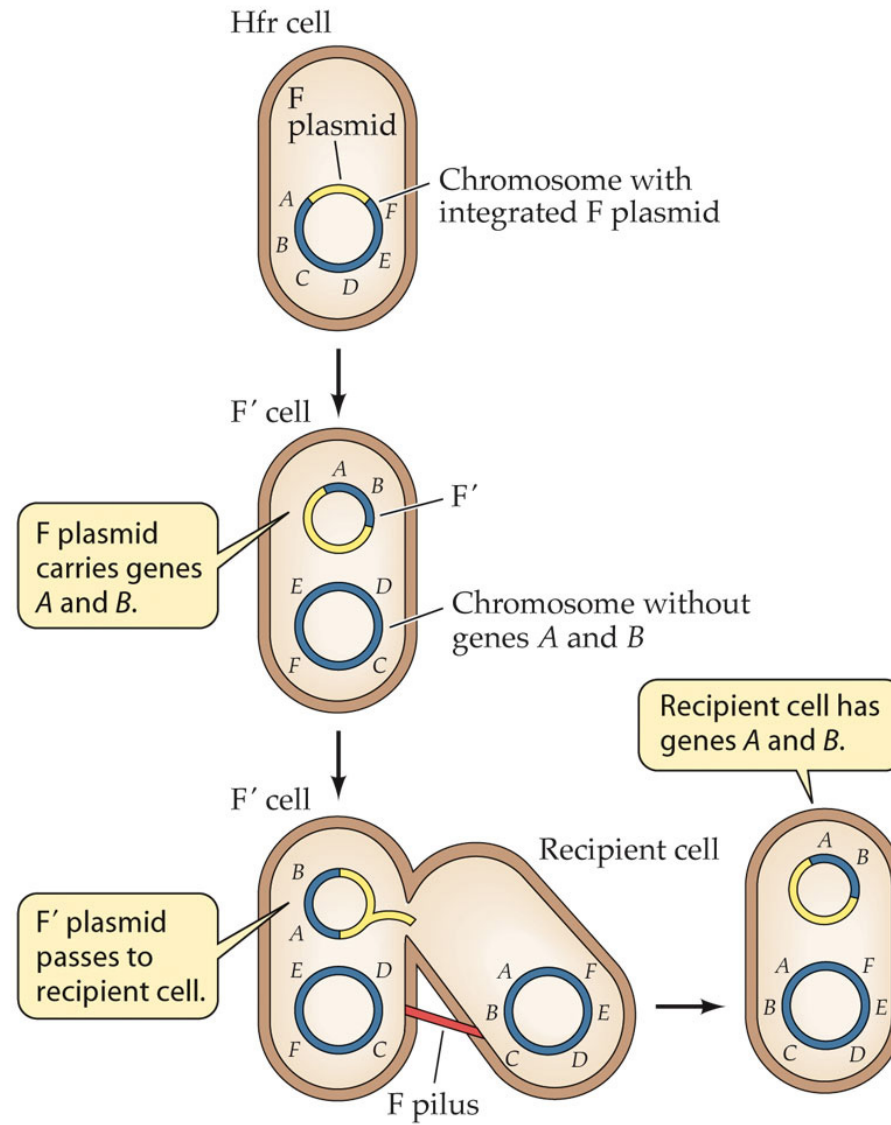




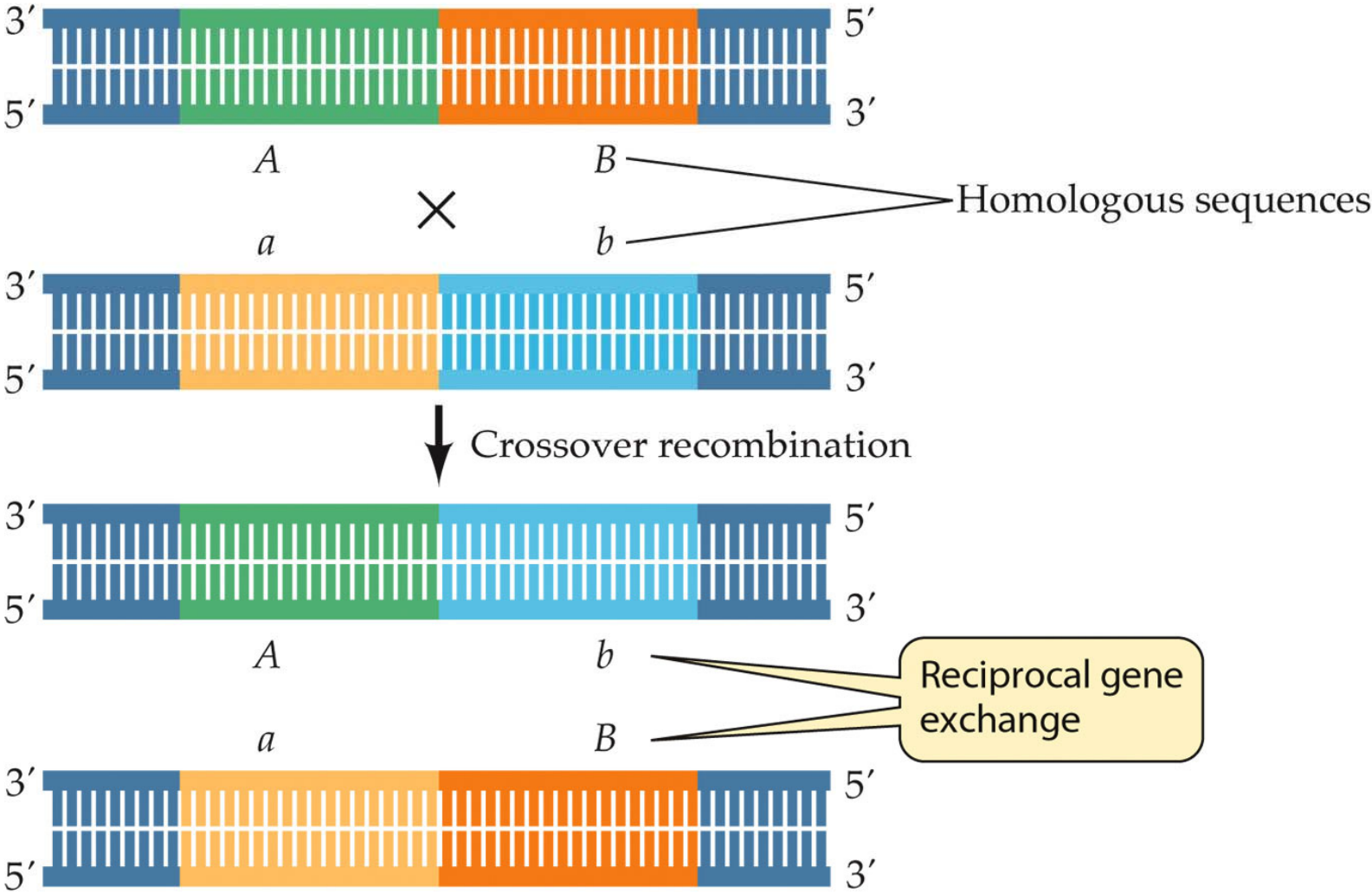
(B)



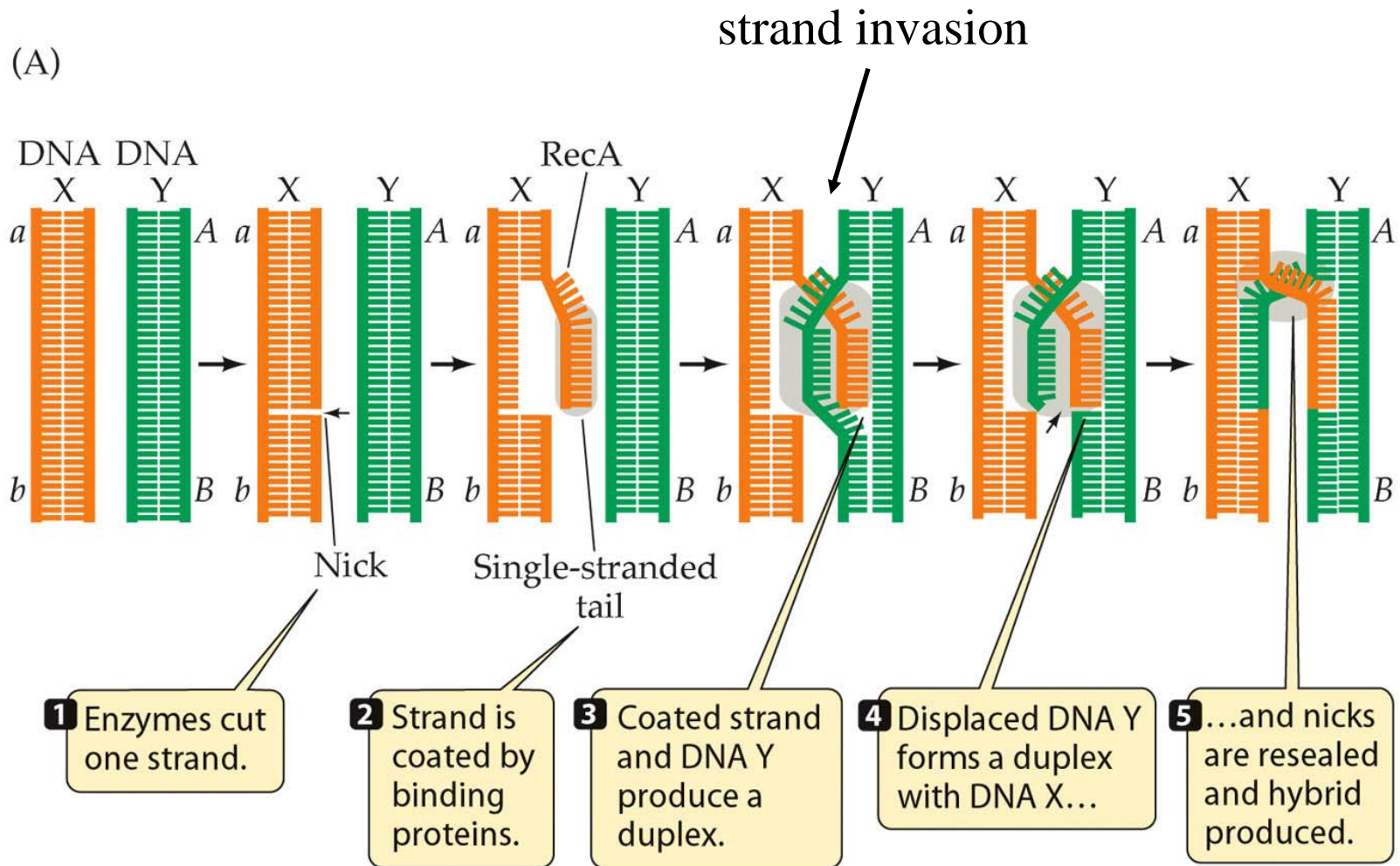
# F' cells



# Homologous Recombination

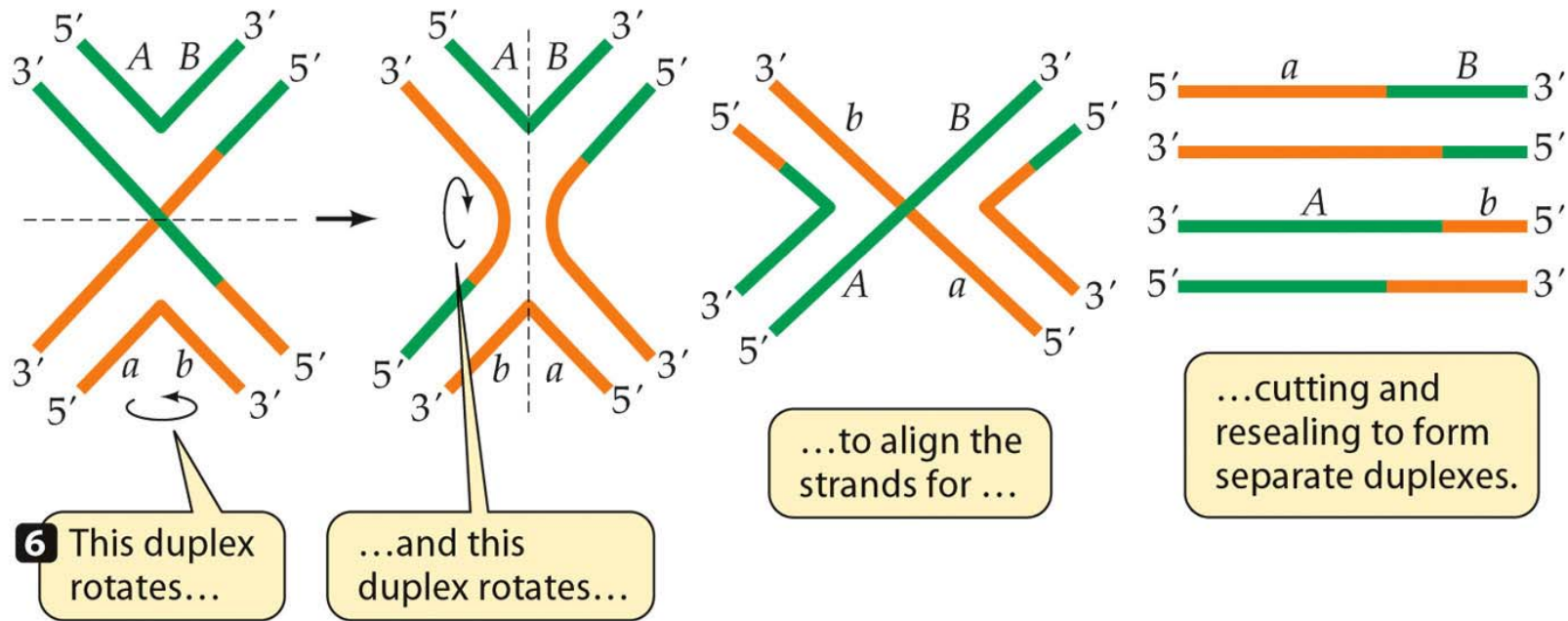


# Homologous Recombination



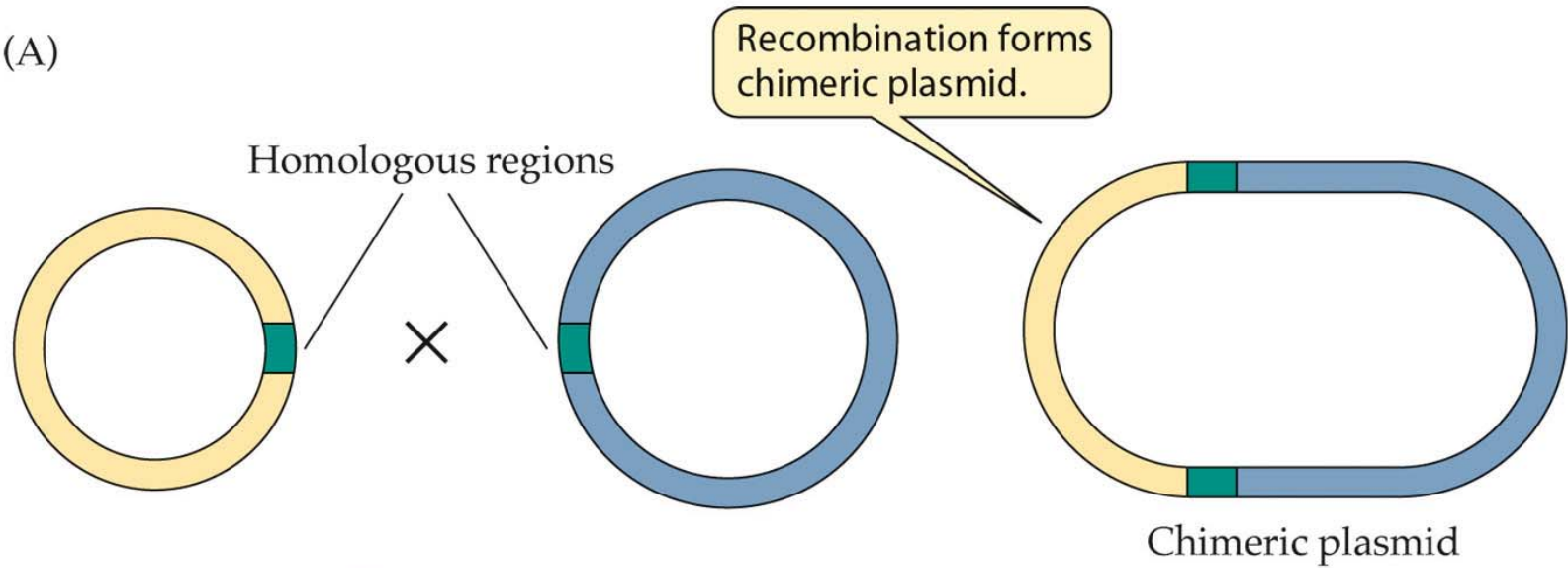
# Homologous Recombination

(B)

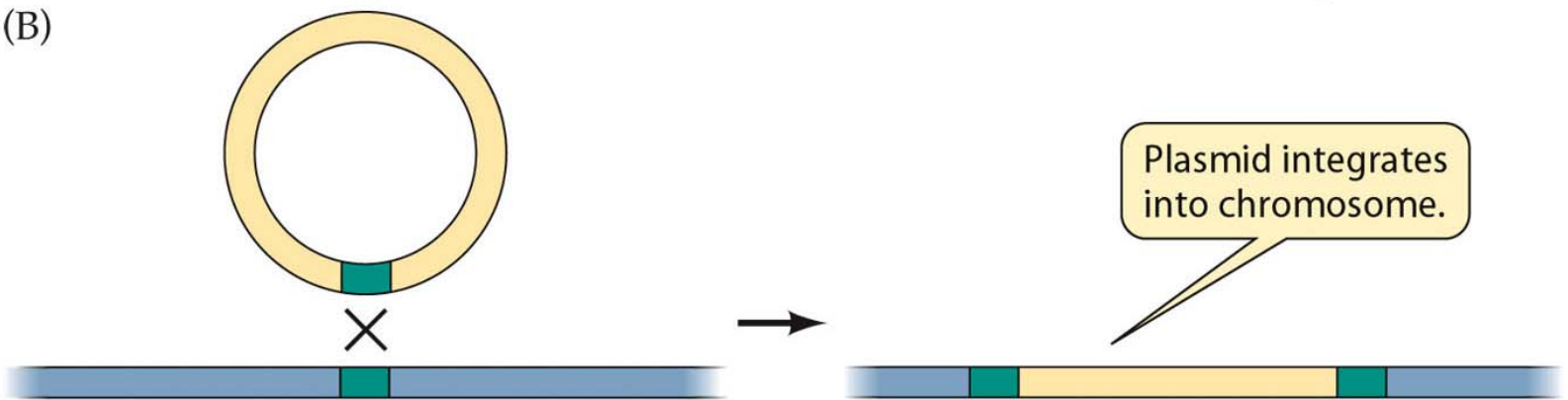


# Types of homologous recombination in bacteria

(A)



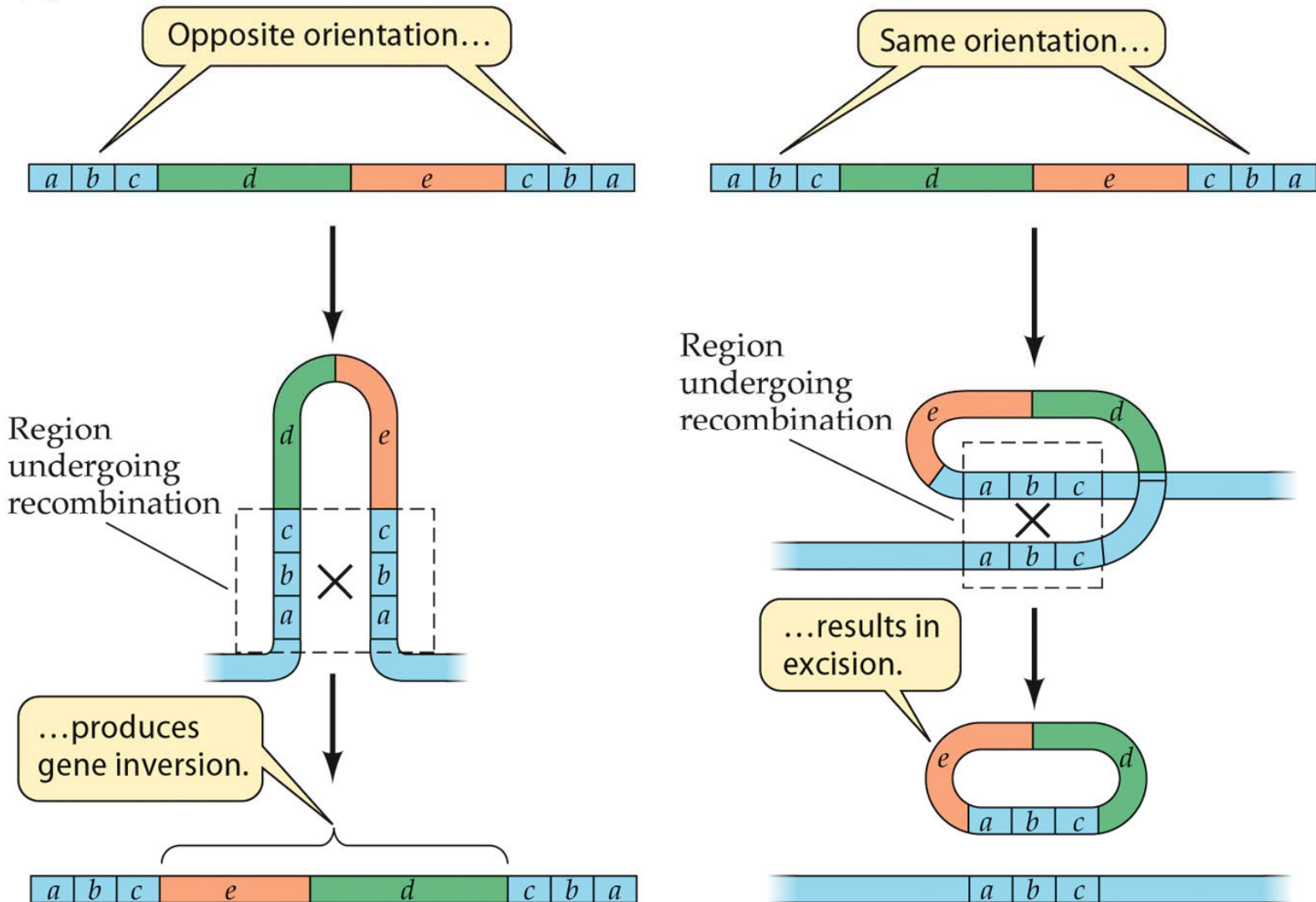
(B)





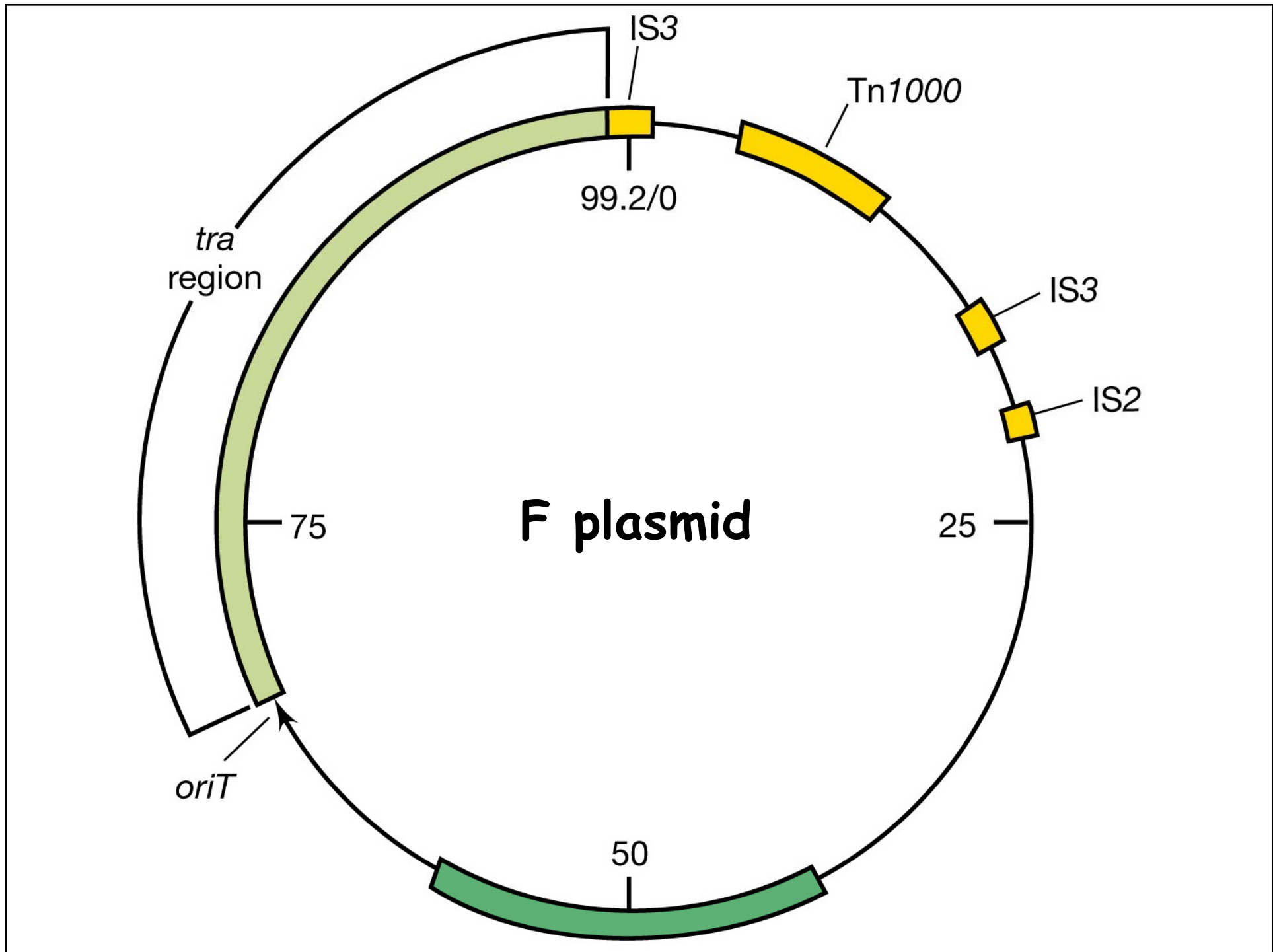
# Types of homologous recombination in bacteria

(C)

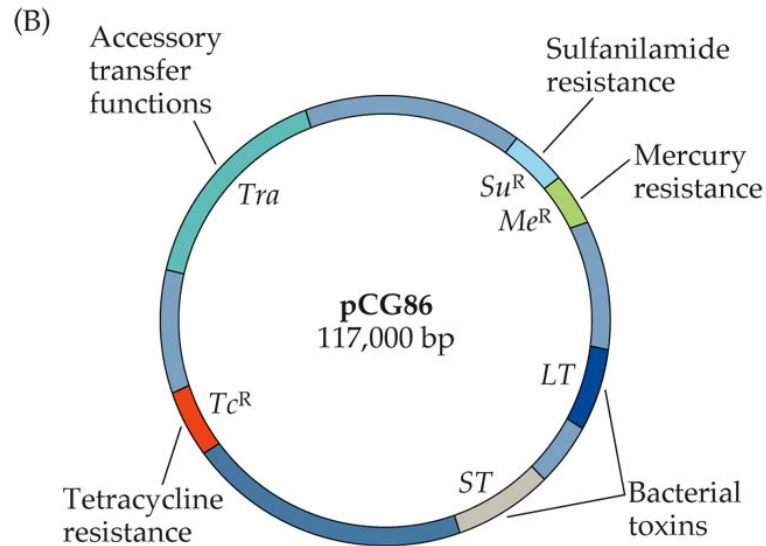
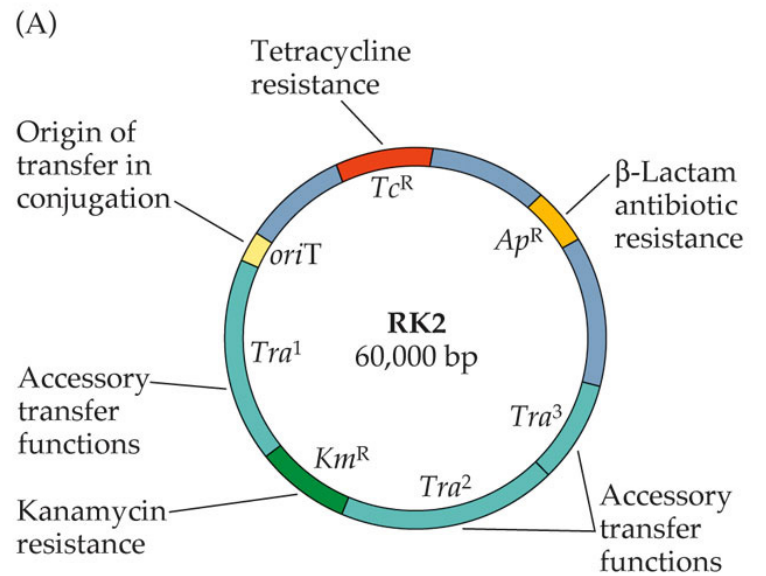


# Plasmids



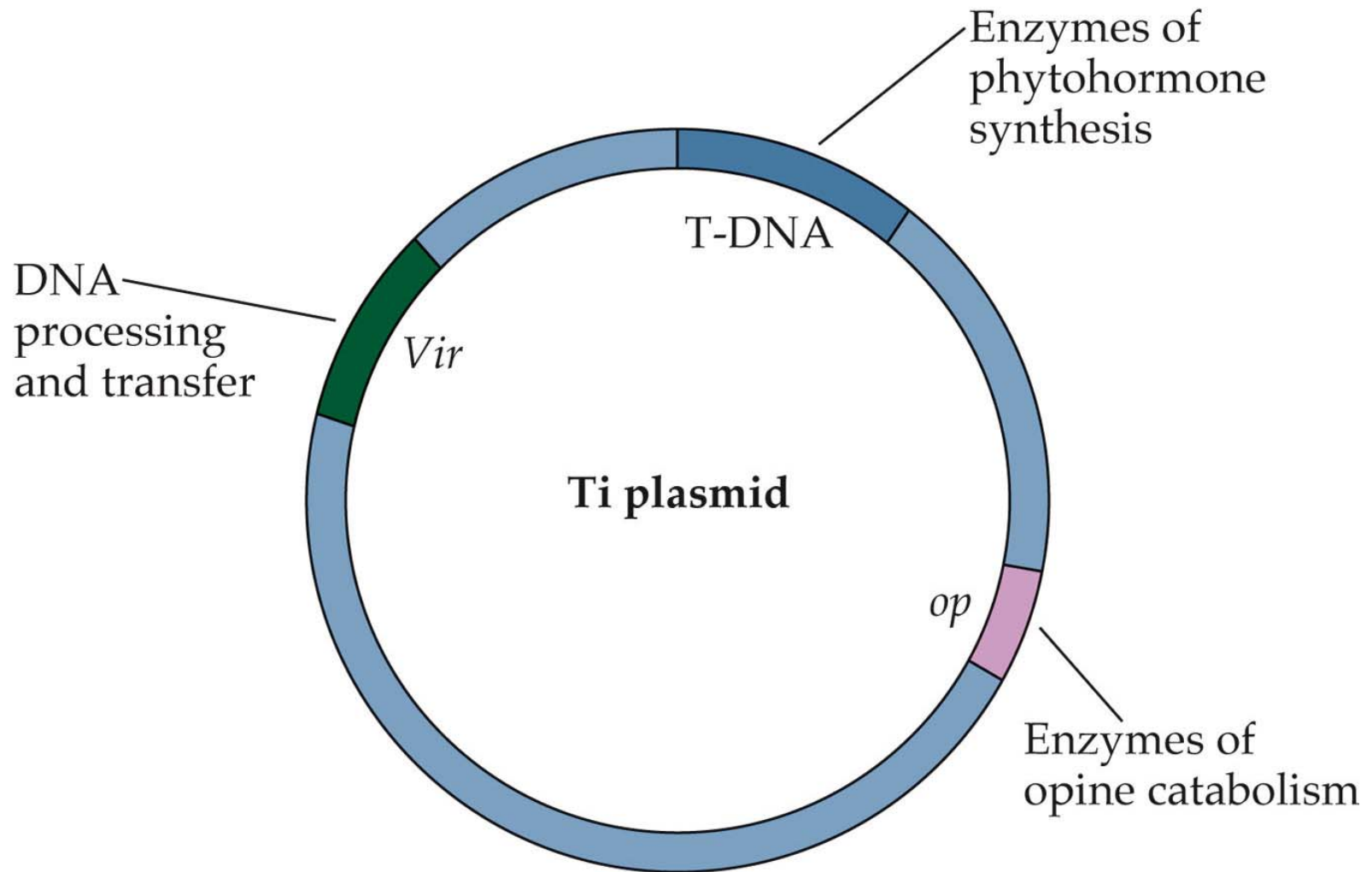


# R plasmids of pathogenic bacteria





# Ti plasmid



**Table 10.3 Some phenotypes conferred by plasmids in prokaryotes**

| Phenotype class <sup>a</sup>  | Organisms <sup>b</sup>  |
|---|---|
| <b>Antibiotic production</b>  | <i>Streptomyces</i>   |
| <b>Conjugation</b>  | <i>Escherichia, Pseudomonas, Rhizobium, Staphylococcus, Streptococcus, Sulfolobus, Vibrio</i> |
| <b>Physiological functions</b>  |   |
| Degradation of octane, camphor, naphthalene                                   | <i>Pseudomonas</i>  |
| Degradation of herbicides   | <i>Alcaligenes</i>  |
| Formation of acetone and butanol (🔗 Section 12.20)                            | <i>Clostridium</i>  |
| Lactose, sucrose or urea utilization and nitrogen fixation                    | Enteric bacteria  |
| Nodulation and symbiotic nitrogen fixation (🔗 Section 19.22)                  | <i>Rhizobium</i>  |
| Pigment production  | <i>Erwinia, Staphylococcus</i>  |
| <b>Resistance</b>   |   |
| Antibiotic resistance (🔗 Section 20.12)                                       | <i>Campylobacter, Enteric bacteria, Neisseria, Staphylococcus</i>                             |
| Resistance to cadmium, cobalt, mercury, nickel, and/or zinc (🔗 Section 19.16) | <i>Acidocella, Alcaligenes, Listeria, Pseudomonas, Staphylococcus</i>                         |
| Bacteriocin resistance (and production)                                       | <i>Bacillus, Enteric bacteria, Lactococcus, Propionibacterium</i>                             |
| <b>Virulence</b>  |   |
| Host cell invasion  | <i>Salmonella, Shigella, Yersinia</i>   |
| Coagulase, hemolysin, enterotoxin (🔗 Sections 21.9 and 21.11)                 | <i>Staphylococcus</i>   |
| Enterotoxin, K antigen (🔗 Sections 12.11 and 21.11)                           | <i>Escherichia</i>  |
| Tumorigenicity in plants (🔗 Section 19.21)                                    | <i>Agrobacterium</i>  |

<sup>a</sup> Only a few of the many phenotypes known to be associated with plasmids are given.

<sup>b</sup> Only a few well-characterized examples are given. All of the organisms given in the list are *Bacteria* except for *Sulfolobus*, which is a member of the *Archaea*.