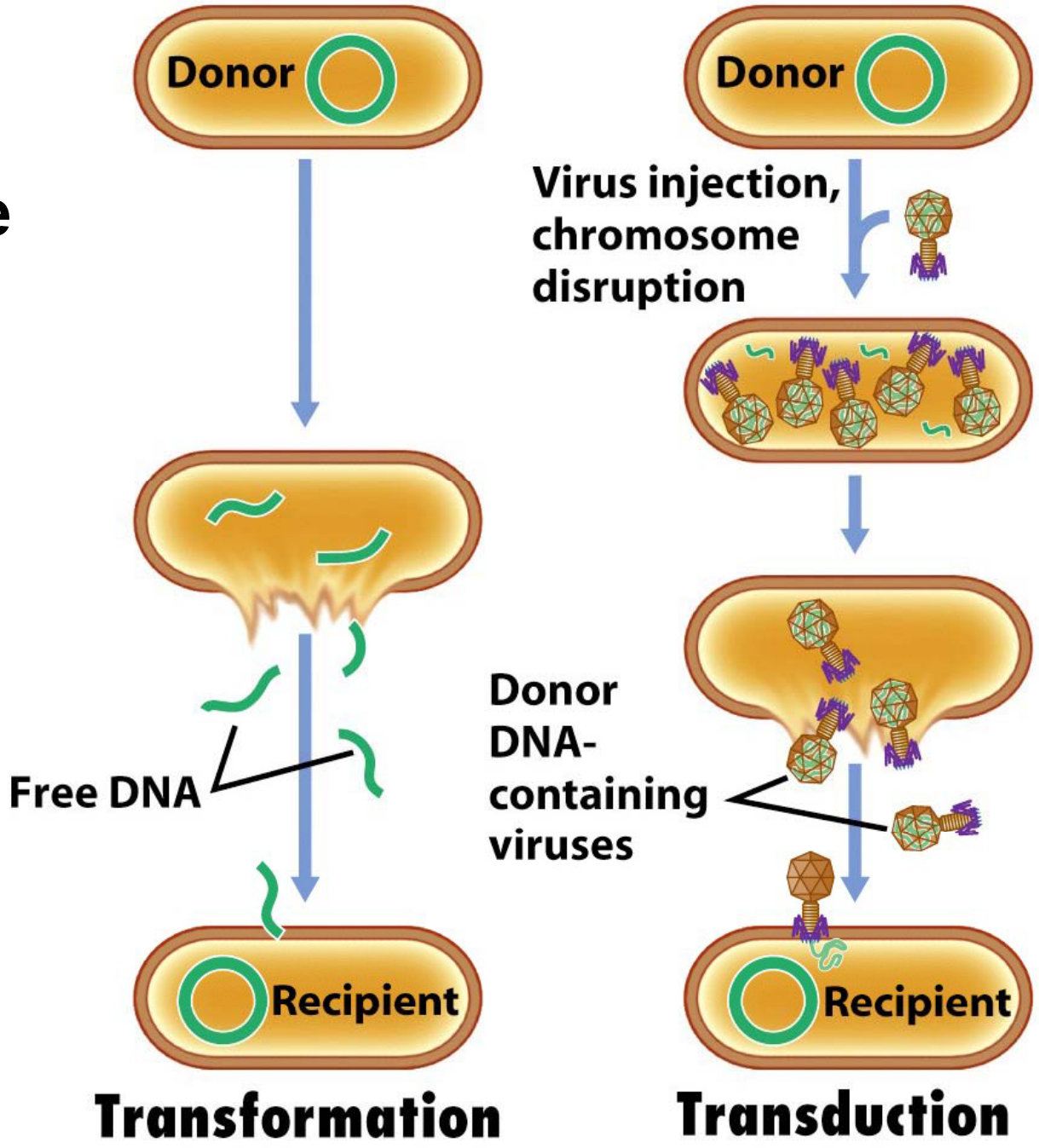
# Microbial Genomics and Chromosome Organization

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

# Microbial Genetic Exchange & Plasmids

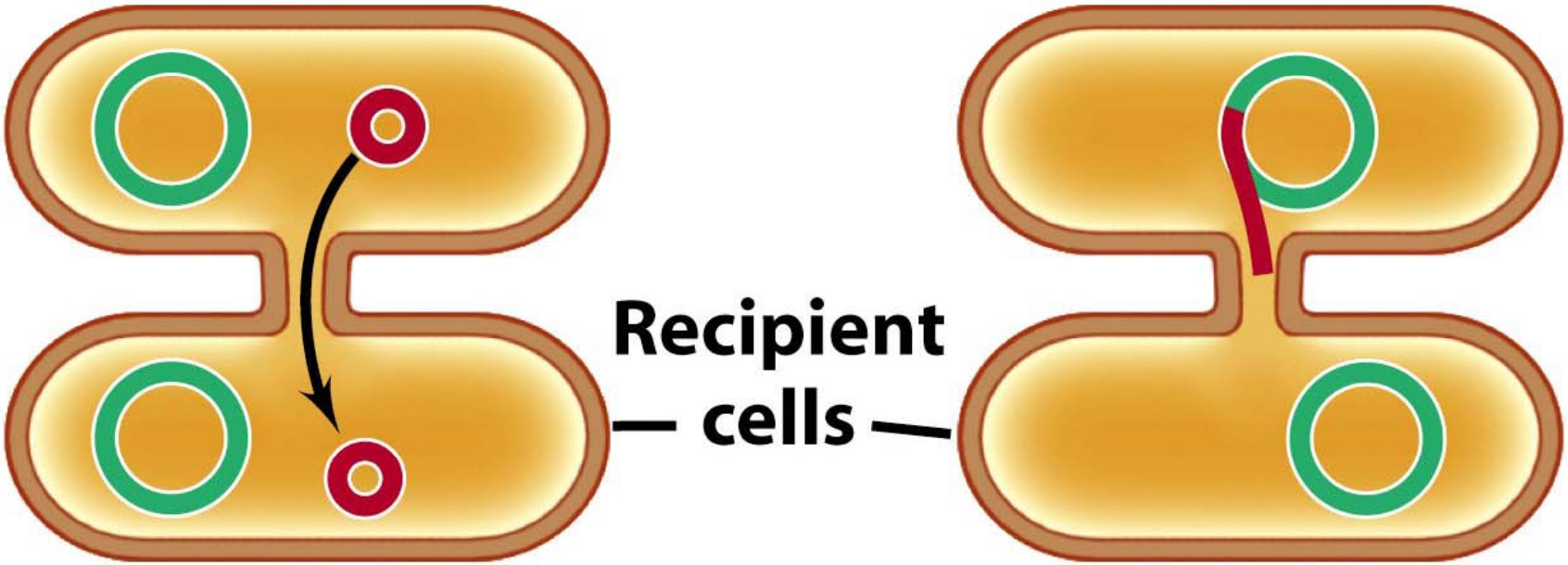
- Microbial Genetic Exchange is unidirectional!
  - Transformation
  - Transduction
  - Conjugation
- Each requires Homologous Recombination
- Types of plasmids

# Microbial Genetic Exchange



Plasmid-containing donor cell

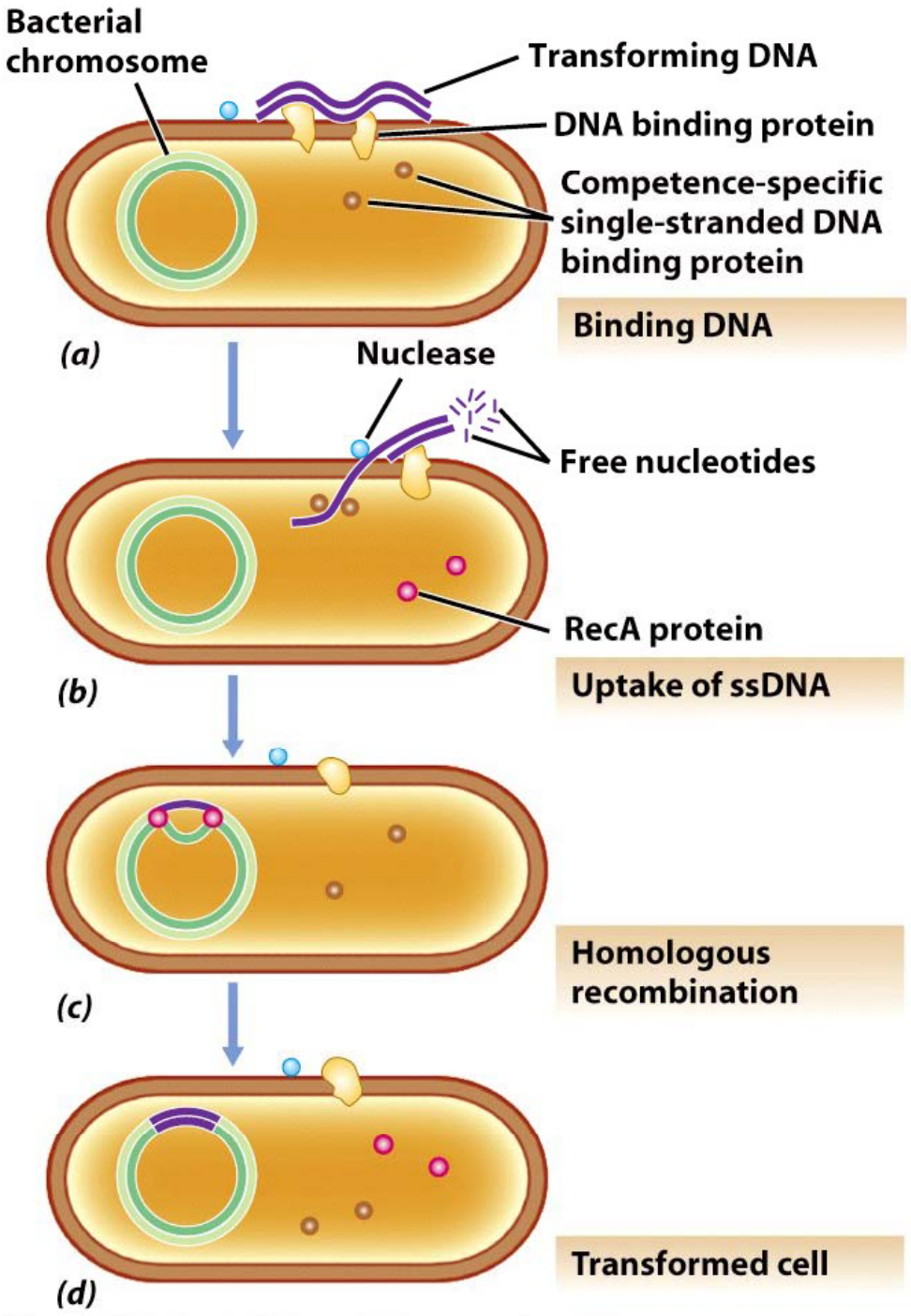
Donor cell with integrated plasmid



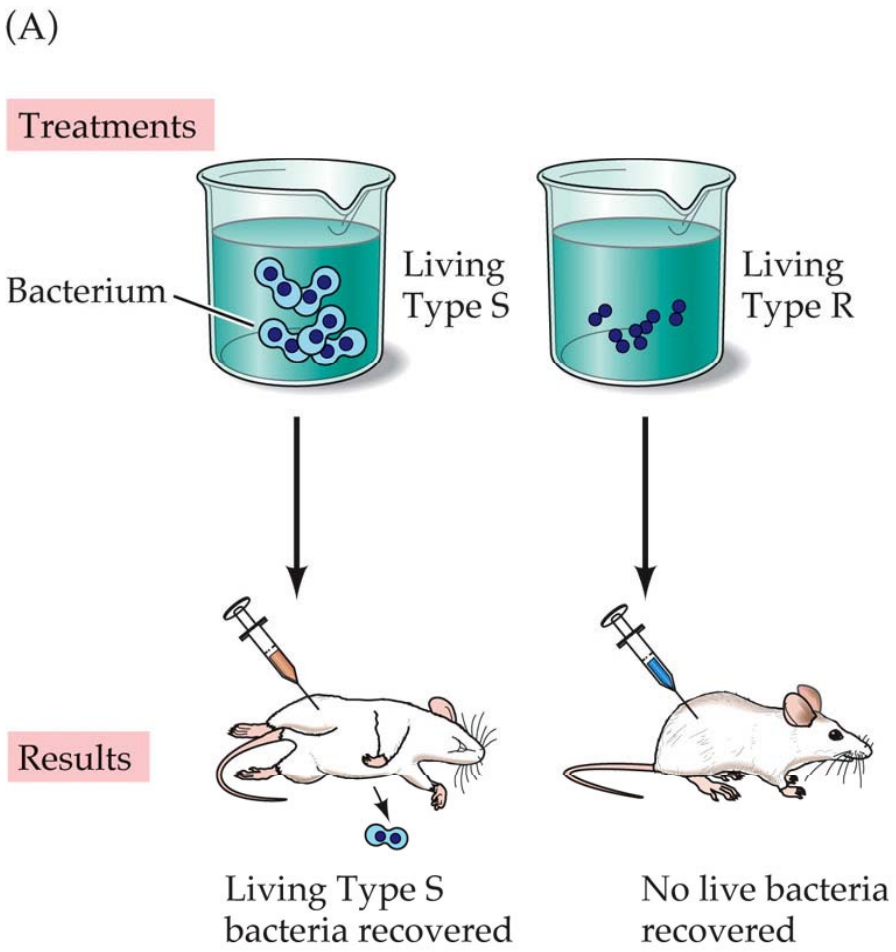

**Conjugation: Plasmid transfer**

**Conjugation: Chromosome transfer**

# Transformation by a Gram + competent cell



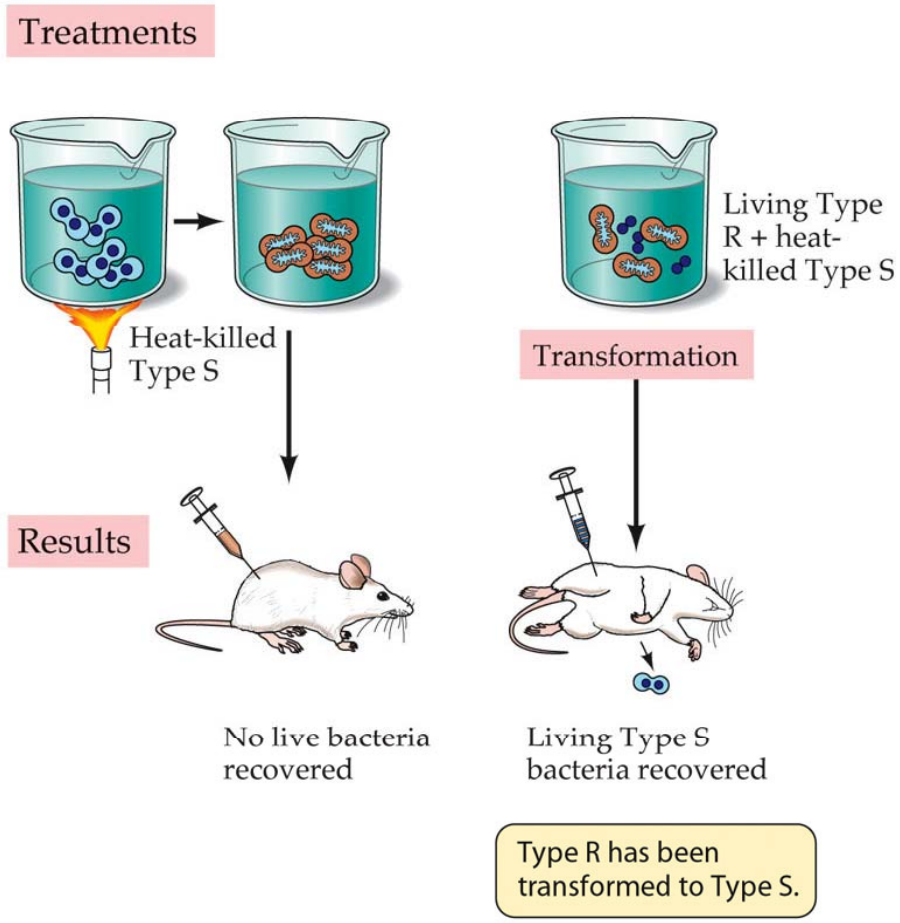
# Demonstration of transformation



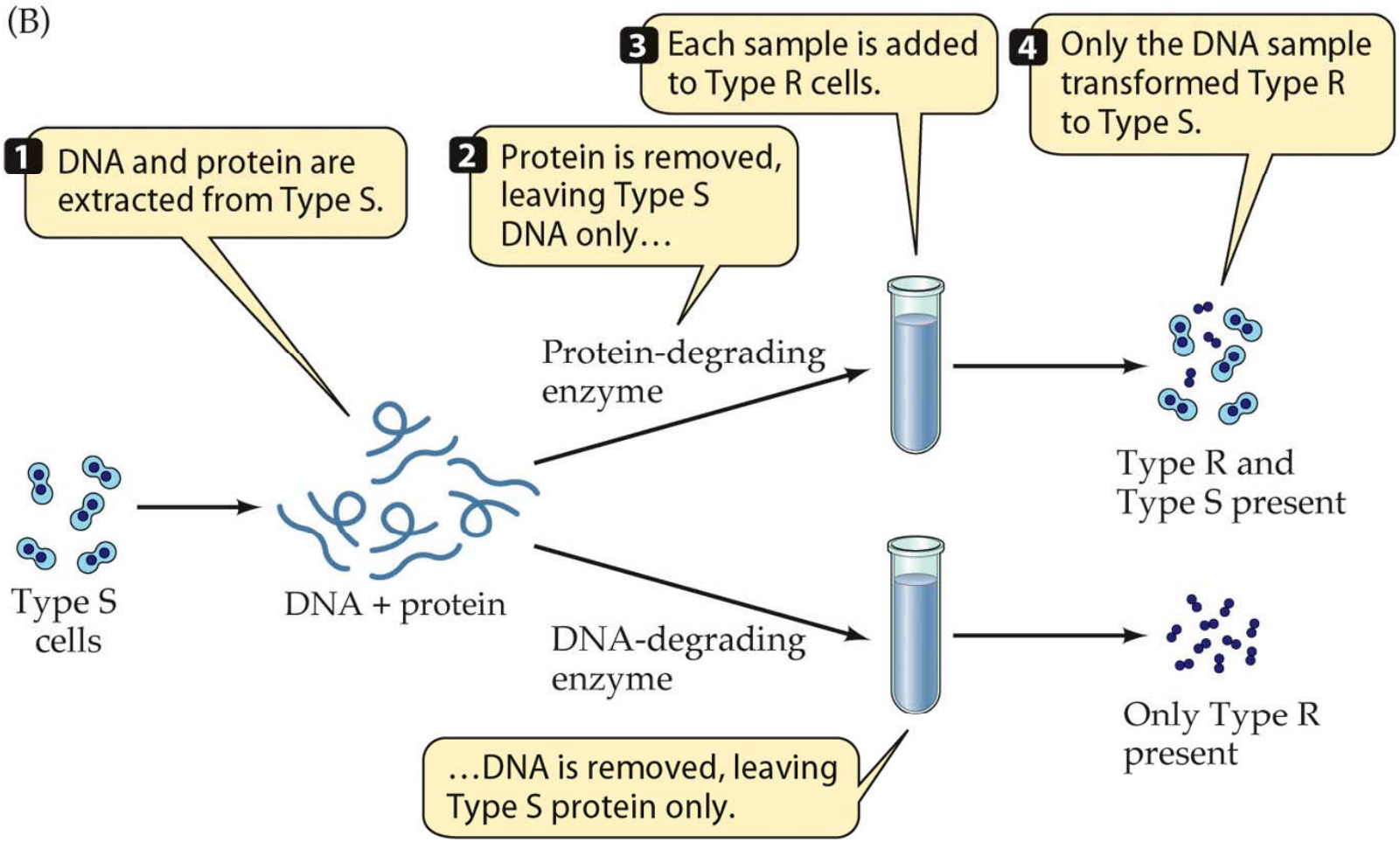

*Streptococcus pneumoniae*

Frederick Griffith, 1928

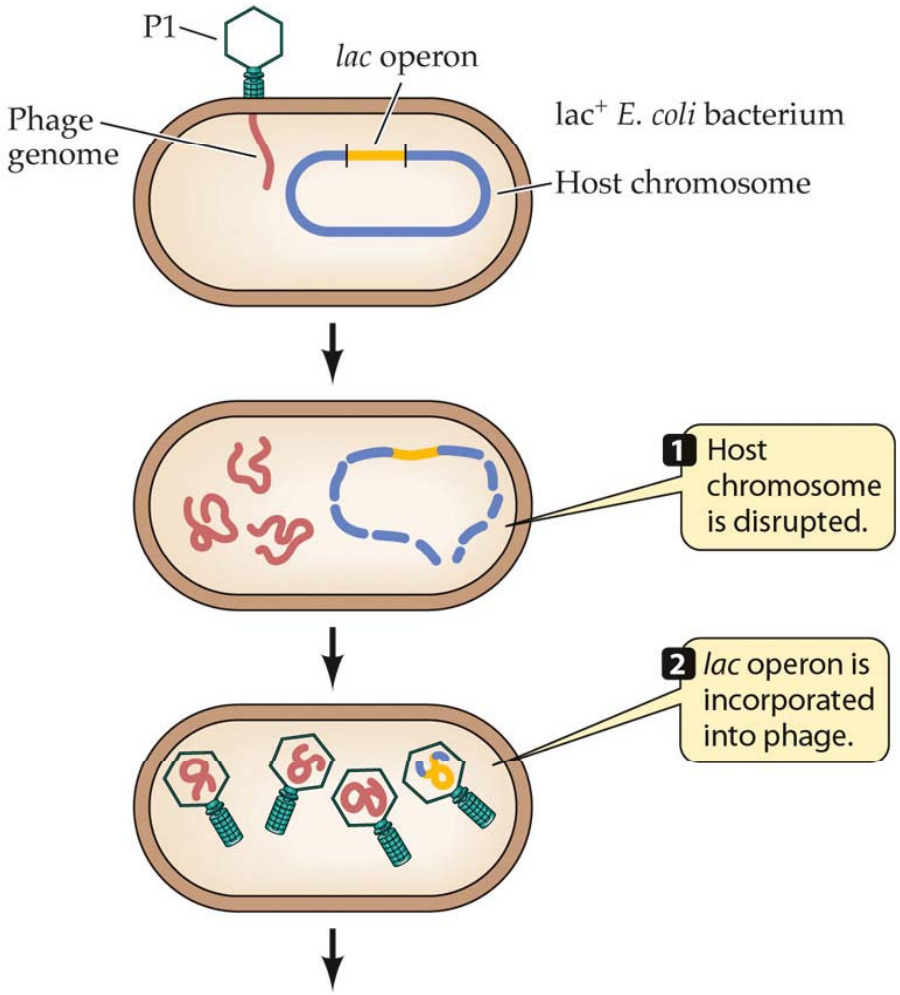
# Demonstration of transformation



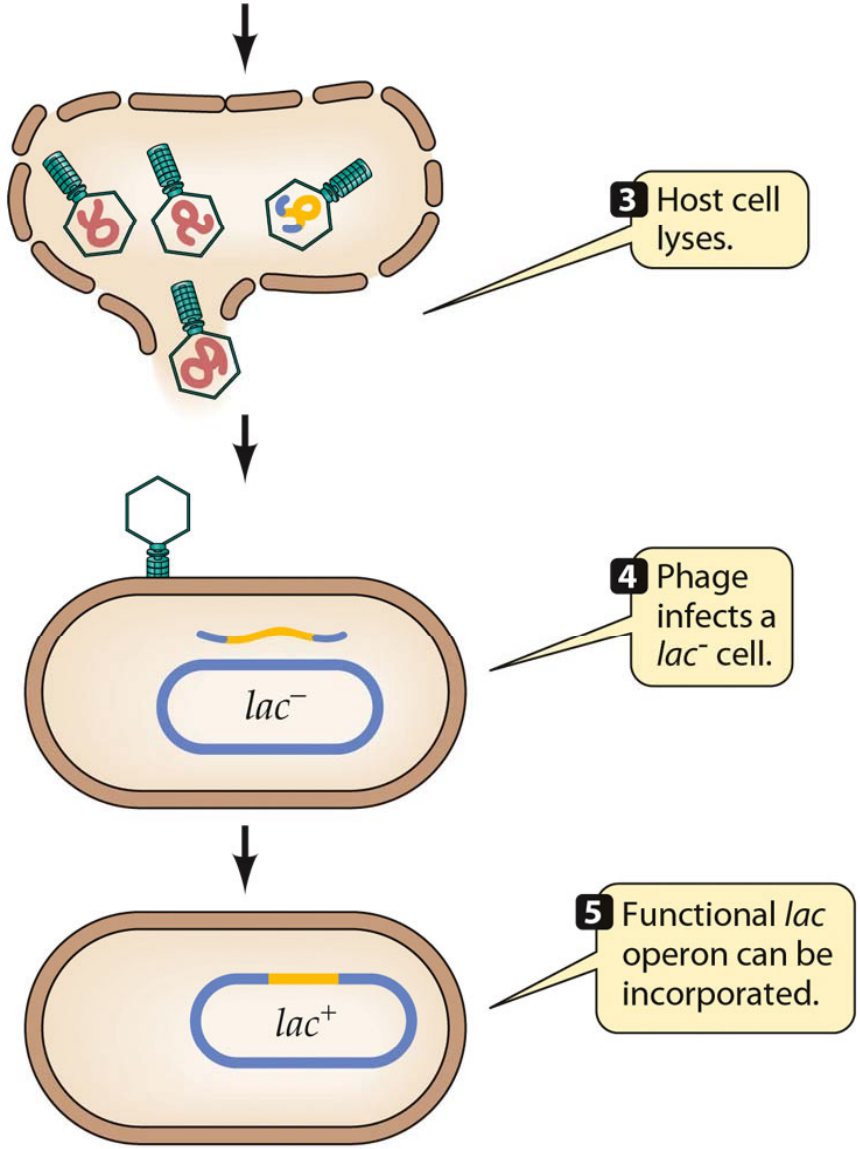

Frederick Griffith, 1928


(B)
1 DNA and protein are extracted from Type S.
2 Protein is removed, leaving Type S DNA only…
3 Each sample is added to Type R cells.
4 Only the DNA sample transformed Type R to Type S.
Type S cells
DNA + protein
Protein-degrading enzyme
DNA-degrading enzyme
Type R and Type S present
Only Type R present
…DNA is removed, leaving Type S protein only.

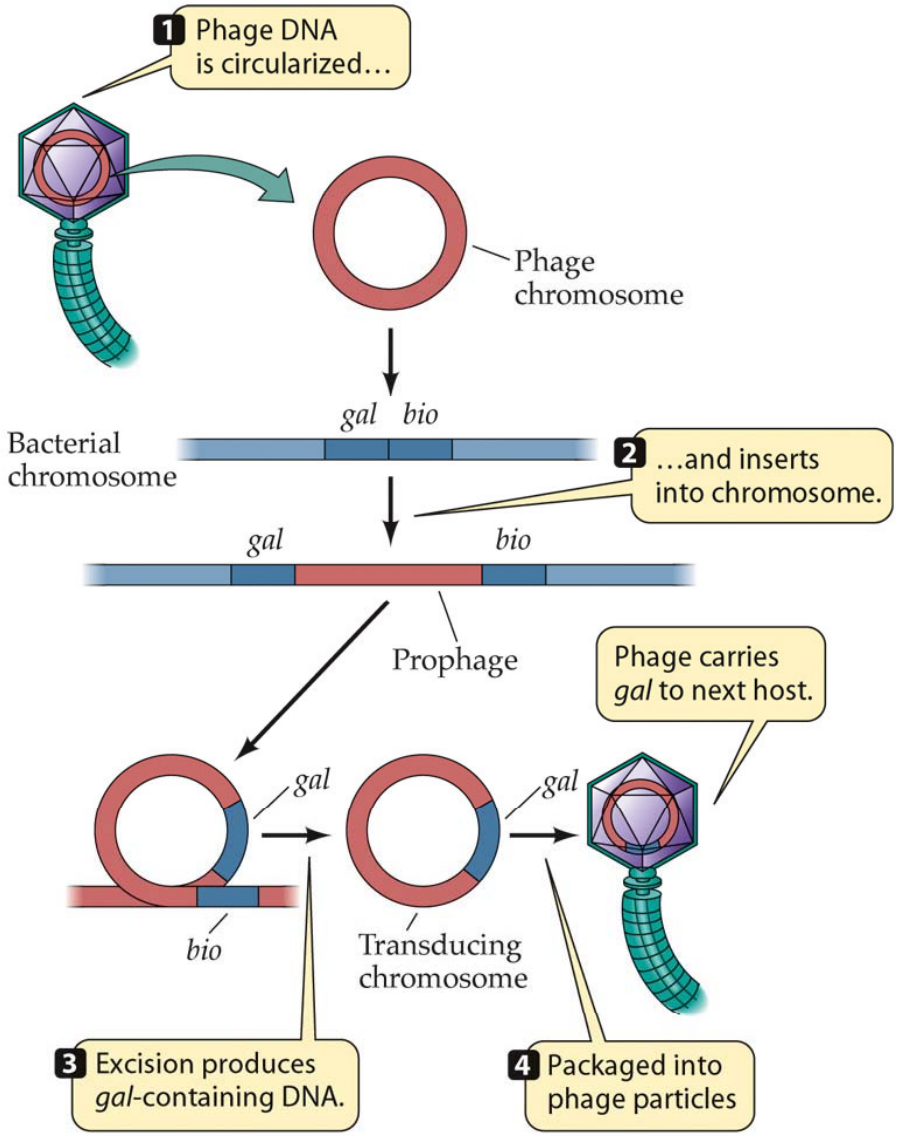
# Generalized transduction

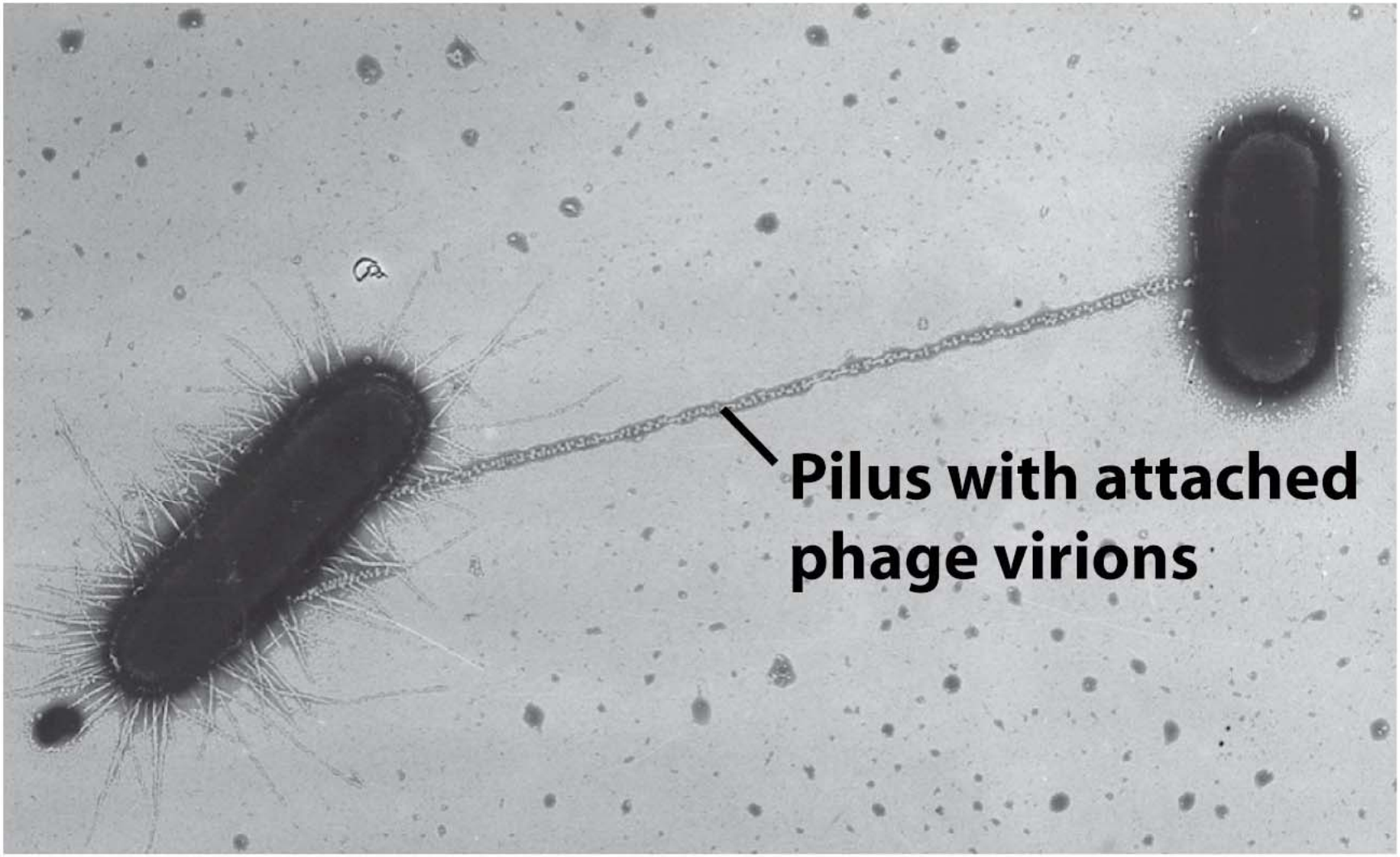


# Generalized transduction (cont.)



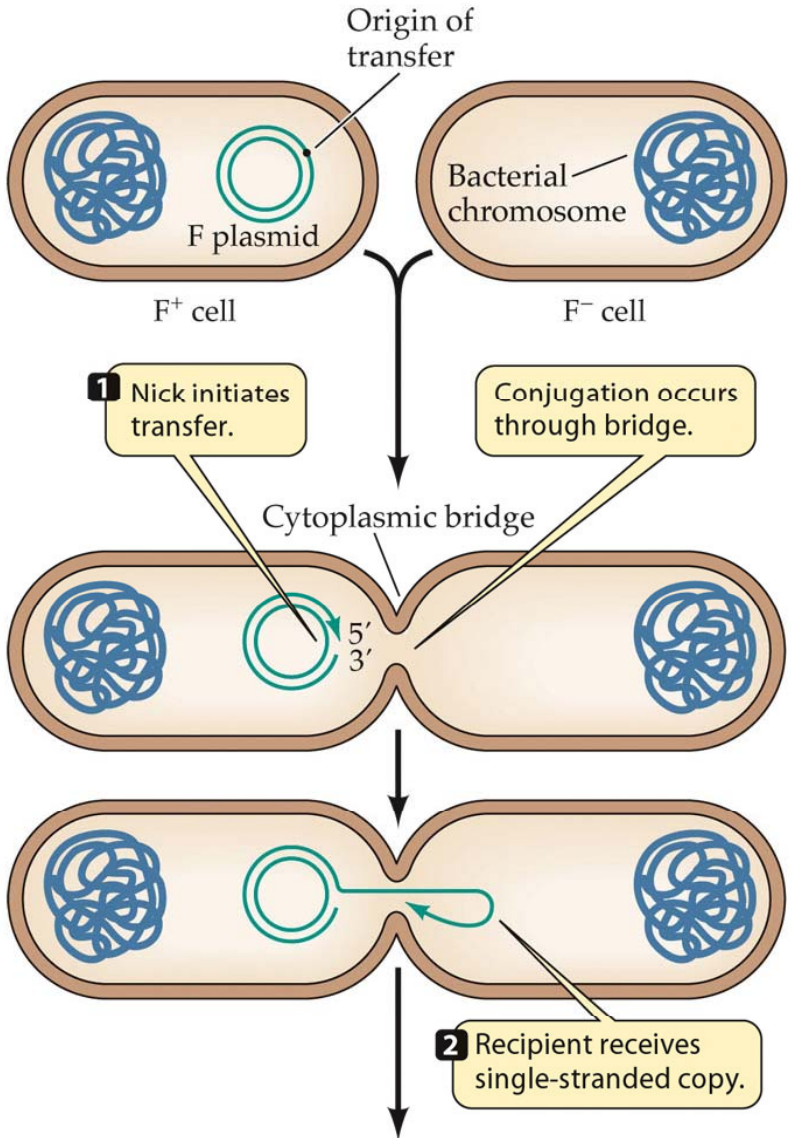
# Specialized transduction




Pilus with attached
phage virions

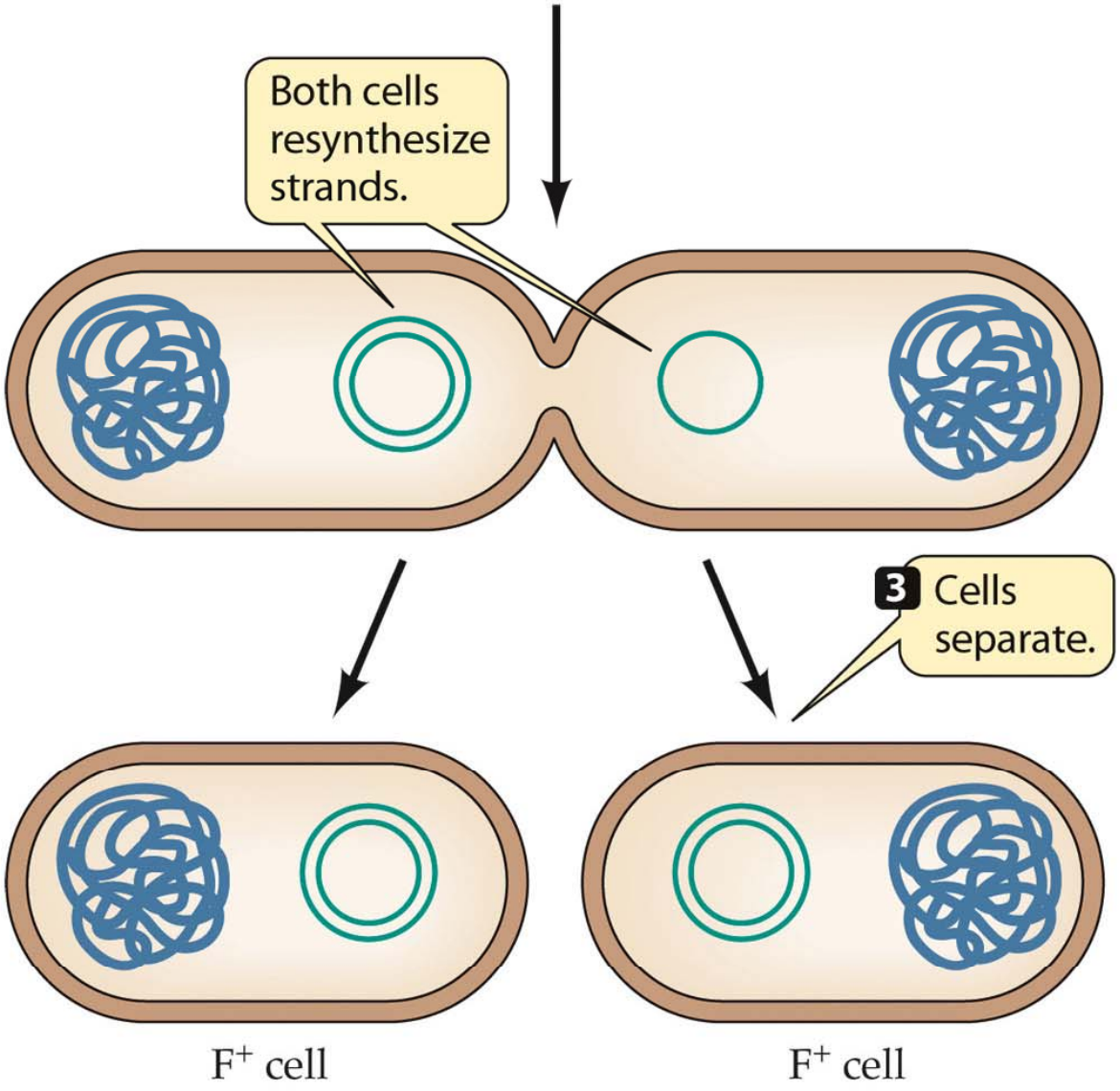
# Bacterial Conjugation

Donor

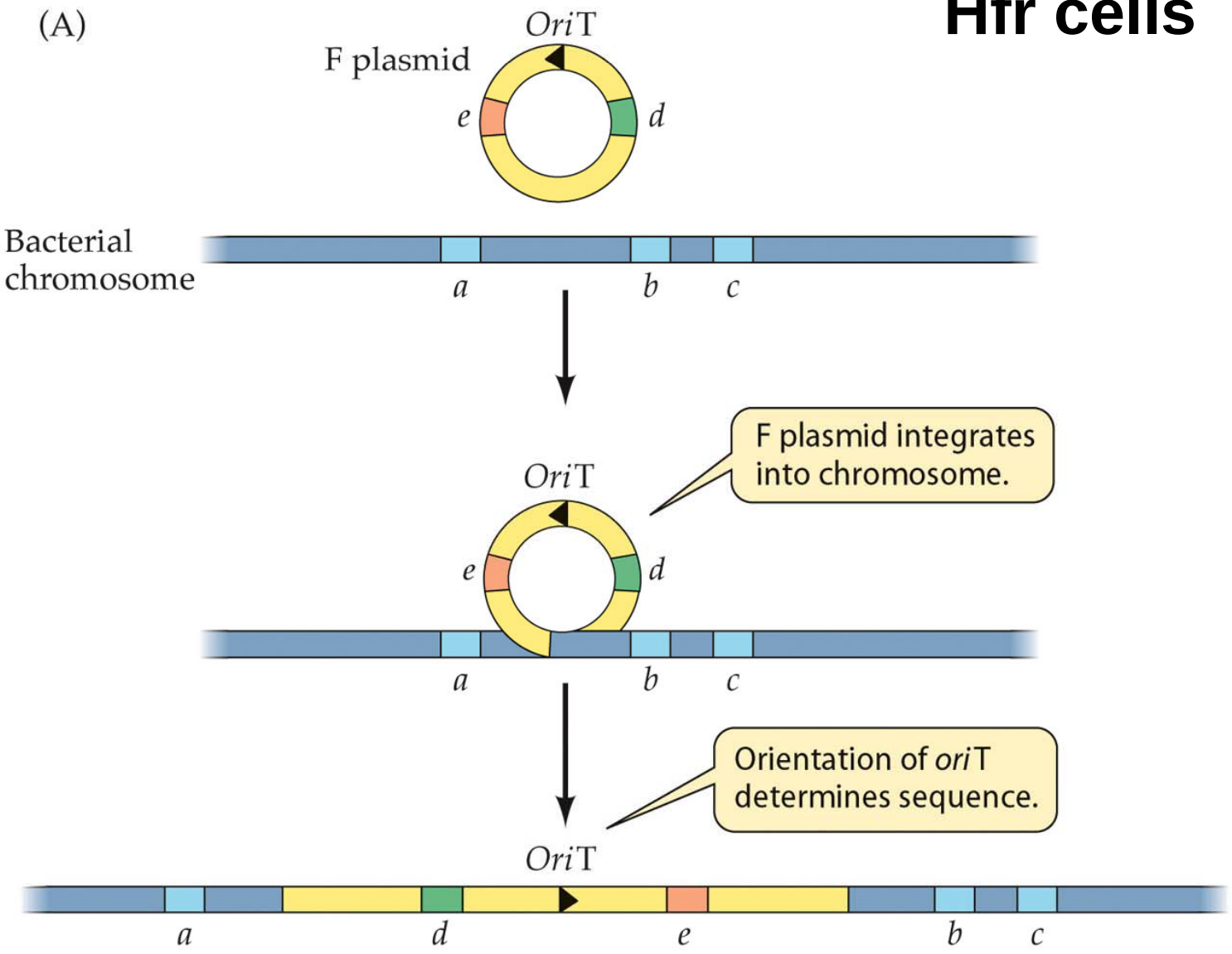



Recipient

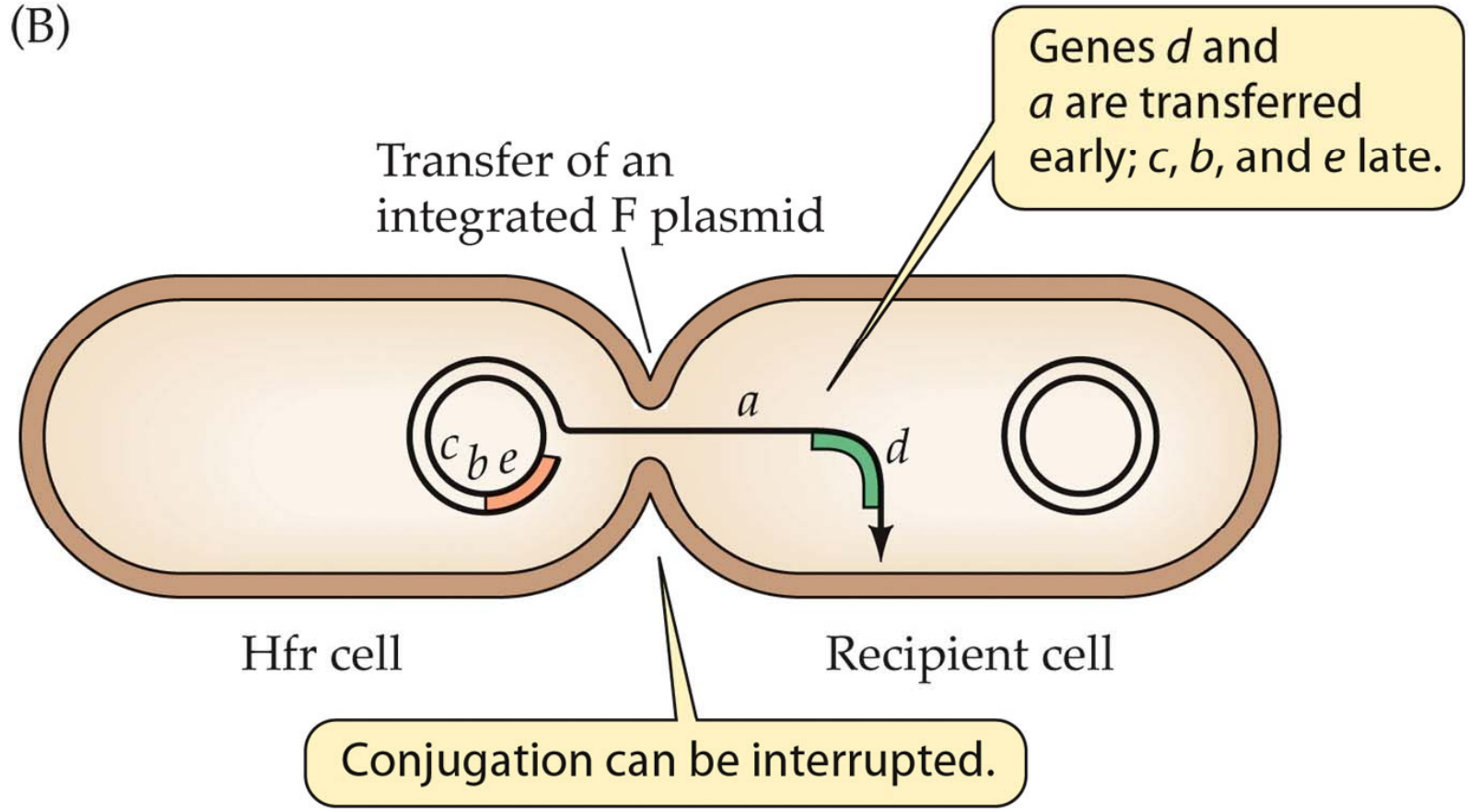
# Bacterial Conjugation



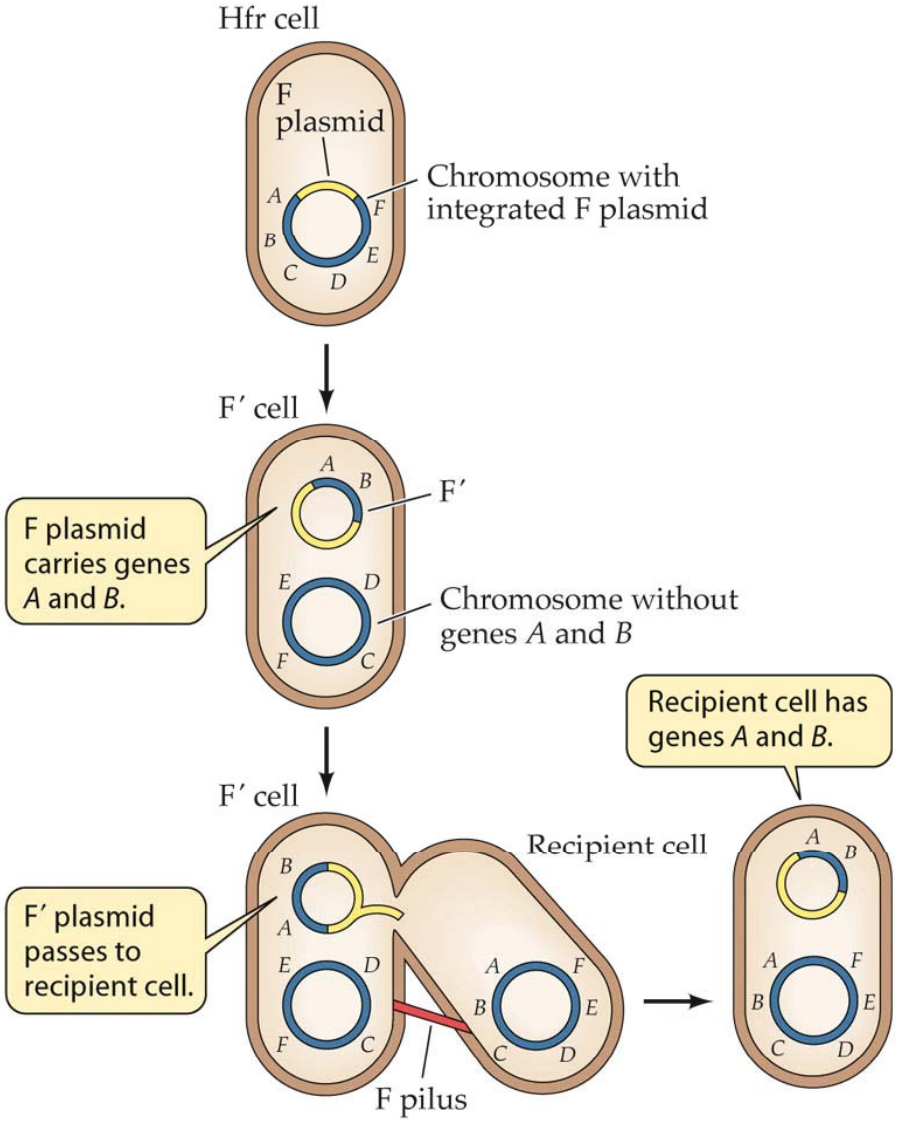
# Hfr cells



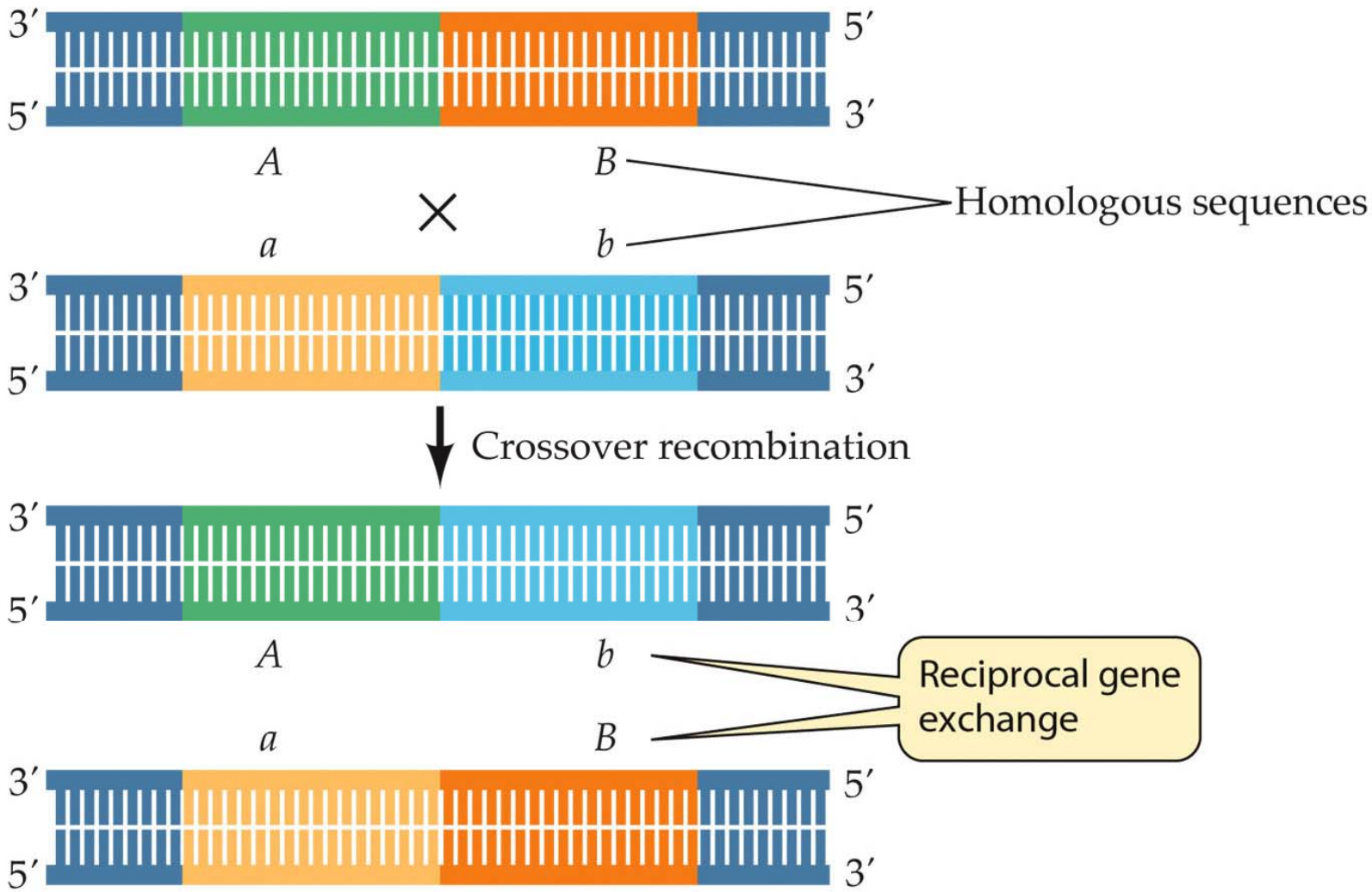

Hfr = High Frequency Recombination


(B)
Genes d and a are transferred early; c, b, and e late.
Transfer of an integrated F plasmid
a
c b e
d
Hfr cell
Recipient cell
Conjugation can be interrupted.

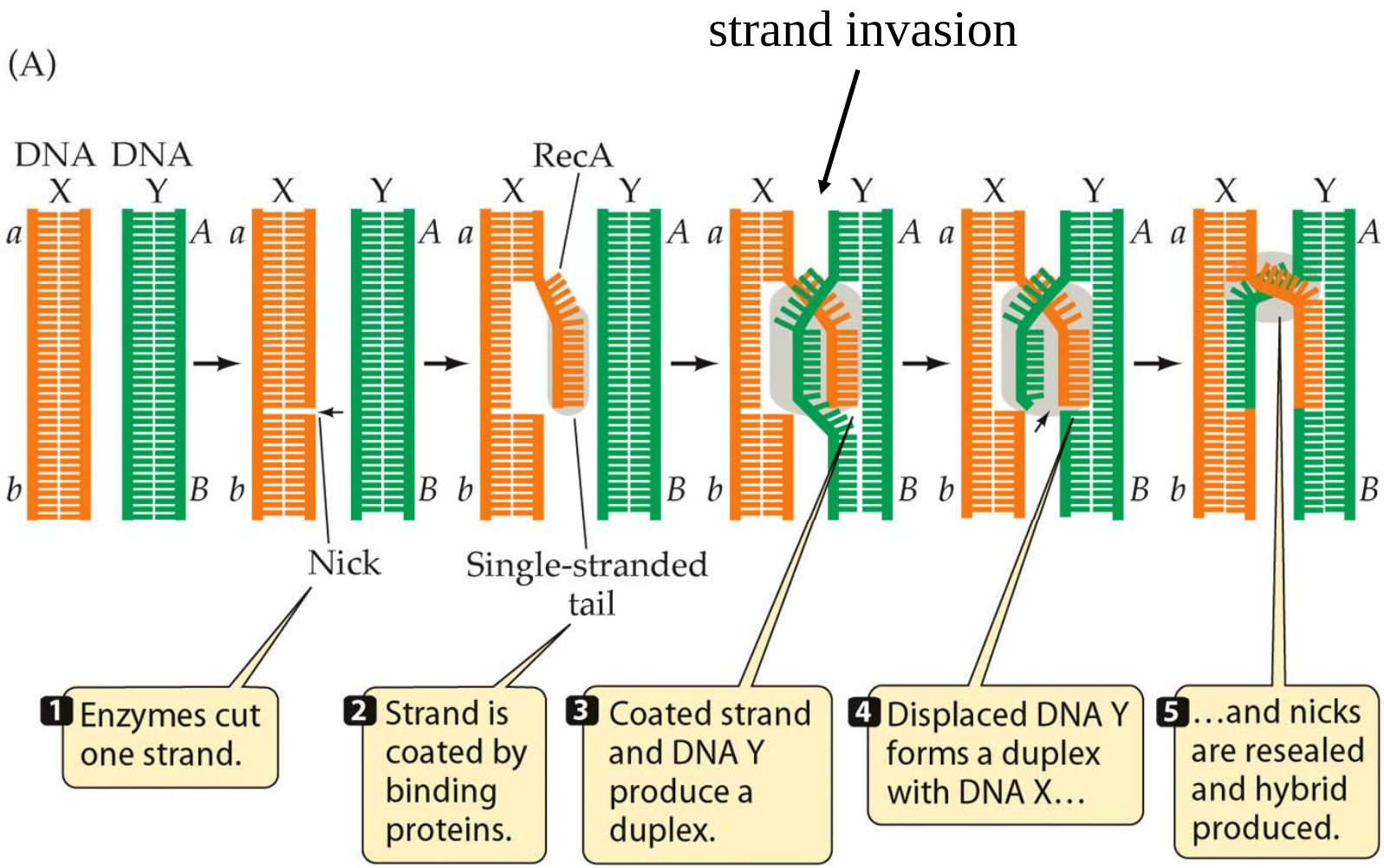
# F’ cells


Hfr cell
F plasmid
Chromosome with integrated F plasmid
A
B
C
D
E
F
F′ cell
F′
F plasmid carries genes A and B.
Chromosome without genes A and B
F′ cell
Recipient cell has genes A and B.
Recipient cell
F′ plasmid passes to recipient cell.
F pilus

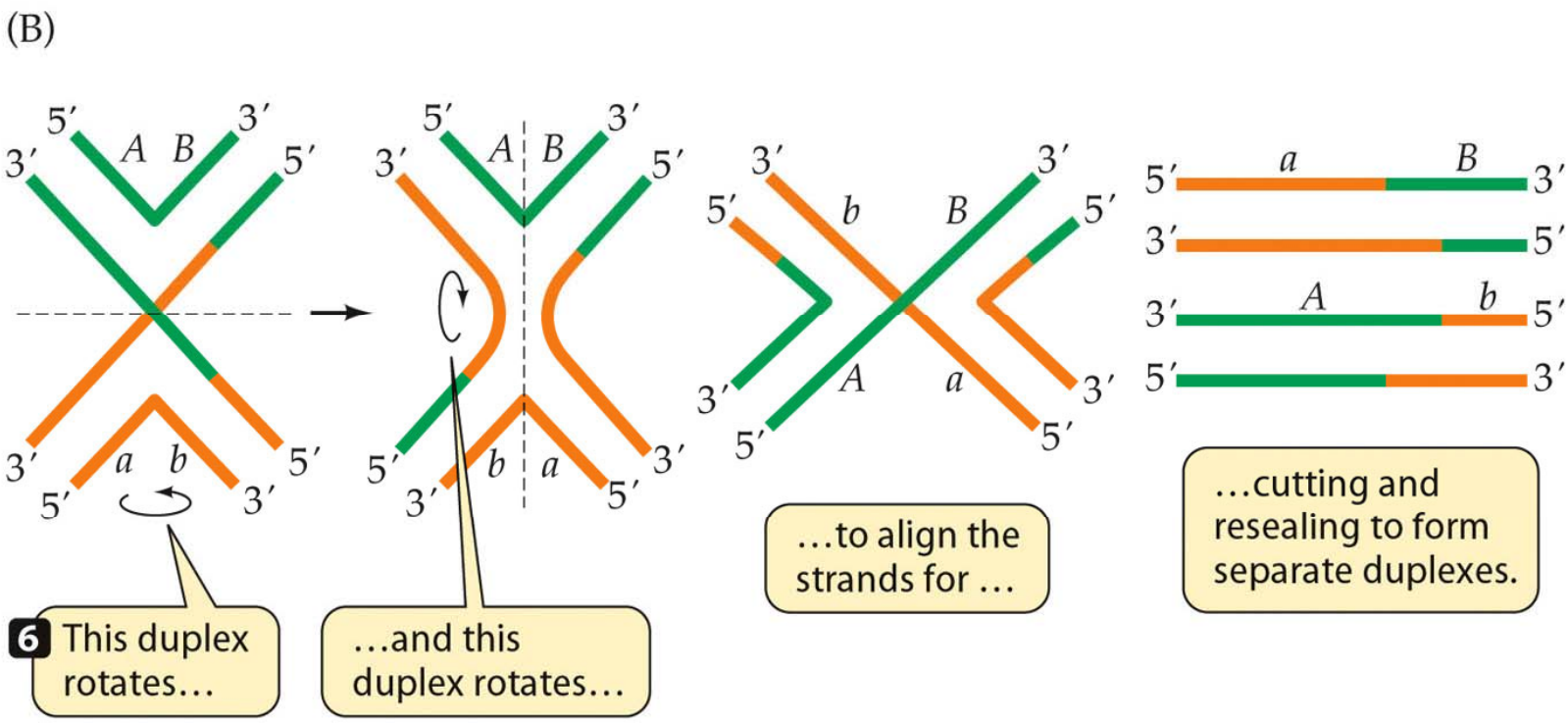
# Homologous Recombination



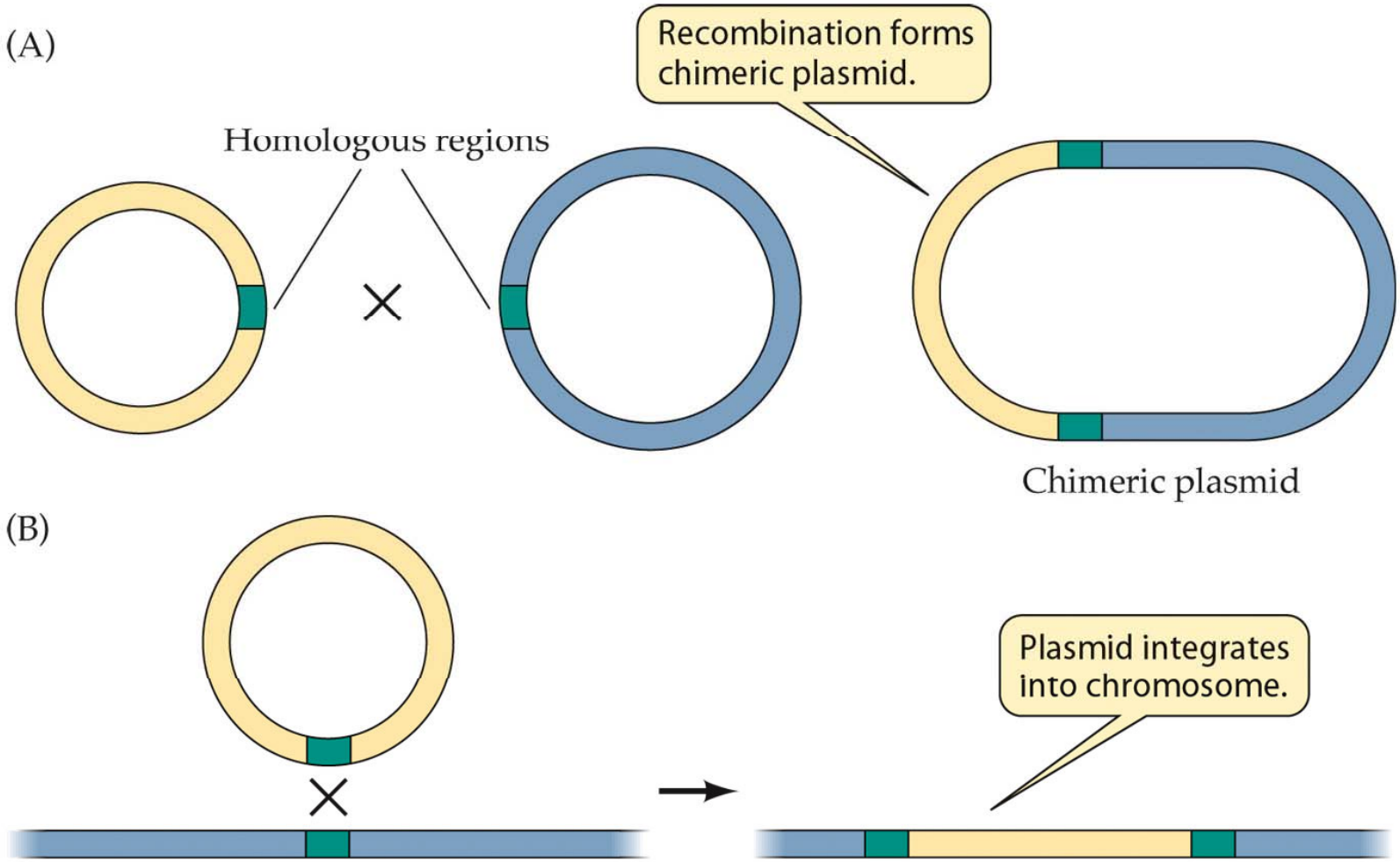
# Homologous Recombination



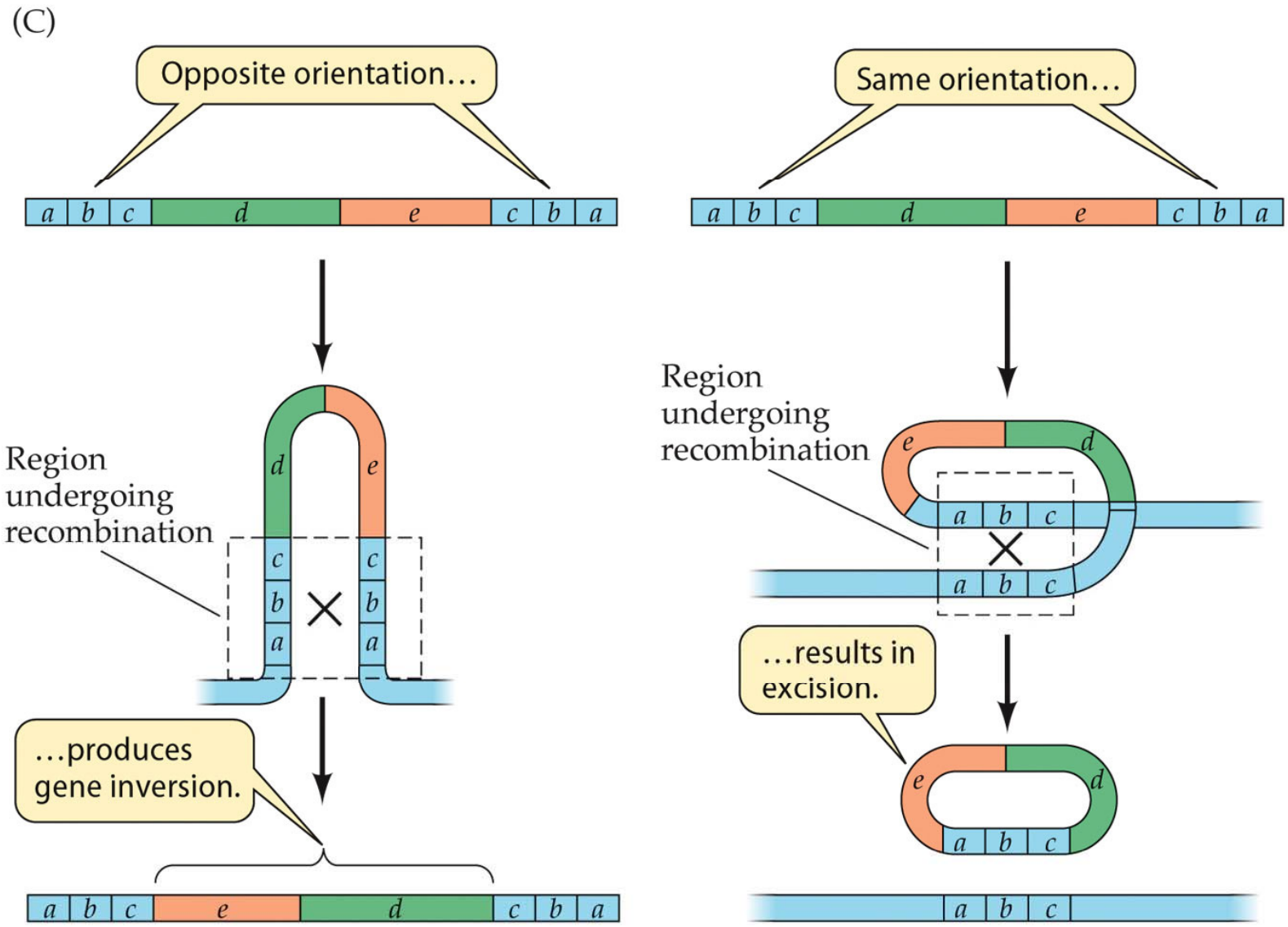
# Homologous Recombination



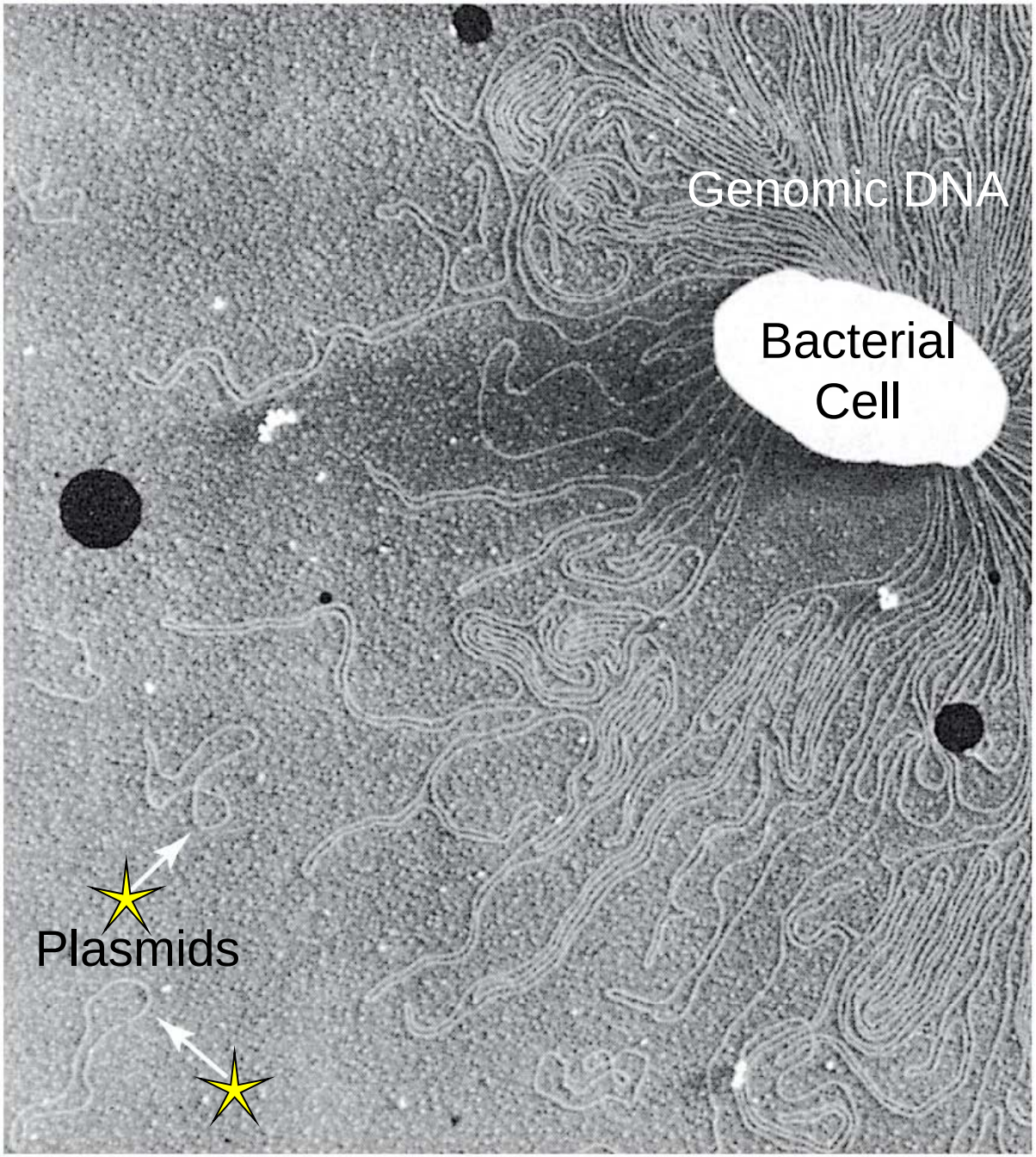
# Types of homologous recombination in bacteria

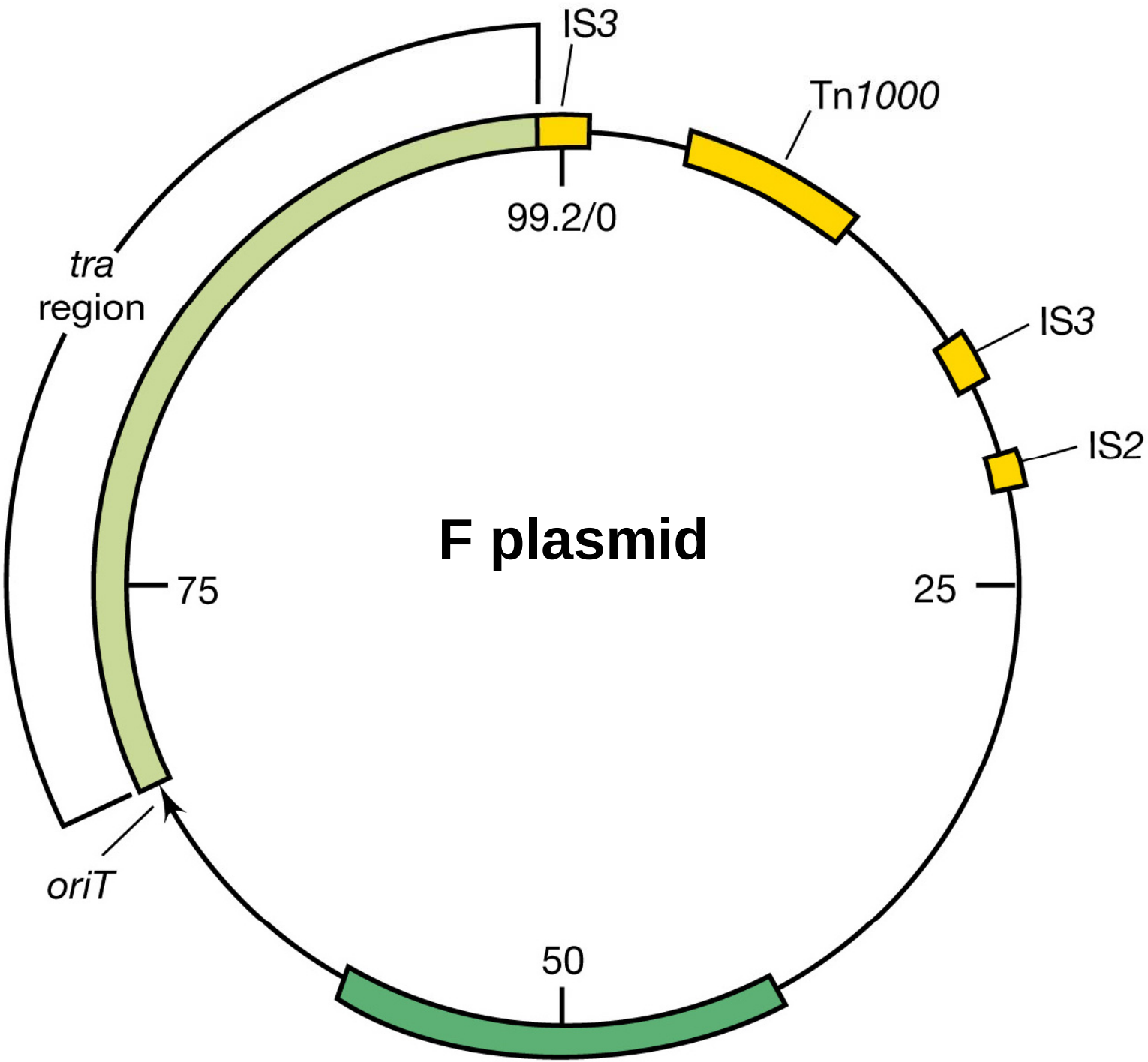


# Types of homologous recombination in bacteria

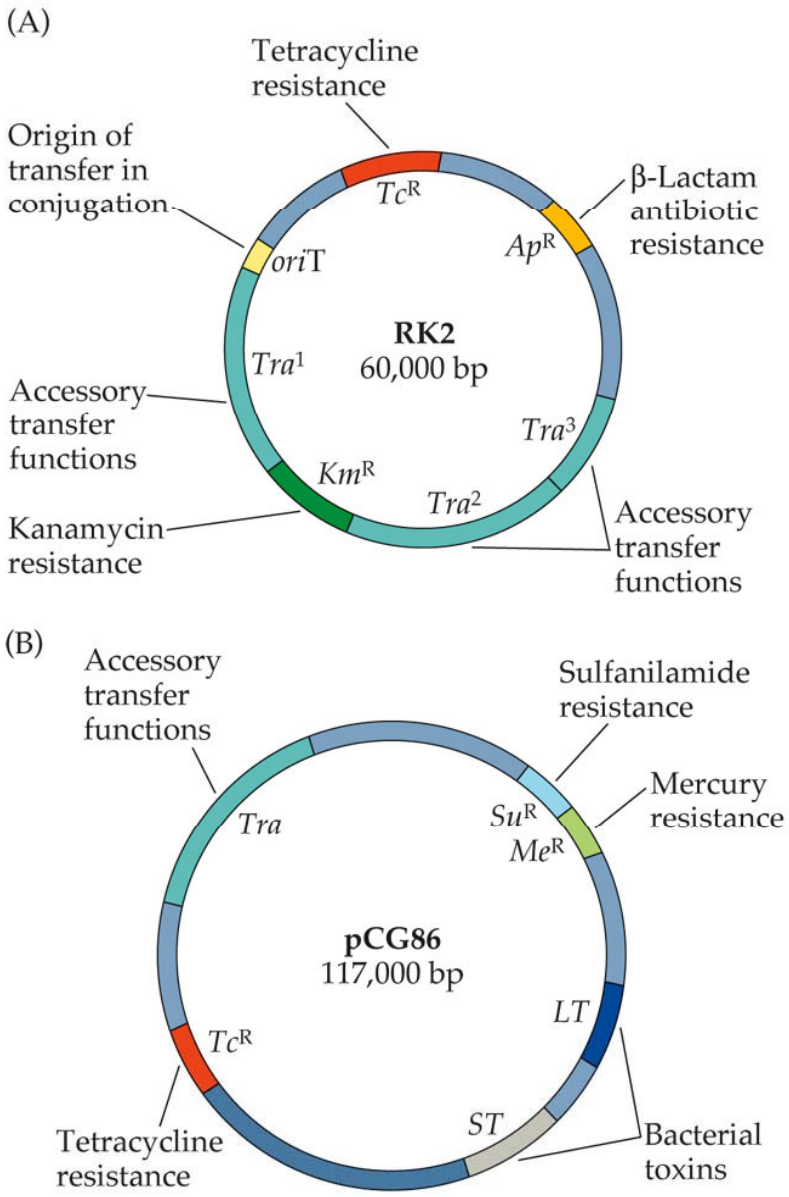


# Plasmids




IS3
Tn1000
99.2/0
tra
region
IS3
IS2
F plasmid
75
25
oriT
50

# R plasmids of pathogenic bacteria



# Ti plasmid


**Figure 25.18** **Crown gall.** Photograph of a crown gall tumor (arrow) on a tobacco plant caused by the crown gall bacterium *Agrobacterium tumefaciens*. The disease usually does not kill the plant but may weaken it and make it more susceptible to drought and diseases.

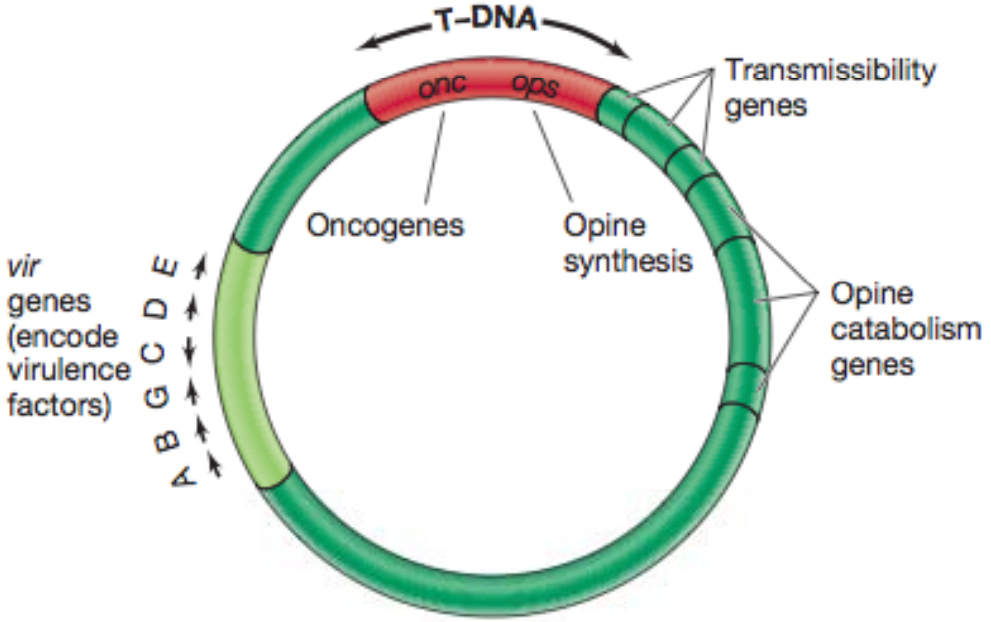



**Figure 25.19** **Structure of the Ti plasmid of *Agrobacterium tumefaciens*.** T-DNA is the region transferred to the plant. Arrows indicate the direction of transcription of each gene. The entire Ti plasmid is about 200 kbp of DNA and the T-DNA is about 20 kbp.

Brock, 13th ed. 2012, page 729

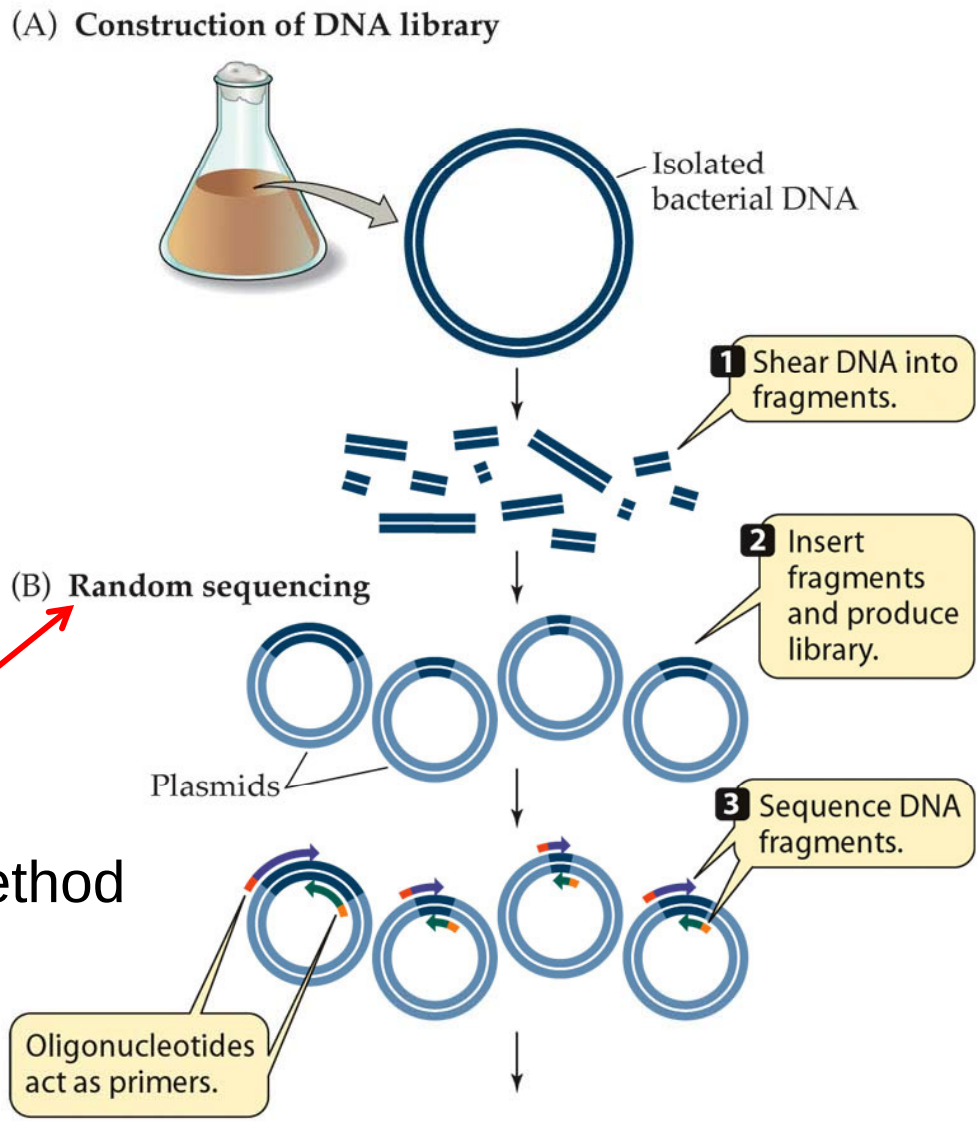
# Chromosomal & nonchromosomal genetic elements

**Table 7.1 Kinds of genetic elements**

| Organism | Element | Description |
|---|---|---|
| **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[a] | 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 |

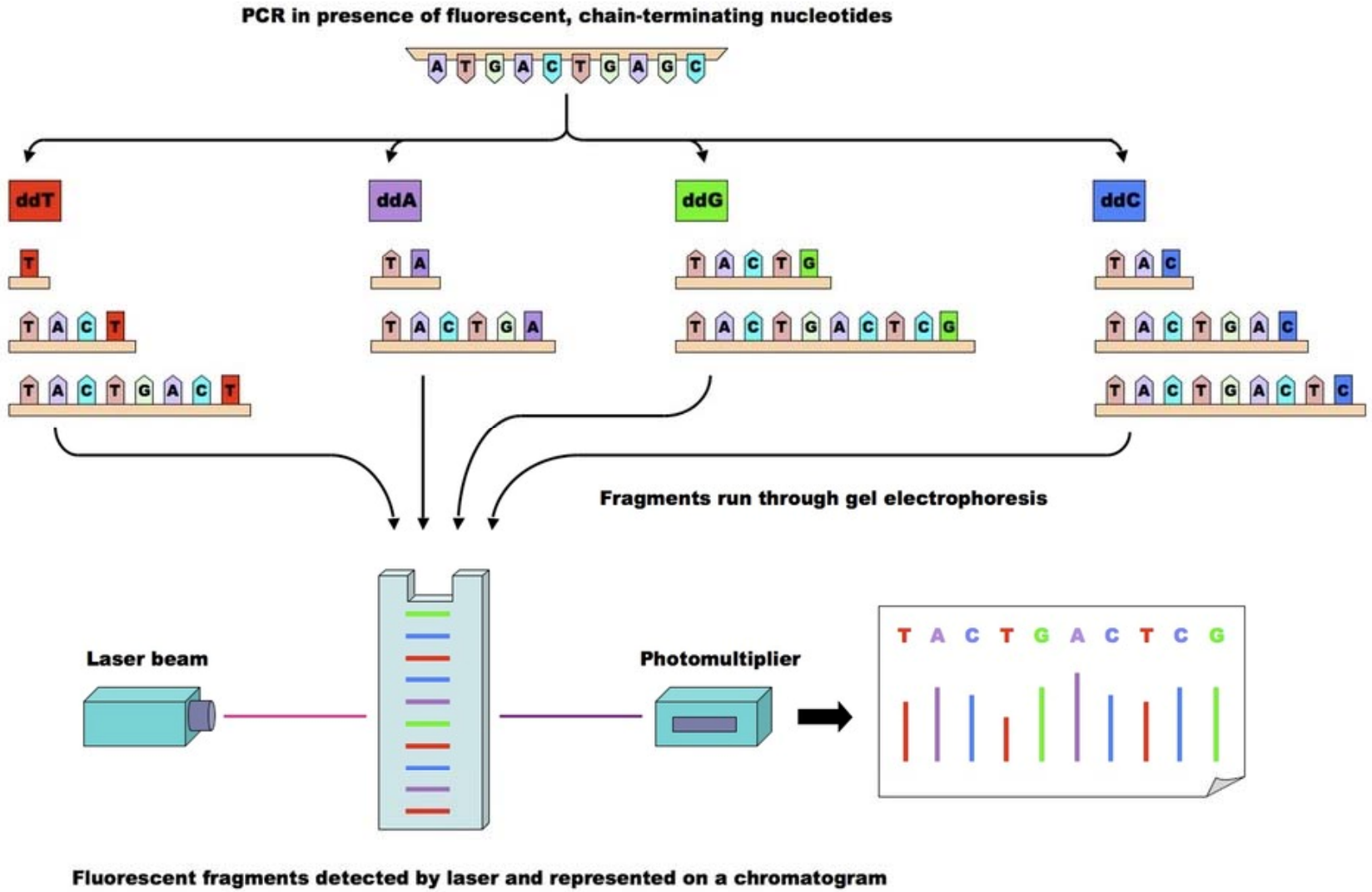
[a]Plasmids are uncommon in eukaryotes.

# Whole-genome shotgun sequencing

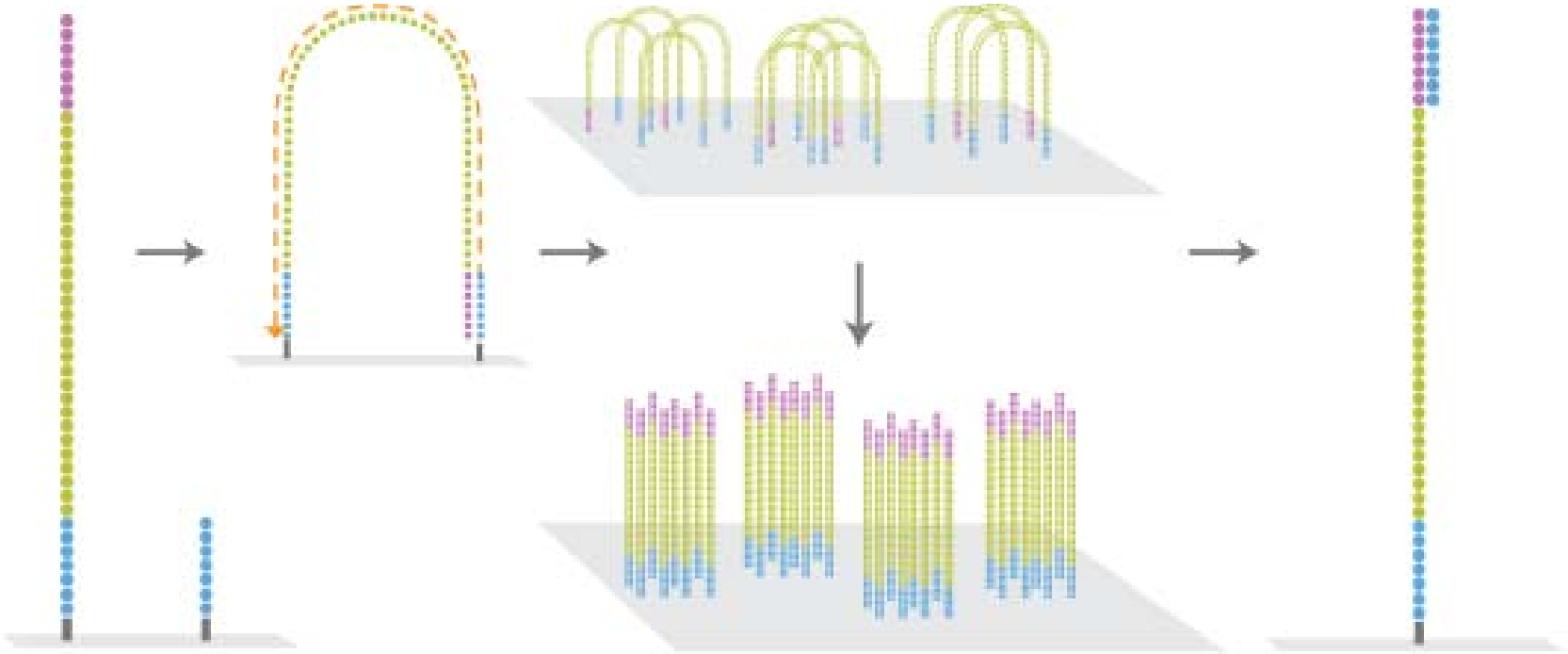



Depends on sequencing method

# Sanger Sequencing

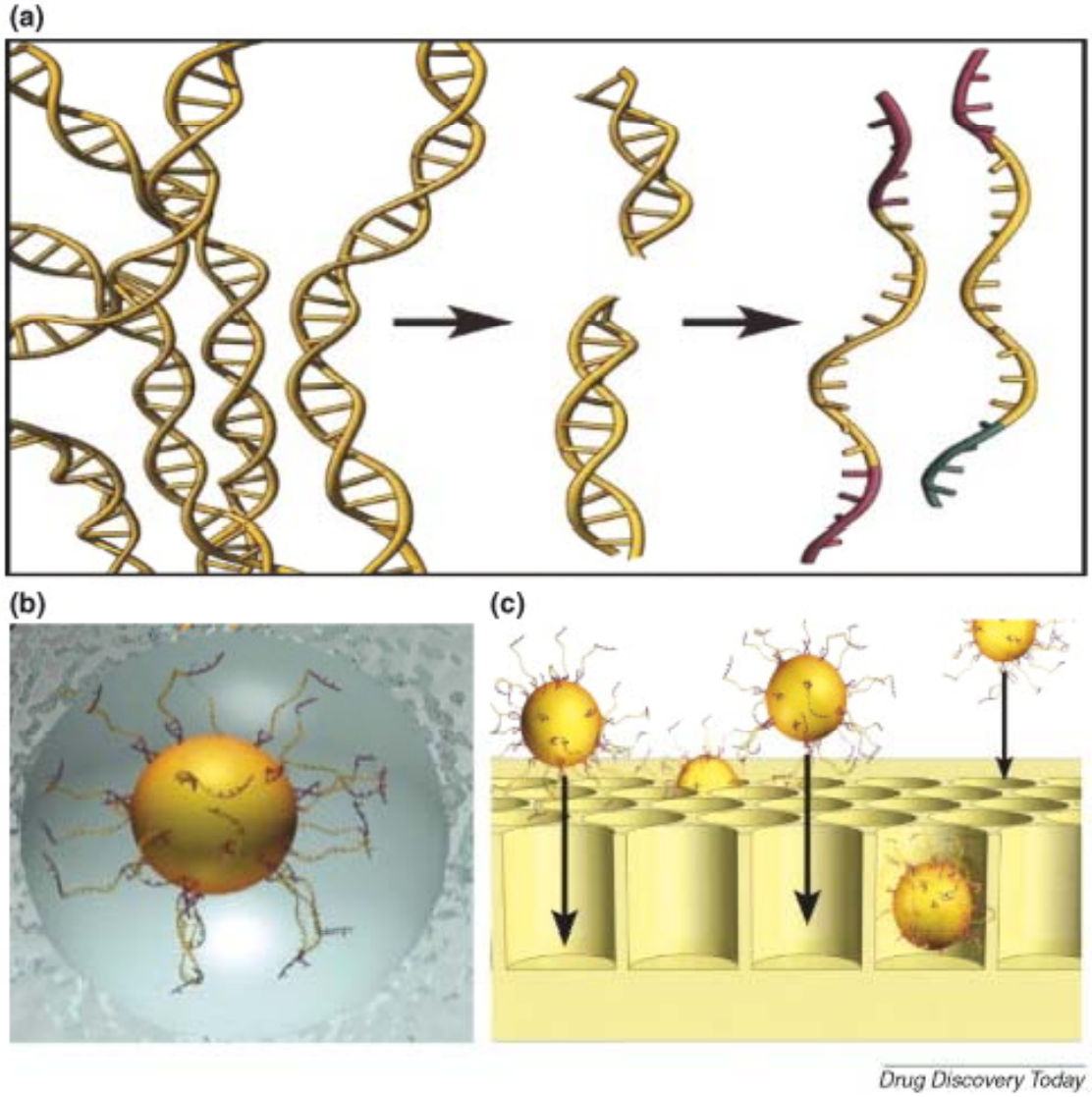


# Next-Generation Sequencing (NGS)

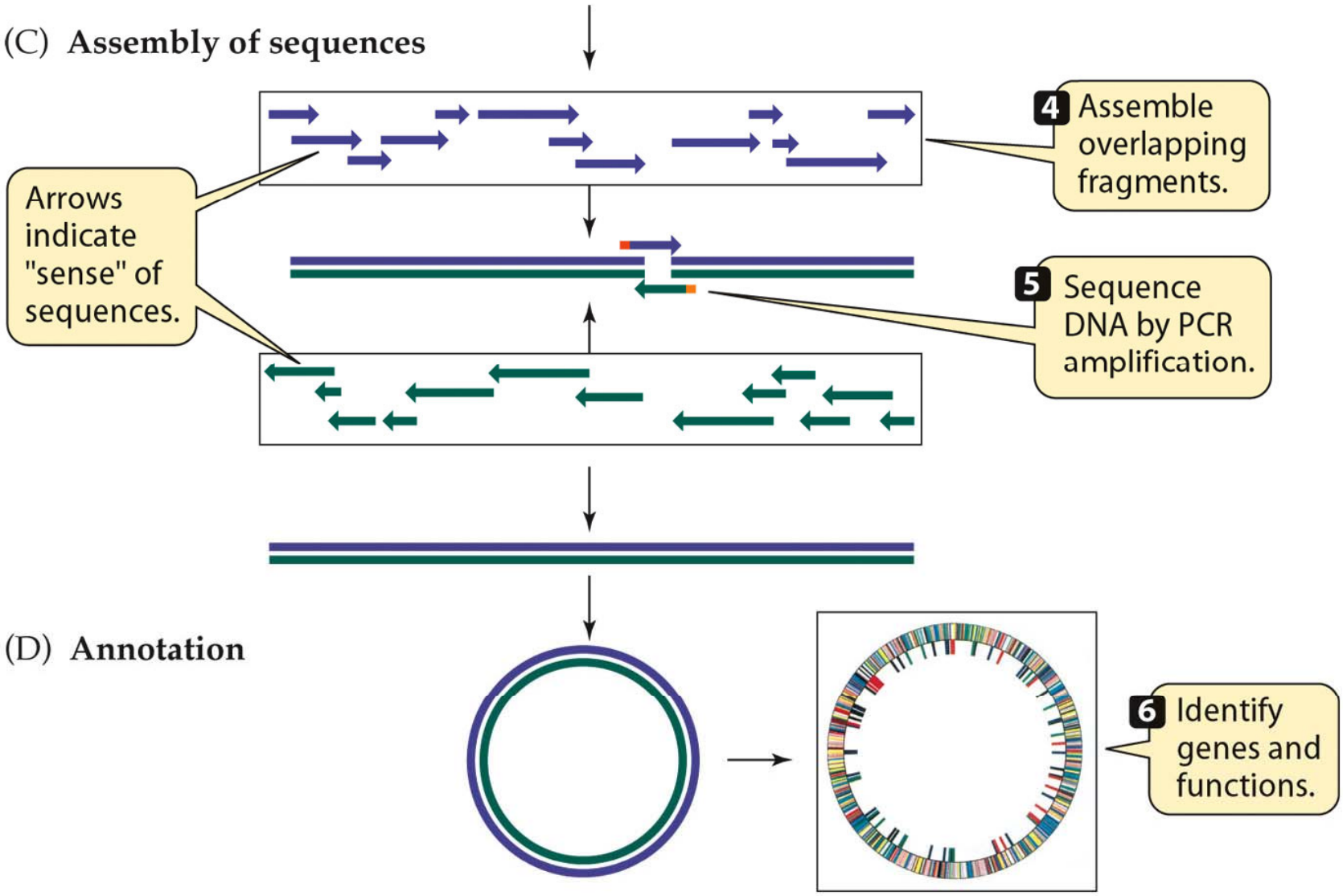


Illumina Platform

# Next-Generation Sequencing (NGS)

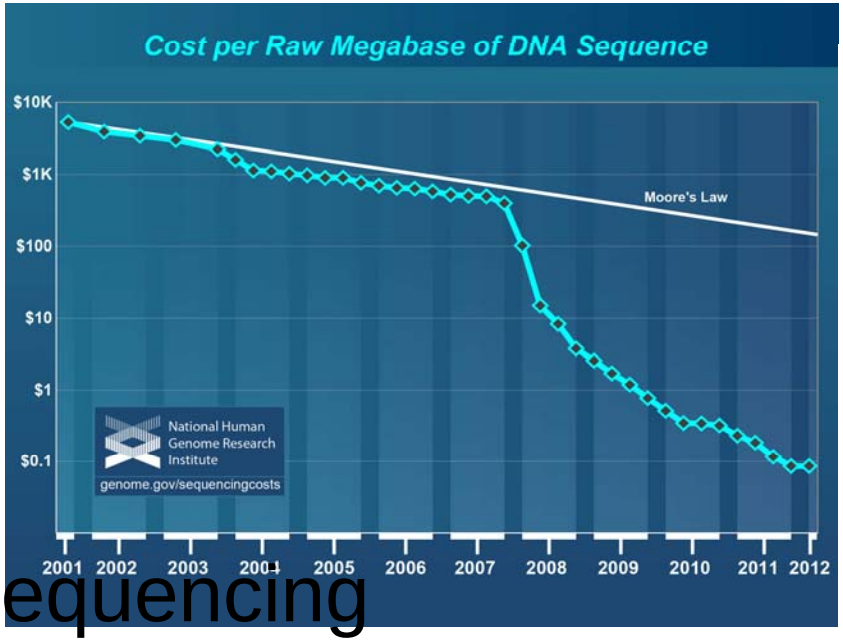



454 Platform

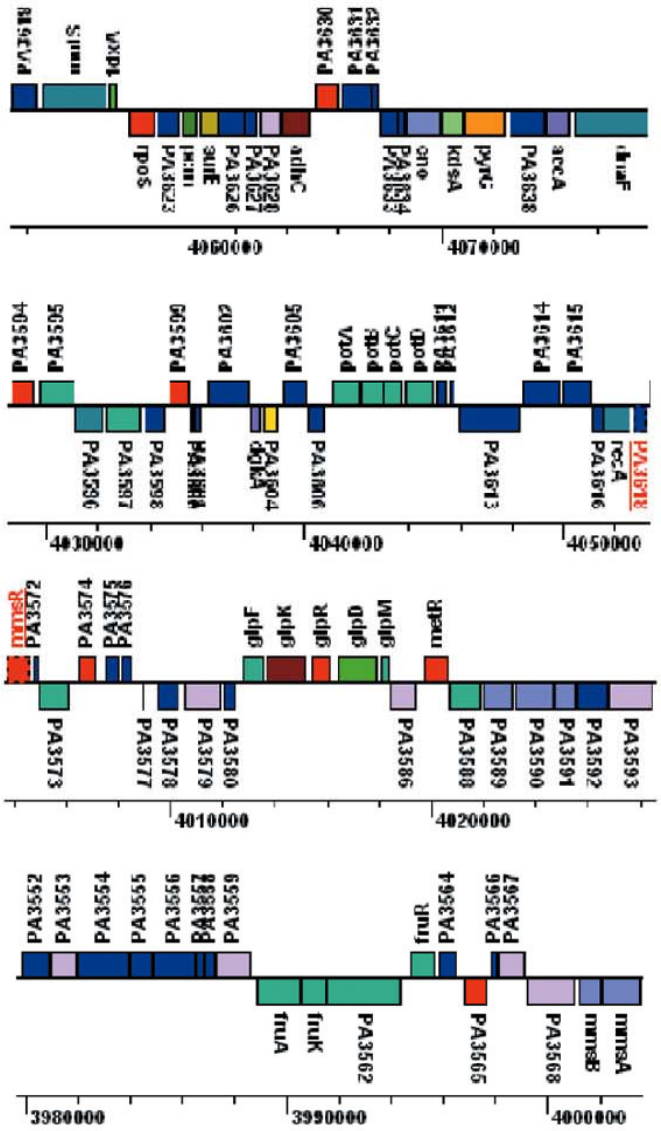

(C) Assembly of sequences
4 Assemble overlapping fragments.
Arrows indicate "sense" of sequences.
5 Sequence DNA by PCR amplification.
(D) Annotation
6 Identify genes and functions.

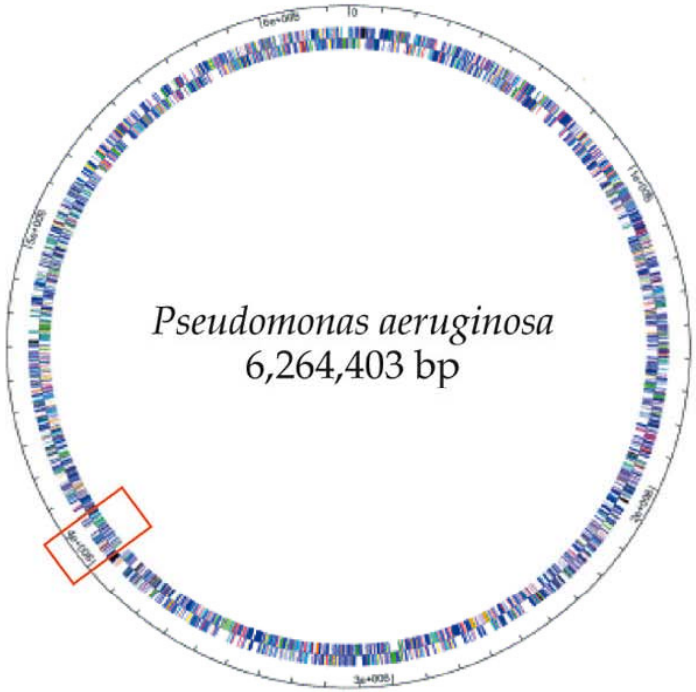
# Sequencing & Annotation

- General outline
  1. Find Sample
  2. Make Library-Depends on sequencing technology
  3. Assemble & Annotate
- Sanger vs. NGS
  - Sanger is expensive
  - NGS is computationally demanding for assembly
- Other methods for DNA sequencing include
  - Ion Torrent, PacBio, SOLiD

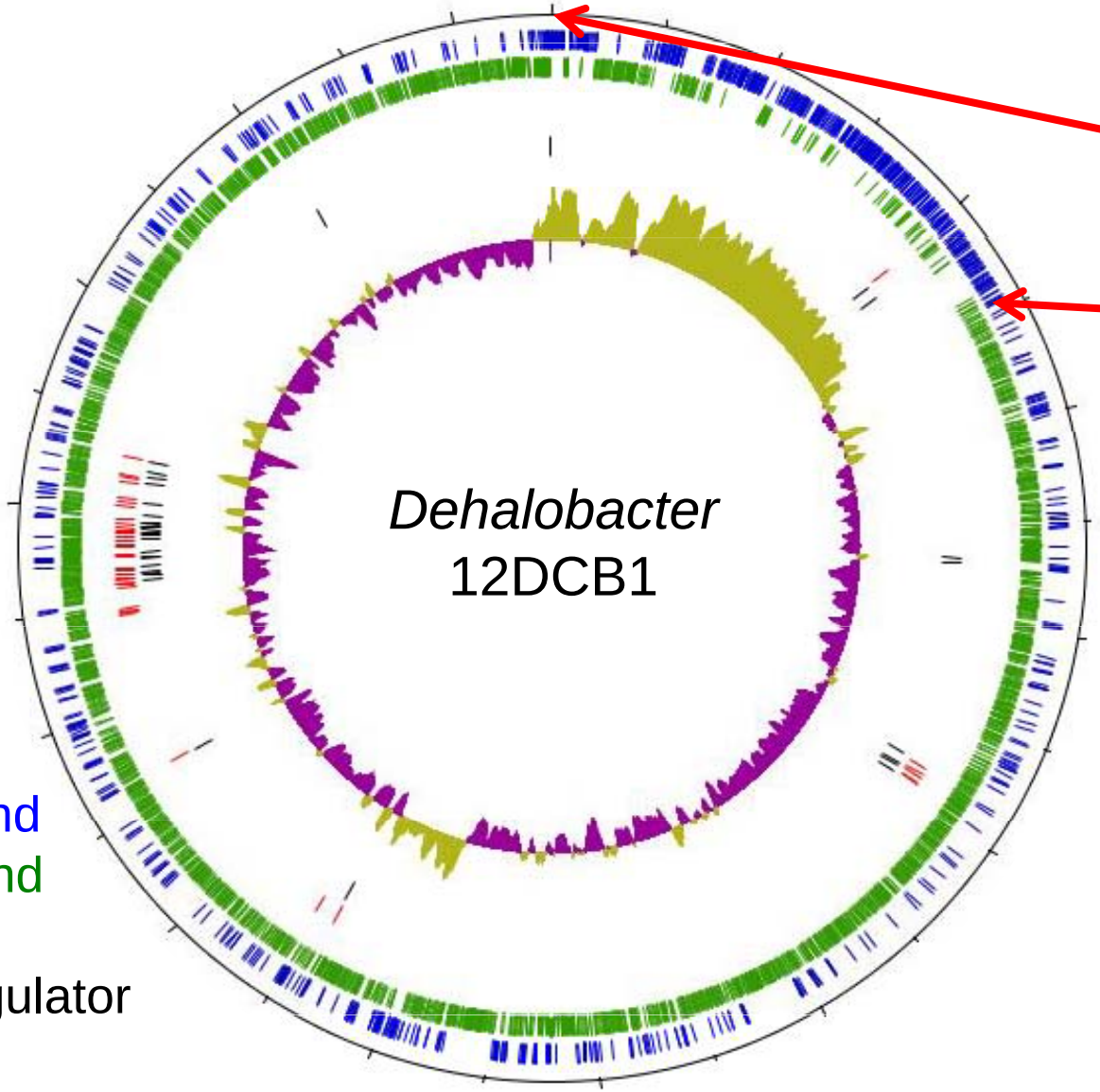


# Genes in a portion of bacterial genome



# Genome Organization




Forward Strand
Reverse Strand
rdh genes
CPR/FNR regulator
GC Skew
an asymmetry between the nucleotide compositions of the leading lagging strand

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

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

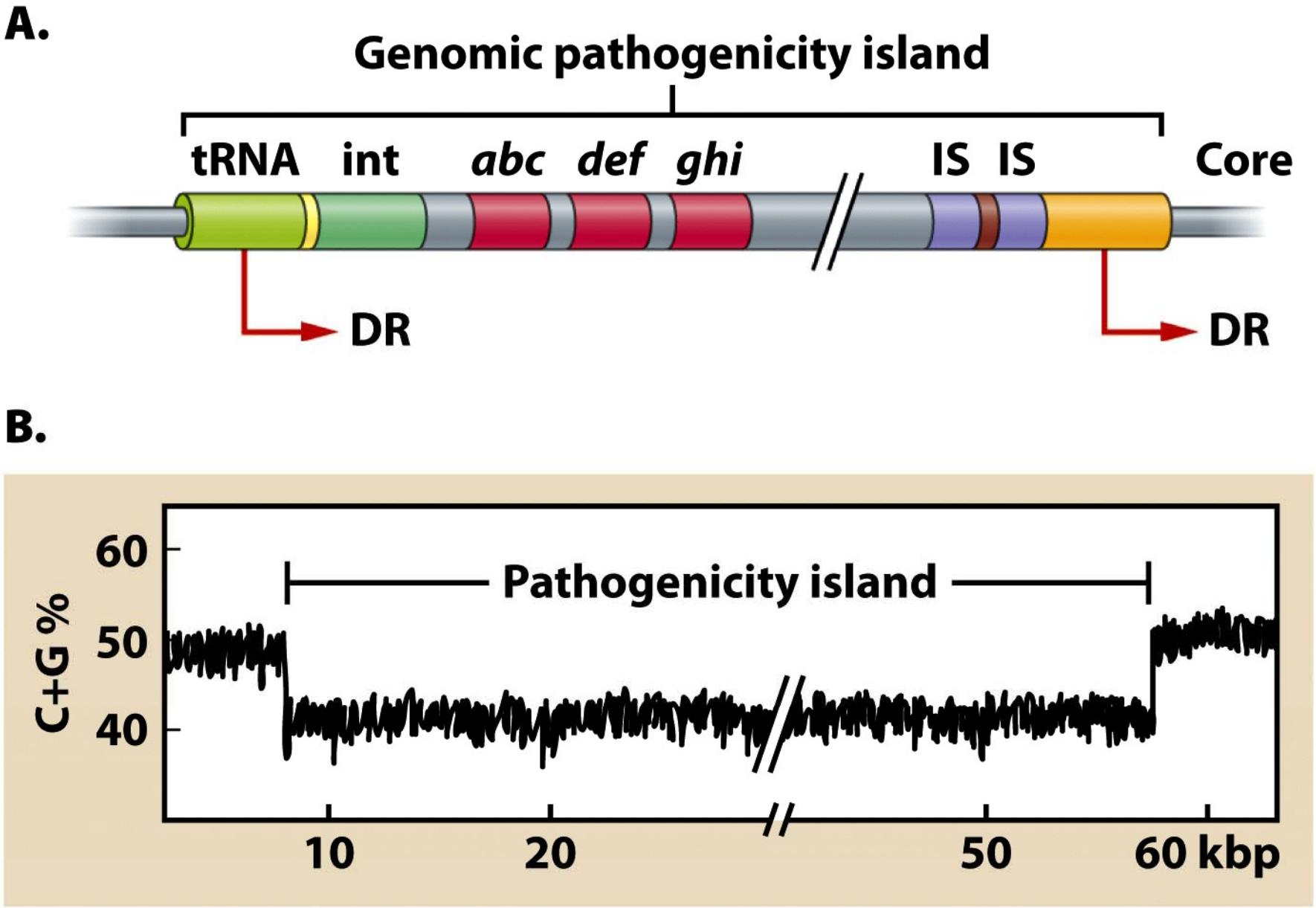



Figure 25.8 Microbiology: An Evolving Science

## Table 15.2 Gene function in bacterial genomes

| Functional categories | Percentage of genes on chromosome in that category | | |
|---|---|---|---|
| | *Escherichia coli* (4.64 Mbp)[a] | *Haemophilus influenzae* (1.83 Mbp)[a] | *Mycoplasma genitalium* (0.58 Mbp)[a] |
| 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 |

[a] Chromosome size. Each organism listed contains only a single circular chromosome.

# Genome size vs. ORFs

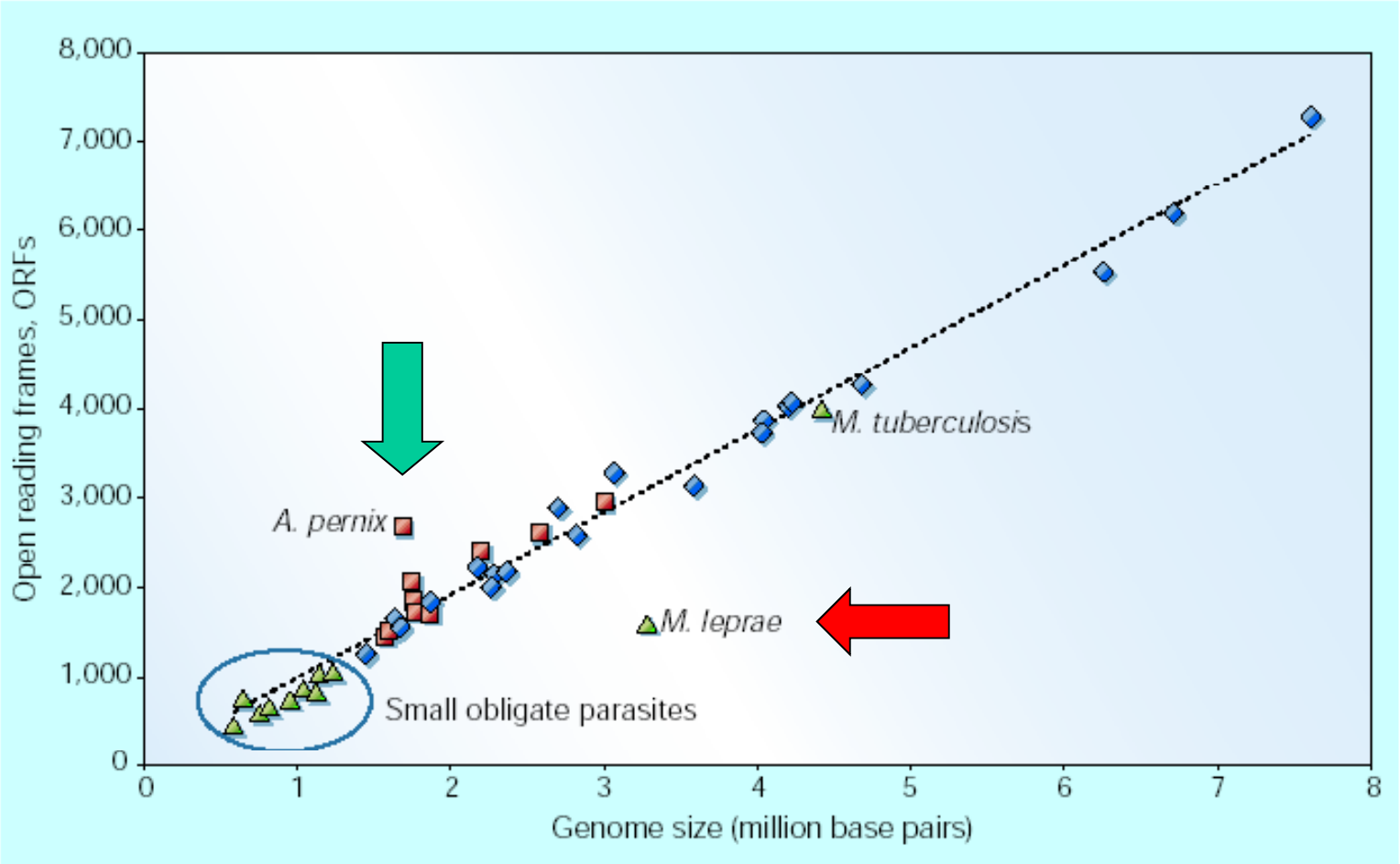



Figure 1 **Number of genes (ORFs) plotted against genome size for 44 fully sequenced genomes, including ten Archaea (squares) and 34 Bacteria. Obligate bacterial parasites are denoted by triangles; all other bacteria are shown as diamonds. *Mycobacterium leprae* is a genome 'in decay' that has a large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

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

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,[1,2*] Scott N. Peterson,[1*†] Steven R. Gill,[1] Robin T. Cline,[1] Owen White,[1] Claire M. Fraser,[1] Hamilton O. Smith,[1‡] J. Craig Venter[1‡§]

*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**

# Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome

Daniel G. Gibson, Gwynedd A. Benders, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B. Stockwell, Anushka Brownley, David W. Thomas, Mikkel A. Algire, Chuck Merryman, Lei Young, Vladimir N. Noskov, John I. Glass, J. Craig Venter, Clyde A. Hutchison III, Hamilton O. Smith*

We have synthesized a 582,970–base pair *Mycoplasma genitalium* genome. This synthetic genome, named *M. genitalium* JCVI-1.0, contains all the genes of wild-type *M. genitalium* G37 except MG408, which was disrupted by an antibiotic marker to block pathogenicity and to allow for selection. To identify the genome as synthetic, we inserted "watermarks" at intergenic sites known to tolerate transposon insertions. Overlapping "cassettes" of 5 to 7 kilobases (kb), assembled from chemically synthesized oligonucleotides, were joined by in vitro recombination to produce intermediate assemblies of approximately 24 kb, 72 k genome"), which were all cloned as bacterial artificial these intermediate clones were sequenced, and clones sequence were identified. The complete synthetic genc associated recombination cloning in the yeast *Sacchar* sequenced. A clone with the correct sequence was ide generally useful for constructing large DNA molecules from combinations of natural and synthetic DNA segm

Science, 2008

Science, 2010

# Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome

Daniel G. Gibson,[1] John I. Glass,[1] Carole Lartigue,[1] Vladimir N. Noskov,[1] Ray-Yuan Chuang,[1] Mikkel A. Algire,[1] Gwynedd A. Benders,[2] Michael G. Montague,[1] Li Ma,[1] Monzia M. Moodie,[1] Chuck Merryman,[1] Sanjay Vashee,[1] Radha Krishnakumar,[1] Nacyra Assad-Garcia,[1] Cynthia Andrews Pfannkoch,[1] Evgeniya A. Denisova,[1] Lei Young,[1] Zhi Qing Qi,[1] Thomas H. Segall-Shapiro,[1] Christopher H. Calvey,[1] Prashanth P. Parmar,[1] Clyde A. Hutchison III,[2] Hamilton O. Smith,[2] J. Craig Venter[1,2]*

We report the design, synthesis, and assembly of the 1.08–mega–base pair *Mycoplasma mycoides* JCVI-syn1.0 genome starting from digitized genome sequence information and its transplantation into a *M. capricolum* recipient cell to create new *M. mycoides* cells that are controlled only by the synthetic chromosome. The only DNA in the cells is the designed synthetic DNA sequence, including "watermark" sequences and other designed gene deletions and polymorphisms, and mutations acquired during the building process. The new cells have expected phenotypic properties and are capable of continuous self-replication.

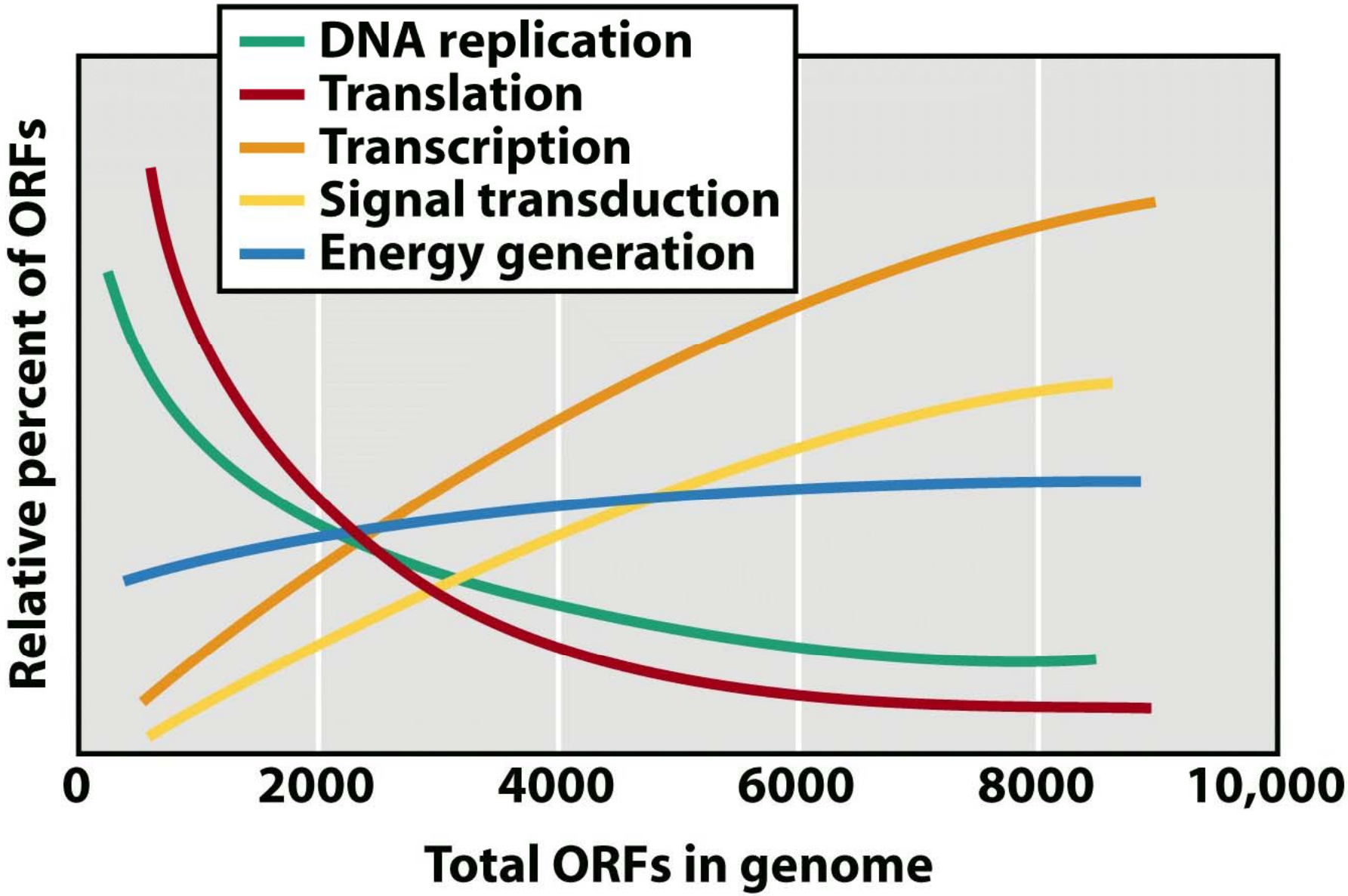
**Table 16.3** **Distribution of genes of unknown function among selected bacterial genomes** *(Part 1)*

| Organism | Genome Size (Mbp) | No. of ORFs (% coding) | | Unknown Function | | Unique ORFs | |
|---|---|---|---|---|---|---|---|
| *Aeropyrum pernix* K1 | 1.67 | 1,885 | (89%) | | | | |
| *A. aeolicus* VF5 | 1.50 | 1,749 | (93%) | 663 | (44%) | 407 | (27%) |
| *A. fulgidus* | 2.18 | 2,437 | (92%) | 1,315 | (54%) | 641 | (26%) |
| *B. subtilis* | 4.20 | 4,779 | (87%) | 1,722 | (42%) | 1,053 | (26%) |
| *B. burgdorferi* | 1.44 | 1,738 | (88%) | 1,132 | (65%) | 682 | (39%) |
| *Chlamydia pneumoniae* AR39 | 1.23 | 1,134 | (90%) | 543 | (48%) | 262 | (23%) |
| *Chlamydia trachomatis* $MoP_n$ | 1.07 | 936 | (91%) | 353 | (38%) | 77 | (8%) |
| *C. trachomatis serovar* D | 1.04 | 928 | (92%) | 290 | (32%) | 255 | (29%) |
| *Deinococcus radiodurans* | 3.28 | 3,187 | (91%) | 1,715 | (54%) | 1,001 | (31%) |
| *E. coli* K-12-MG1655 | 4.60 | 5,295 | (88%) | 1,632 | (38%) | 1,114 | (26%) |
| *H. influenzae* | 1.83 | 1,738 | (88%) | 595 | (35%) | 237 | (14%) |
| *H. pylori* 26695 | 1.66 | 1,589 | (91%) | 744 | (45%) | 539 | (33%) |
| *Methanobacterium thermotautotrophicum* | 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 | |
|---|---|---|---|---|---|---|---|
| *Methanococcus jannaschii* | 1.66 | 1,783 | (87%) | 1,076 | (62%) | 525 | (30%) |
| *M. tuberculosis* CSU#93 | 4.41 | 4,275 | (92%) | 1,521 | (39%) | 606 | (15%) |
| *M. genitalium* | 0.58 | 483 | (91%) | 173 | (37%) | 7 | (2%) |
| *M. pneumoniae* | 0.81 | 680 | (89%) | 248 | (37%) | 67 | (10%) |
| *N. meningitidis* MC58 | 2.24 | 2,155 | (83%) | 856 | (40%) | 517 | (24%) |
| *Pyrococcus horikoshii* OT3 | 1.74 | 1,994 | (91%) | 589 | (42%) | 453 | (22%) |
| *Rickettsia prowazekii* Madrid E | 1.11 | 878 | (75%) | 311 | (37%) | 209 | (25%) |
| *Synechocystis* sp. | 3.57 | 4,003 | (87%) | 2,384 | (75%) | 1,426 | (45%) |
| *T. maritma* MSB8 | 1.86 | 1,879 | (95%) | 863 | (46%) | 373 | (26%) |
| *T. pallidum* | 1.14 | 1,039 | (93%) | 461 | (44%) | 280 | (27%) |
| *Vibrio cholerae* El Tor N1696 | 4.03 | 3,890 | (88%) | 1,806 | (46%) | 934 | (24%) |
| **Totals:** | 50.60 | 52,462 | (89%) | 22.358 | (43%) | 12,161 | (23%) |

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


DNA replication
Translation
Transcription
Signal transduction
Energy generation
Relative percent of ORFs
0
2000
4000
6000
8000
10,000
Total ORFs in genome

# Lessons from full genomes

- Size range 600Kb to 12Mb
- Vast number of putative genes with no known function
- Pathogenicity can be 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

# Metagenomics

- Genomic analysis of pooled DNA from an environmental sample containing organisms that have not been isolated
- Describes the functional capacities of the community

# Community structure and metabolism through reconstruction of microbial genomes from the environment

**Gene W. Tyson[1], Jarrod Chapman[3,4], Philip Hugenholtz[1], Eric E. Allen[1], Rachna J. Ram[1], Paul M. Richardson[4], Victor V. Solovyev[4], Edward M. Rubin[4], Daniel S. Rokhsar[3,4] & Jillian F. Banfield[1,2]**

[1]*Department of Environmental Science, Policy and Management,* [2]*Department of Earth and Planetary Sciences, and* [3]*Department of Physics, University of California, Berkeley, California 94720, USA*
[4]*Joint Genome Institute, Walnut Creek, California 94598, USA*

**Microbial communities are vital in the functioning of all ecosystems; however, most microorganisms are uncultivated, and their roles in natural systems are unclear. Here, using random shotgun sequencing of DNA from a natural acidophilic biofilm, we report reconstruction of near-complete genomes of *Leptospirillum* group II and *Ferroplasma* type II, and partial recovery of three other genomes. This was possible because the biofilm was dominated by a small number of species populations and the frequency of genomic rearrangements and gene insertions or deletions was relatively low. Because each sequence read came from a different individual, we could determine that single-nucleotide polymorphisms are the predominant form of heterogeneity at the strain level. The *Leptospirillum* group II genome had remarkably few nucleotide polymorphisms, despite the existence of low-abundance variants. The *Ferroplasma* type II genome seems to be a composite from three ancestral strains that have undergone homologous recombination to form a large population of mosaic genomes. Analysis of the gene complement for each organism revealed the pathways for carbon and nitrogen fixation and energy generation, and provided insights into survival strategies in an extreme environment.**

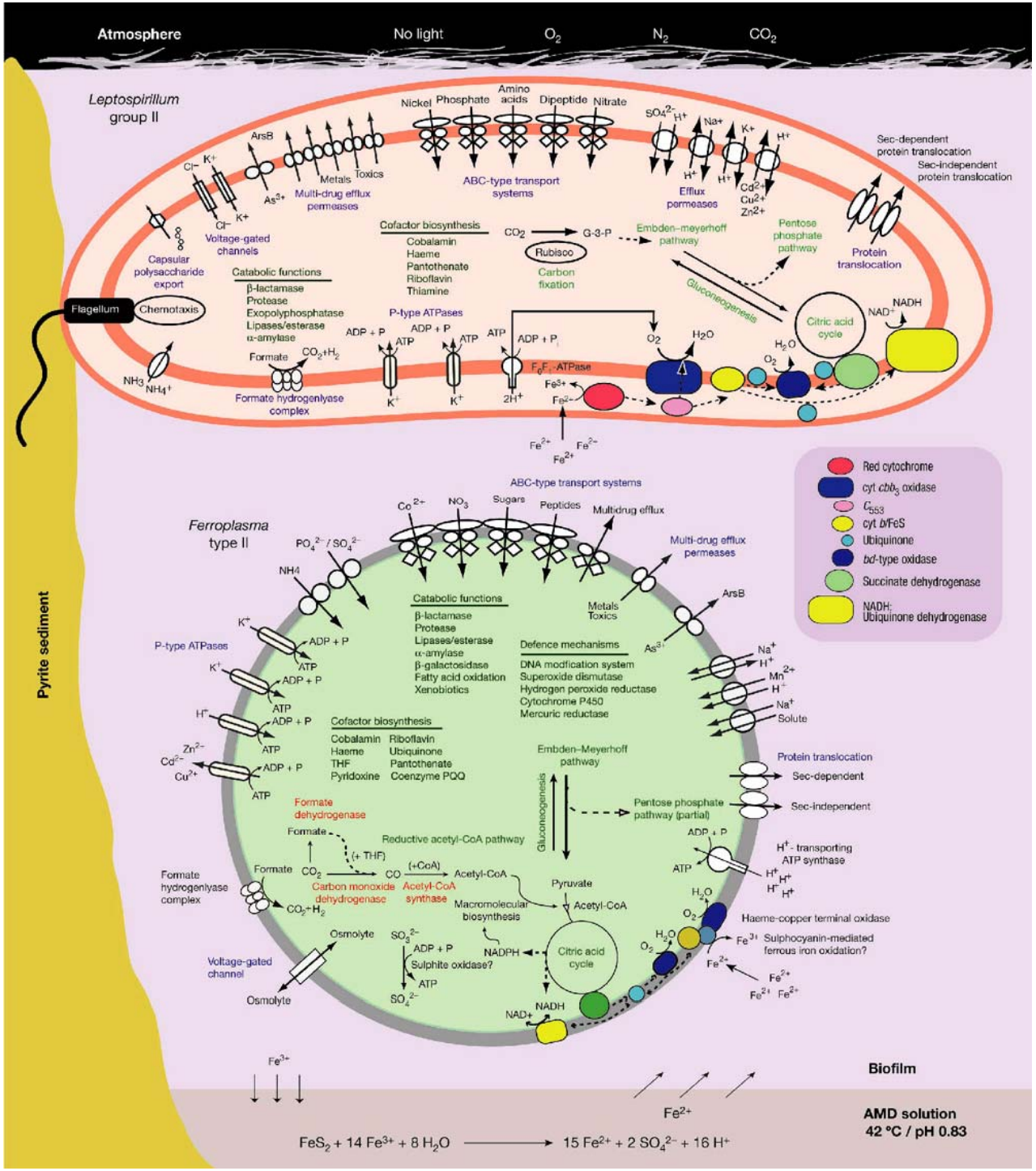



Figure 4: The cell cartoons are shown within a biofilm that is attached to the surface of an acid mine drainage stream (viewed in cross-section). Tight coupling between ferrous iron oxidation, pyrite dissolution and acid generation is indicated. Rubisco, ribulose 1,5-bisphosphate carboxylase–oxygenase. THF, tetrahydrofolate.

Tyson *et al* 2004 Nature

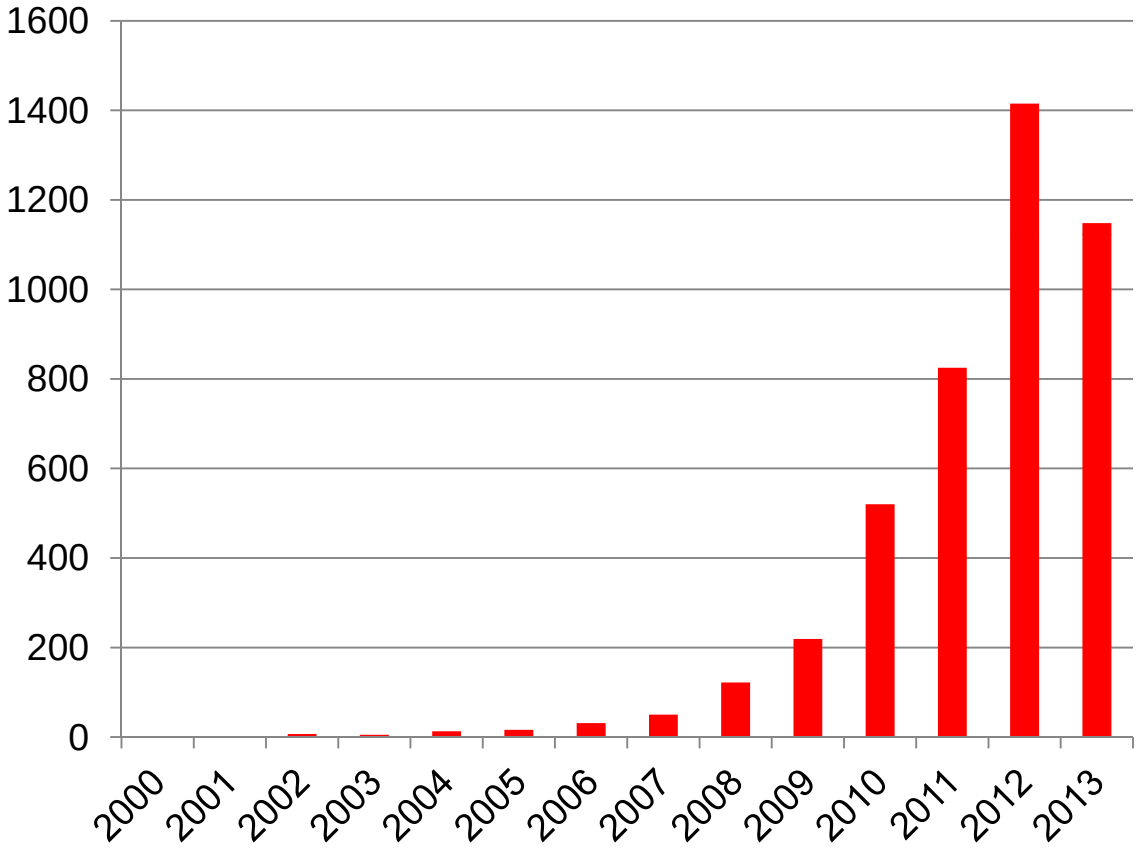
# Human Microbiome Project


NIH HUMAN MICROBIOME PROJECT

To demonstrate hypothesized correlations between the microbiome and human health and disease. These projects will leverage advances made by the HMP's large scale sequencing efforts to examine the relationship between changes in the human microbiome and diseases of interest.

http://www.hmpdacc.org/

# Human Microbiome




Implicated in

- Obesity
- Mental Illnesses
- Bacterial Vaginosis
- Various Cancers
- Inflammatory Bowel Disease
- Autism
- Arthritis
- Psoriasis
- Asthma

Pubmed: Human Microbiome Literature Search