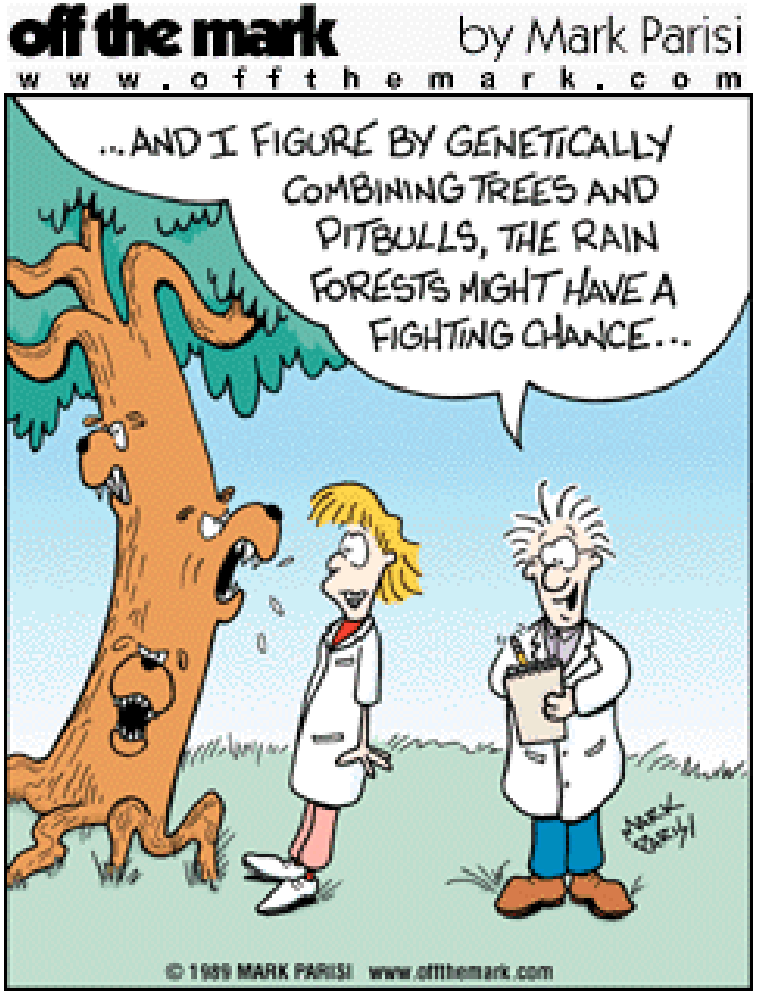
# Recombinant DNA Technology



# Definitions

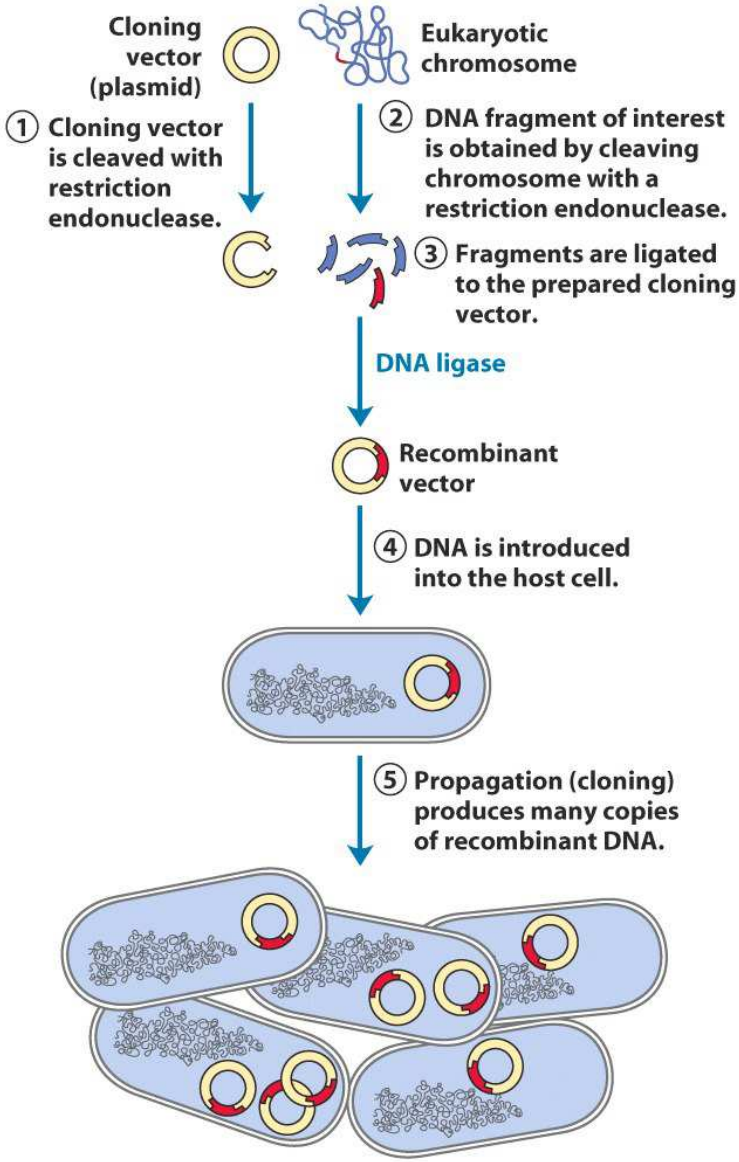
Recombinant DNA, a DNA construct created by fusing different fragments of DNA

Genetic Engineering, the deliberate alteration of DNA through the creation of recombinant DNA

Genetically Modified Organism, a living entity modified through genetic engineering

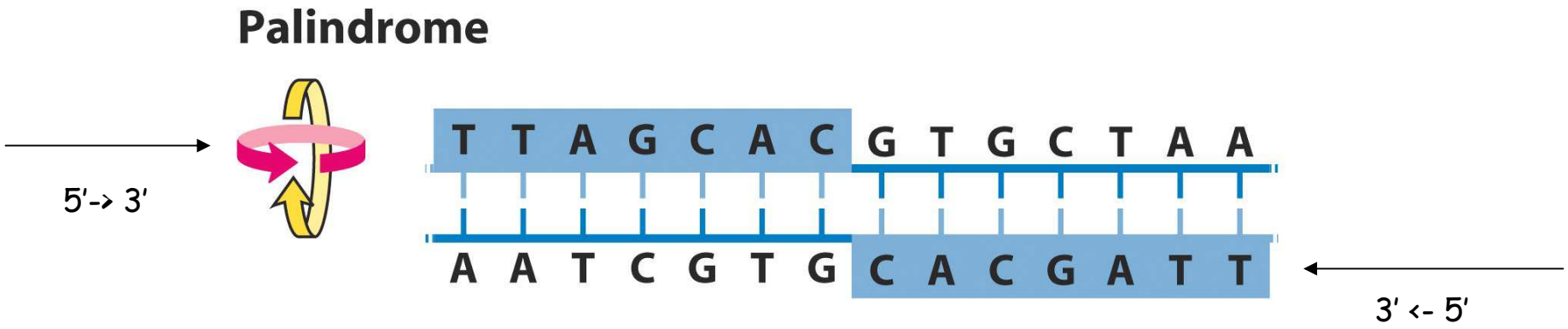
Transgenic, a genetically modified organism containing DNA from another source

# Recombinant DNA Technology




Clones -> Cells or organisms with identical DNA

# Restriction endonucleases



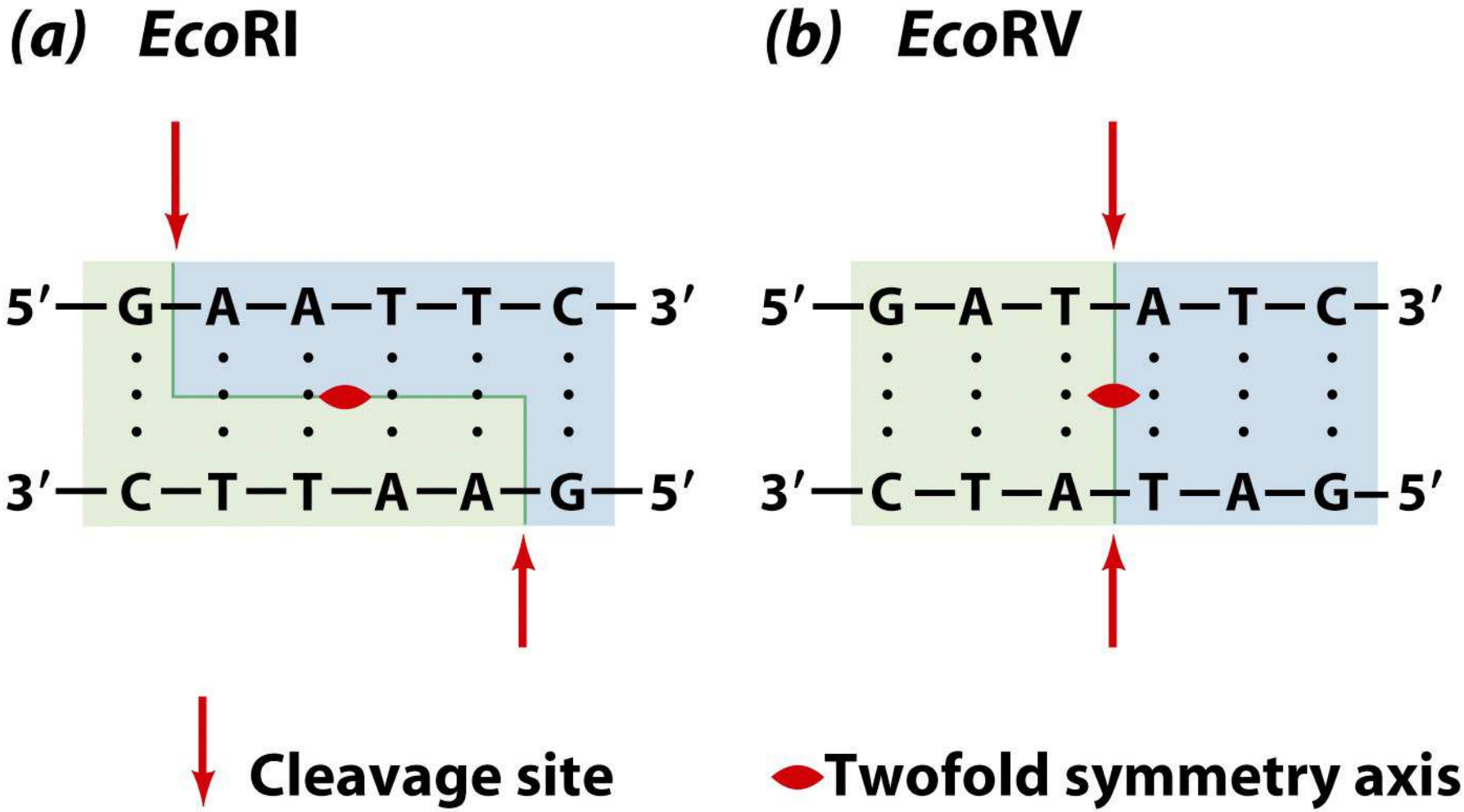
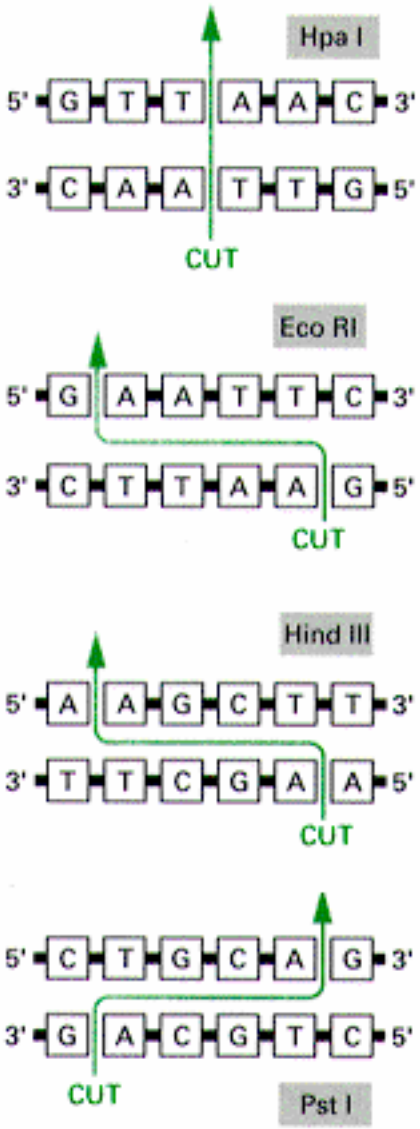



Figure 3-16 Fundamentals of Biochemistry, 2/e

How frequently will cut *Eco*RI in a random DNA sequence?



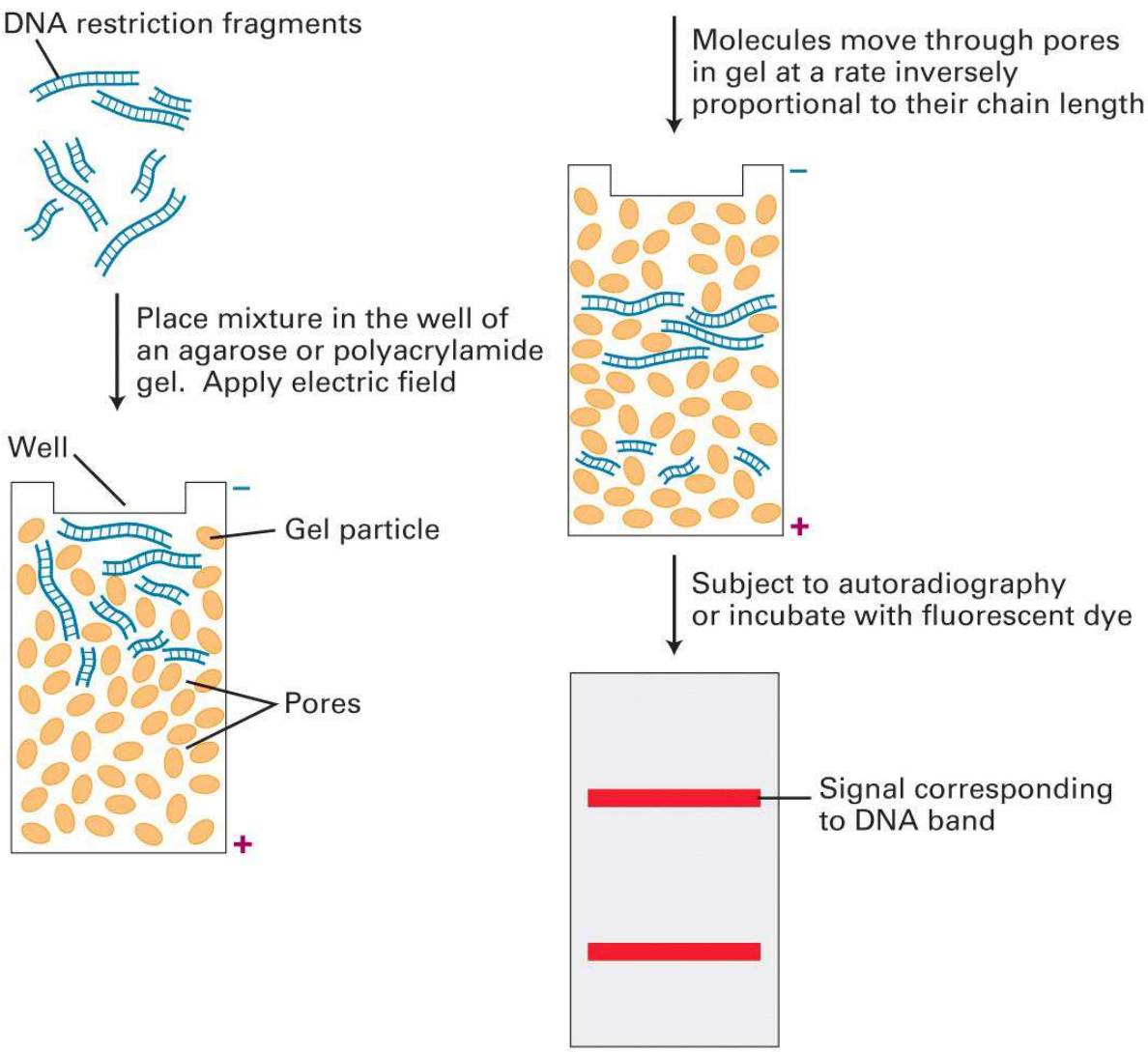
## Table 3-2 Recognition and Cleavage Sites of Some Restriction Enzymes

| Enzyme | Recognition Sequence[a] | Microorganism |
|---|---|---|
| *Alu*I | AG↓CT | *Arthrobacter luteus* |
| *Bam*HI | G↓GATCC | *Bacillus amyloliquefaciens* H |
| *Bgl*I | GCCNNNNN↓NGGC | *Bacillus globigii* |
| *Bgl*II | A↓GATCT | *Bacillus globigii* |
| *Eco*RI | G↓AATTC | *Escherichia coli* RY13 |
| *Eco*RII | ↓CC($\frac{A}{T}$)GG | *Escherichia coli* R245 |
| *Eco*RV | GAT↓ATC | *Escherichia coli* J62 pLG74 |
| *Hae*II | RGCGC↓Y | *Haemophilus aegyptius* |
| *Hae*III | GG↓CC | *Haemophilus aegyptius* |
| *Hin*dIII | A↓AGCTT | *Haemophilus influenzae* $R_d$ |
| *Hpa*II | C↓CGG | *Haemophilus parainfluenzae* |
| *Msp*I | C↓CGG | *Moraxella* species |
| *Pst*I | CTGCA↓G | *Providencia stuartii* 164 |
| *Pvu*II | CAG↓CTG | *Proteus vulgaris* |
| *Sal*I | G↓TCGAC | *Streptomyces albus* G |
| *Taq*I | T↓CGA | *Thermus aquaticus* |
| *Xho*I | C↓TCGAG | *Xanthomonas holcicola* |

[a]The recognition sequence is abbreviated so that only one strand, reading 5′ to 3′, is given. The cleavage site is represented by an arrow (↓). R, Y, and N represent a purine nucleotide, a pyrimidine nucleotide, and any nucleotide, respectively.

*Source:* Roberts, R.J. and Macelis, D., REBASE—the restriction enzyme database, http://rebase.neb.com.

# Gel Electrophoresis



# Gel Electrophoresis

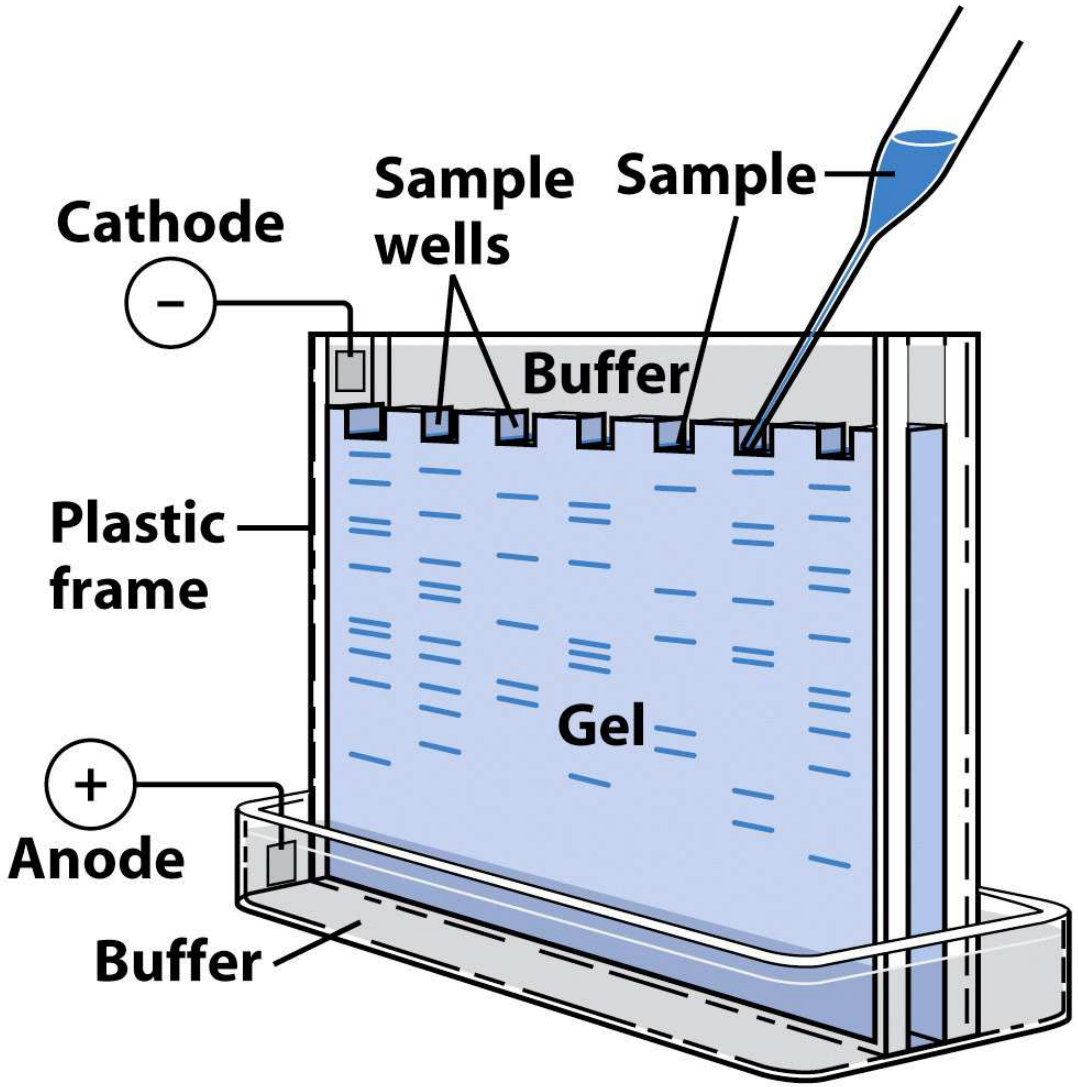
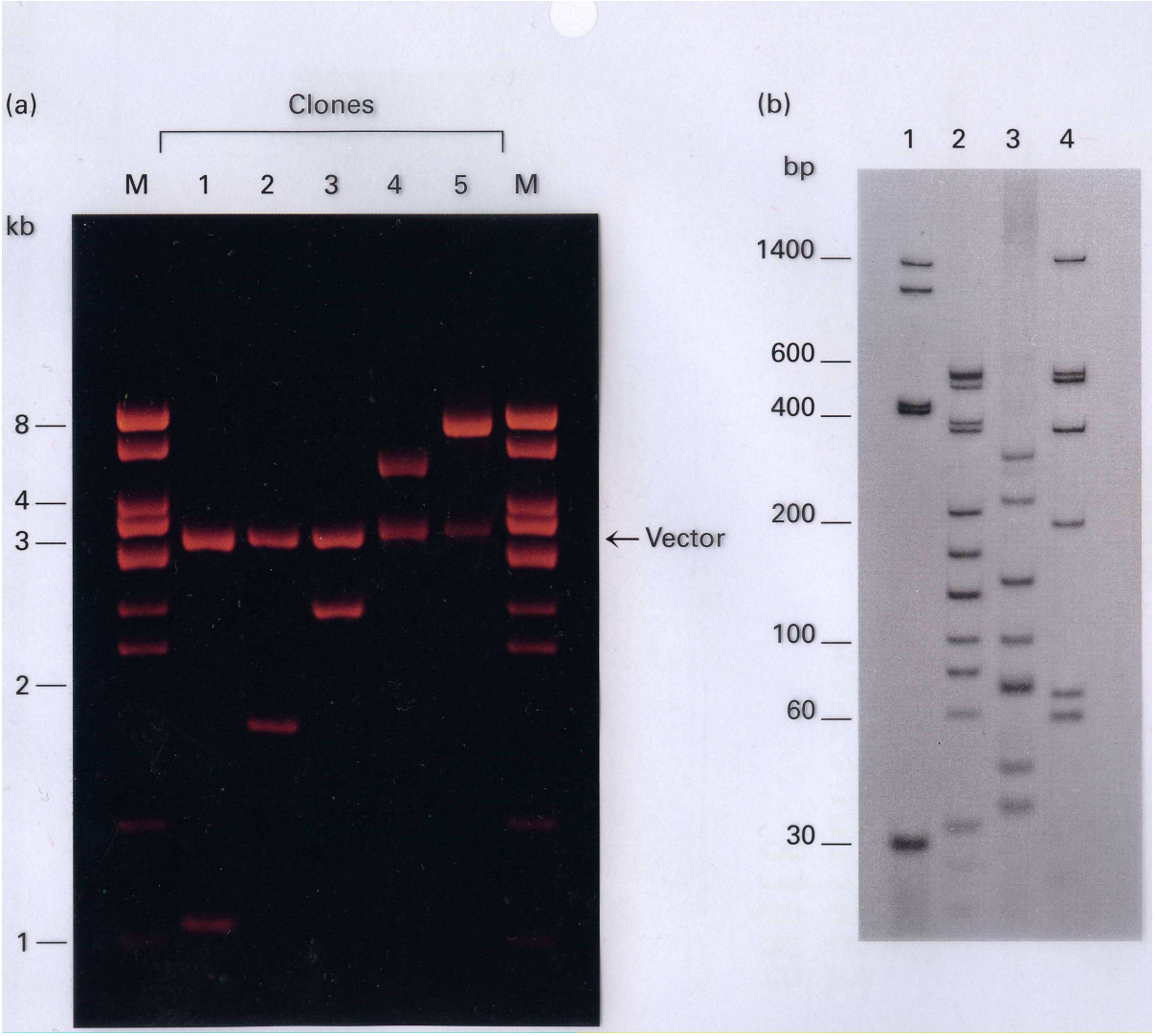
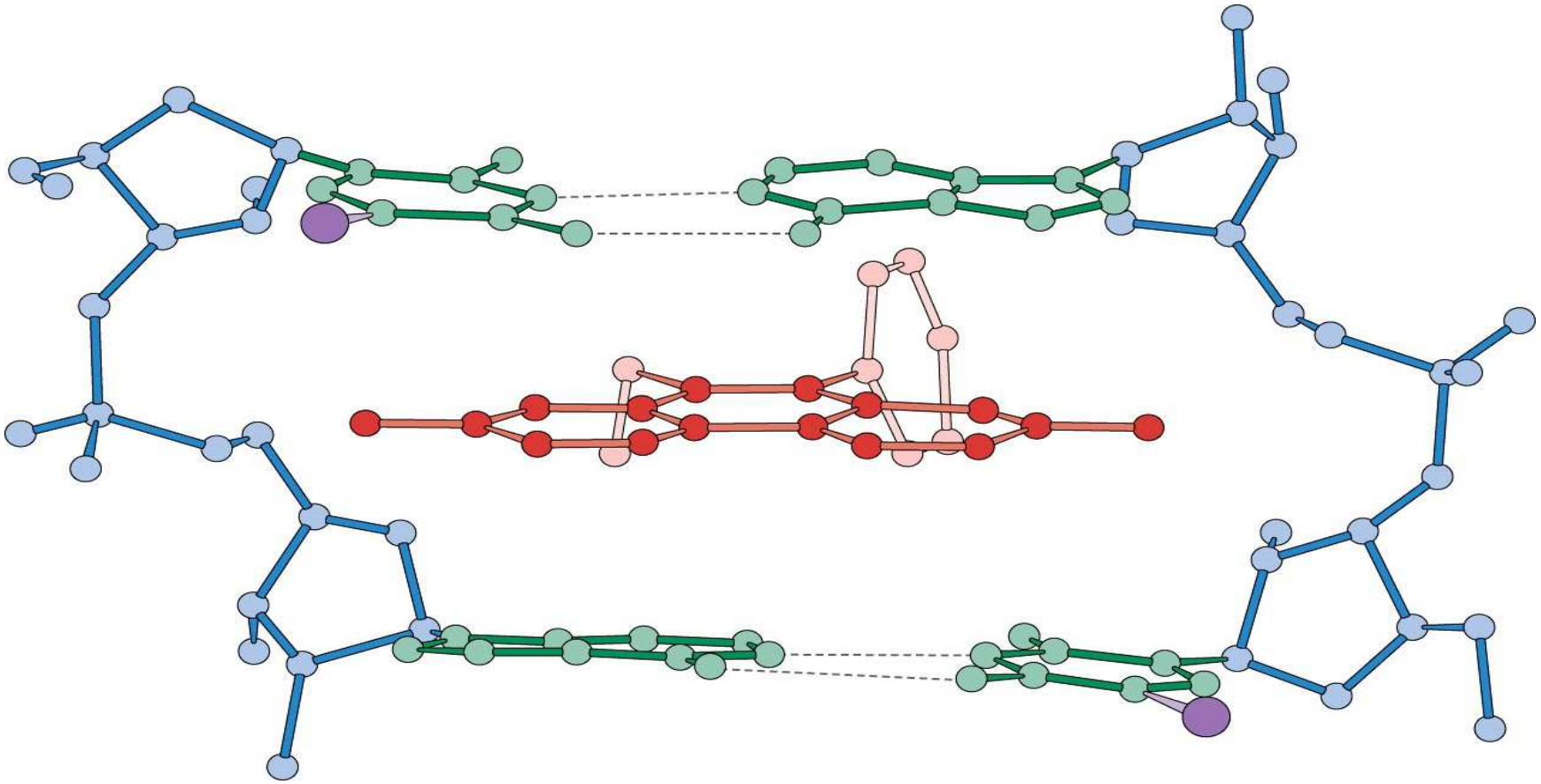



Figure 3-17 Fundamentals of Biochemistry, 2/e

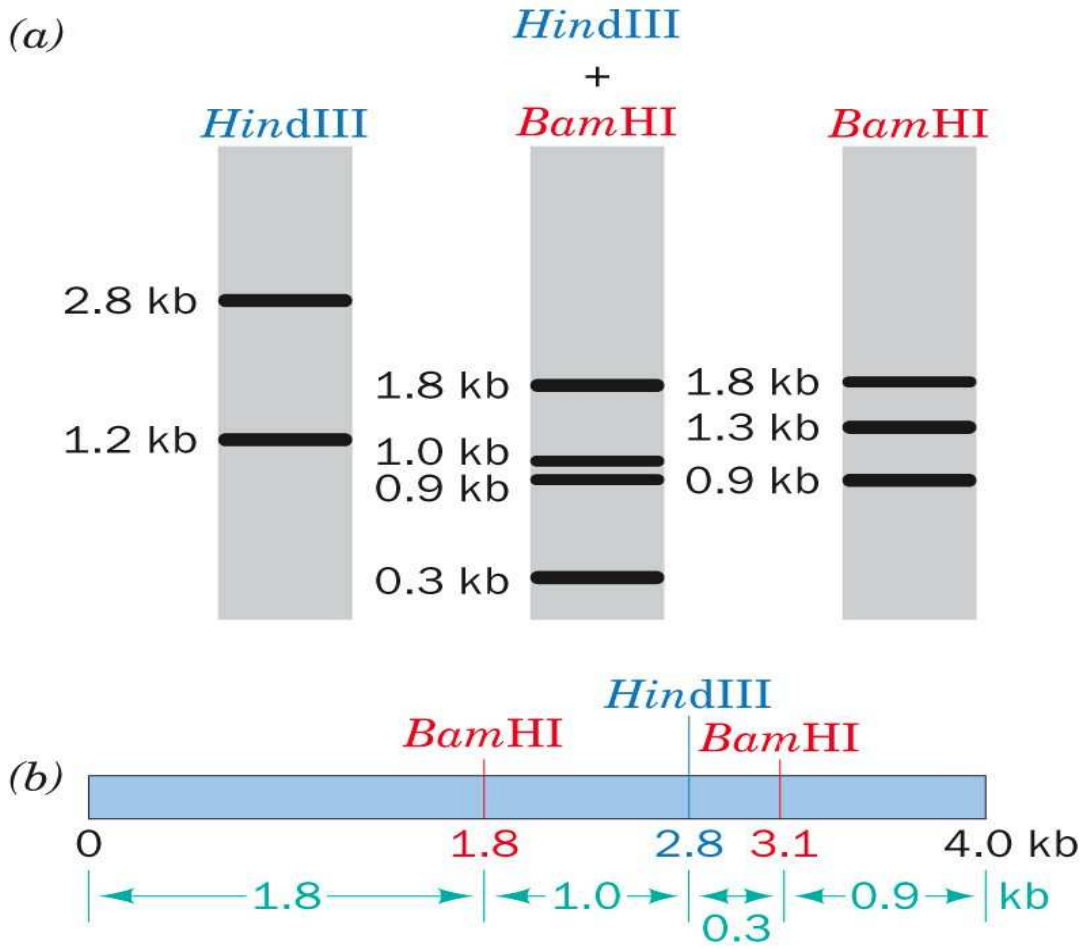
# Gel Electrophoresis



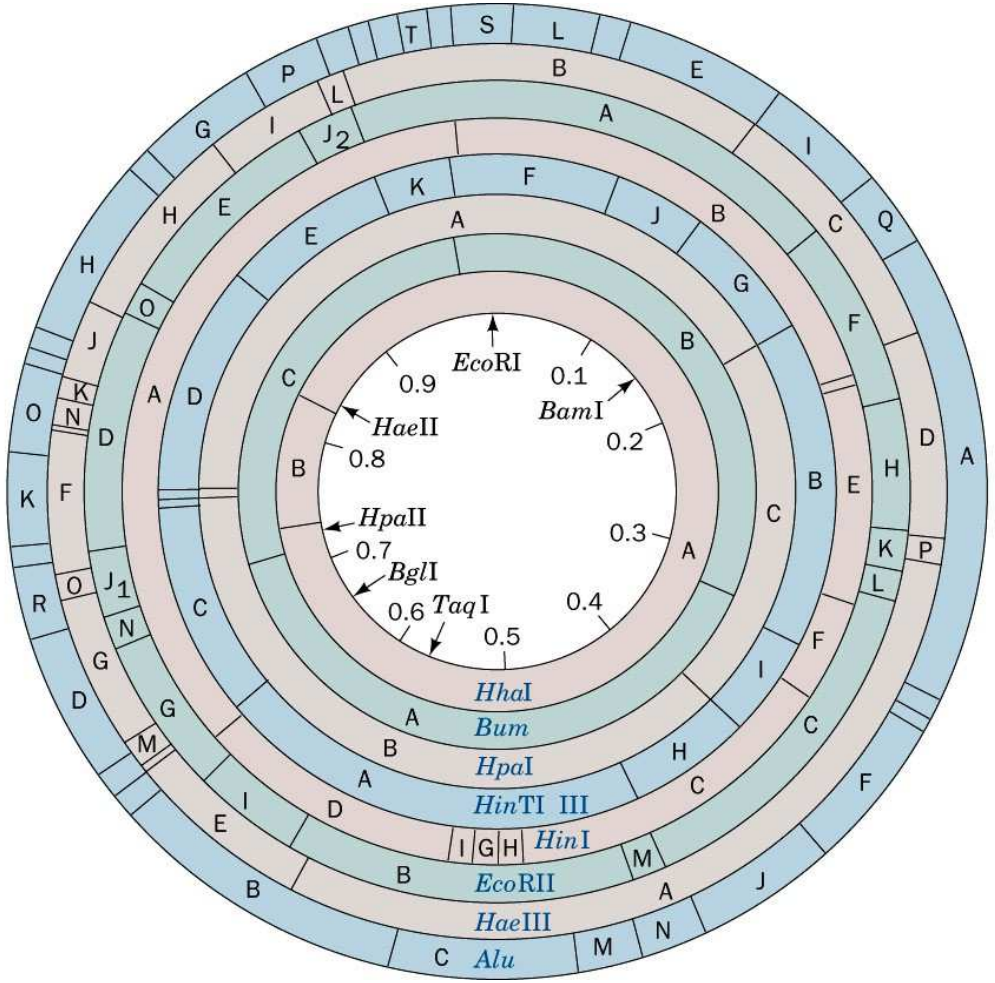
# X-Ray structure of a complex of ethidium bromid with DNA.

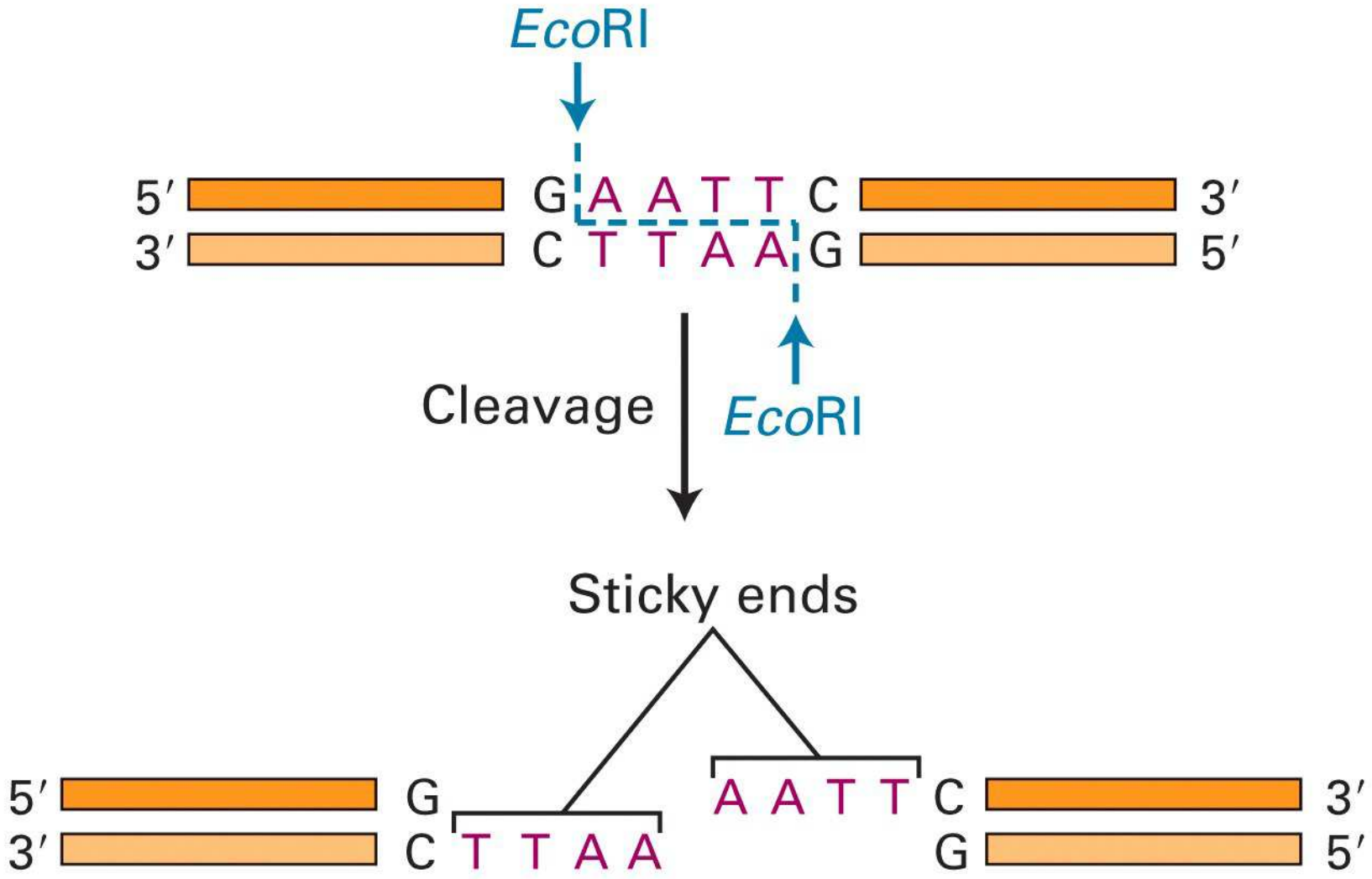


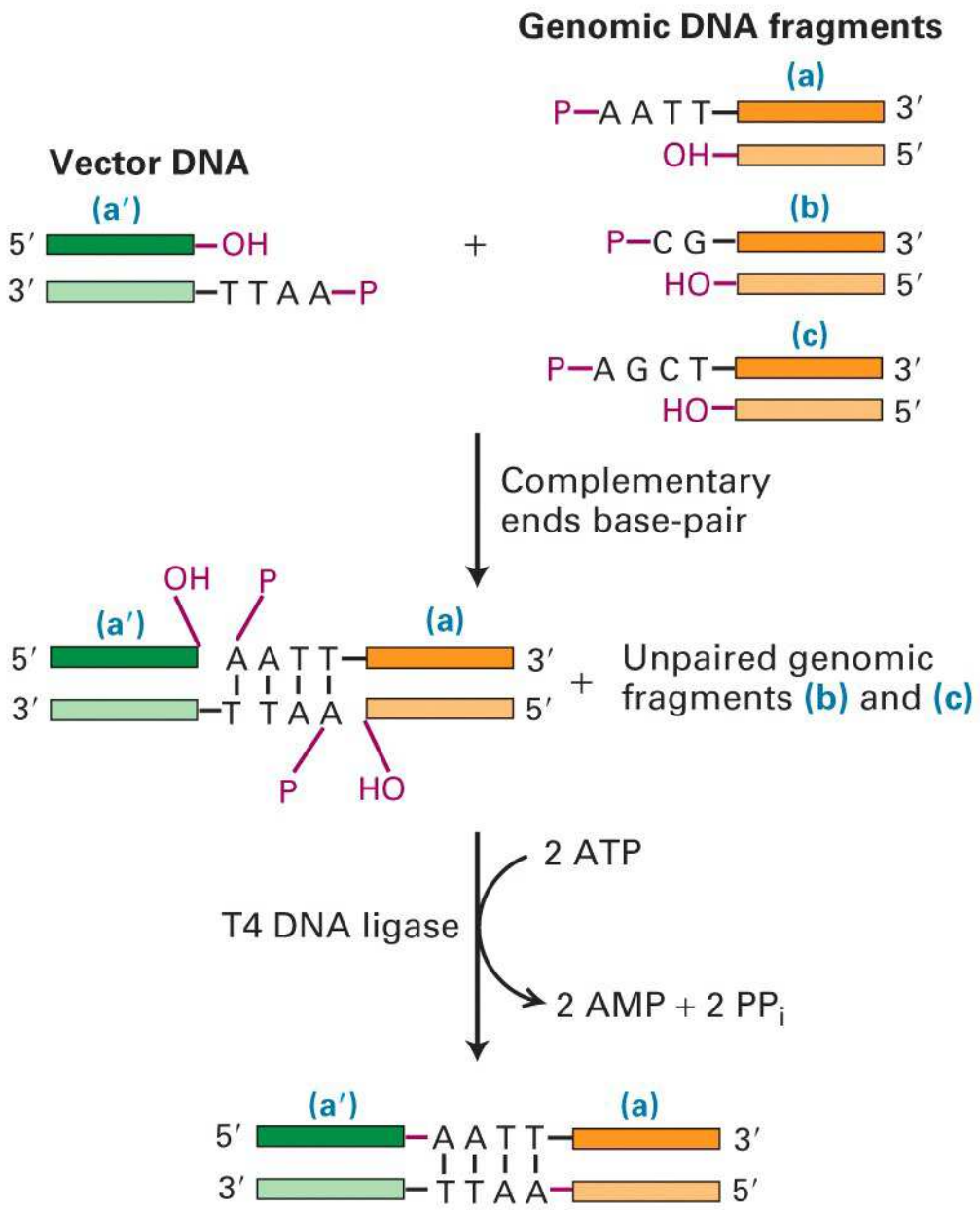
# Construction of a restriction map.

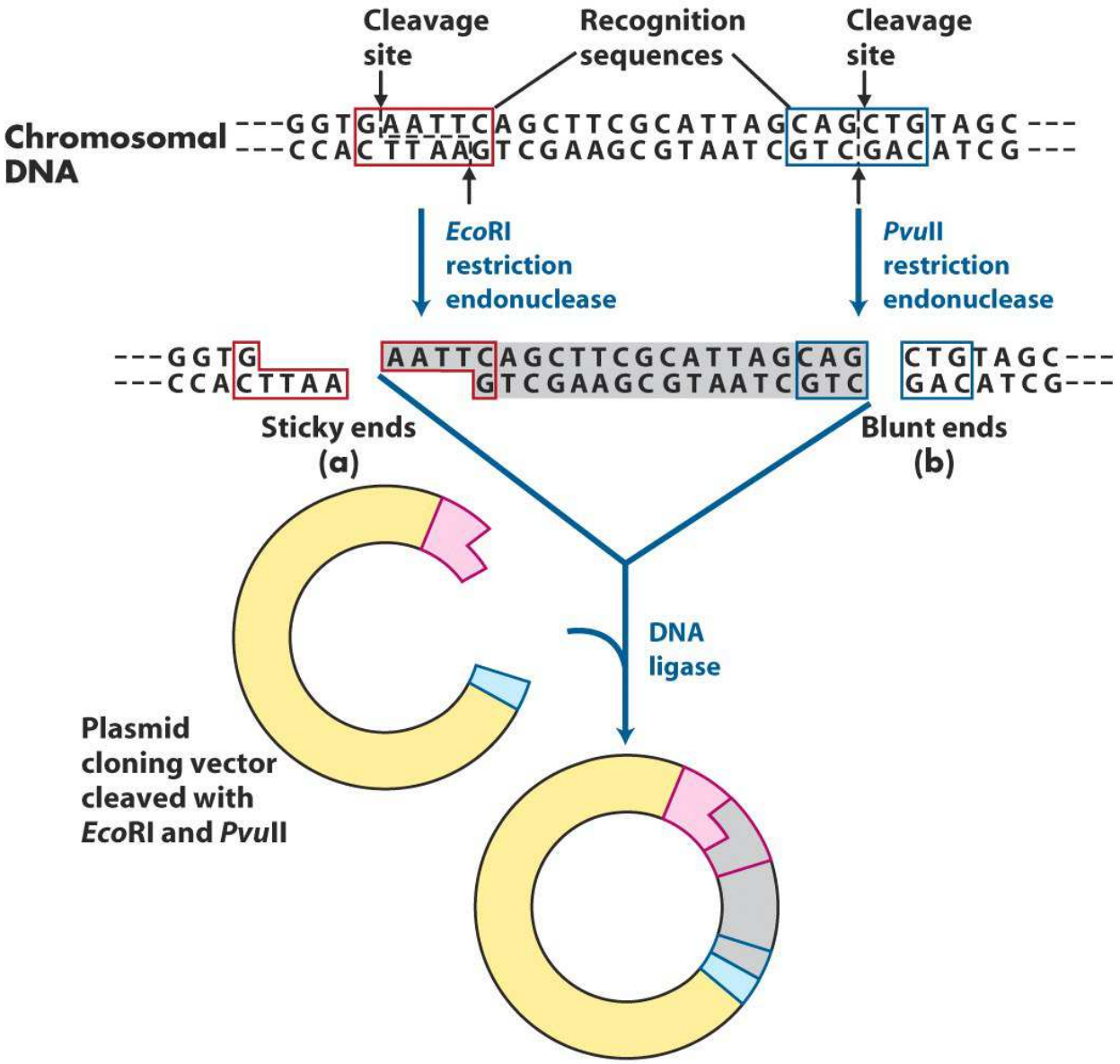


# Restriction map for the 5243-bp circular DNA of SV40.

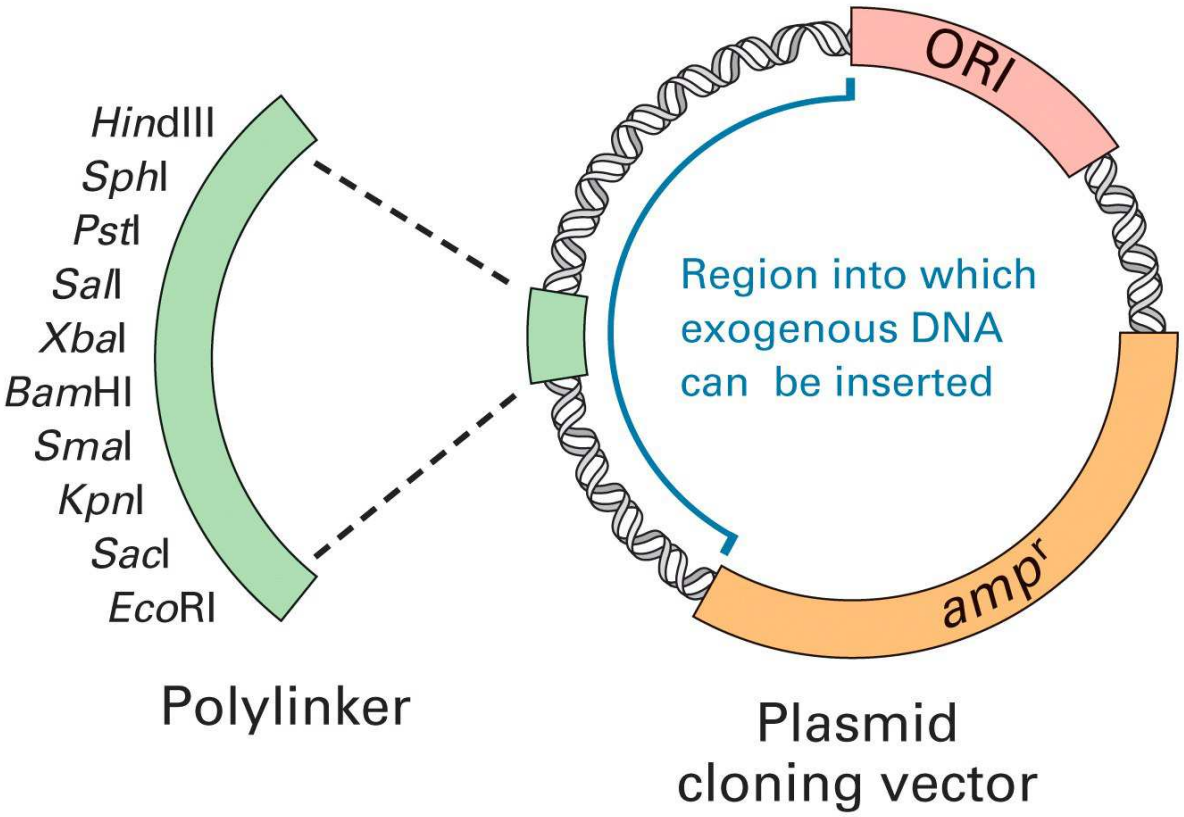



EcoRI
5′
3′
G A A T T C
C T T A A G
3′
5′
Cleavage
EcoRI
Sticky ends
5′
3′
G
C T T A A
A A T T C
G
3′
5′


Genomic DNA fragments
(a)
P—A A T T—3′
OH—5′
Vector DNA
(a′)
5′—OH
3′—T T A A—P
+
(b)
P—C G—3′
HO—5′
(c)
P—A G C T—3′
HO—5′
Complementary ends base-pair
OH
P
(a′)
(a)
5′ A A T T 3′
3′ T T A A 5′
+
Unpaired genomic fragments (b) and (c)
P
HO
2 ATP
T4 DNA ligase
2 AMP + 2 PPi
(a′)
(a)
5′ A A T T 3′
3′ T T A A 5′


Cleavage site
Recognition sequences
Cleavage site
Chromosomal DNA
---GGTGAATTCAGCTTCGCATTAGCAGCTGTAGC---
---CCACTTAAGTCGAAGCGTAATCGTCGACATCG---
EcoRI restriction endonuclease
PvuII restriction endonuclease
---GGTG
---CCACTTAA
AATTCAGCTTCGCATTAGCAG
GTCGAAGCGTAATCGTC
CTGTAGC---
GACATCG---
Sticky ends (a)
Blunt ends (b)
DNA ligase
Plasmid cloning vector cleaved with EcoRI and PvuII

# Plasmid Cloning Vectors



# Plasmid Cloning Vectors

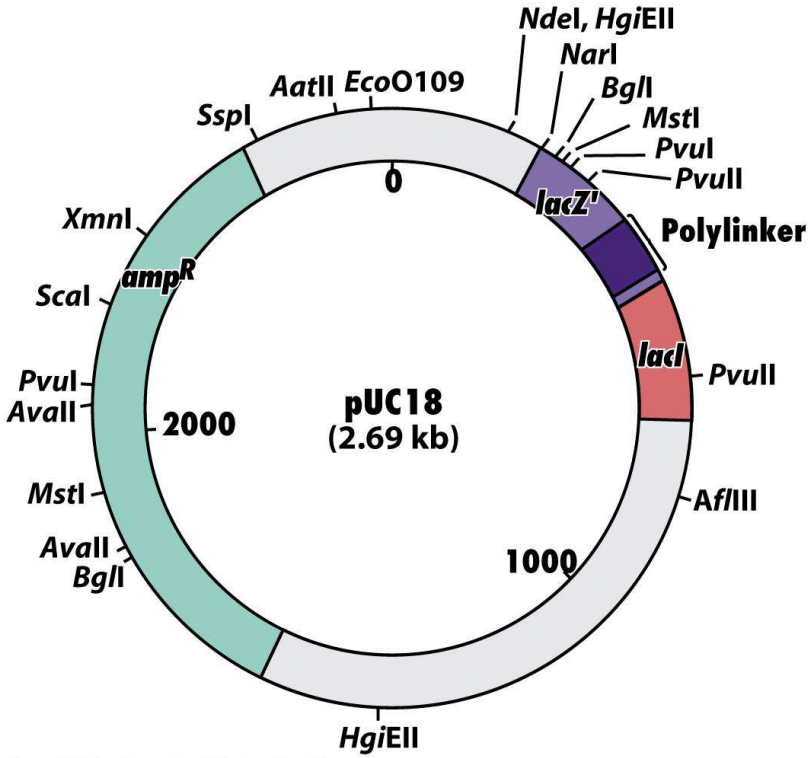
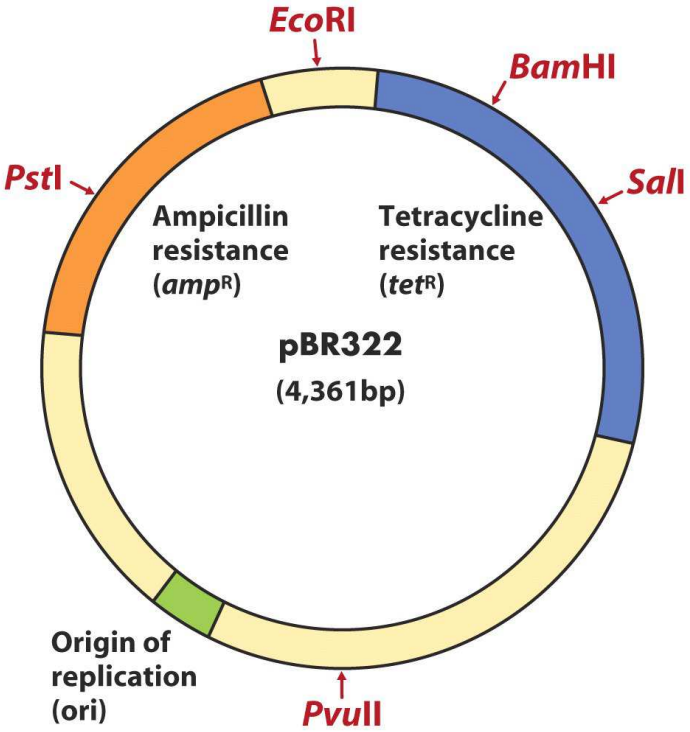


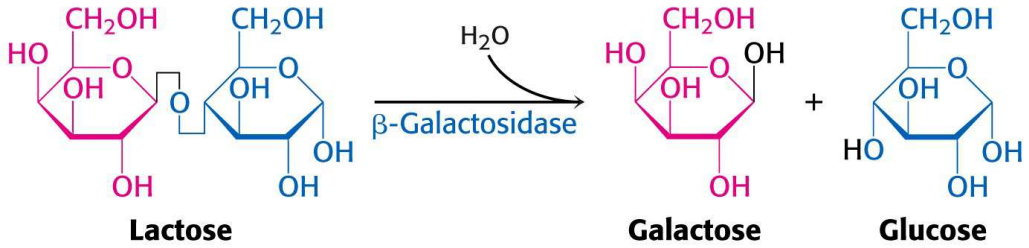

Figure 3-25 Fundamentals of Biochemistry, 2/e

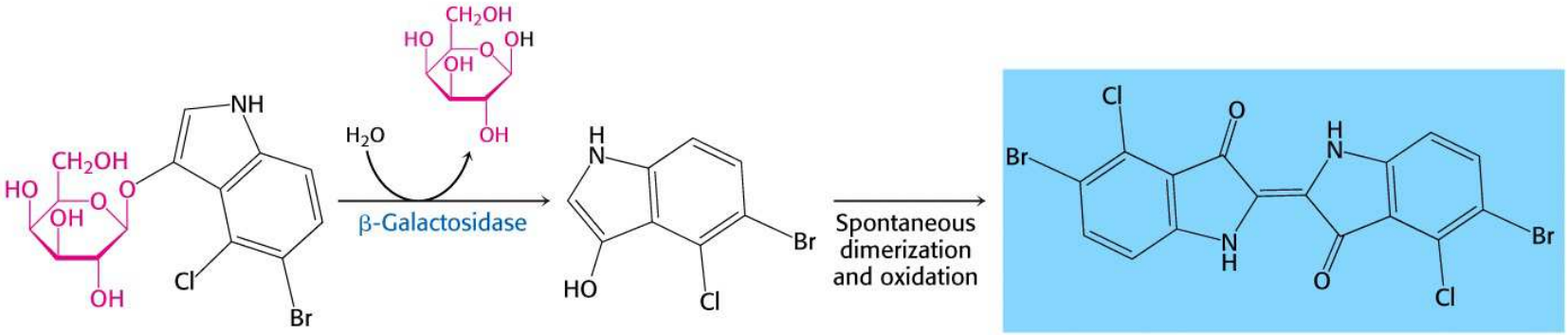
# Insertional inactivation

## Gene in cloning site:

- *LacZ* -> pUC18 (*lacZ* complements the host defect in *lacZ*)
  -> pUC18 into host organism -> active *lacZ* (β-galactosidase) from plasmid-> cleavage of X-gal (blue colonies)
  -> gene cloned into polylinker -> *lacZ* gene disrupted -> no cleavage of X-gal (white colonies)

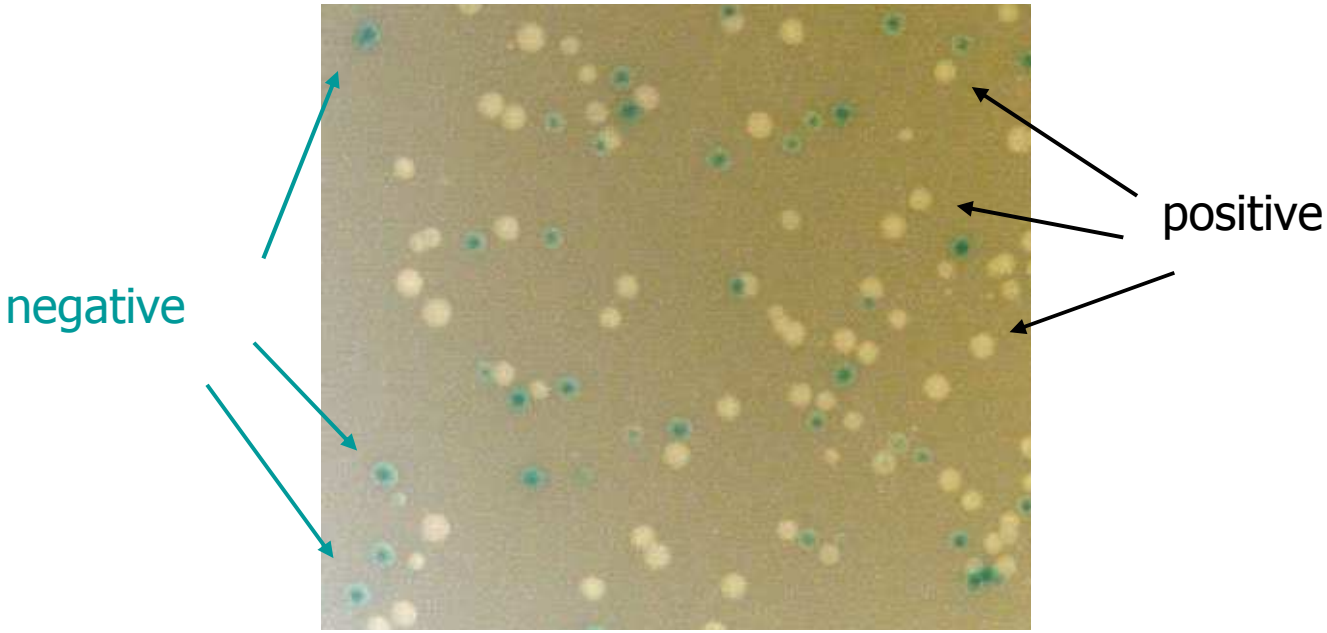



**Lactose** **Galactose** **Glucose**




**X-Gal** **5,5′-Dibromo-4,4′-dichloro-indigo**

# Blue/White Selection



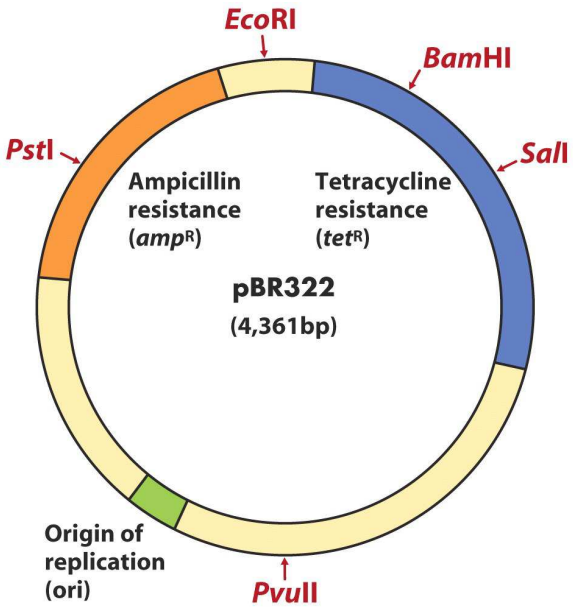


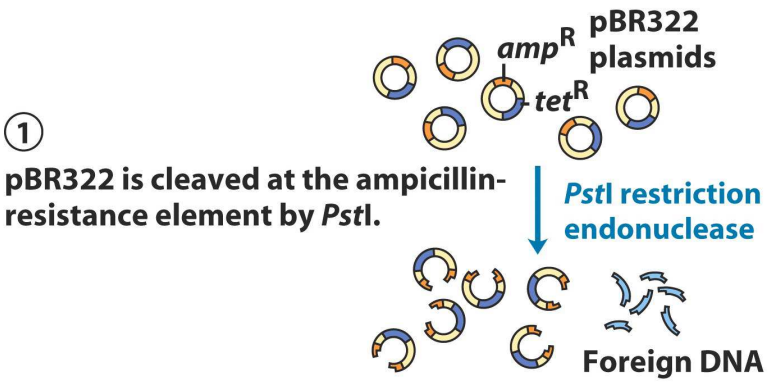

http://www.bio.davidson.edu/Courses/genomics/method/reporters.html

# Insertional inactivation

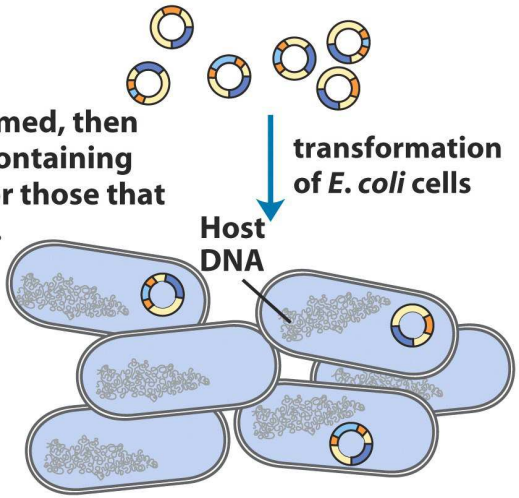
Gene in cloning site:

- Resistance marker -> pBR322 (cloning sites within antibiotics resistence marker)
  -> plasmid into host -> resistance against 2 antibiotics
  -> gene cloned within one resistance marker -> gene for antibiotic resistance marker disrupted -> sensitive against one antibiotic

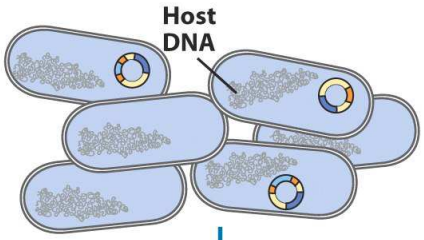


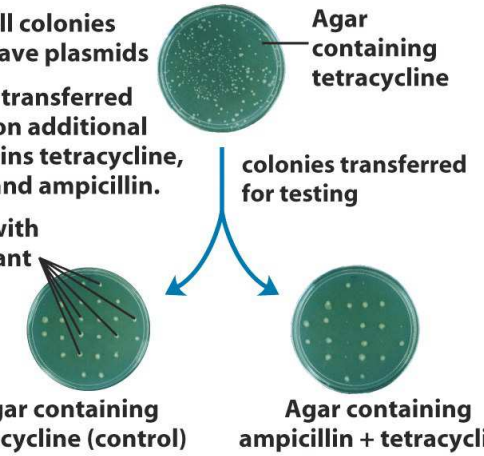




② Foreign DNA is ligated to cleaved pBR322. Where ligation is successful, the ampicillin-resistance element is disrupted. The tetracycline-resistance element remains intact.

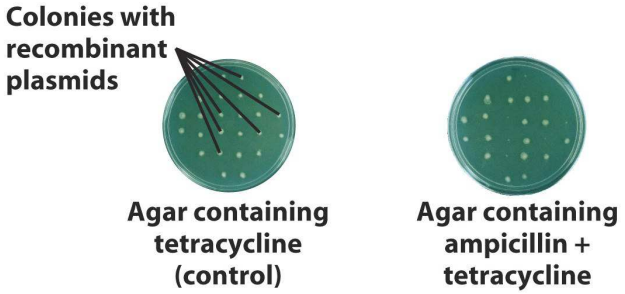



③ *E. coli* cells are transformed, then grown on agar plates containing tetracycline to select for those that have taken up plasmid.



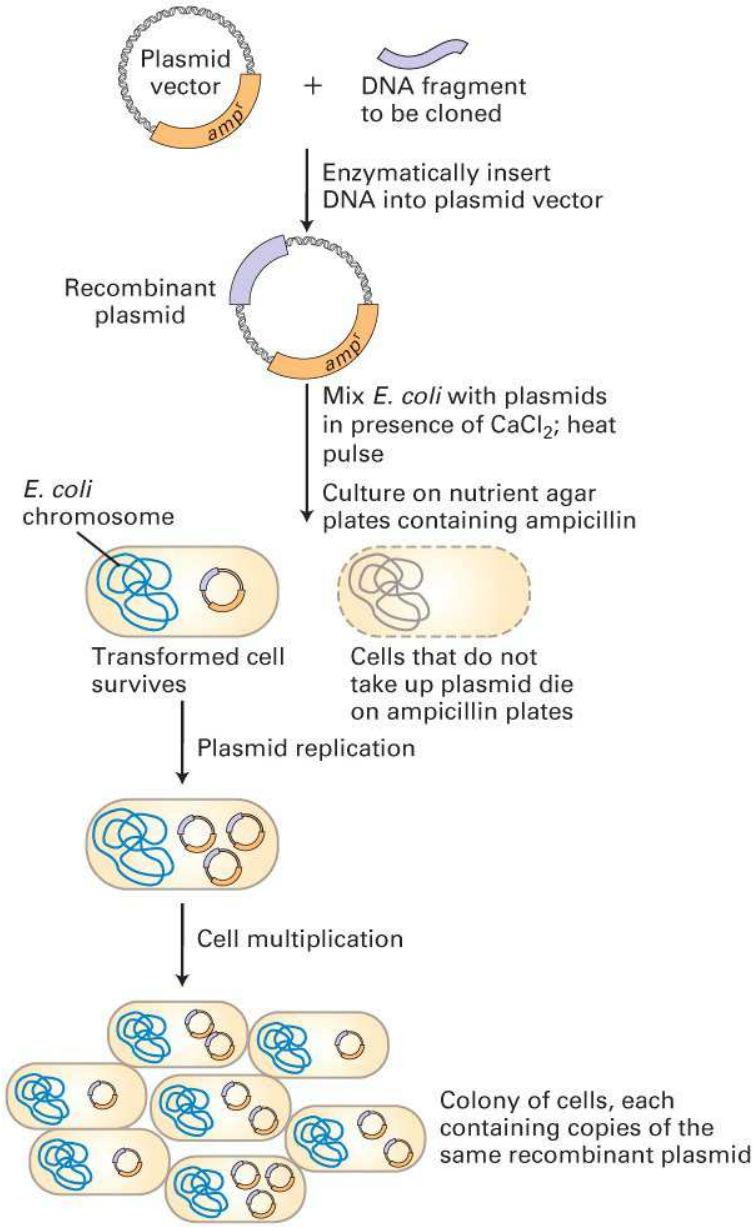




④ Individual colonies are transferred to matching positions on additional plates. One plate contains tetracycline, the other tetracycline and ampicillin.




⑤ Cells that grow on tetracycline but not on tetracycline + ampicillin contain recombinant plasmids with disrupted ampicillin resistance, hence the foreign DNA. Cells with pBR322 without foreign DNA retain ampicillin resistance and grow on both plates.

# Transformation and Selection



# Vectors

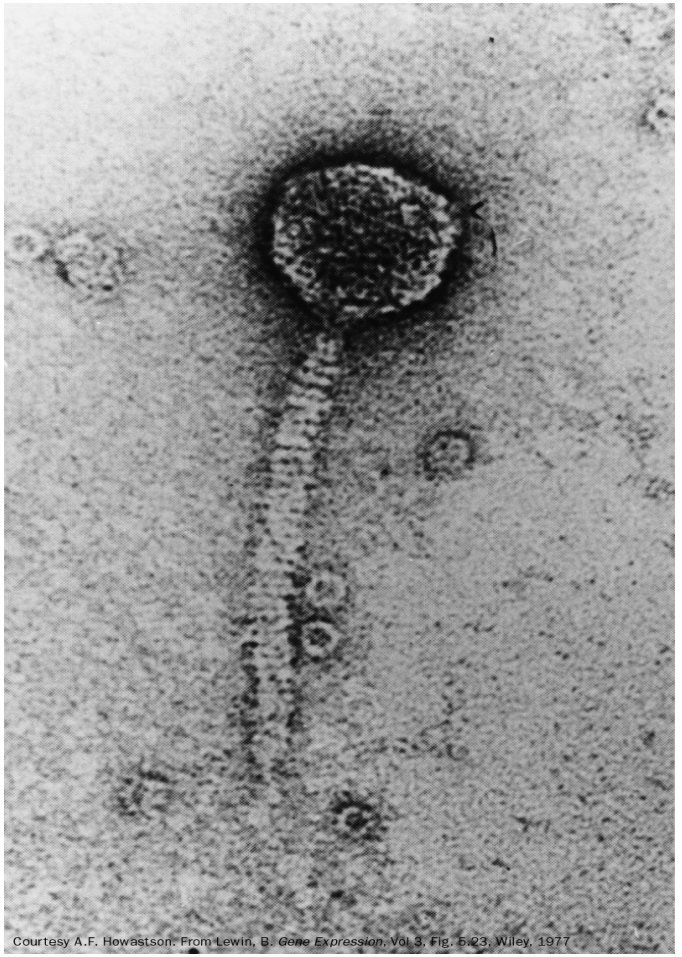
*Table 4.6* Insert capacities of some commonly used vector systems

| Vector system | Host cell | Insert capacity (kb) |
|---|---|---|
| Plasmid | *E. coli* | 0.1–10 |
| Bacteriophage λ | λ/*E. coli* | 10–20 |
| Cosmid | *E. coli* | 35–45 |
| Bacteriophage P1 | *E. coli* | 80–100 |
| BAC | *E. coli* | 50–300 |
| P1 bacteriophage-derived artificial chromosome | *E. coli* | 100–300 |
| Yeast artificial chromosome | Yeast | 100–2,000 |
| Human artificial chromosome | Cultured human cells | >2,000 |

Similar principle of cloning, suitable for different purposes

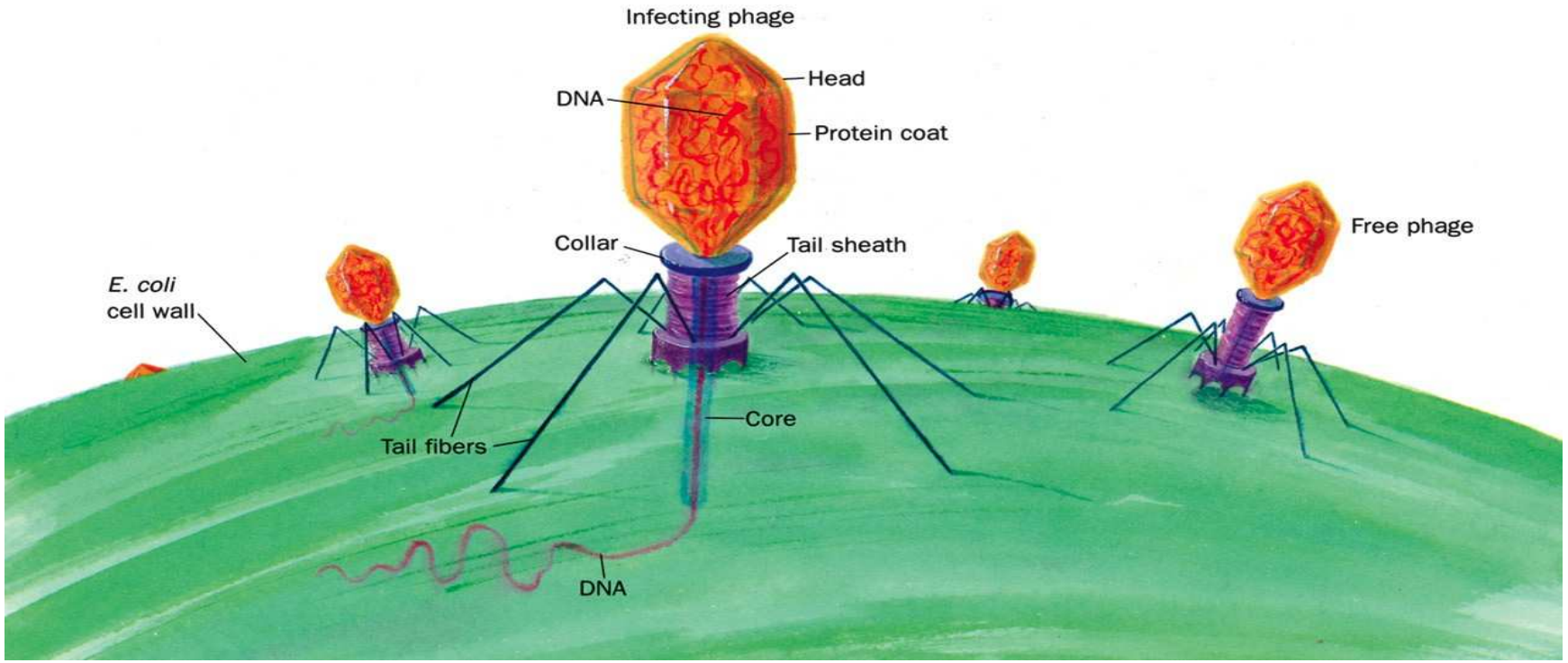
# Bacteriophages


Electron micrograph of the filamentous bacteriophage M13.



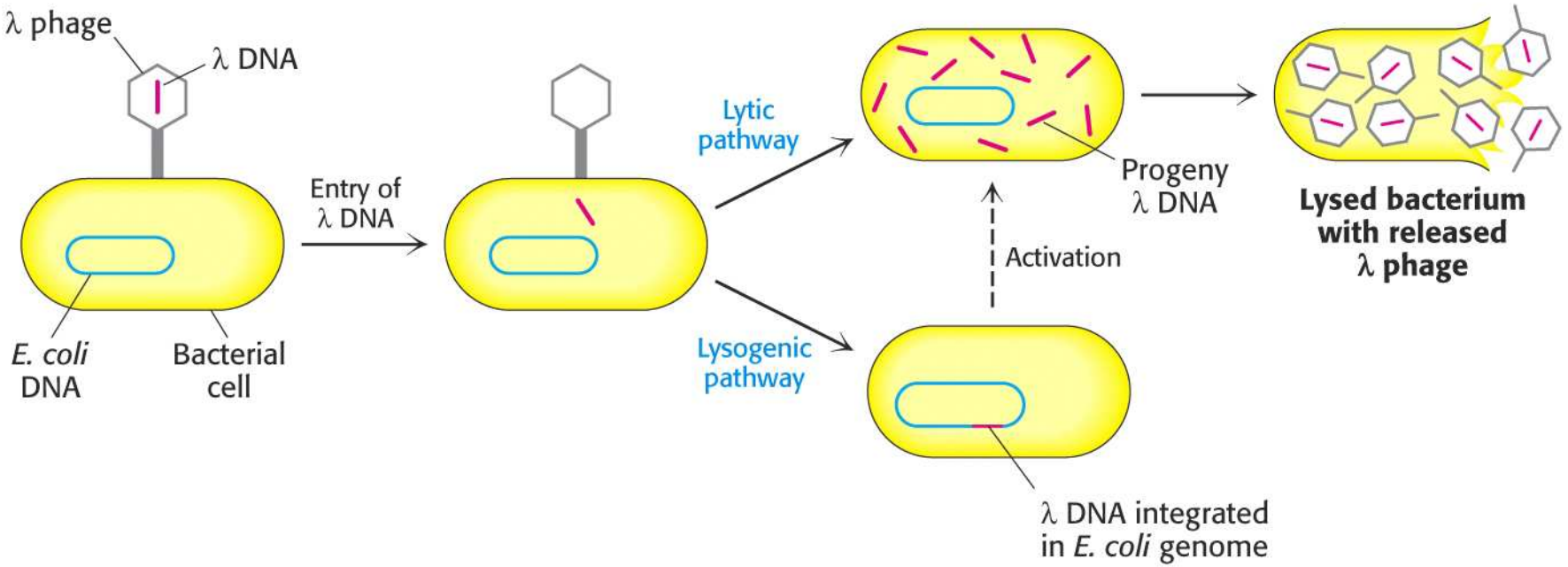

Electron micrograph of bacteriophage λ.

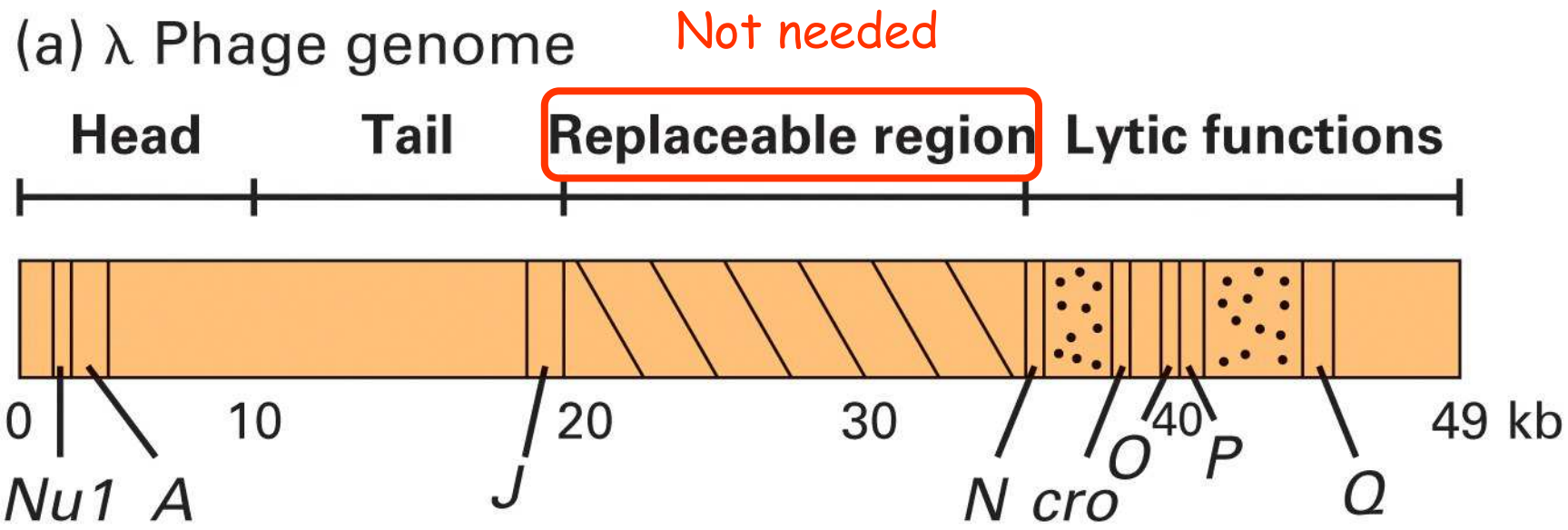
# Bacteriophage T2 injecting its DNA into an *E. coli*




Page 84

# Life Cycle of Bacteriophage λ






One of First Genome Sequences

# Replication of bacteriophage upon infection of a cell

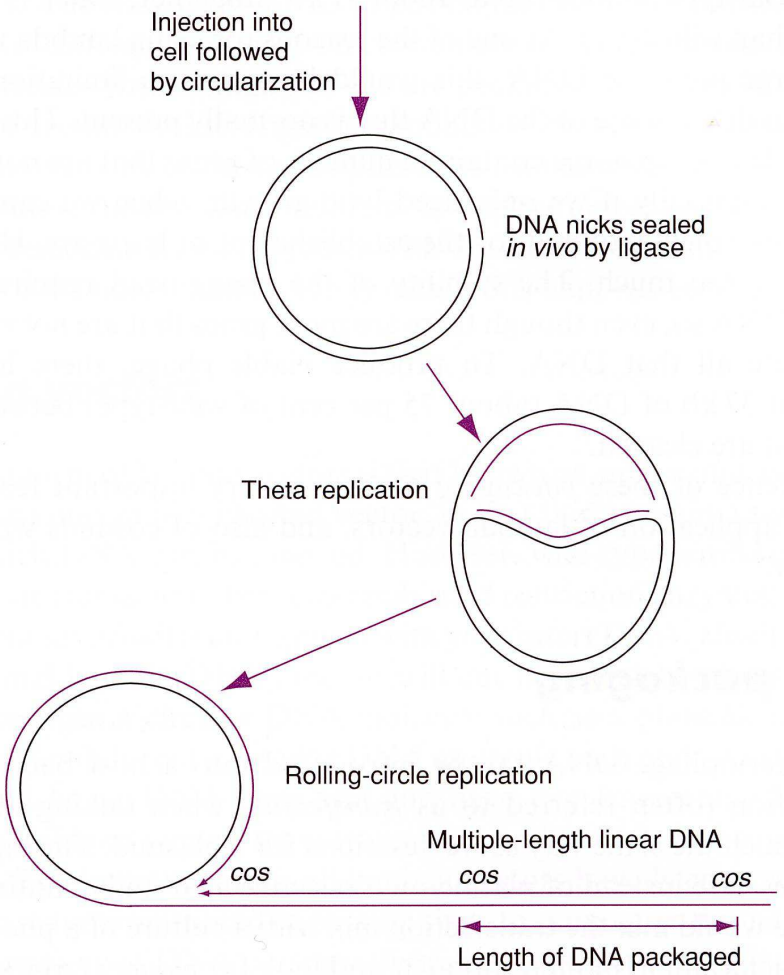
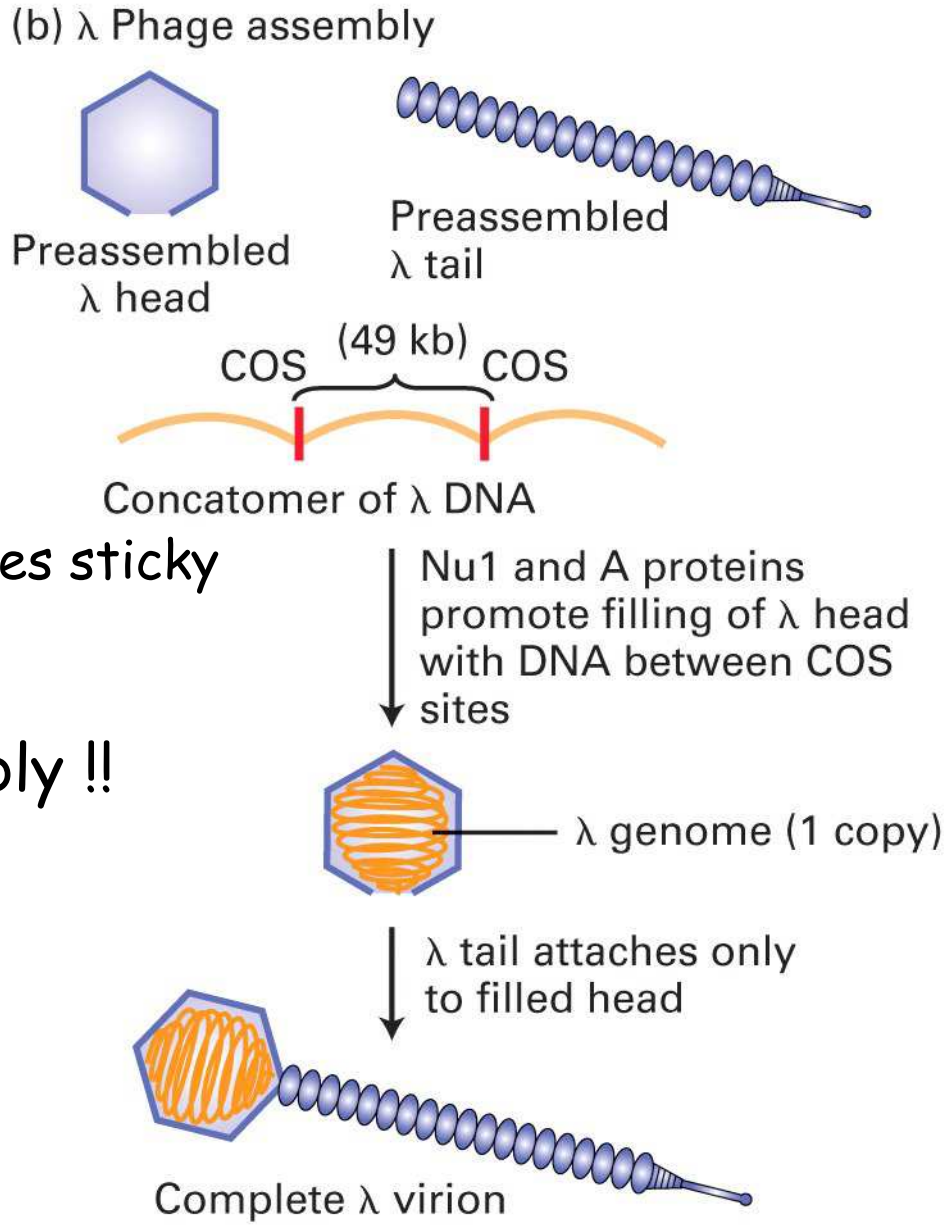
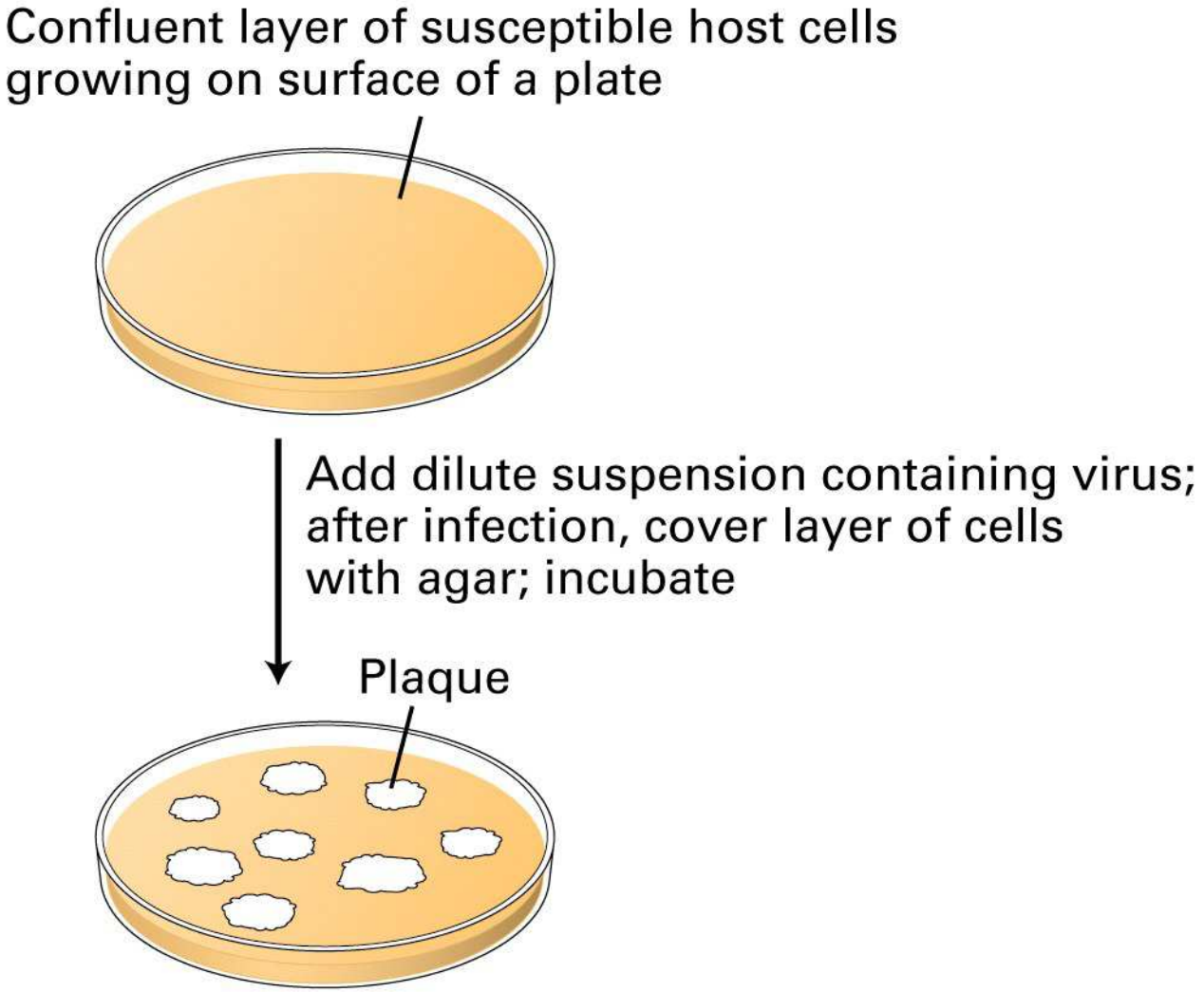



**Figure 6.6** Replication of bacteriophage lambda DNA

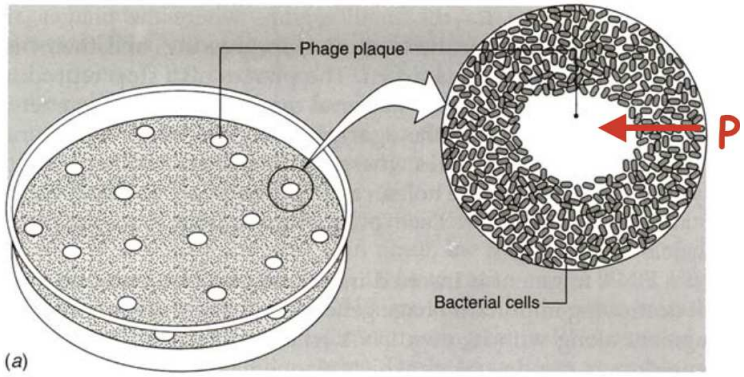


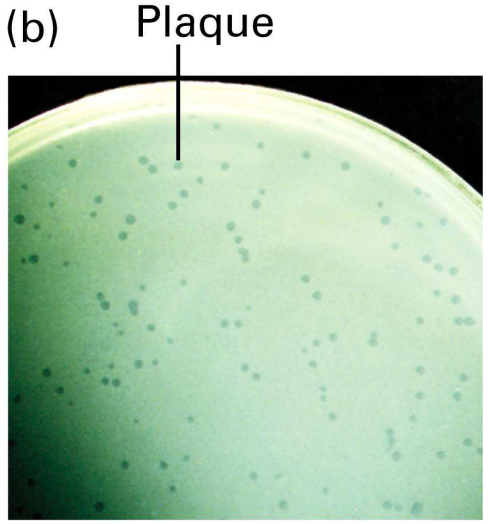

Cos site 12 bases sticky singlestrand

Self-assembly !!

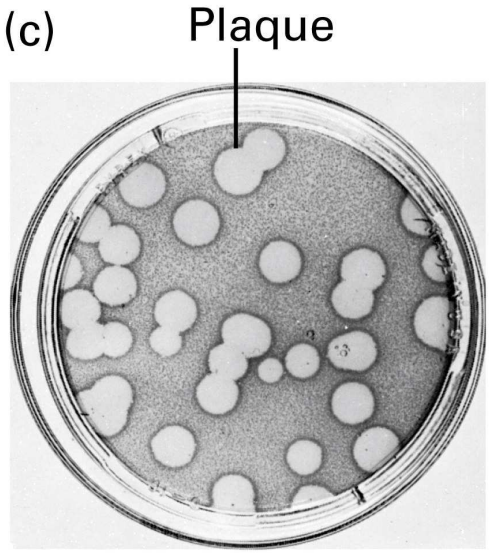



Each plaque represents cell lysis initiated by one viral particle (agar restricts movement so that virus can infect only contiguous cells)



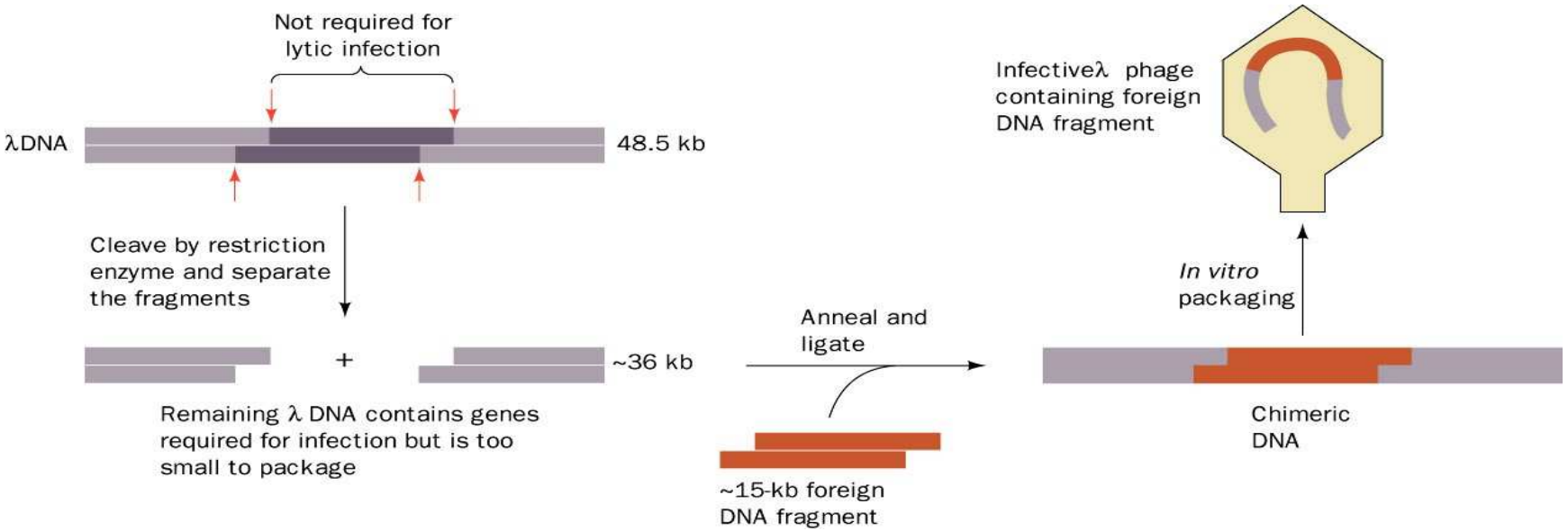



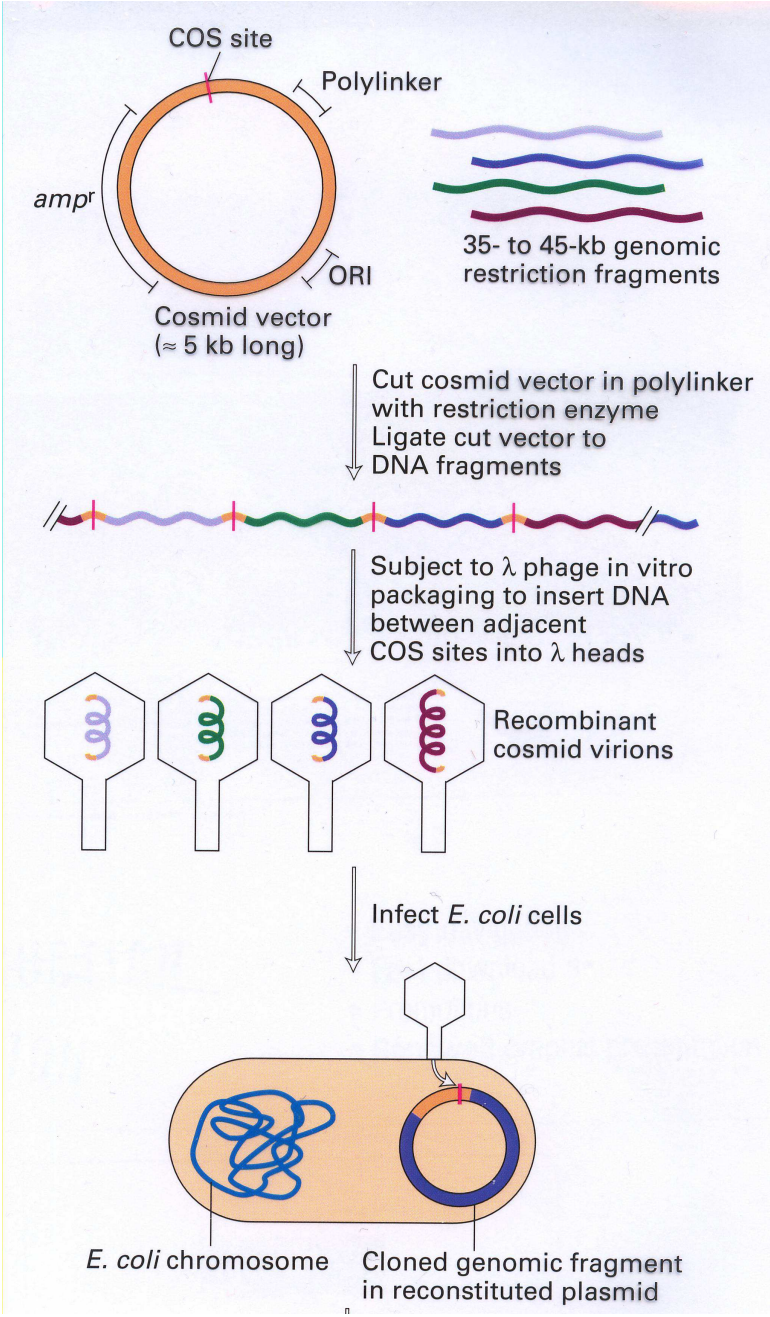

1. Each Plaque is a Virus Clone Representing One Viral Infection!




2. Selectable Marker is Bacterial Cell Destruction & Plaque Formation

# Cloning of foreign DNA in λ phages.

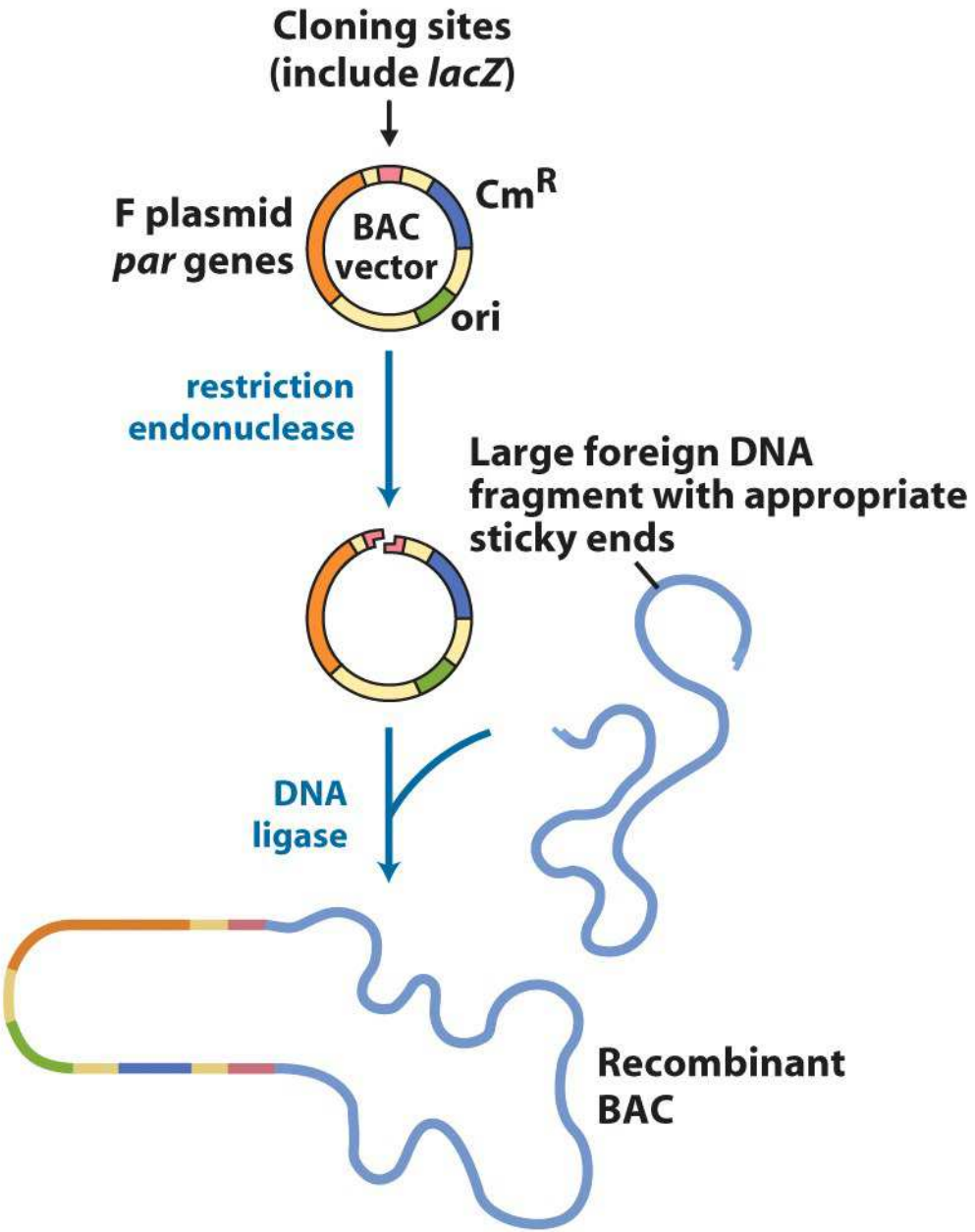


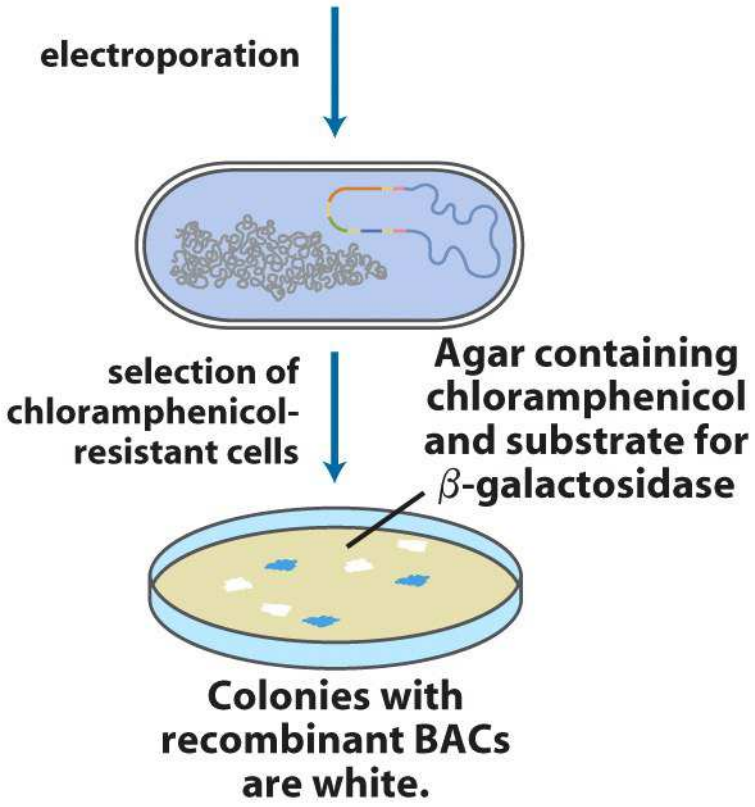



Cosmid = Cos – Plasmid

Behave as plasmids in E. coli after infection – ORI, antibiotic resistance

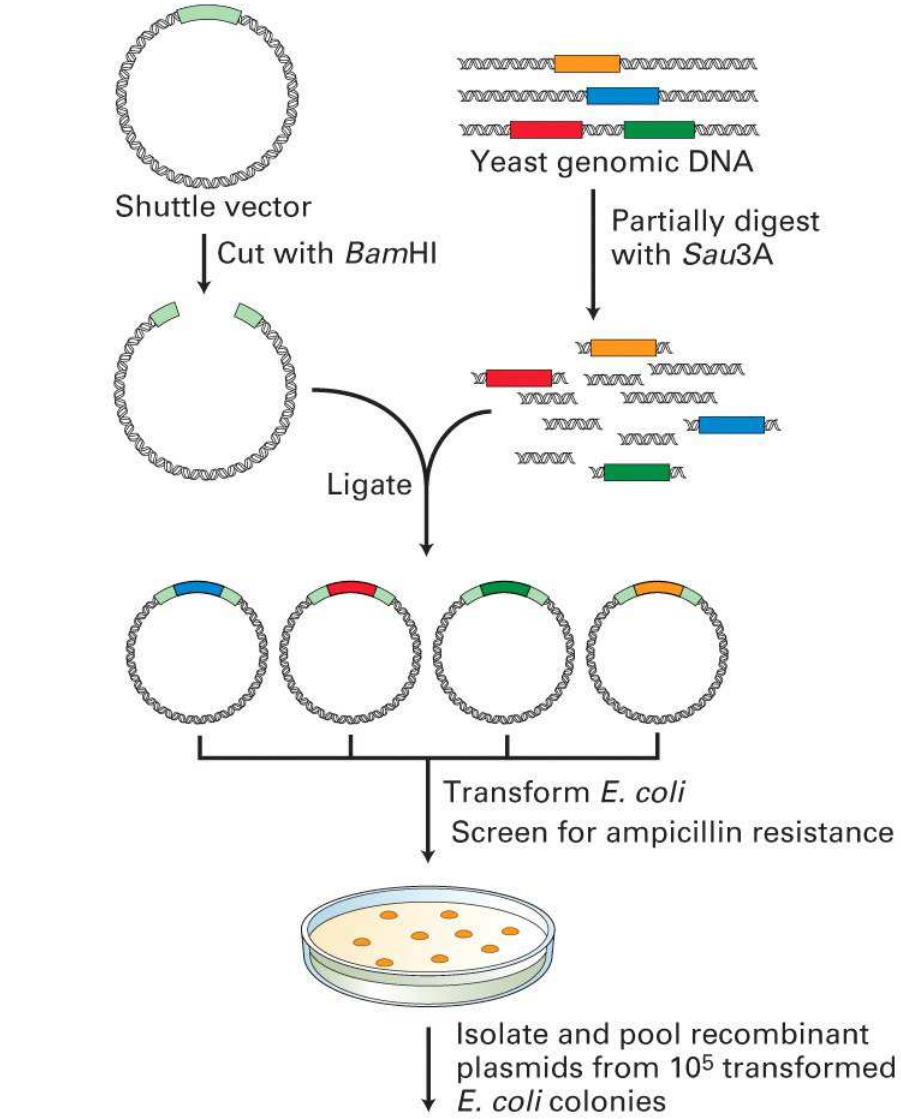
No need for head and tail genes, no plaques




BAC Bacterial artificial chromosome based on F (sex) plasmid (1-2 copies per cell

# What is a genome library ?

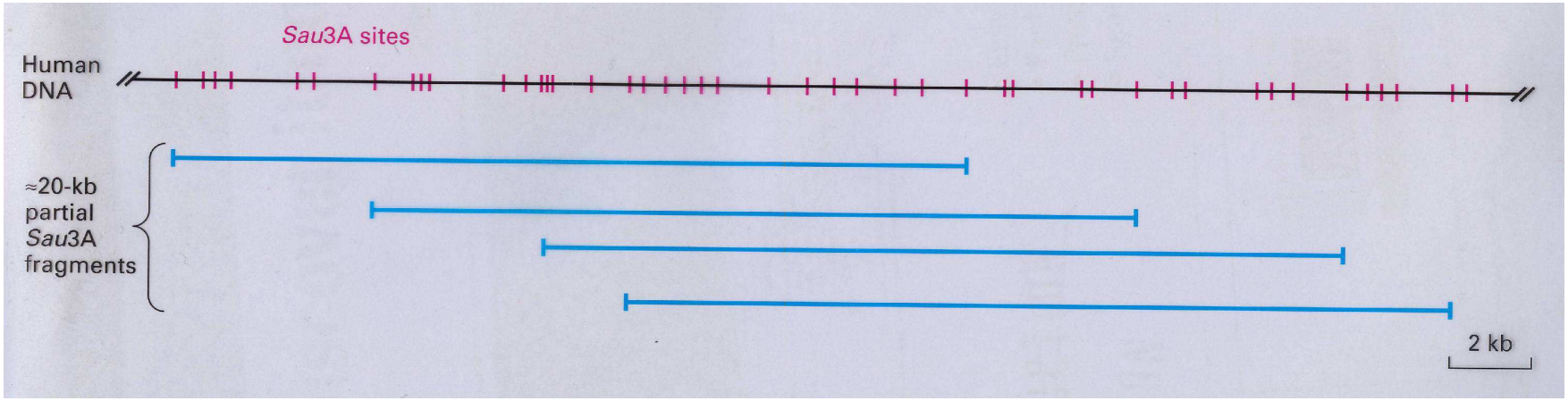



A collection of clones made from a set of randomly generated overlapping DNA fragments representing the entire genome of an organism.

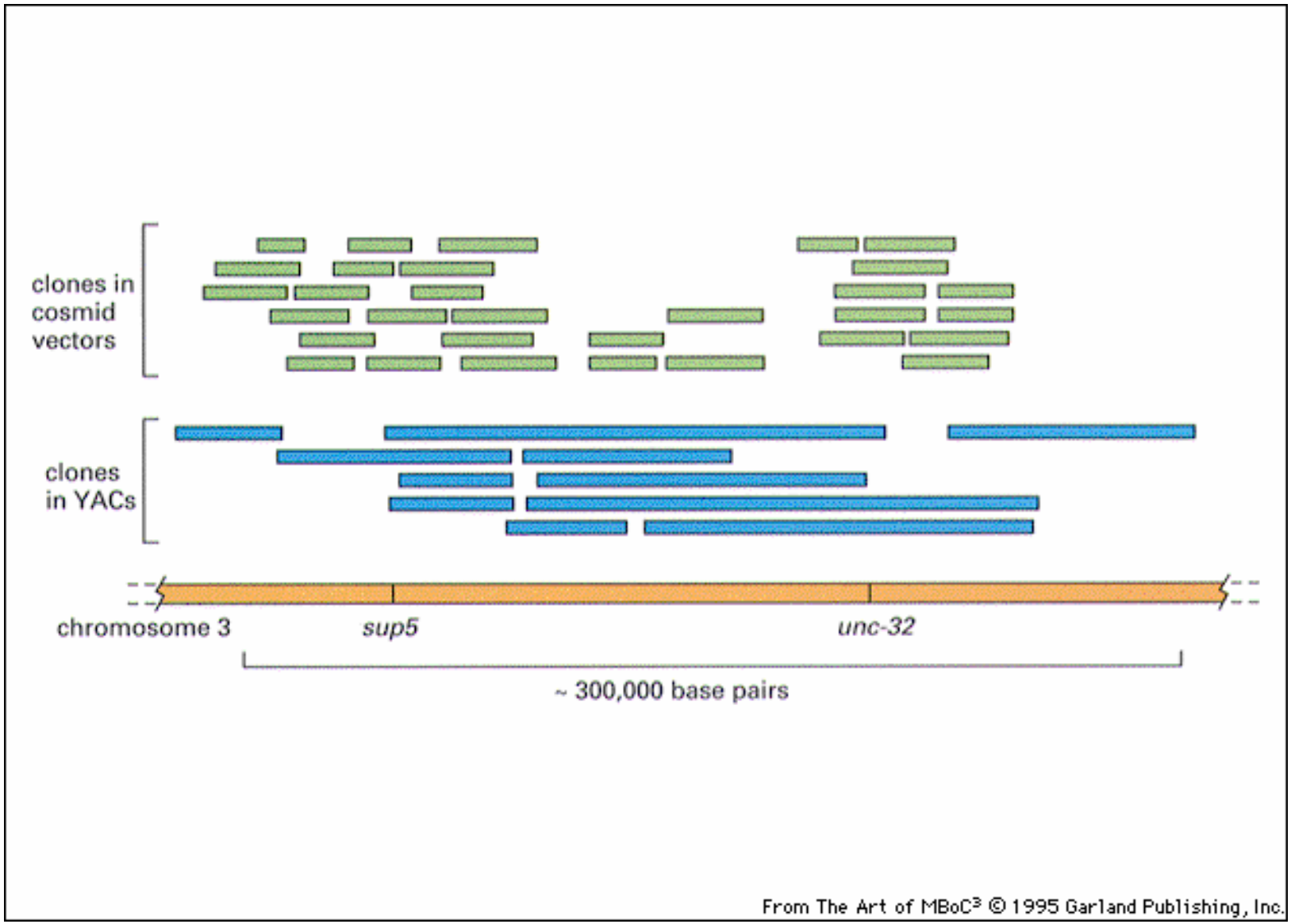
Partially ~ overlapping

# Fragmentation of genomic DNA

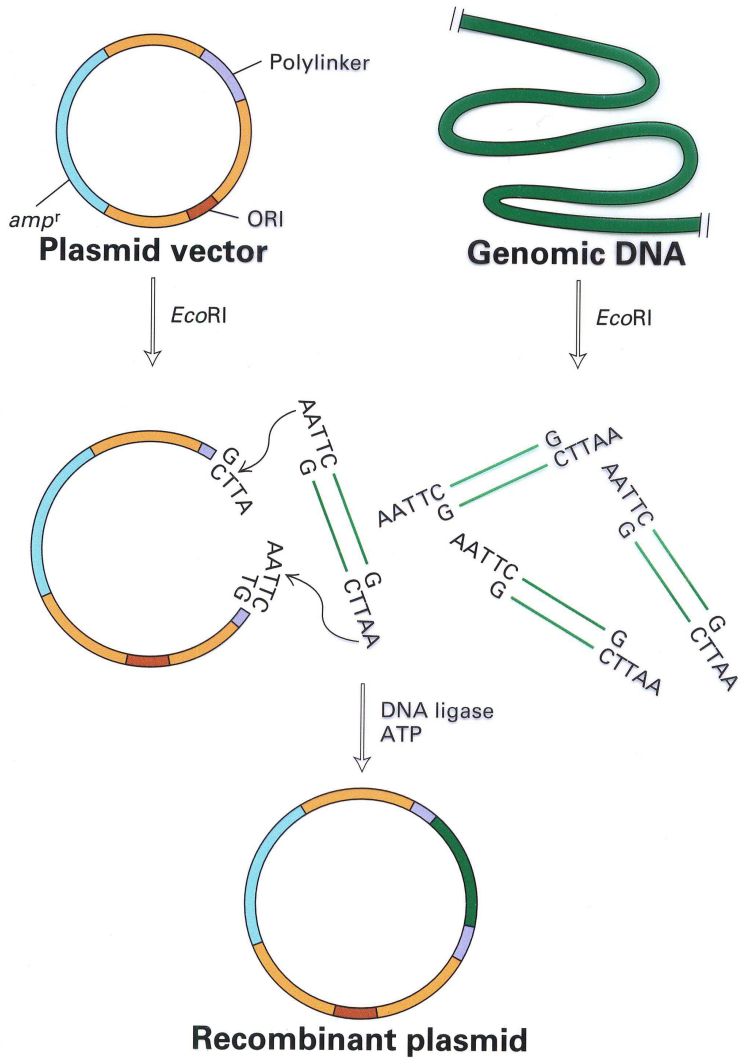
*Sau3A* GATC

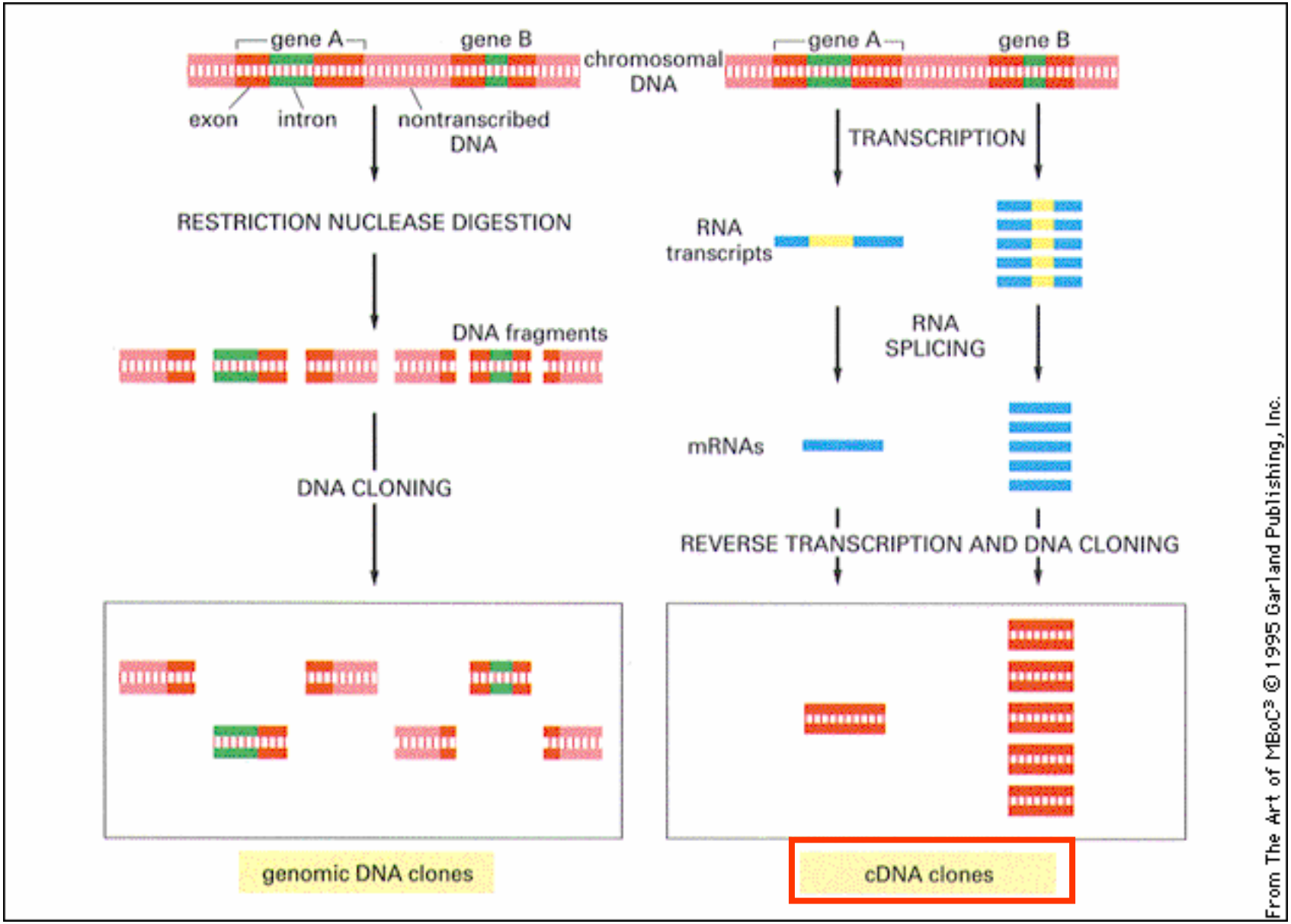



Partially ~ overlapping

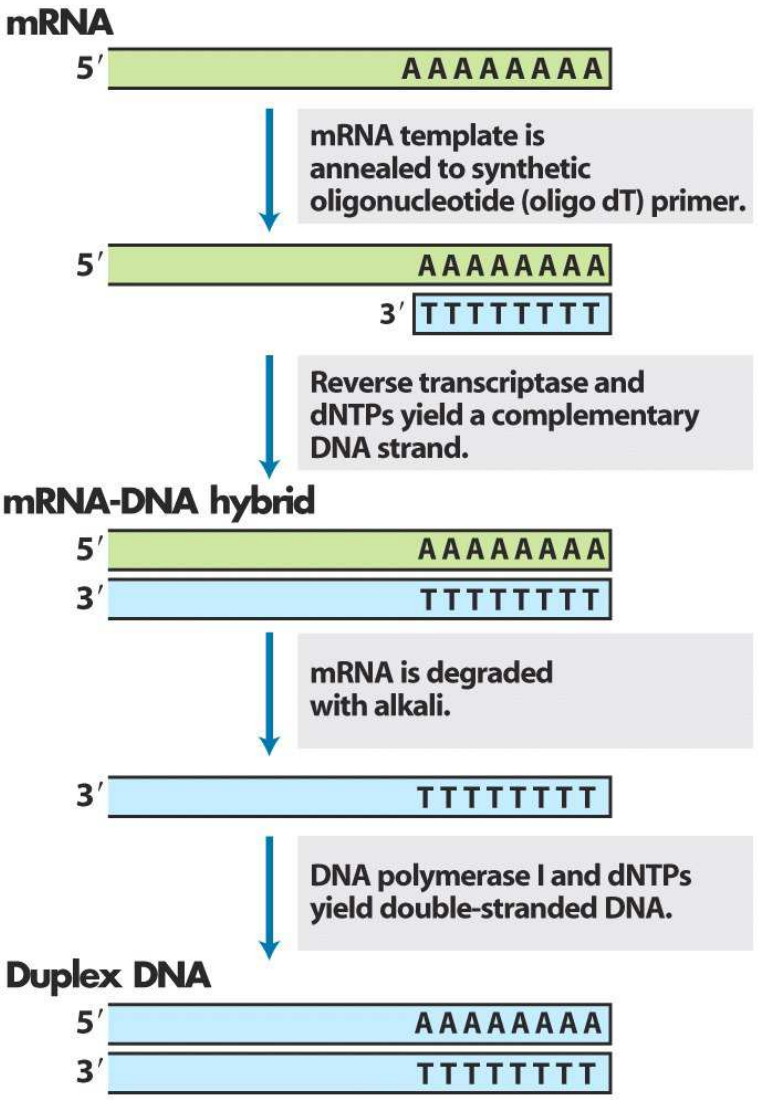


# Creation of Libraries

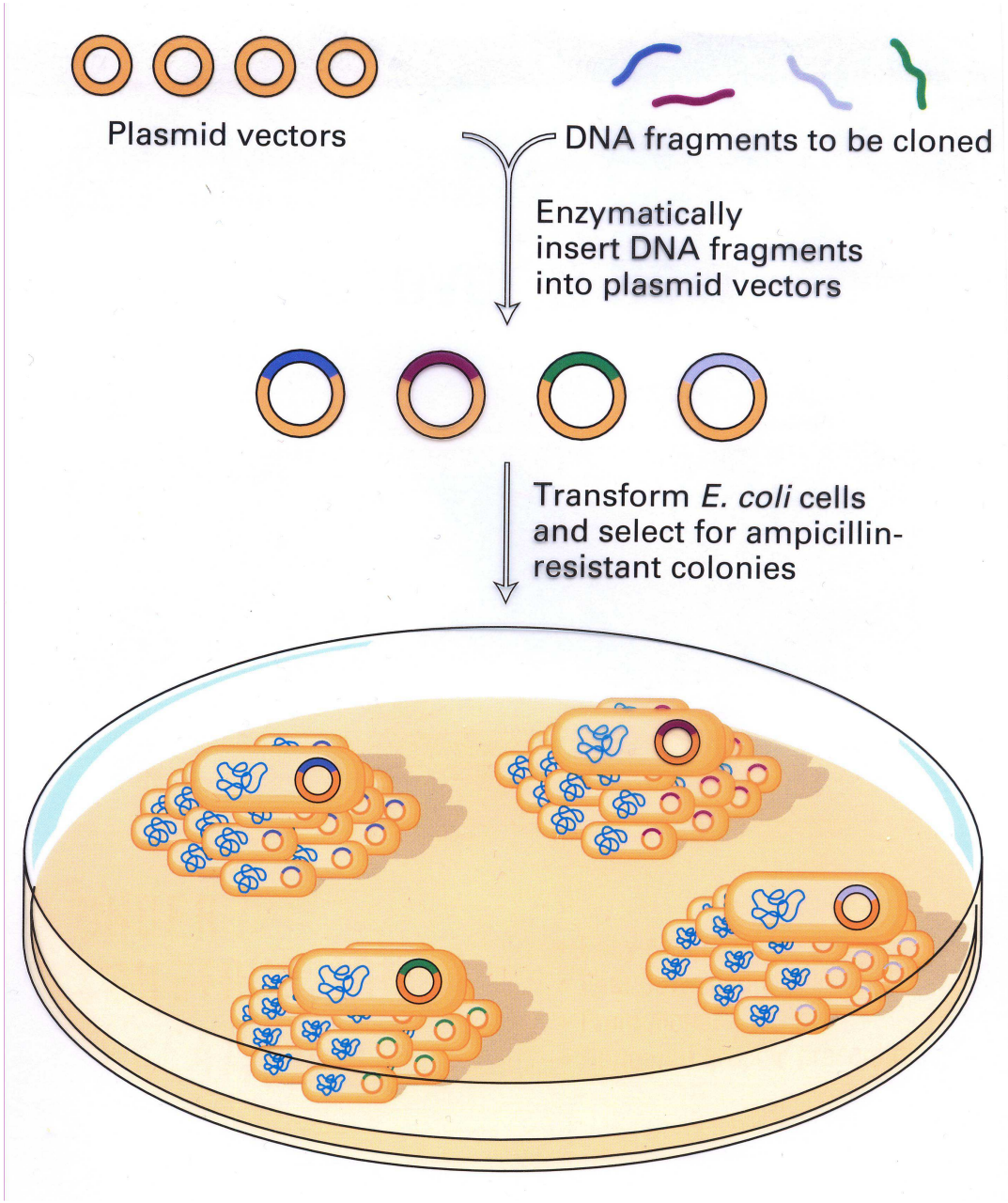




# cDNA synthesis

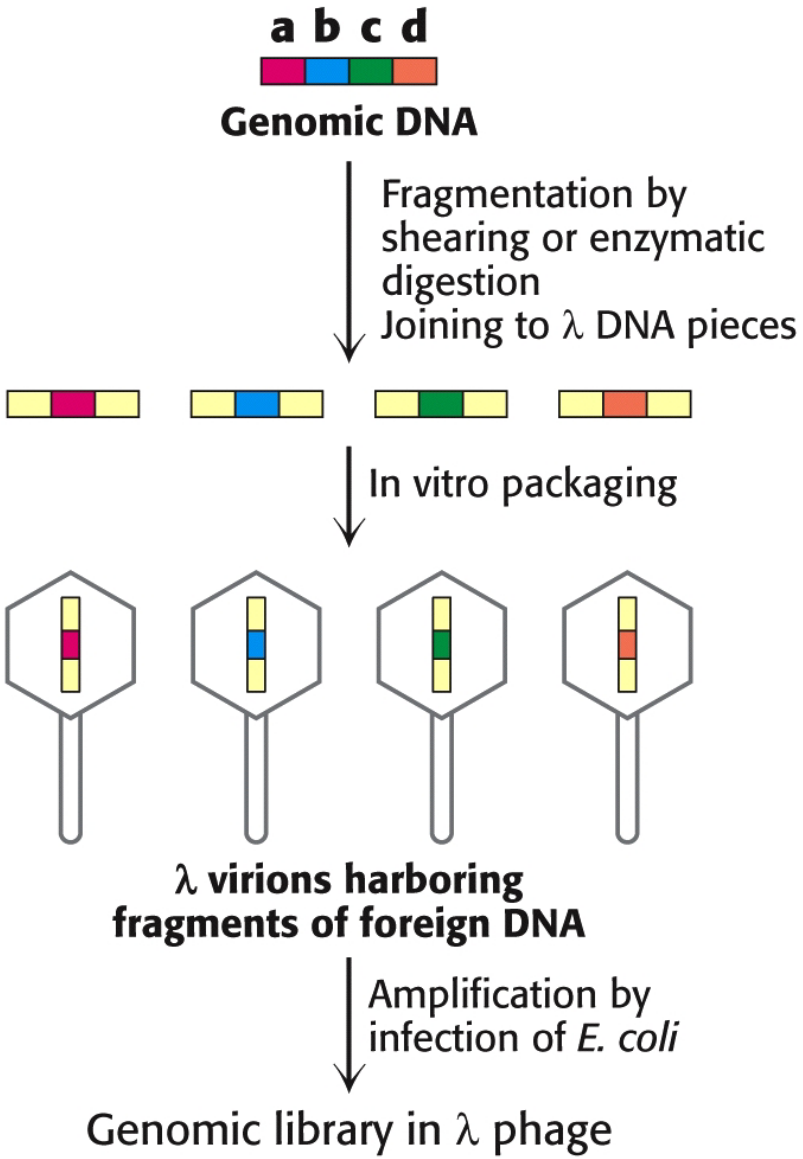


# DNA Library




Clones -> genetically identical

# Genomic phage library



# Evaluation of library

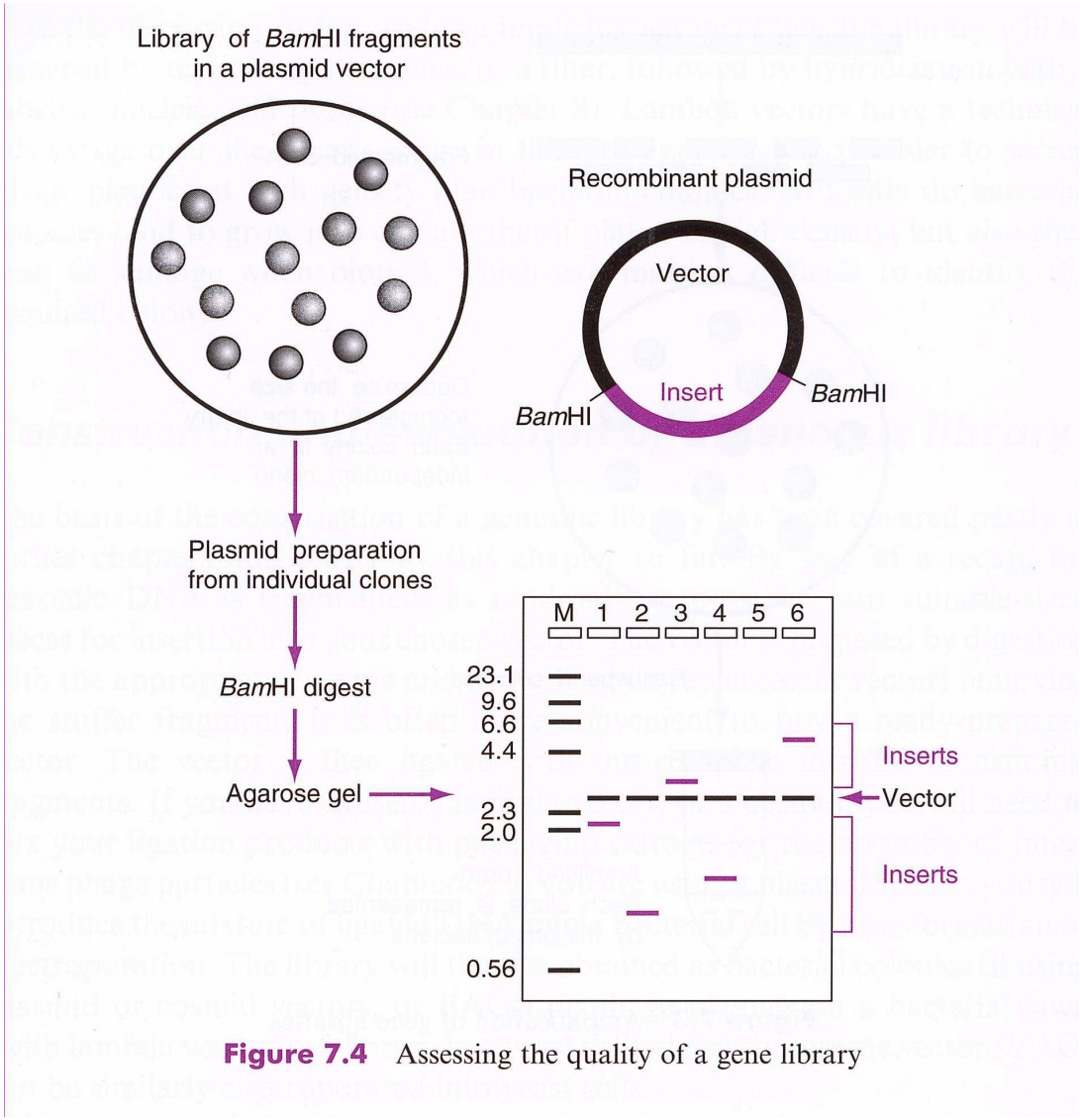



**Figure 7.4** Assessing the quality of a gene library

# Ordered library

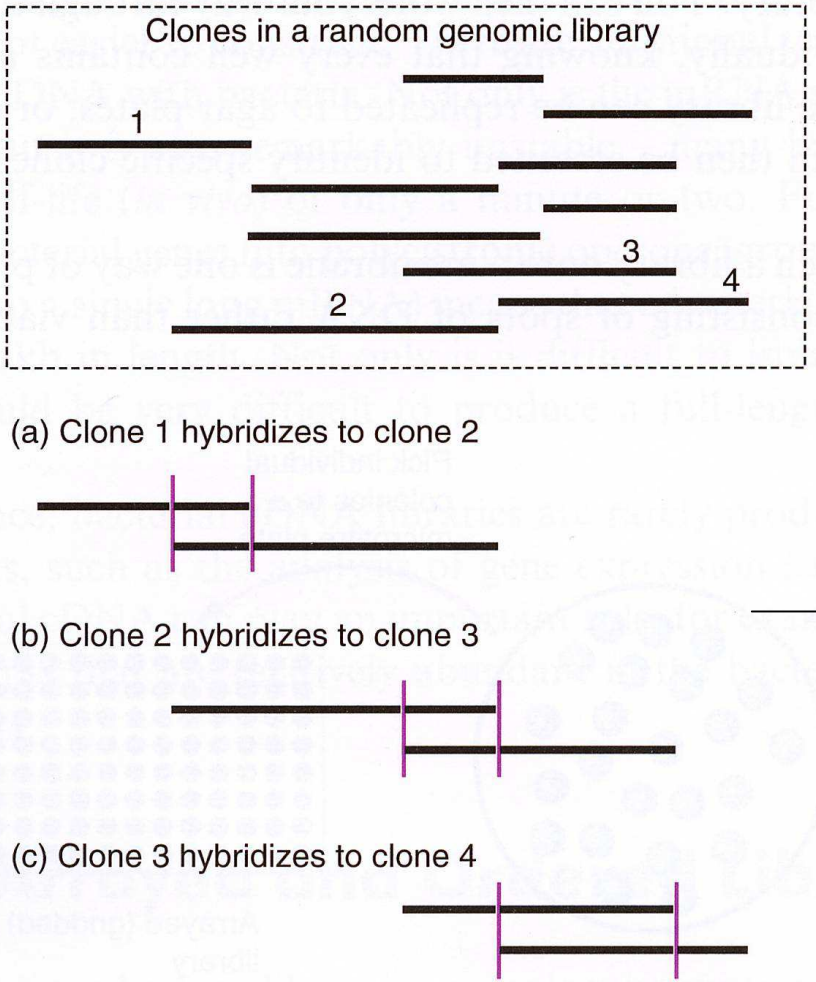



**Figure 7.11** Production of an ordered library

"Chromosome Walking"
-> also used in "Human Genome Project"

Screening of the library - later

# Evaluation of library

**Box 7.1 Estimates of the required size of genomic libraries**

| Organism | Genome size | Vector type | Insert size | *P* | Library size |
|---|---|---|---|---|---|
| Bacterium | $4 \times 10^6$ bases | plasmid | 4 kb | 0.99 | $4.6 \times 10^3$ |
| | | lambda replacement | 18 kb | 0.99 | $1.0 \times 10^3$ |
| | | cosmid | 40 kb | 0.99 | 458 |
| | | BAC | 300 kb | 0.99 | 59 |
| | | | | | |
| Mammal | $3 \times 10^9$ bases | plasmid | 4 kb | 0.99 | $3.5 \times 10^6$ |
| | | lambda replacement | 18 kb | 0.99 | $7.7 \times 10^5$ |
| | | cosmid | 40 kb | 0.99 | $3.5 \times 10^5$ |
| | | BAC | 300 kb | 0.99 | $4.6 \times 10^4$ |

The values shown for the genome sizes of bacteria and mammals are examples for the purpose of this calculation. The actual genome sizes vary quite widely from one organism to another. The insert sizes for specific vectors will also vary.

Why must be the size of the library much larger than the size of The particular genome?

# Sizes of Some DNA Molecules.

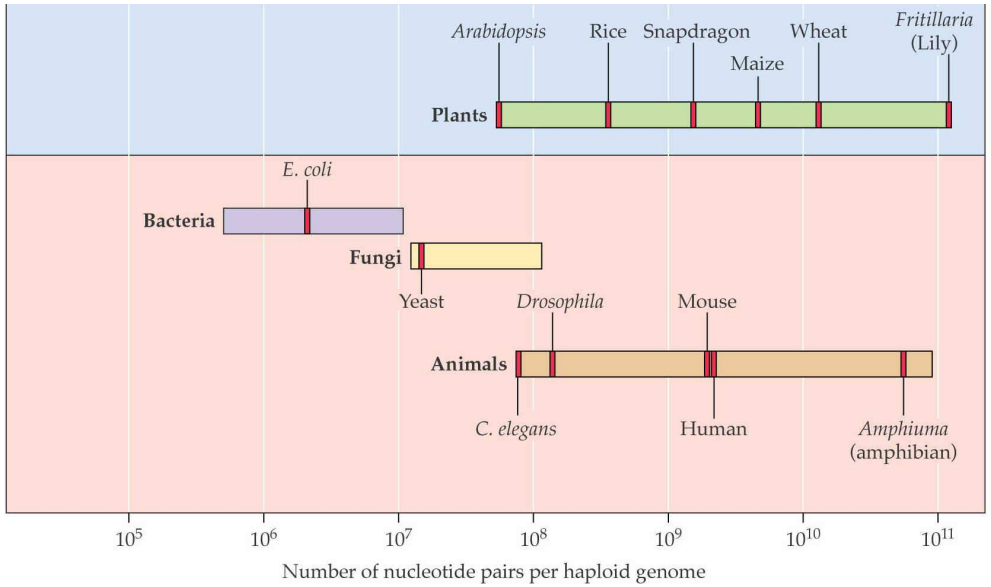
| Organism | Number of base pairs (kb)[a] | Contour length (μm) |
|---|---|---|
| ***Viruses*** | | |
| Polyoma, SV40 | 5.2 | 1.7 |
| λ Bacteriophage | 48.6 | 17 |
| T2, T4, T6 bacteriophage | 166 | 55 |
| Fowlpox | 280 | 193 |
| ***Bacteria*** | | |
| *Mycoplasma hominis* | 760 | 260 |
| *Escherichia coli* | 4,600 | 1,600 |
| ***Eukaryotes*** | | |
| Yeast (in 17 haploid chromosomes) | 12,000 | 4,100 |
| *Drosophila* (in 4 haploid chromosomes) | 180,000 | 61,000 |
| Human (in 23 haploid chromosomes) | 3,200,000 | 1,100,000 |
| Lungfish (in 19 haploid chromosomes) | 102,000,000 | 35,000,000 |

[a]kb = kilobase pair = 1000 base pairs (bp).

*Source:* Mainly Kornberg, A. and Baker, T.A., *DNA Replication* (2nd ed.), *p.* 20, Freeman (1992).

# How big genomes are?

*Amoeba dubia* 200x
the size of human genome




*Paris japonica* 50x
octoploid


Lungfish 40x

# Plant genome sizes

**54 Mbp – *Cardamine amara***

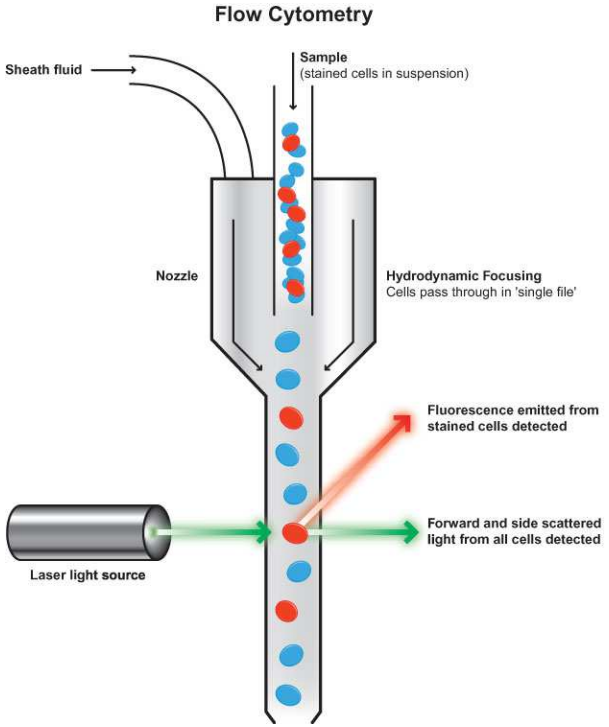
**124 852 Mbp - *Fritillaria***

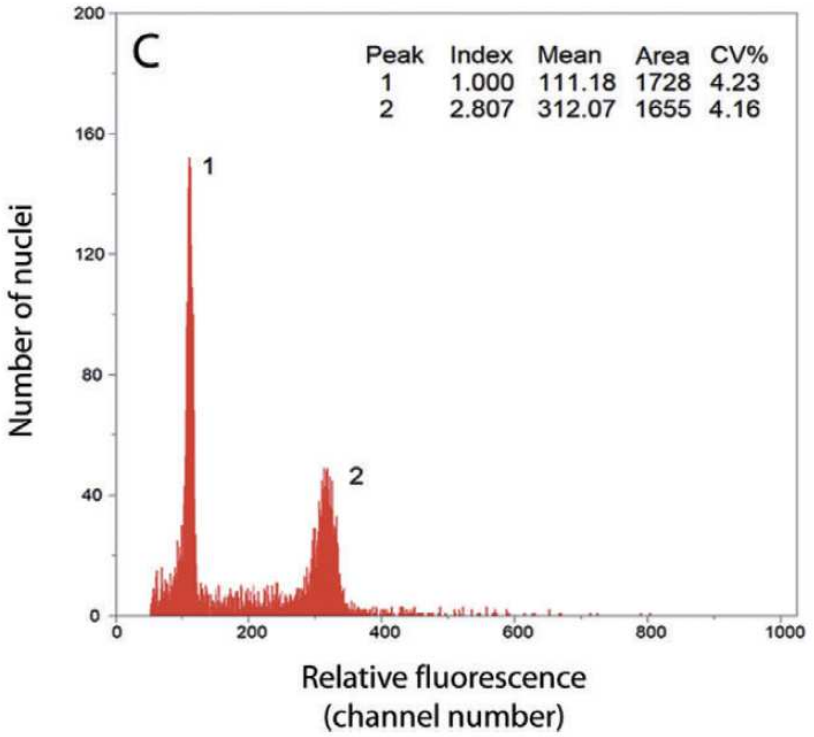
**149 000 Mbp - *Paris japonica***

- ***currently the largest***
**(not only plant)**

http://data.kew.org/cvalues/

# How to determine genome size?




Stained DNA – fluorescent – amount of fluorescence from single cell

# C-value

C-value = size of genome in non-replicated gamete
- normal diploid cell 2C

genome size (bp) = (0.910 x $10^9$) x DNA content (pg)

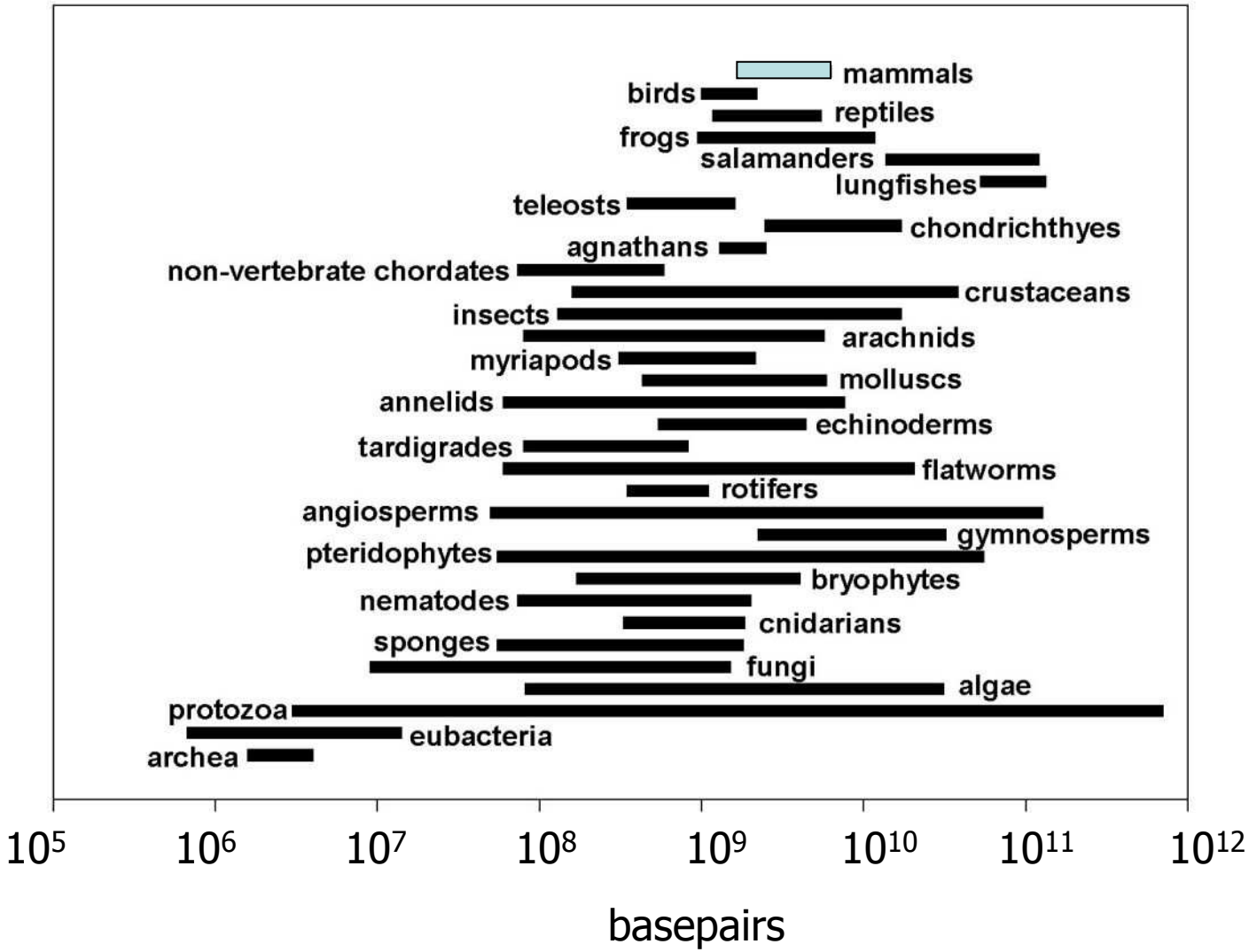
DNA content (pg) = genome size (bp) / (0.910 x $10^9$)

1 pg = cca 910 Mbp; MW (1 bp) = cca 660 Da

# C-value paradox

there is no strong correlation between complexity of an organism and the size of its genome

# Amount of DNA in a Genome Does Not Correlate with Complexity



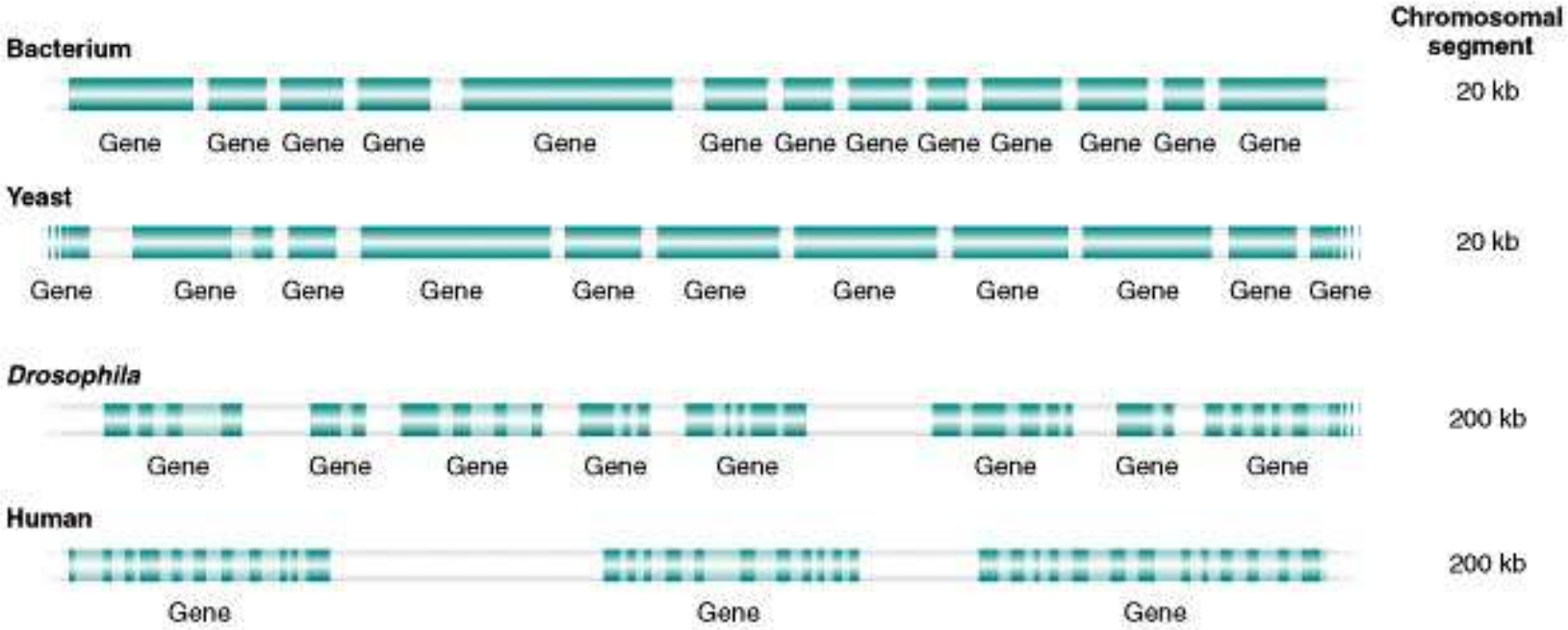
## TABLE 17.1

### Representative Sequenced Genomes

| ORGANISM | HAPLOID GENOME SIZE (Mb) | NUMBER OF GENES | PROTEIN-CODING SEQUENCE |
|---|---|---|---|
| Bacteria | | | |
| *M. genitalium* | 0.58 | 485 | 88% |
| *H. influenzae* | 1.8 | 1,738 | 89% |
| *E. coli* | 4.6 | 4,377 | 88% |
| Yeasts | | | |
| *S. cerevisiae* | 12.5 | 5,770 | 70% |
| *S. pombe* | 12.5 | 4,929 | 60% |
| Plants | | | |
| *A. thaliana* | 115 | 28,000 | 25% |
| Rice | 390 | 37,544 | 12% |
| Animals | | | |
| *C. elegans* | 100 | 19,427 | 25% |
| *D. melanogaster* | 123 | 13,379 | 13% |
| Pufferfish | 342 | 27,918 | 10% |
| Chicken | 1,130 | 25,000 | 3% |
| Human | 3,300 | 24,000 | 1.2% |

Mb = millions of base pairs

# Gene Spacing in Various Species



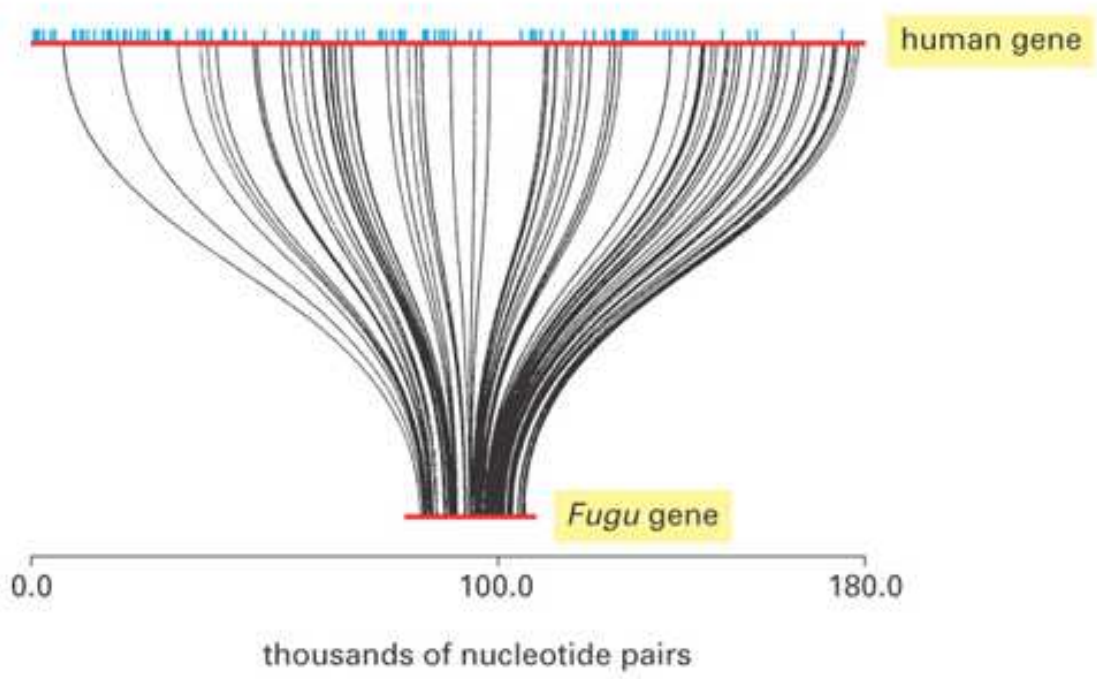



http://www.mun.ca/biology/desmid/brian/BIOL2250/Week_Two/genespac.jpg

# Log-log plot of the total number of annotated proteins in genomes submitted to GeneBank as a function of genome size

# Even the Amount of DNA a Gene Spans Differs Amongst Species

Genome
3.7 Gb




0.39 Gb

Figure 9-21 Essential Cell Biology, 2/e. (© 2004 Garland Science)

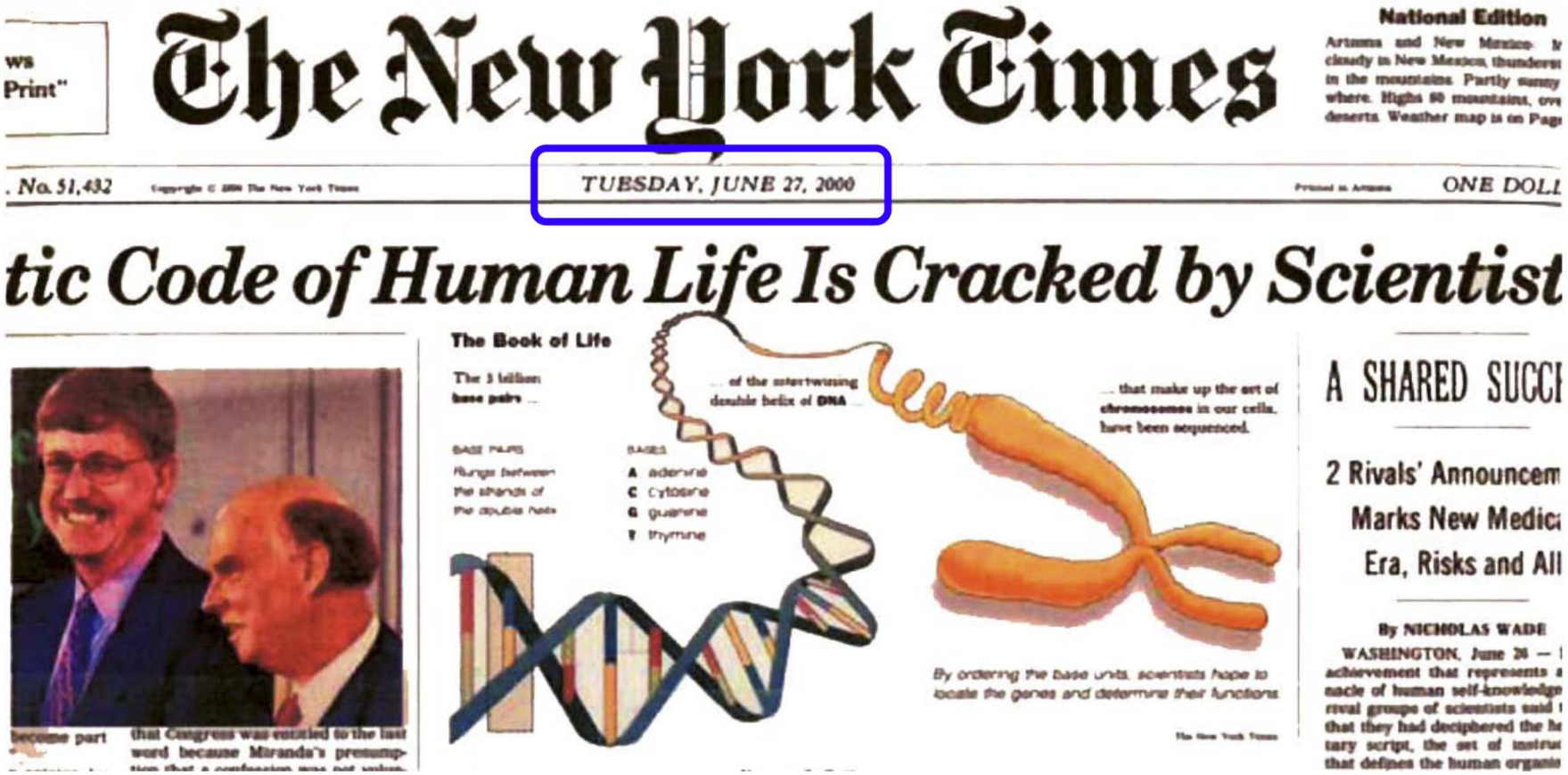
# Genome sequencing

- Determination of the order of nucleotides in chromosome(s)
- Genes, non-coding DNA, regulatory DNA, structural RNA….

- Not part of the lecture

# The Human Genome Was Sequenced Ten Years Ago!

## The Human Genome Project



ws
Print"

The New York Times

National Edition

. No. 51,432 TUESDAY, JUNE 27, 2000 ONE DOLL

tic Code of Human Life Is Cracked by Scientist

The Book of Life

The 3 billion base pairs ... of the intertwining double helix of DNA ... that make up the set of chromosomes in our cells, have been sequenced.

BASE PAIRS Rungs between the strands of the double helix

BASES
A adenine
C cytosine
G guanine
T thymine

By ordering the base units, scientists hope to locate the genes and determine their functions

A SHARED SUCCI

2 Rivals' Announcem
Marks New Medic
Era, Risks and All

By NICHOLAS WADE

WASHINGTON, June 26 —

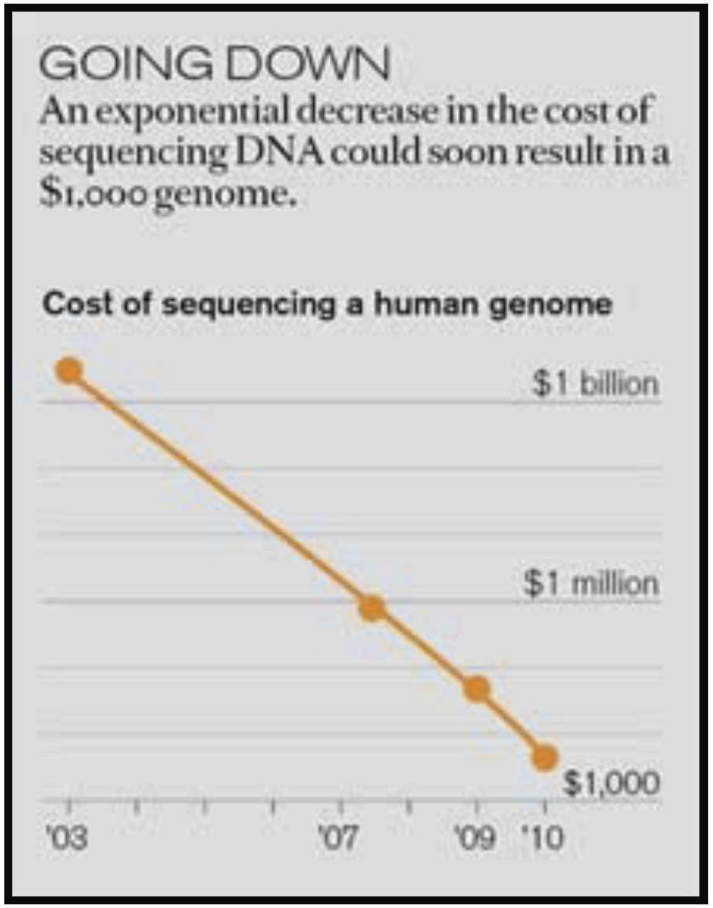
become part that Congress was entitled to the last word because Miranda's presump-

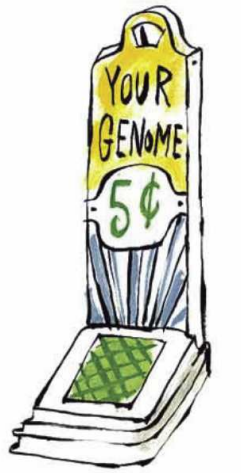
achievement that represents a
nacle of human self-knowledge
rival groups of scientists said
that they had deciphered the he
tary script, the set of instru
that defines the human organis

## Public & Private Effort Using Different Strategies - A Race!

## 3 Billion Dollars & Took 15 Years

# Cost of Sequencing is Going Down Precipitously!

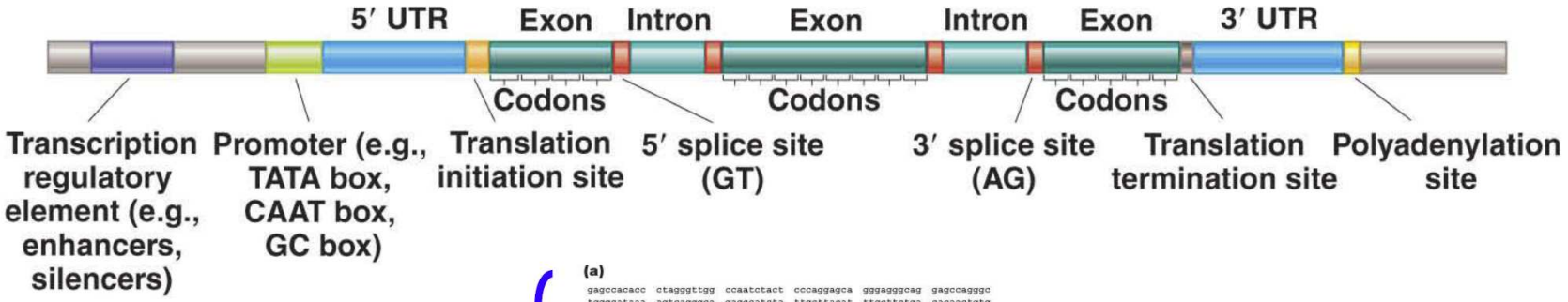





23andMe

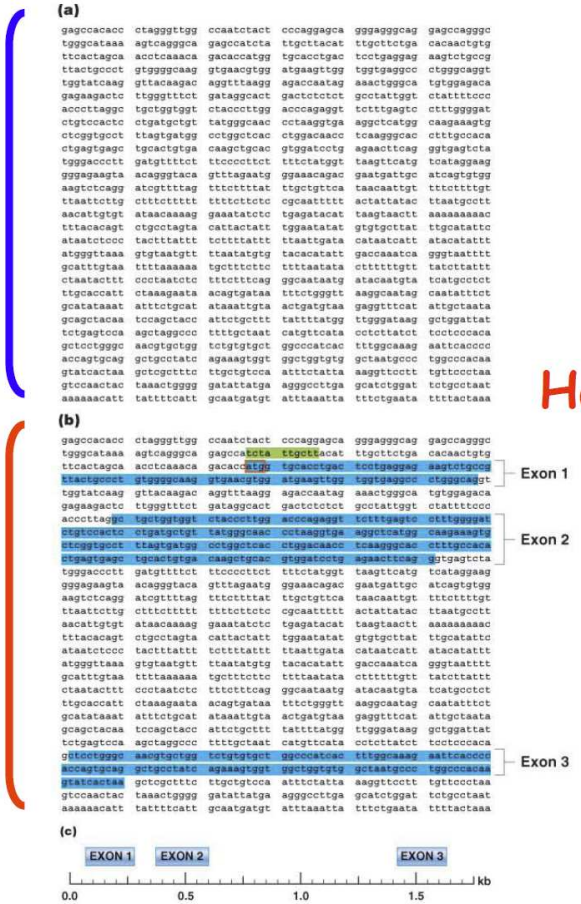
Knome

# Finding and Annotating Genes Within Genome Sequences




Unannotated

Annotated




A HUGE Job!
Especially if Genes
Have Unknown Functions


PHASE TWO: INTERPRETATION
I THINK I FOUND A CORNER PIECE.
3 BILLION PIECES
GENOME