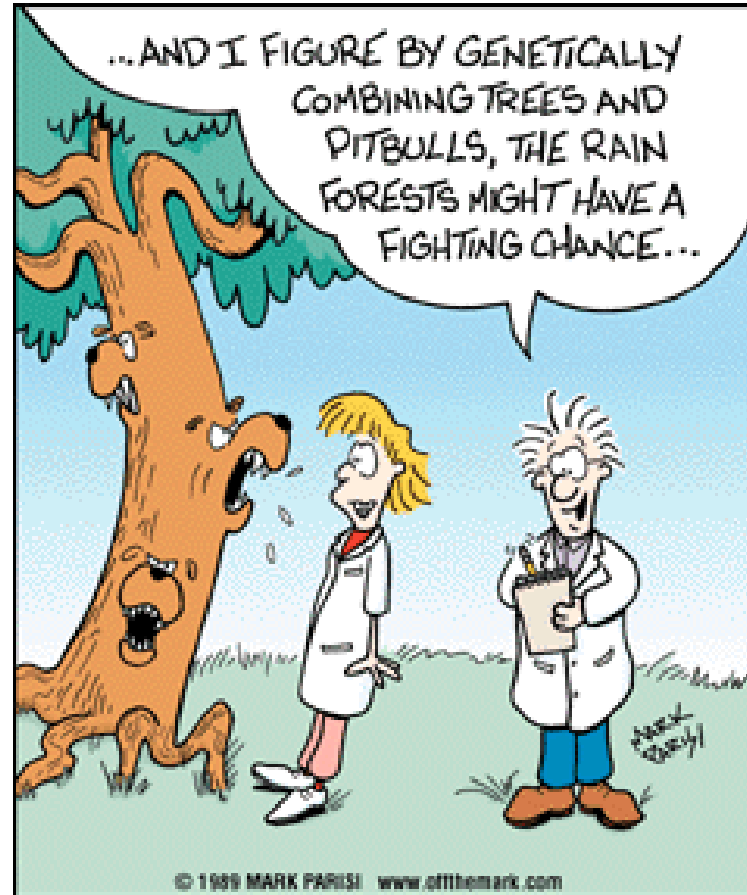


# Recombinant DNA Technology

**off the mark** by Mark Parisi  
www.offthemark.com



# Definitions

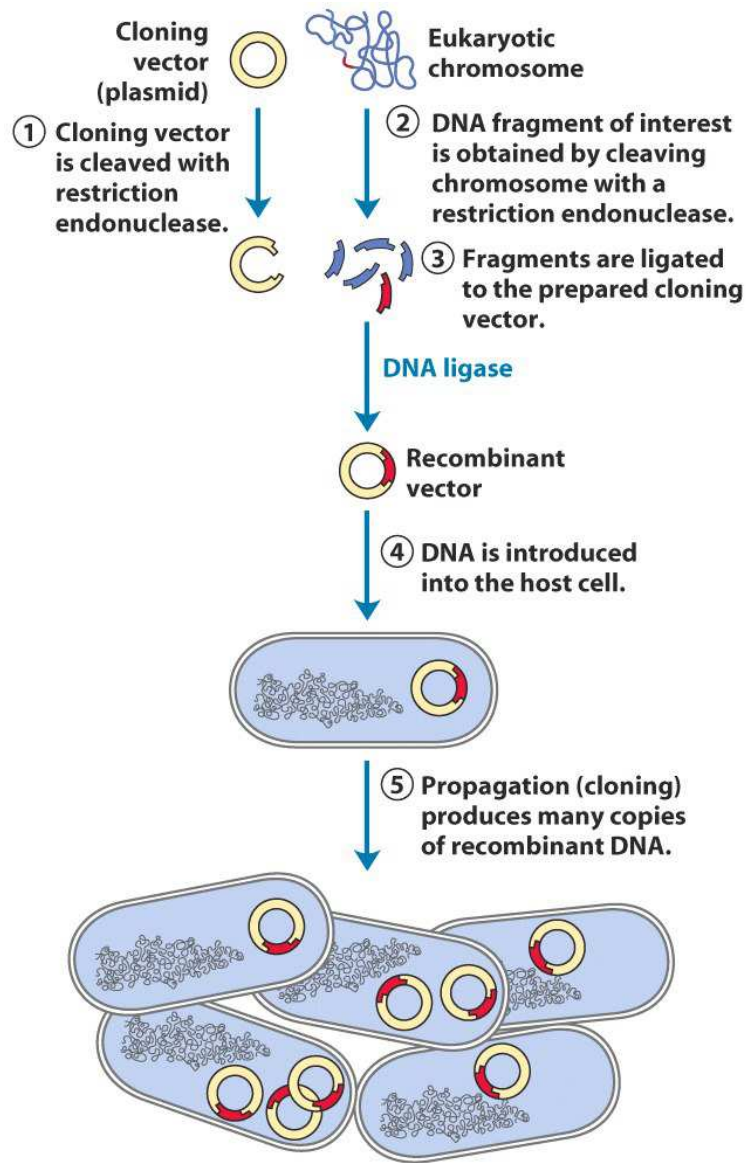
**Recombinant DNA**, a DNA construct created by fusing different fragments of DNA

**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

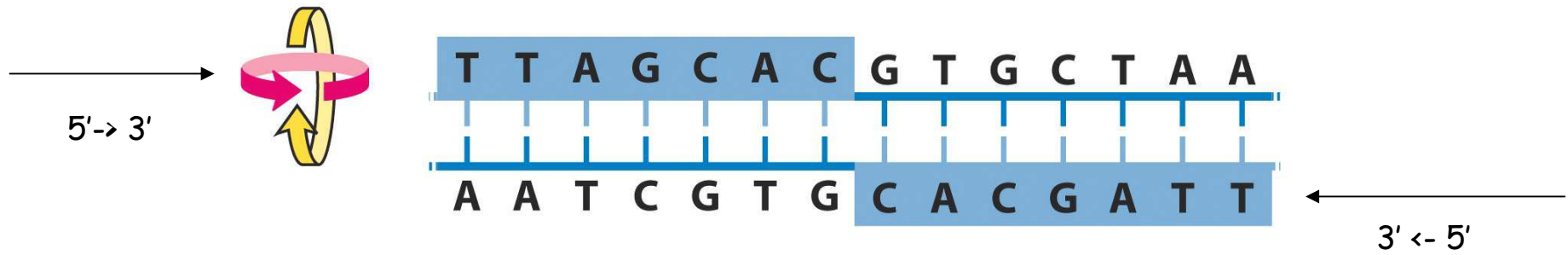
# Recombinant DNA Technology



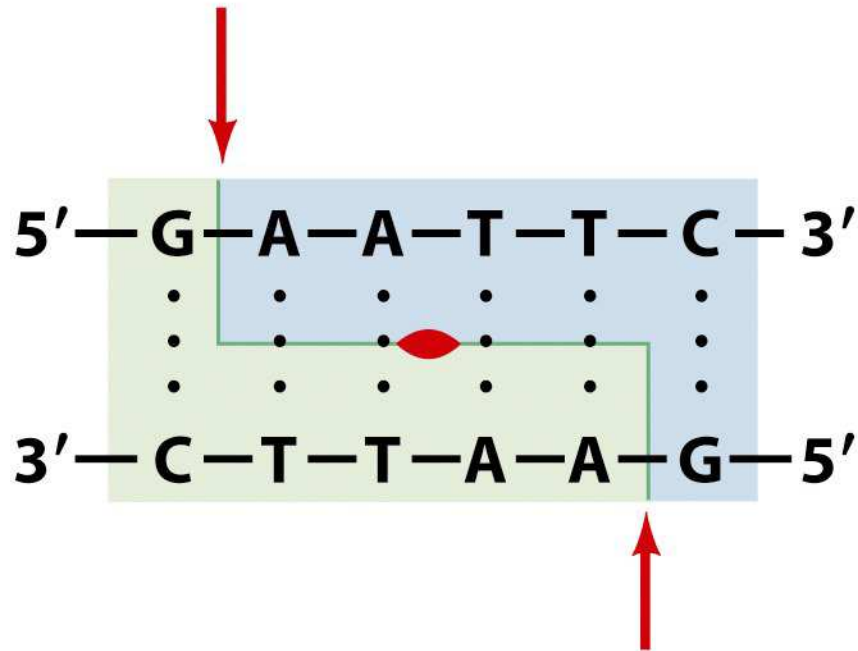
**Clones** -> Cells or organisms with identical DNA

# Restriction endonucleases

## Palindrome

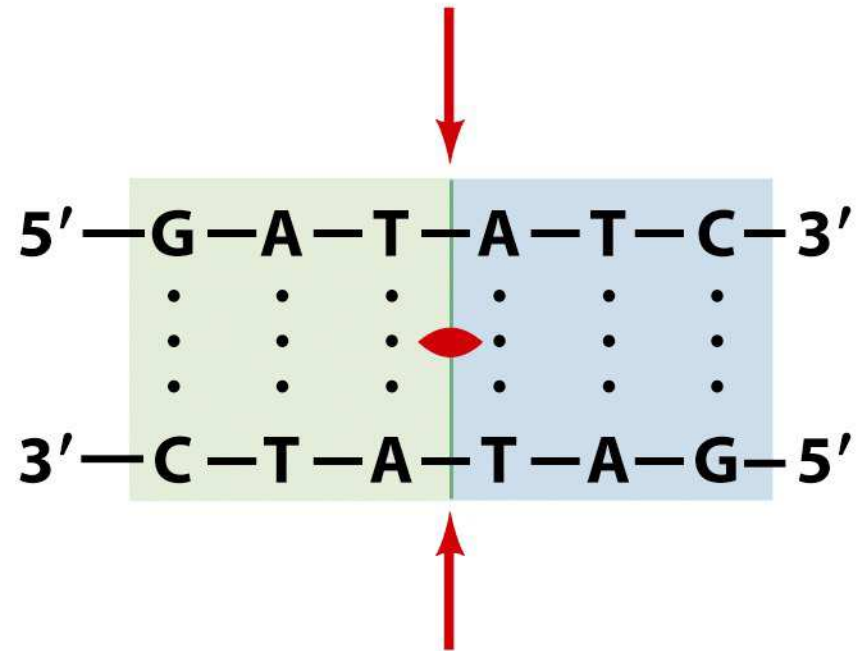


**(a) *EcoRI***



 **Cleavage site**

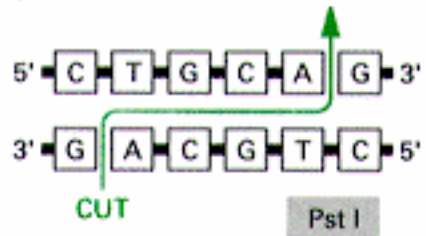
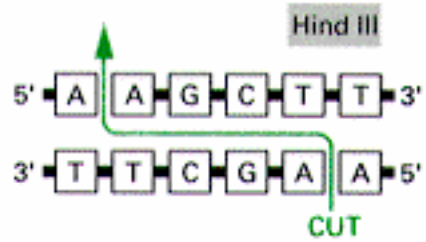
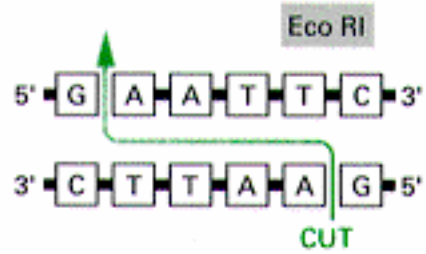
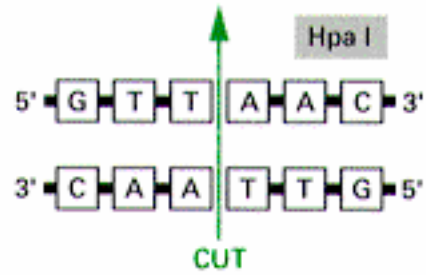
**(b) *EcoRV***



 **Twofold symmetry axis**

Figure 3-16 Fundamentals of Biochemistry, 2/e  
© 2006 John Wiley & Sons

How frequently will cut *EcoRI* in a random DNA sequence?



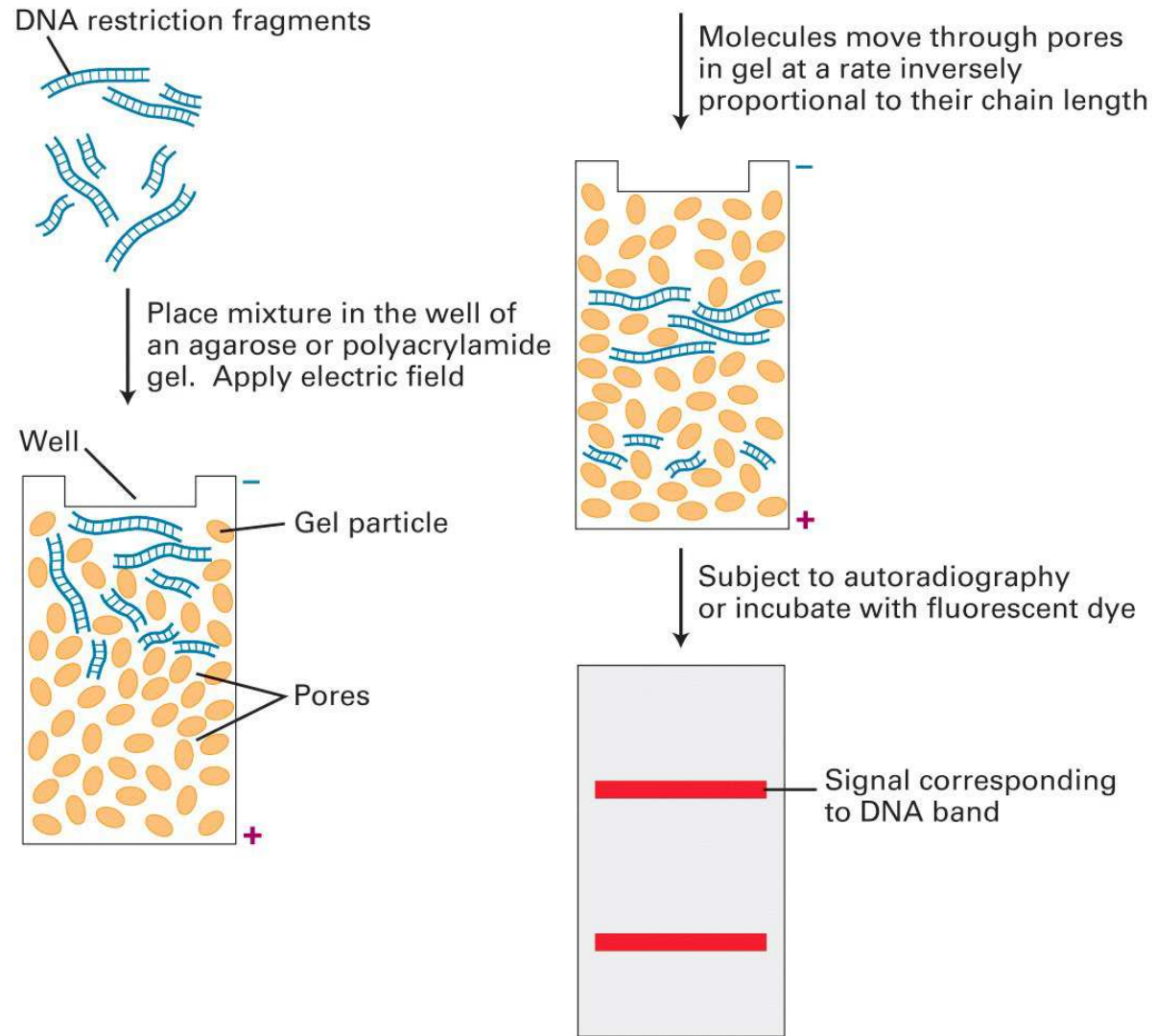
**Table 3-2** Recognition and Cleavage Sites of Some Restriction Enzymes

Enzyme	Recognition Sequence <sup>a</sup>	Microorganism
<i>Alu</i> I	AG↓CT	<i>Arthrobacter luteus</i>
<i>Bam</i> HI	G↓GATCC	<i>Bacillus amyloliquefaciens</i> H
<i>Bgl</i> I	GCCNNNNN↓NGGC	<i>Bacillus globigii</i>
<i>Bgl</i> II	A↓GATCT	<i>Bacillus globigii</i>
<i>Eco</i> RI	G↓AATTC	<i>Escherichia coli</i> RY13
<i>Eco</i> RII	↓CC(↑)GG	<i>Escherichia coli</i> R245
<i>Eco</i> RV	GAT↓ATC	<i>Escherichia coli</i> J62 pLG74
<i>Hae</i> II	RGCGC↓Y	<i>Haemophilus aegyptius</i>
<i>Hae</i> III	GG↓CC	<i>Haemophilus aegyptius</i>
<i>Hind</i> III	A↓AGCTT	<i>Haemophilus influenzae</i> R <sub>d</sub>
<i>Hpa</i> II	C↓CGG	<i>Haemophilus parainfluenzae</i>
<i>Msp</i> I	C↓CGG	<i>Moraxella</i> species
<i>Pst</i> I	CTGCA↓G	<i>Providencia stuartii</i> 164
<i>Pvu</i> II	CAG↓CTG	<i>Proteus vulgaris</i>
<i>Sal</i> I	G↓TCGAC	<i>Streptomyces albus</i> G
<i>Taq</i> I	T↓CGA	<i>Thermus aquaticus</i>
<i>Xho</i> I	C↓TCGAG	<i>Xanthomonas holcicola</i>

<sup>a</sup>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

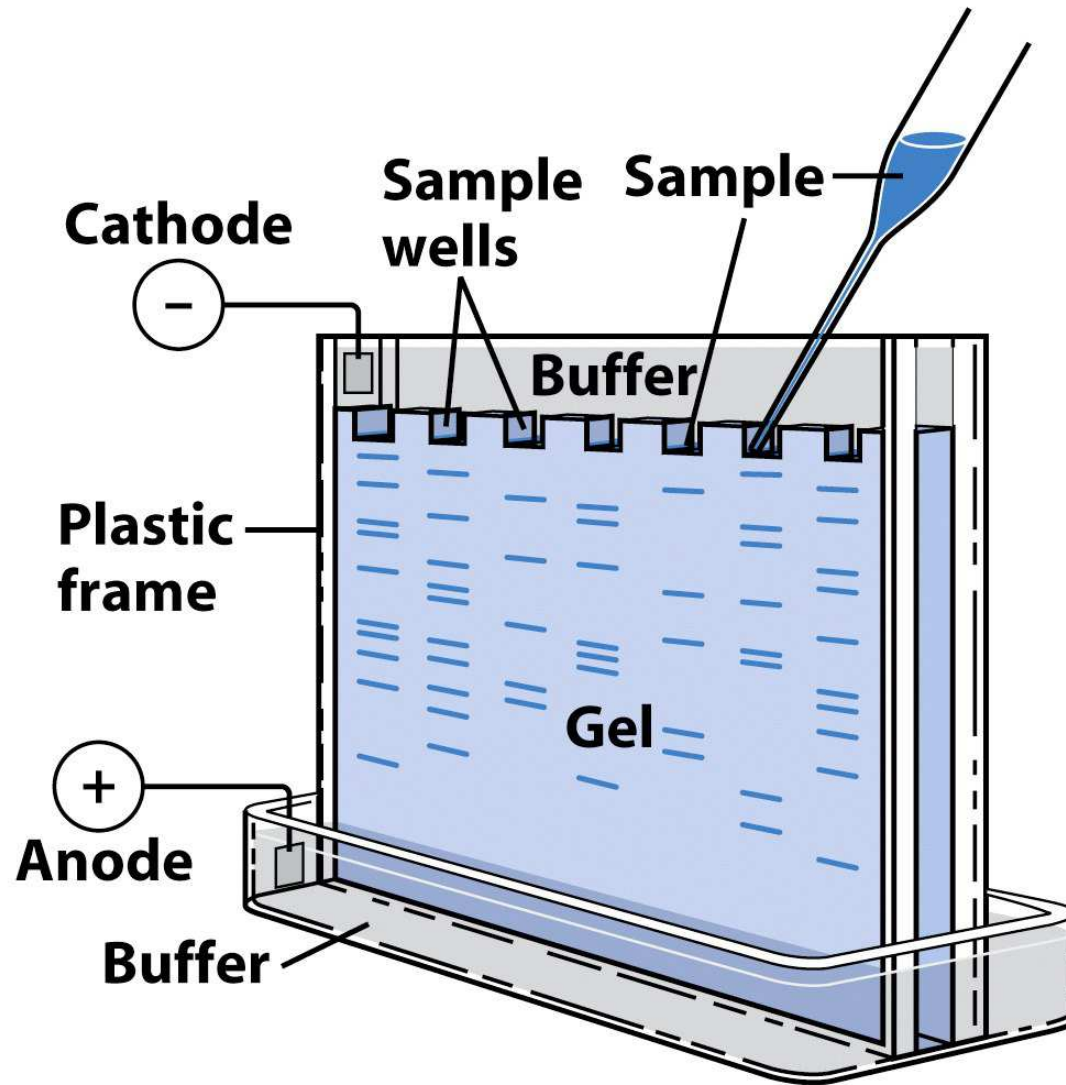
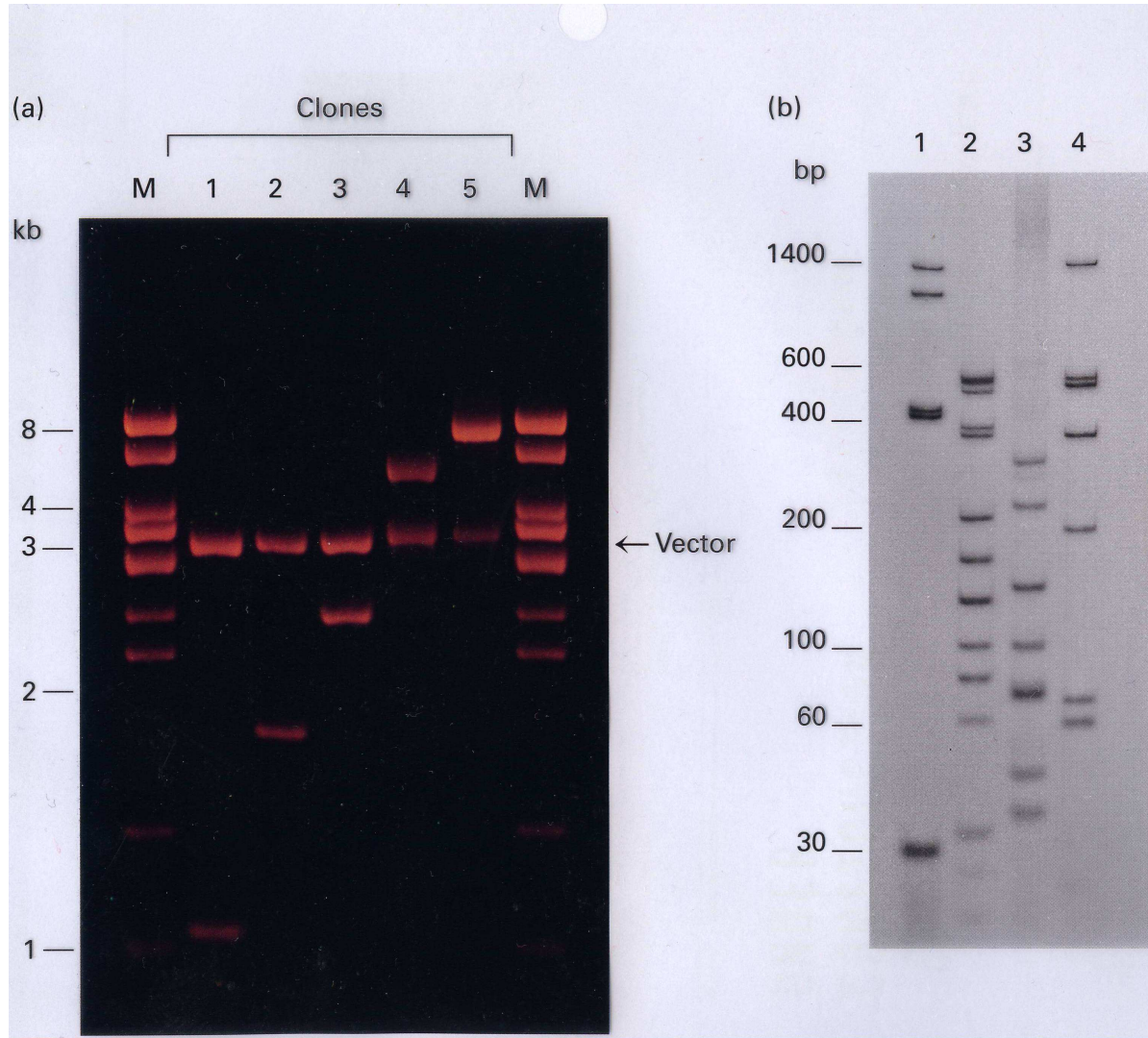
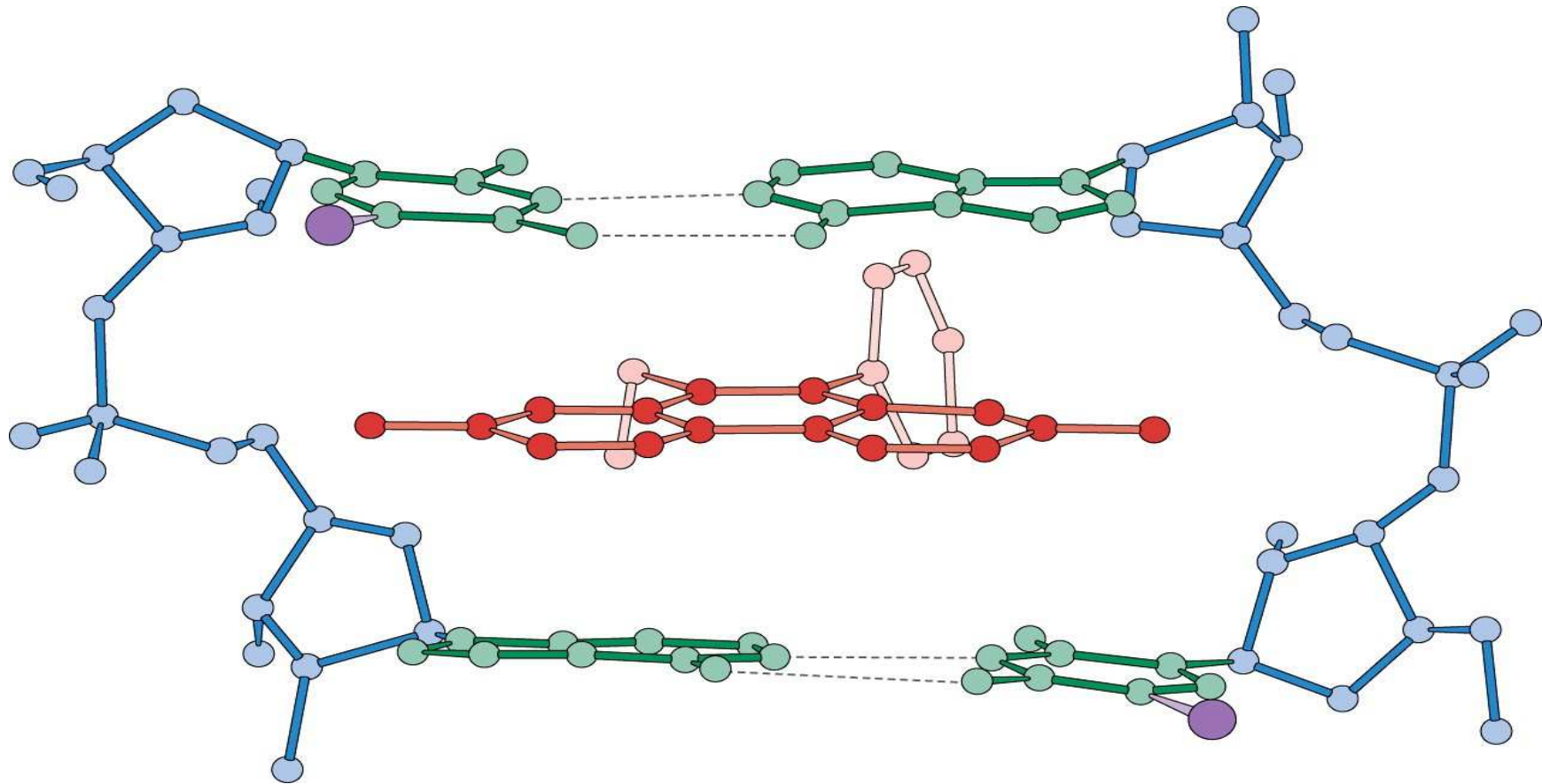


Figure 3-17 Fundamentals of Biochemistry, 2/e  
© 2006 John Wiley & Sons

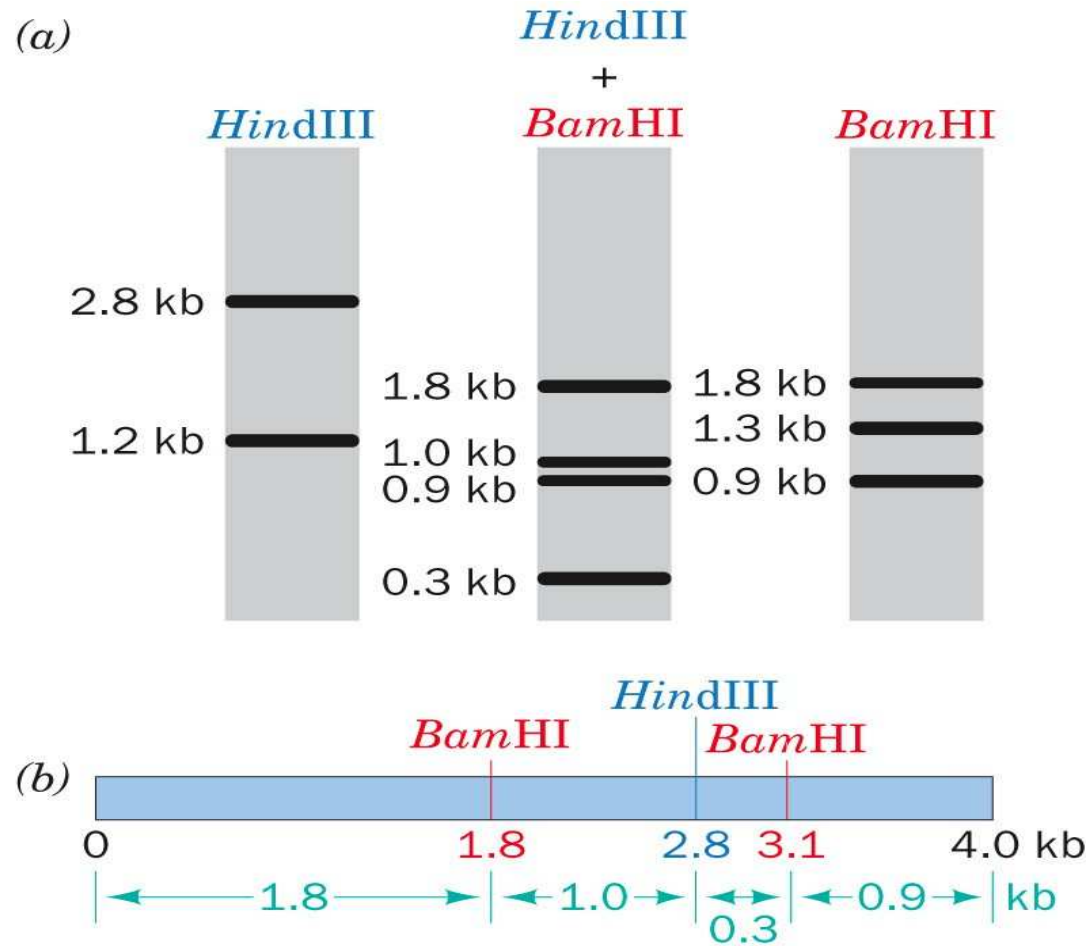
# Gel Electrophoresis



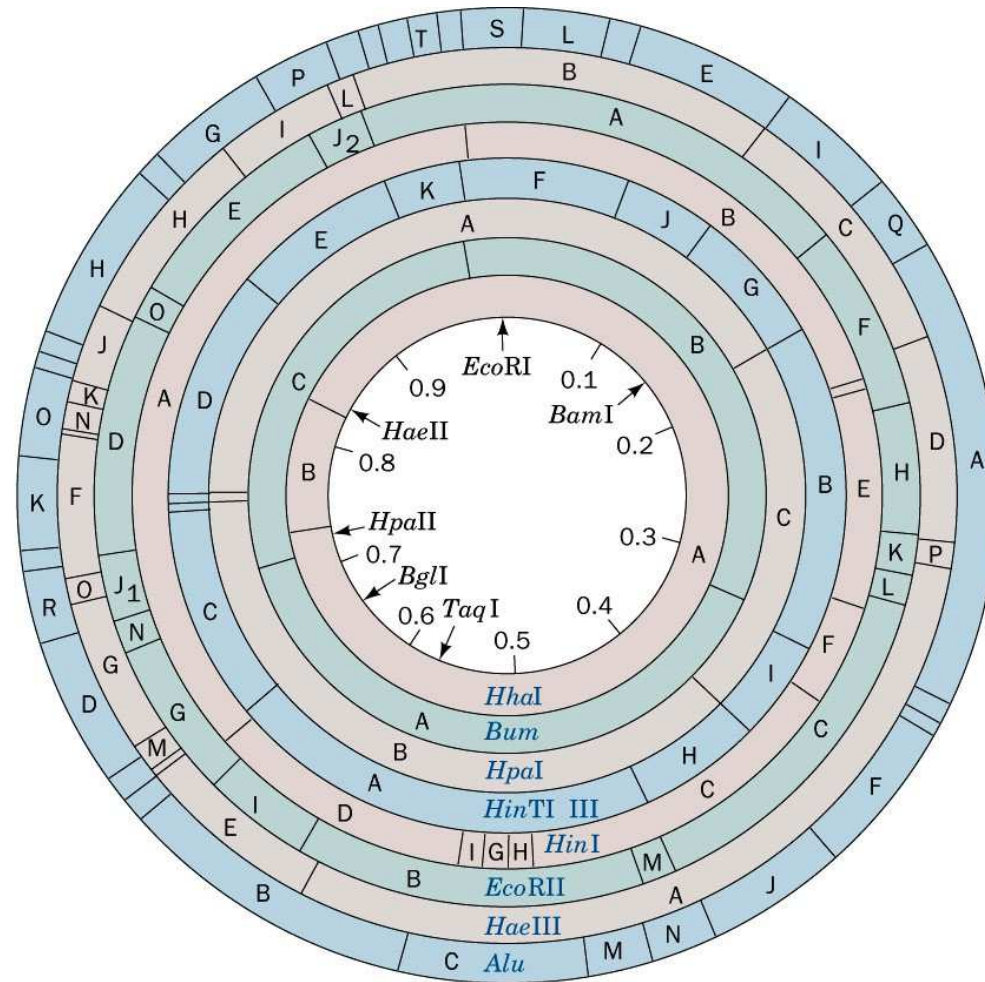
# X-Ray structure of a complex of ethidium bromide with DNA.

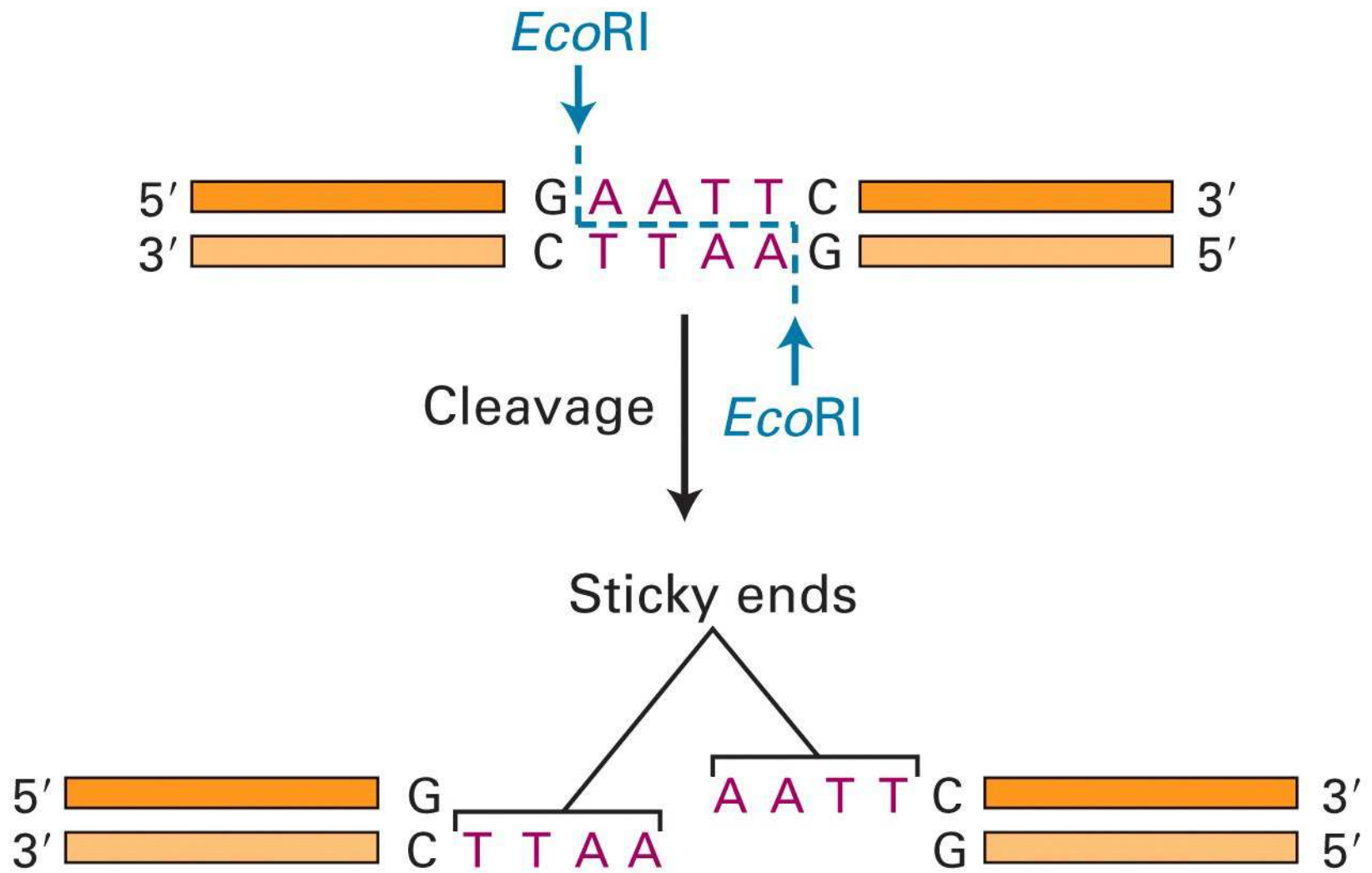


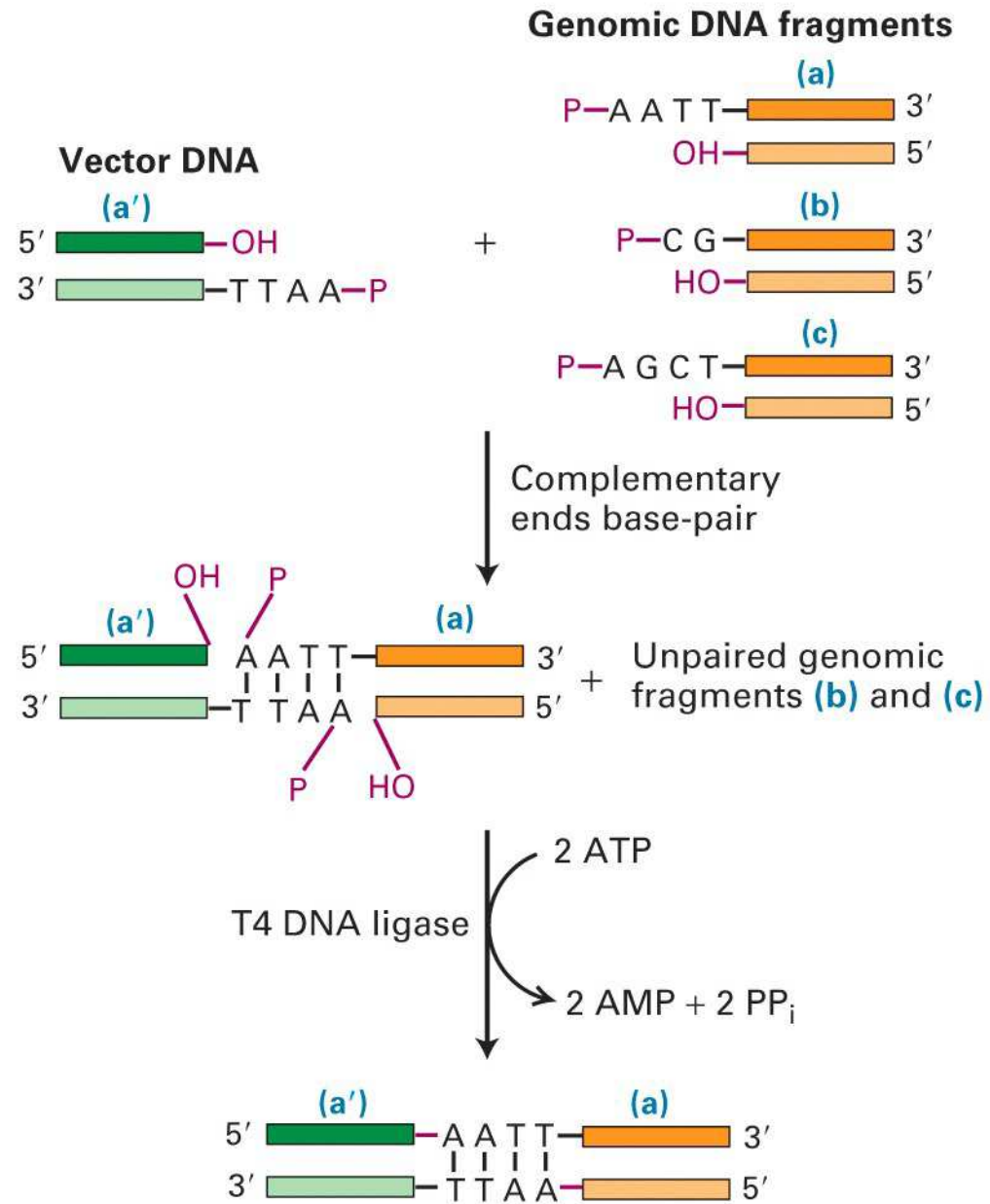
# Construction of a restriction map.

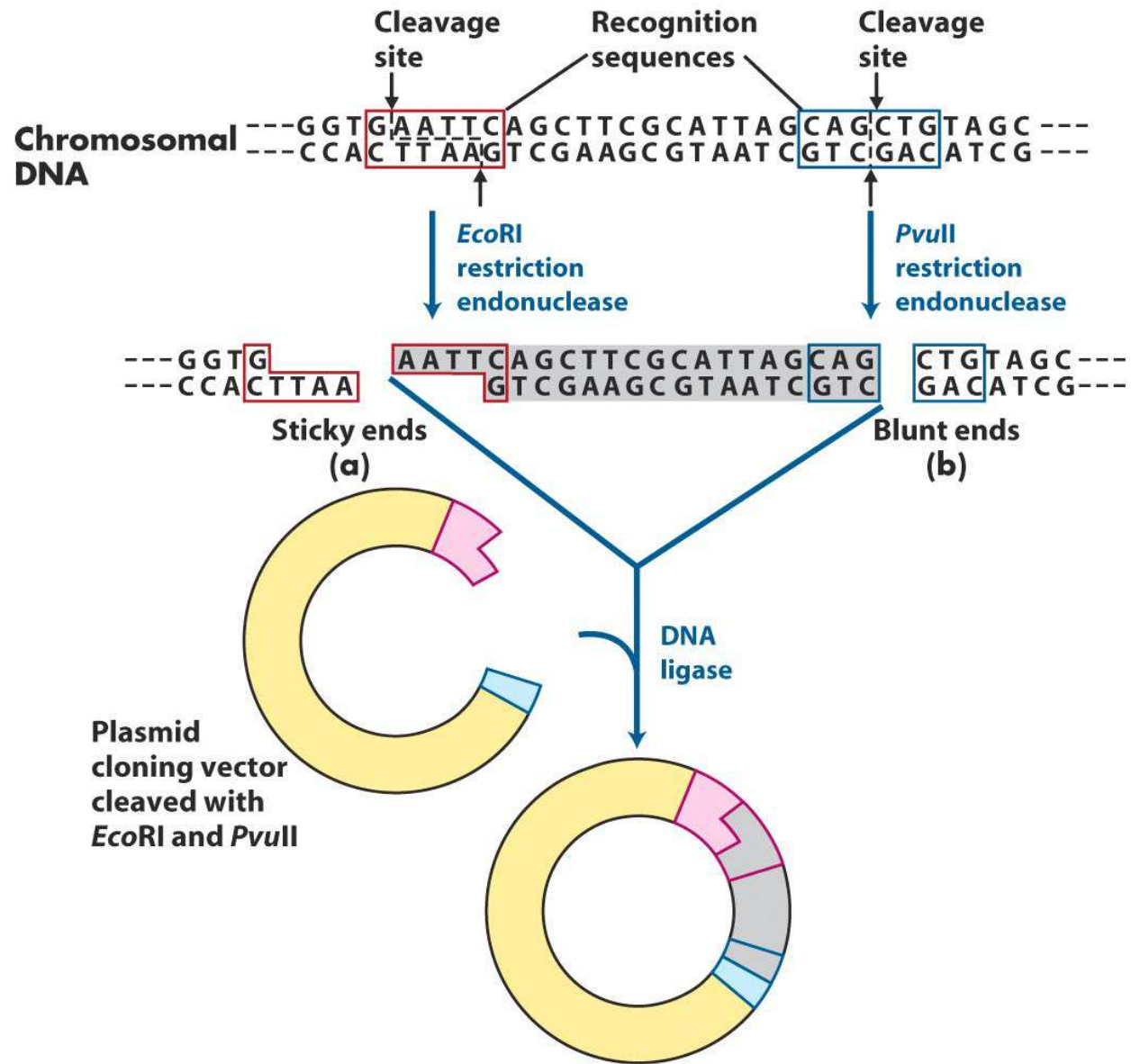


# Restriction map for the 5243-bp circular DNA of SV40.

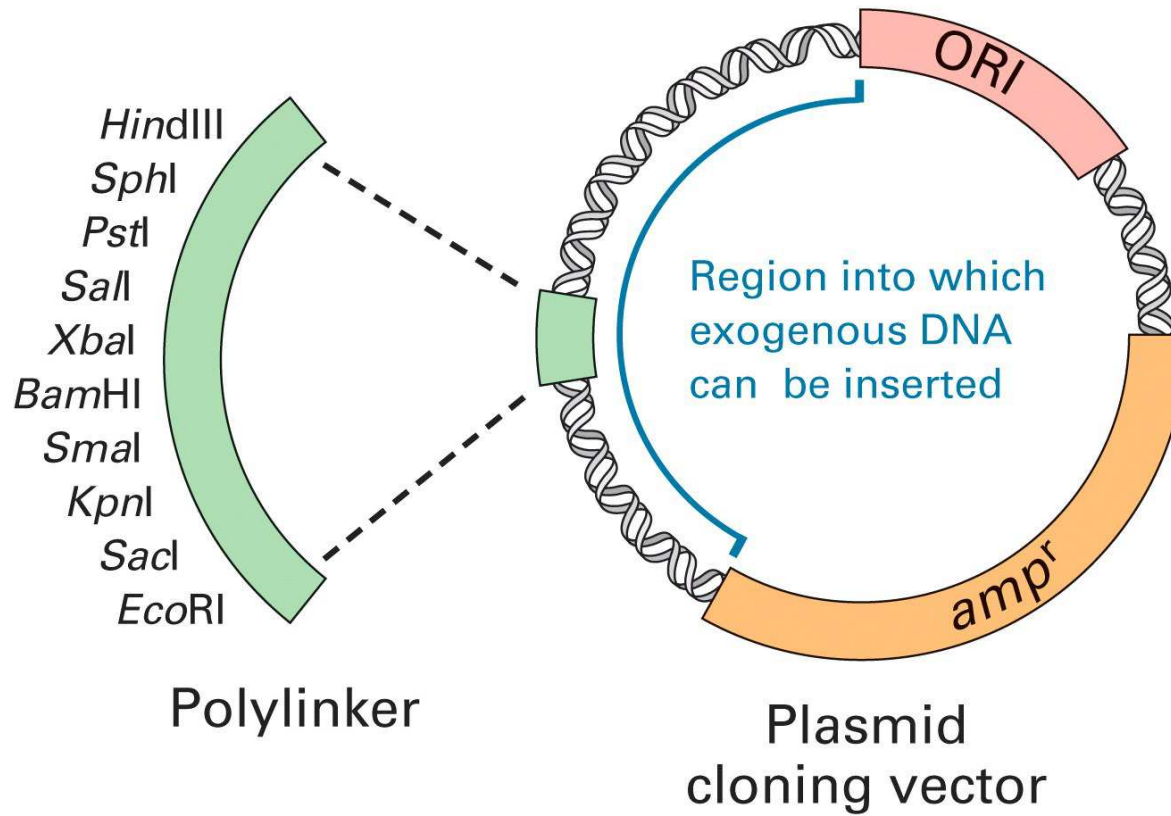








# Plasmid Cloning Vectors



# Plasmid Cloning Vectors

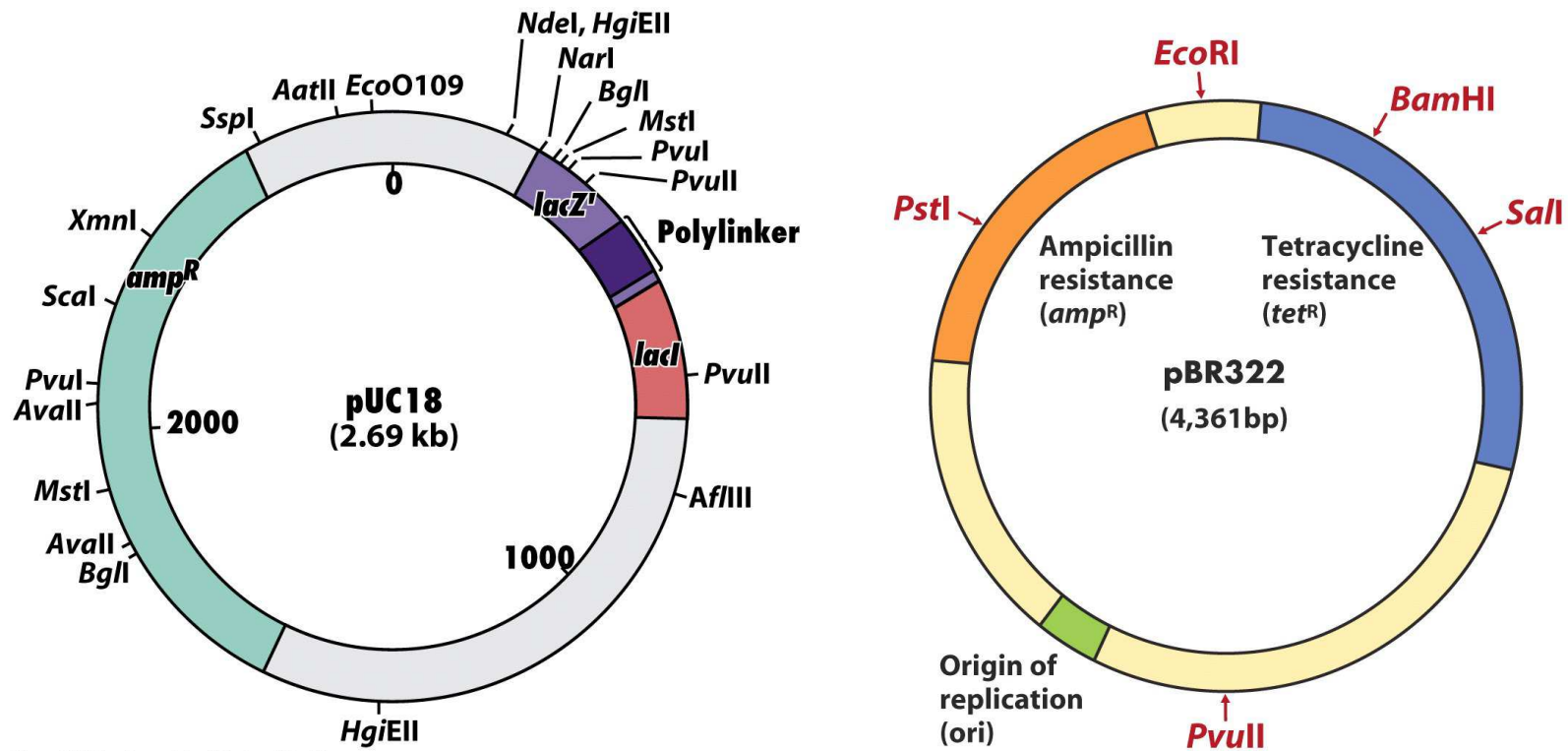
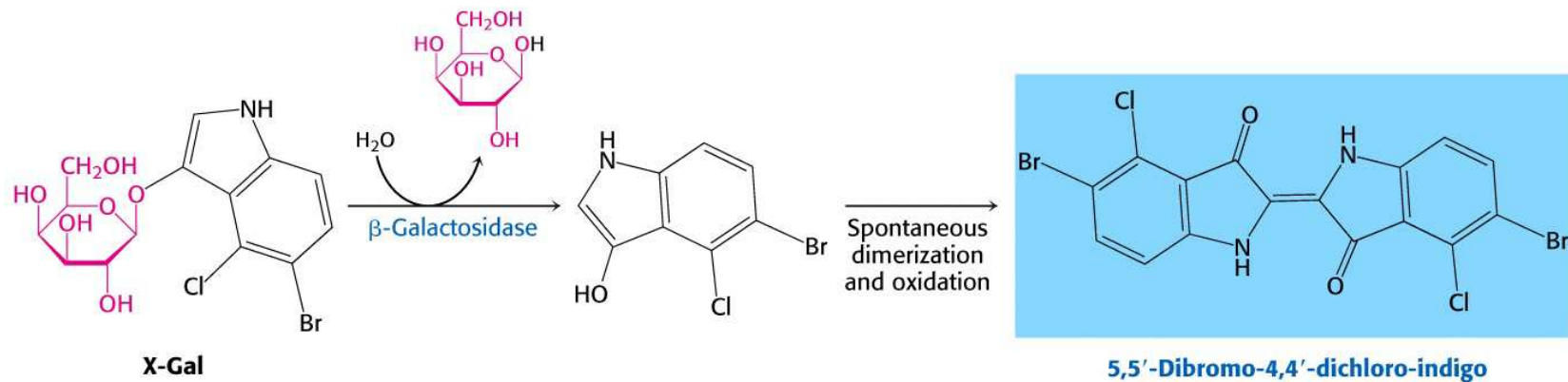
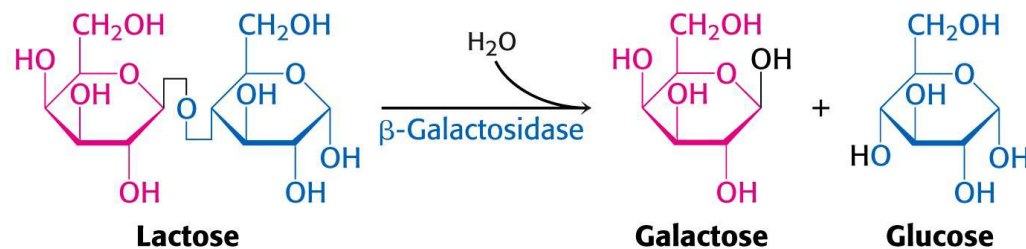


Figure 3-25 Fundamentals of Biochemistry, 2/e  
© 2006 John Wiley & Sons

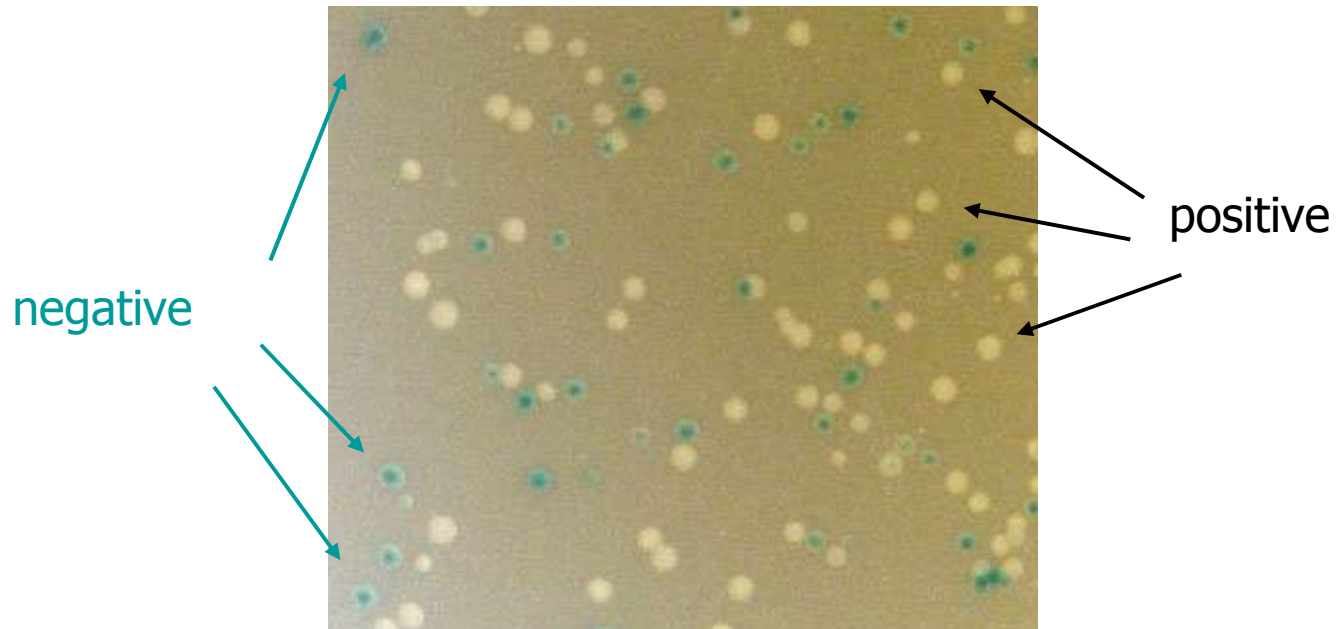
# Insertional inactivation

## Gene in cloning site:

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



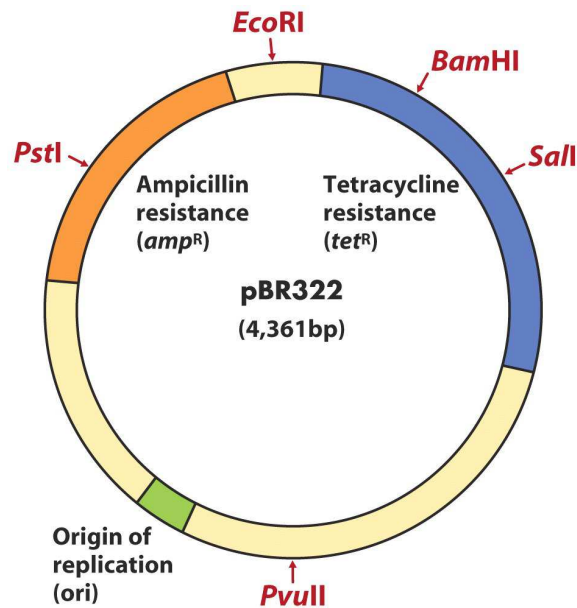
# Blue/White Selection



# Insertional inactivation

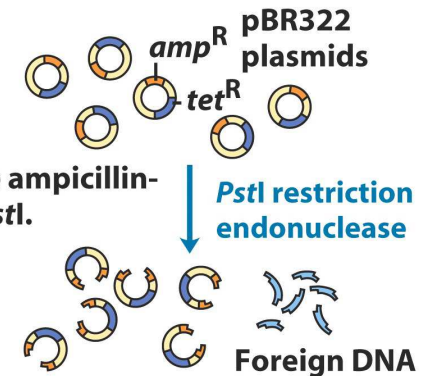
## Gene in cloning site:

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



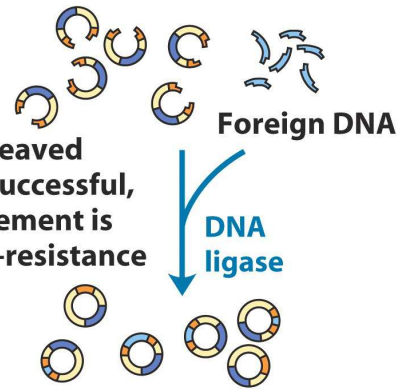
①

pBR322 is cleaved at the ampicillin-resistance element by *Pst*I.



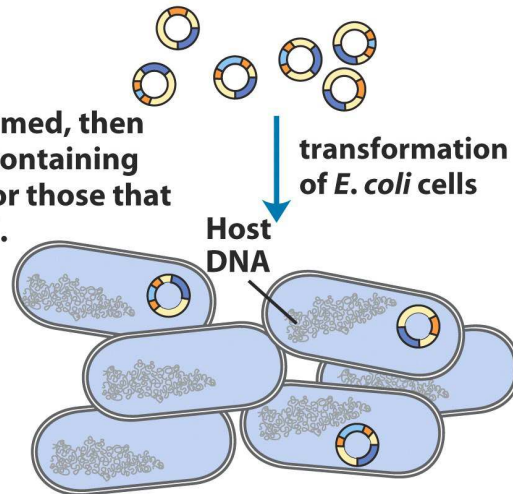
②

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



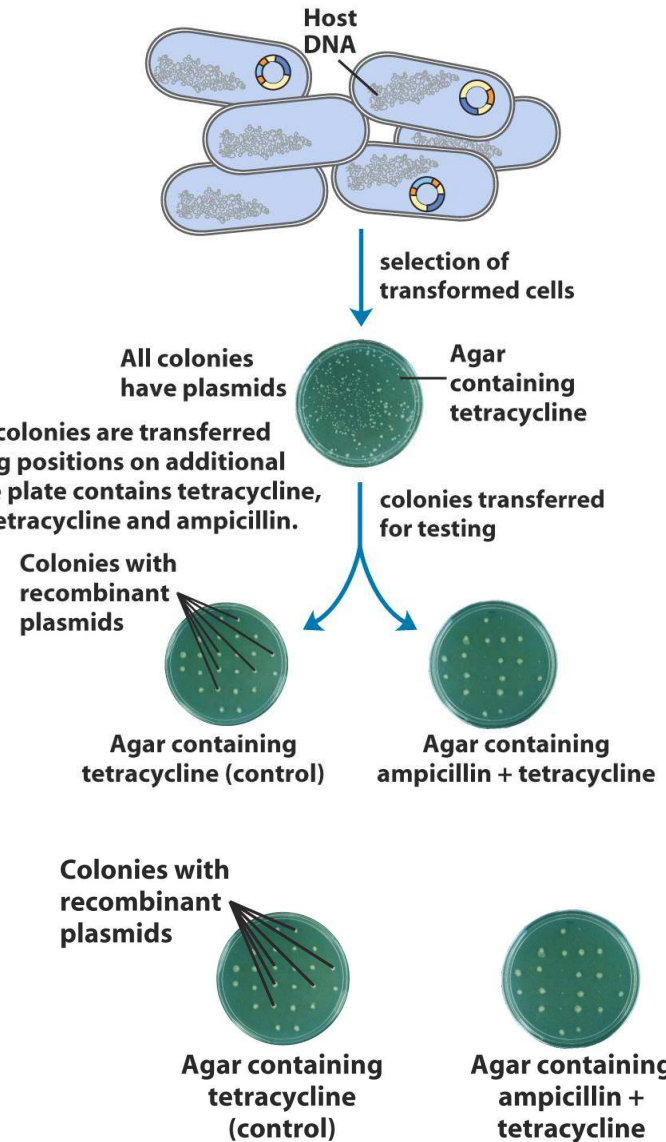
③

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



④

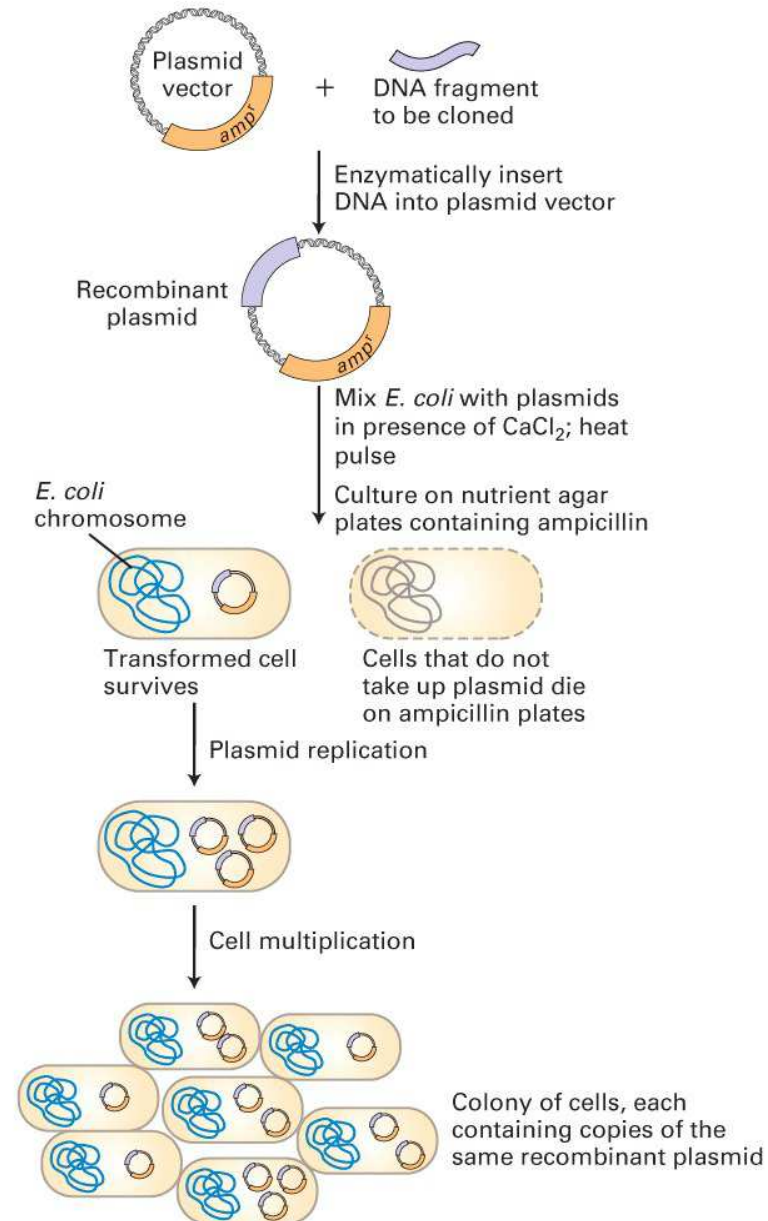
Individual colonies are transferred to matching positions on additional plates. One plate contains tetracycline, the other tetracycline and ampicillin.



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	<i>E. coli</i>	0.1–10
Bacteriophage $\lambda$	$\lambda$ / <i>E. coli</i>	10–20
Cosmid	<i>E. coli</i>	35–45
Bacteriophage P1	<i>E. coli</i>	80–100
BAC	<i>E. coli</i>	50–300
P1 bacteriophage-derived artificial chromosome	<i>E. coli</i>	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

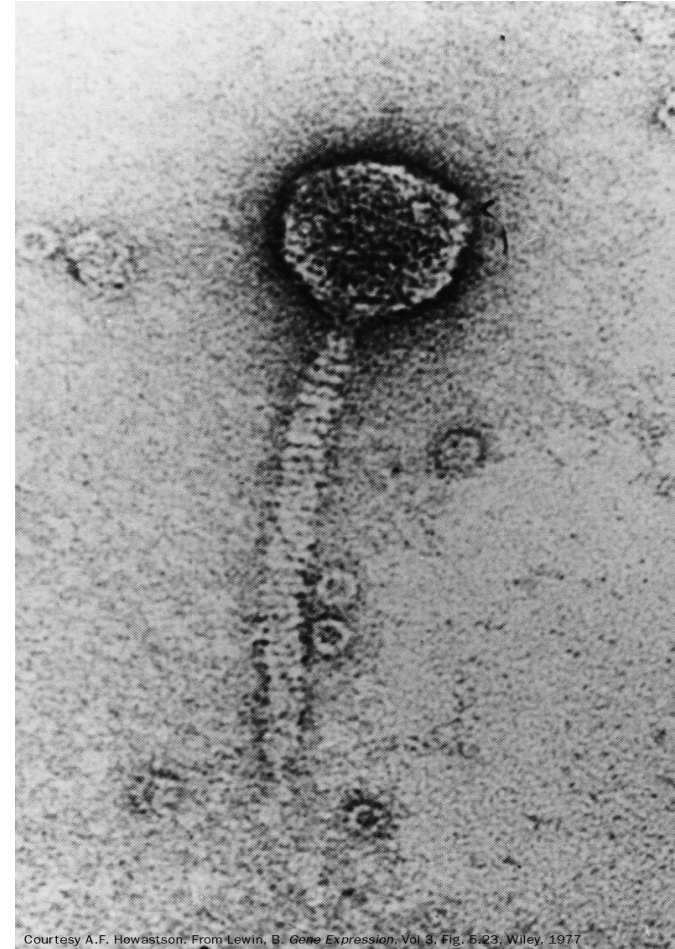
# Bacteriophages



Courtesy of Robley Williams, Stanford University, Emeritus, and Harold Fisher, University of Rhode Island

Page 107

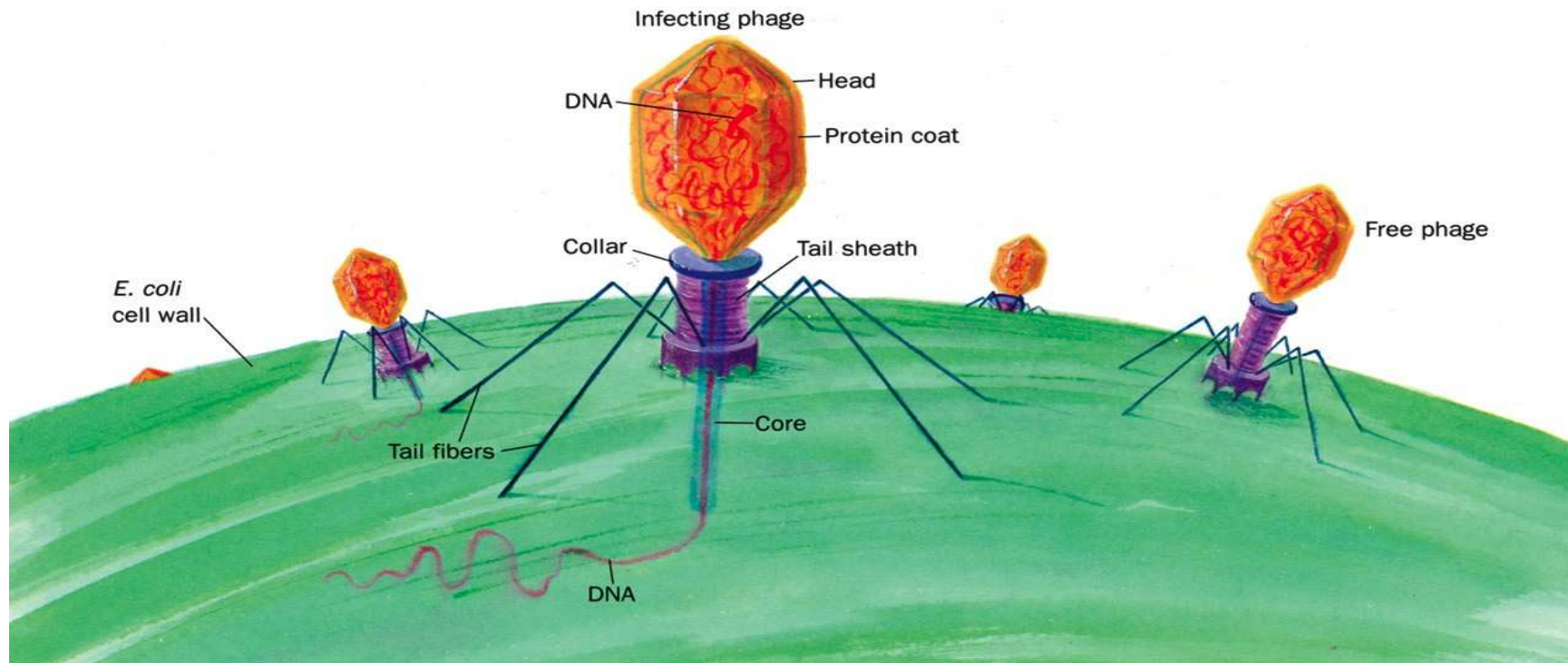
Electron micrograph of the filamentous bacteriophage M13.



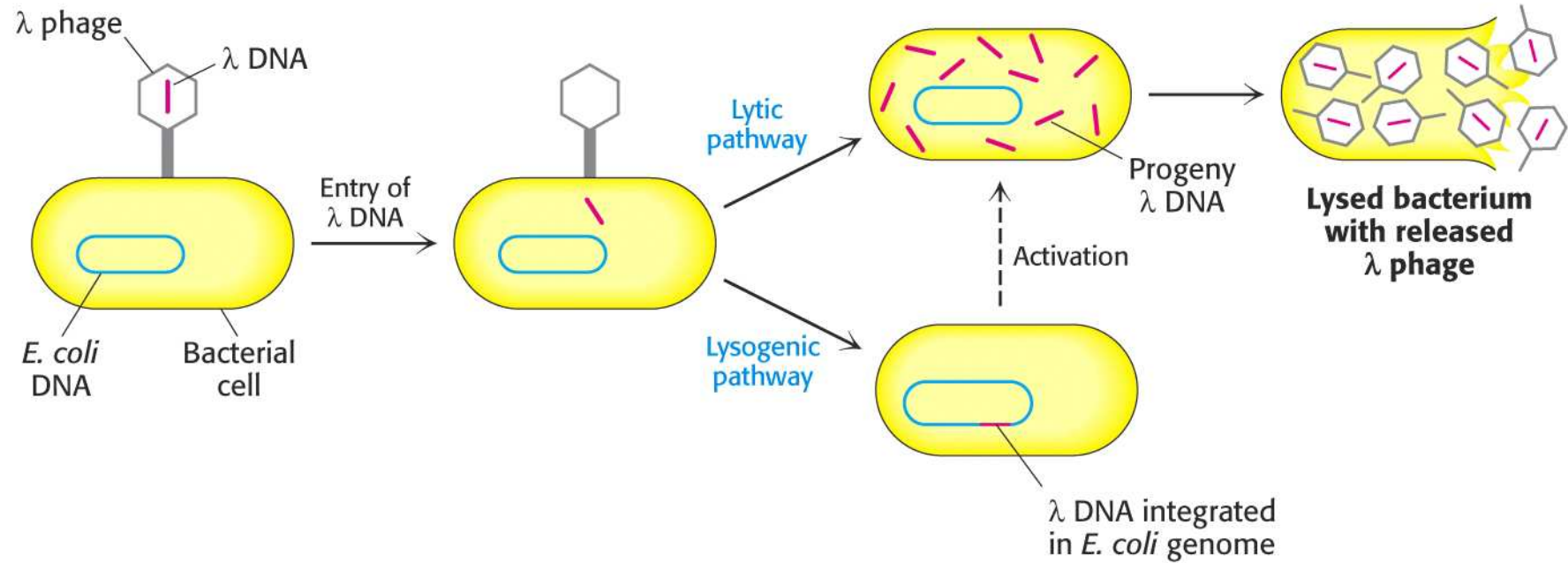
Courtesy A.F. Howatson. From Lewin, B. *Gene Expression*, Vol. 3, Fig. 5.23, Wiley, 1977

Electron micrograph of bacteriophage  $\lambda$ .

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

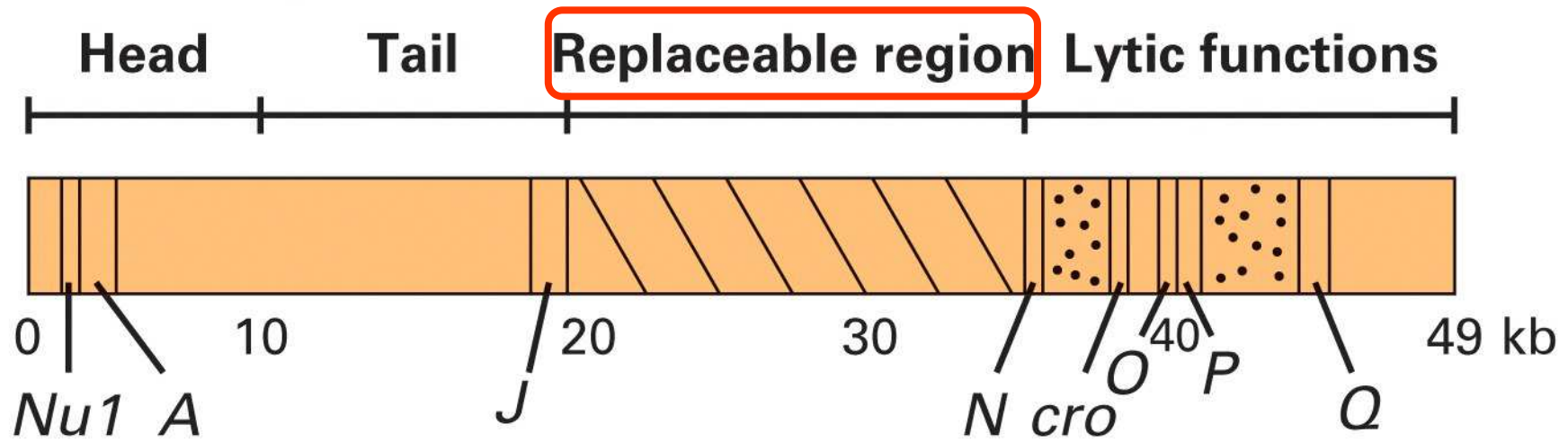


# Life Cycle of Bacteriophage $\lambda$



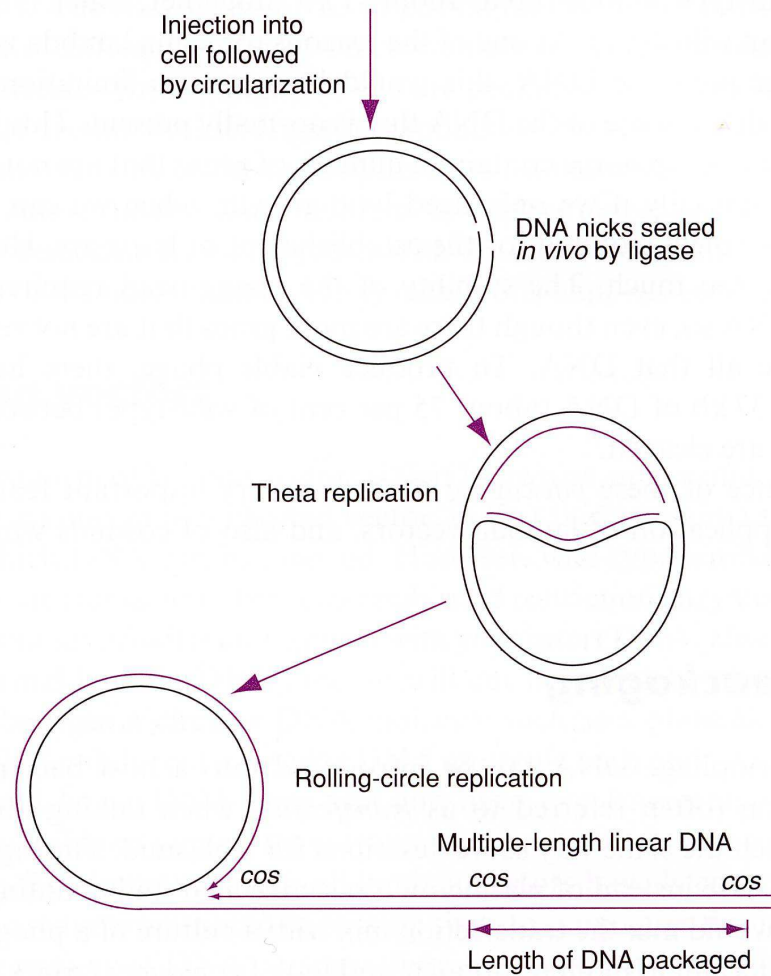
(a)  $\lambda$  Phage genome

Not needed



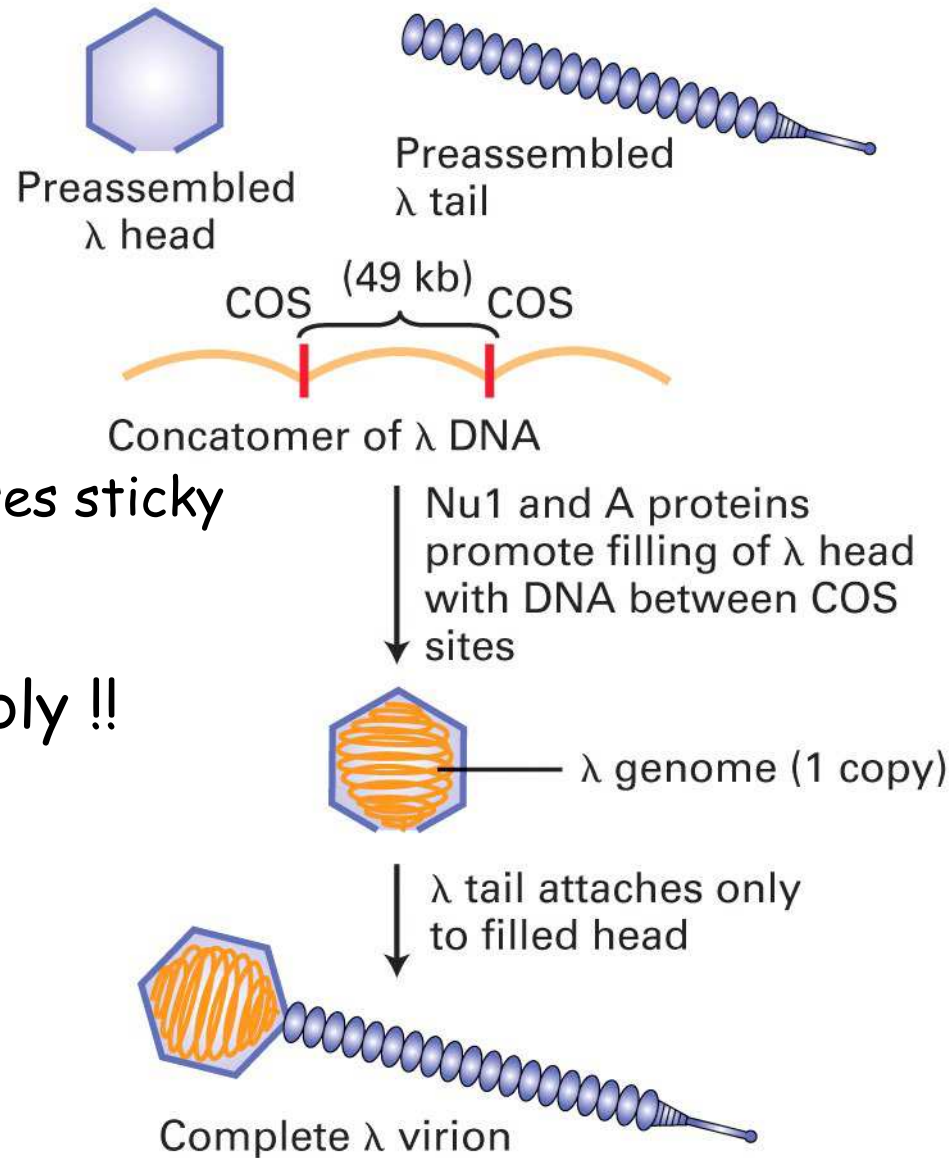
One of First Genome Sequences

# Replication of bacteriophage upon infection of a cell



**Figure 6.6** Replication of bacteriophage lambda DNA

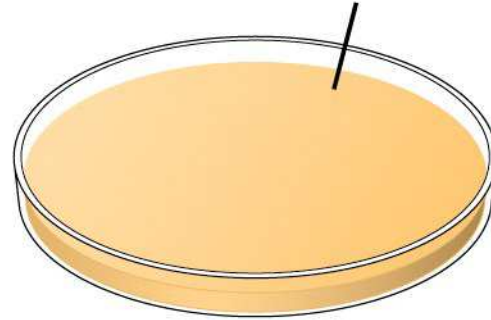
(b)  $\lambda$  Phage assembly



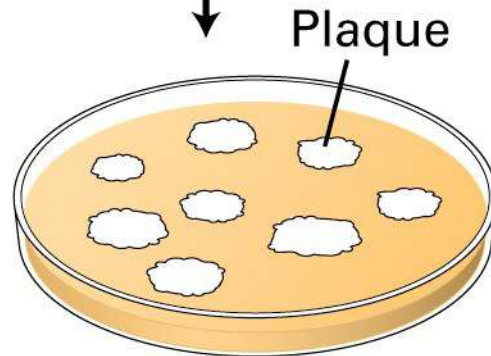
Cos site 12 bases sticky  
singlestrand

Self-assembly !!

Confluent layer of susceptible host cells  
growing on surface of a plate

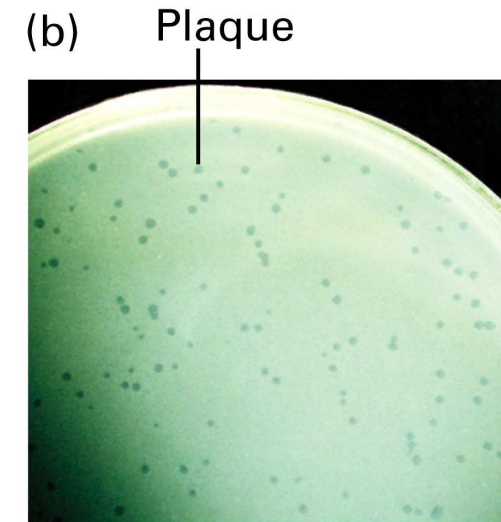
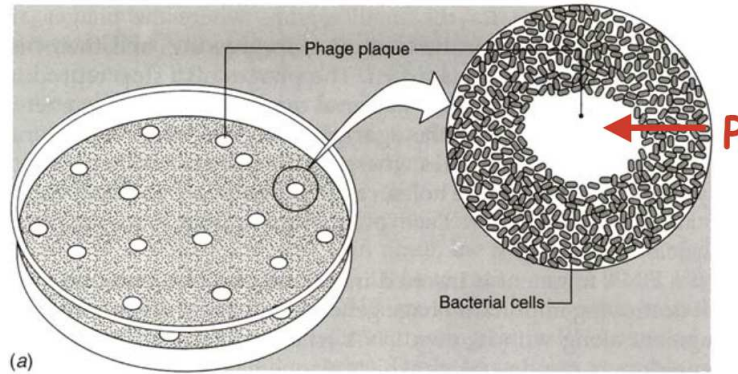


Add dilute suspension containing virus;  
after infection, cover layer of cells  
with agar; incubate



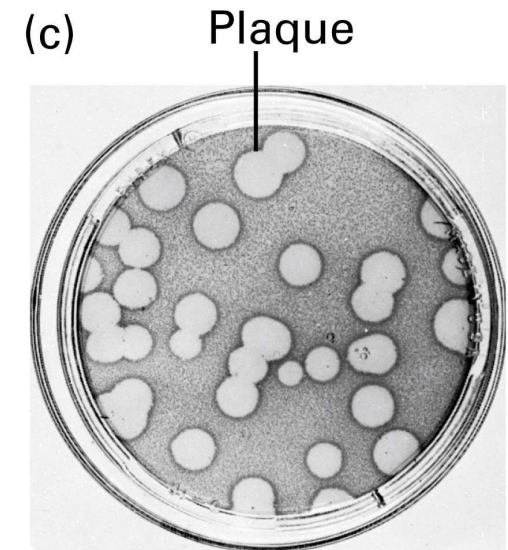
Each plaque represents cell lysis initiated by one viral  
particle (agar restricts movement so that virus can  
infect only contiguous cells)

## Molecular genetics and bacteriophage

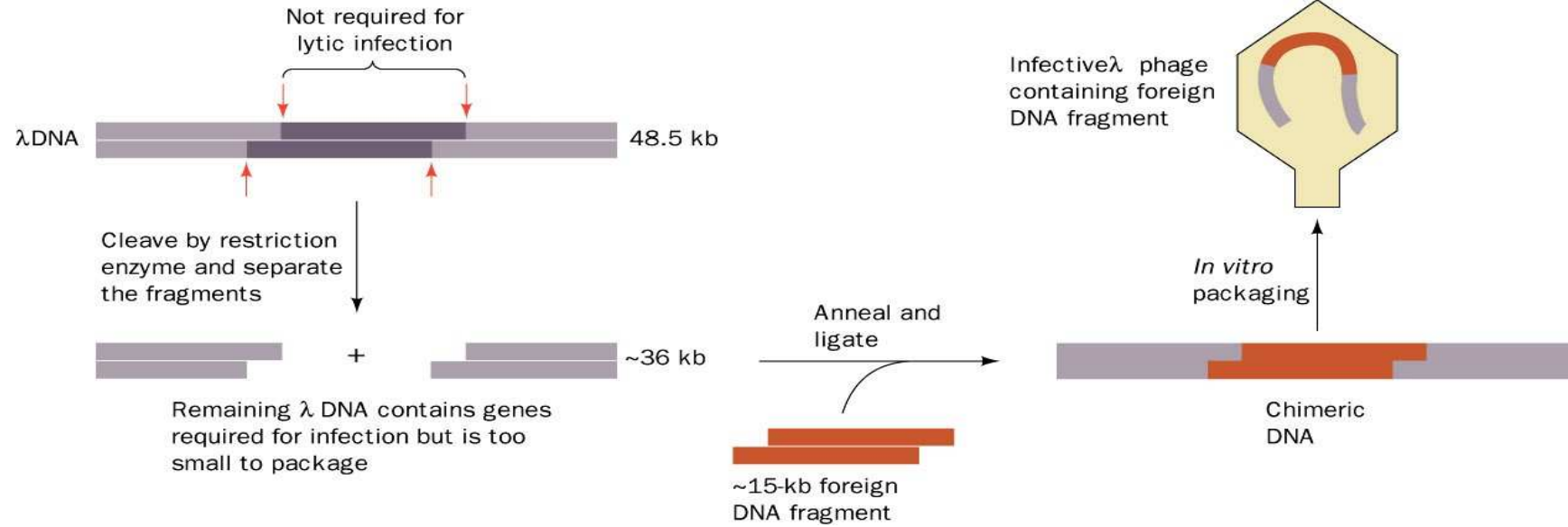


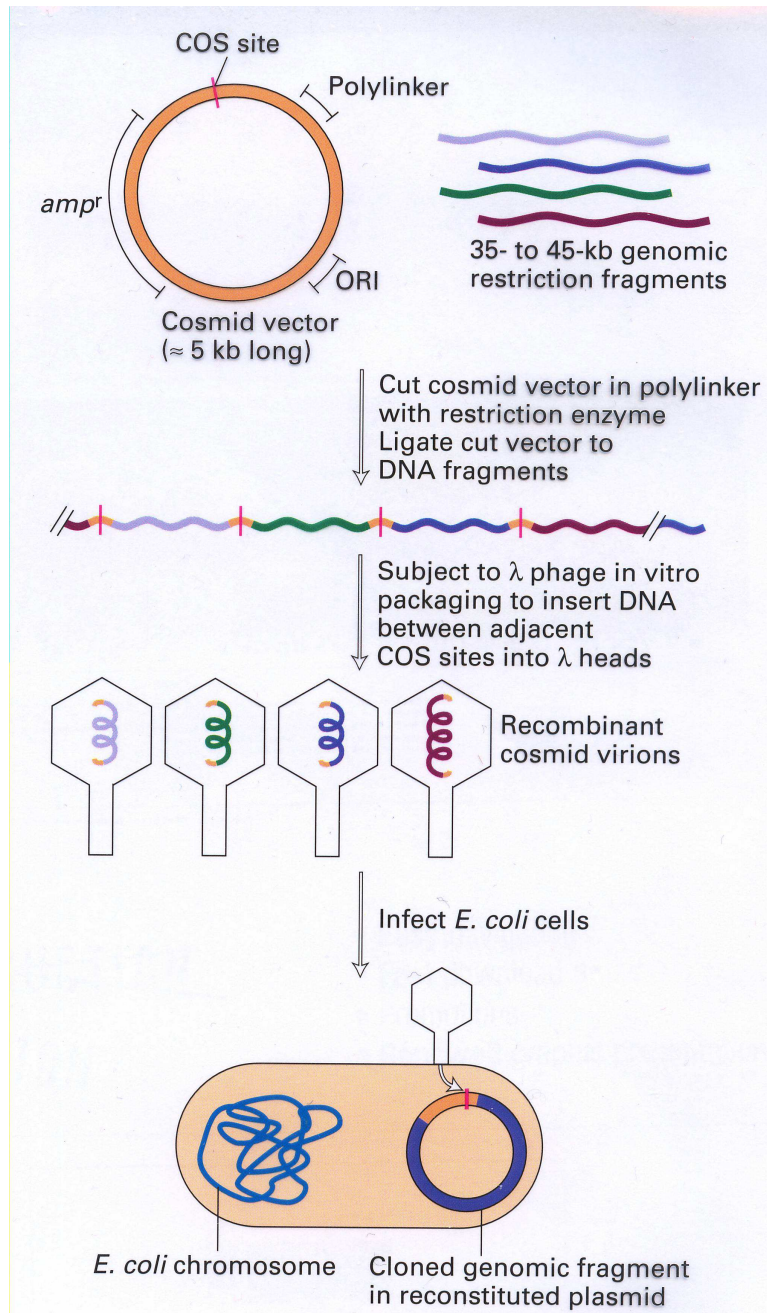
1. Each Plaque is a  
Virus Clone  
Representing One  
Viral Infection!

2. Selectable Marker is  
Bacterial Cell Destruction  
& Plaque Formation



# Cloning of foreign DNA in $\lambda$ phages.

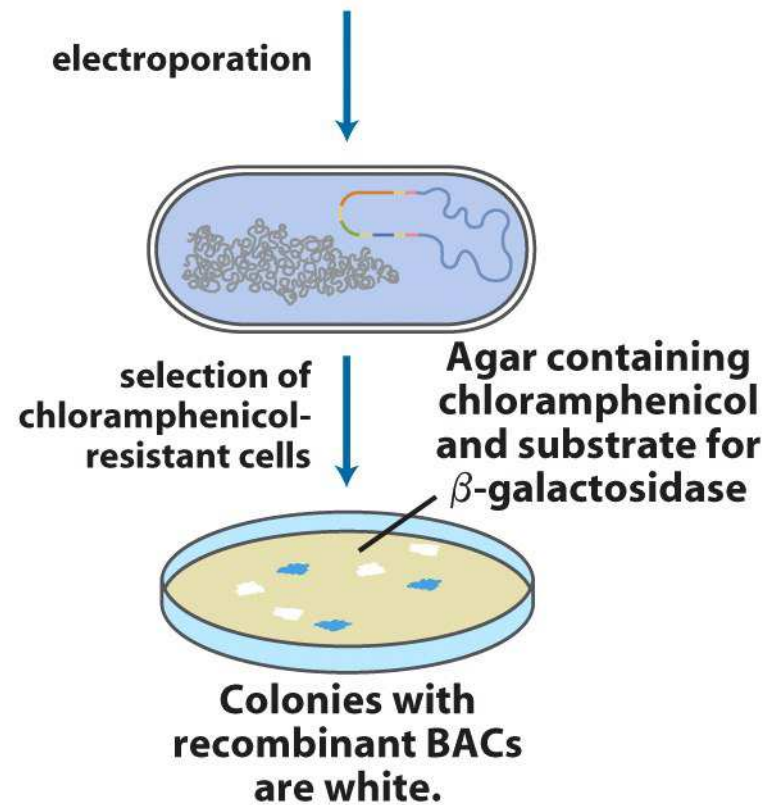
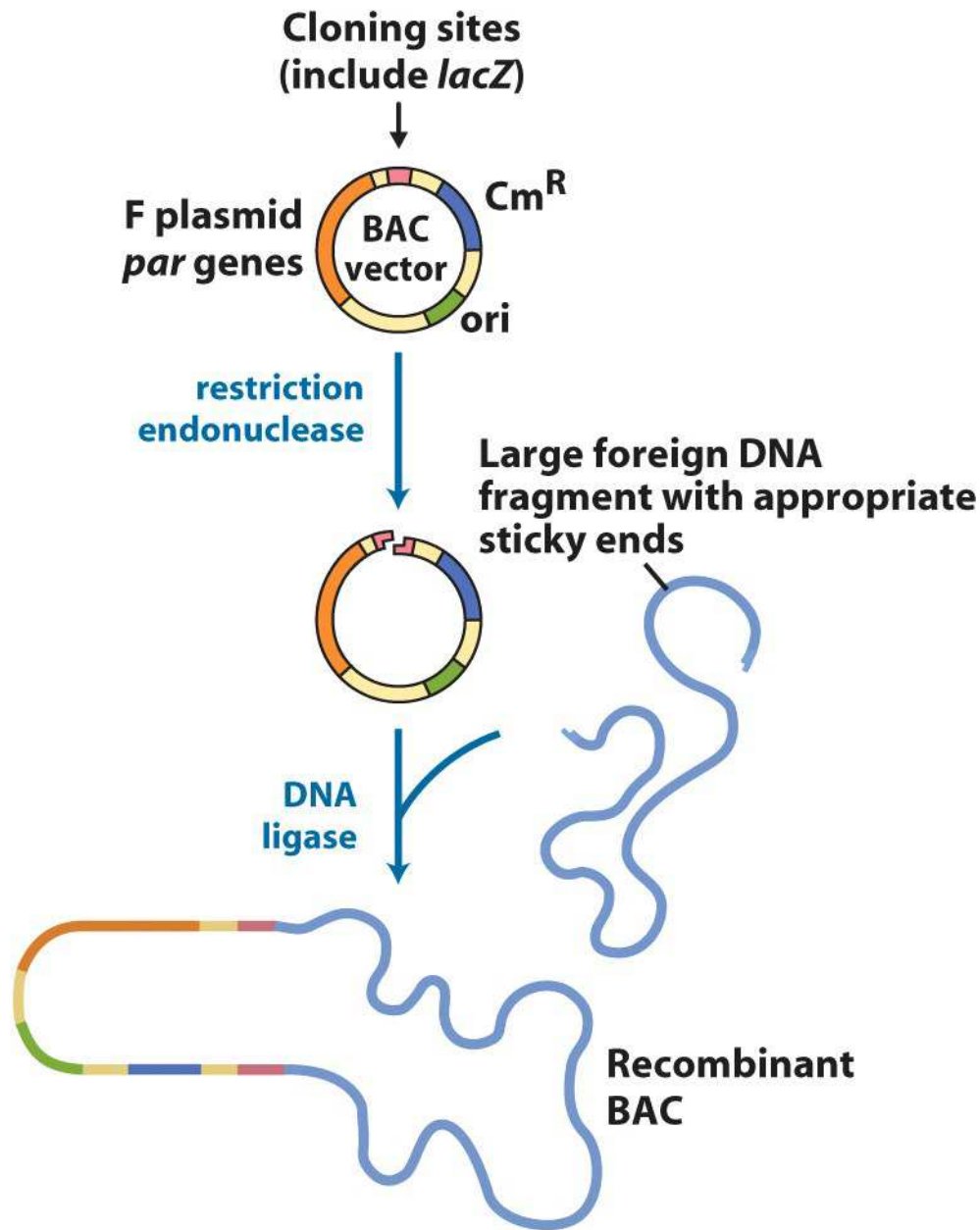




Cosmid = Cos - Plasmid

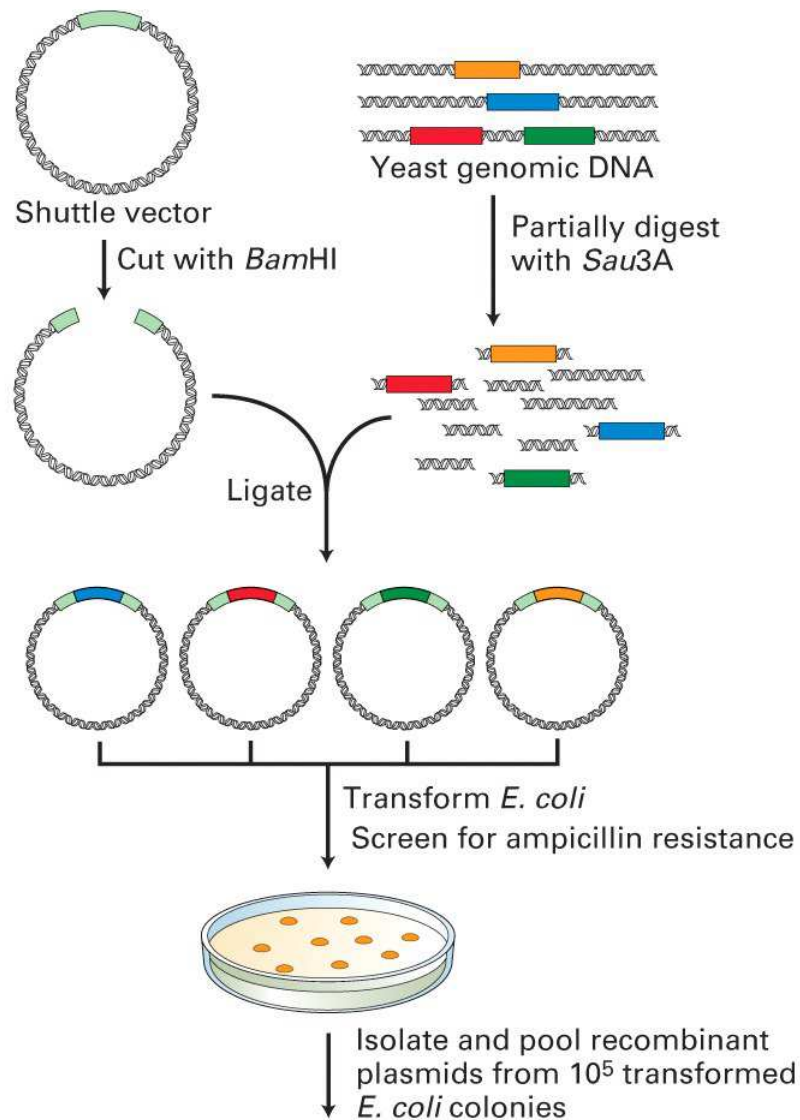
Behave as plasmids in *E. coli* after infection - ORI, antibiotic resistance

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 ?



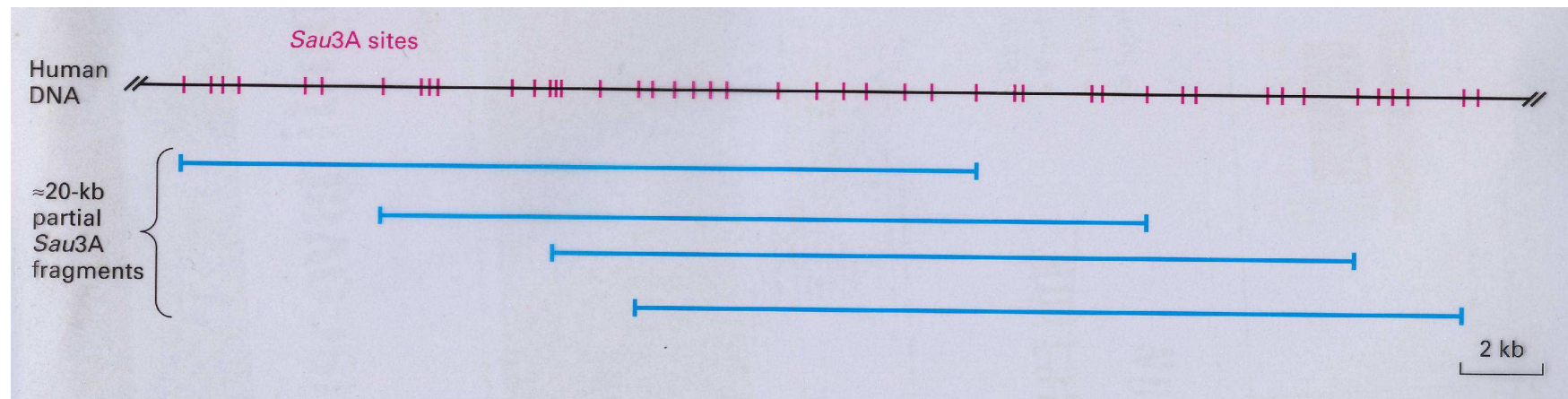
A collection of clones made from a set of randomly generated overlapping DNA fragments representing the entire genome of an organism.

Partially ~ overlapping

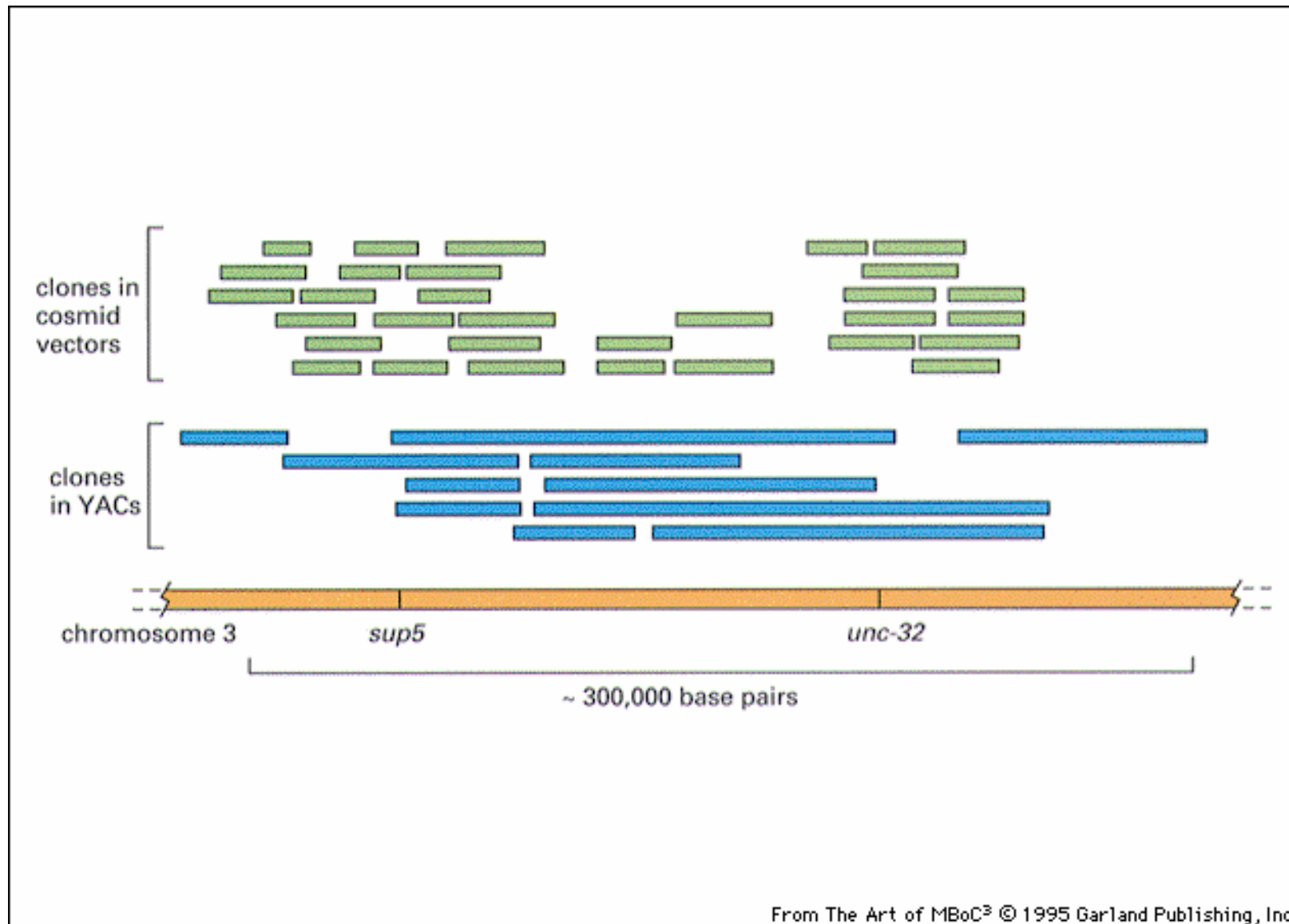
Assay yeast genomic library by functional complementation

# Fragmentation of genomic DNA

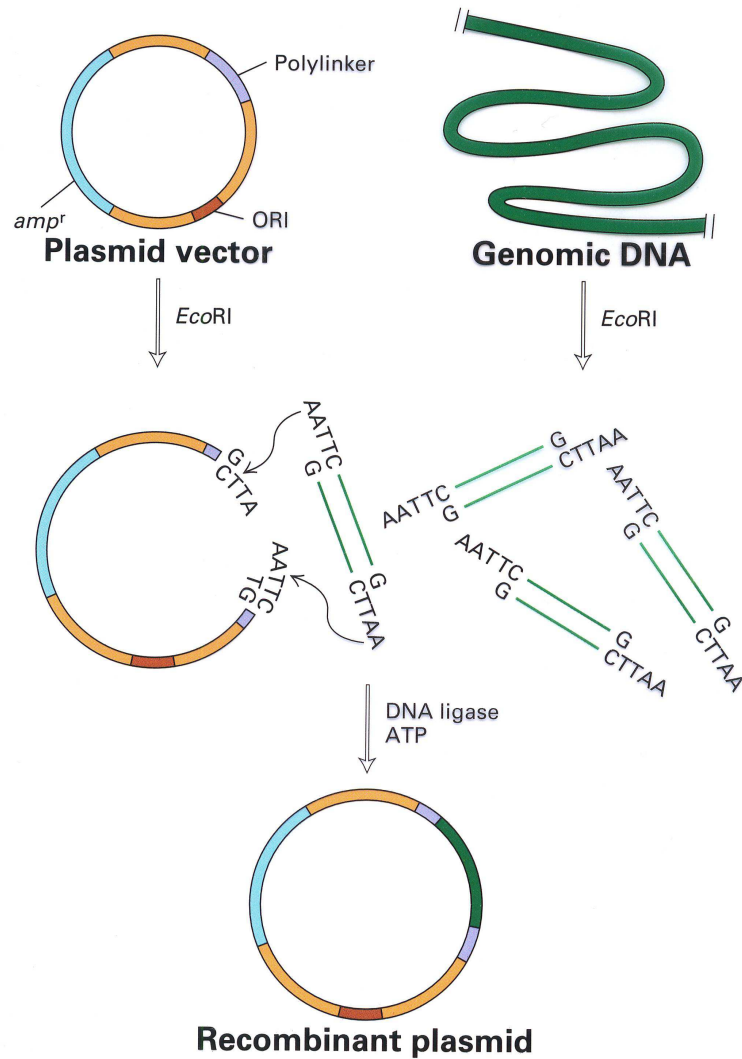
*Sau3A* GATC

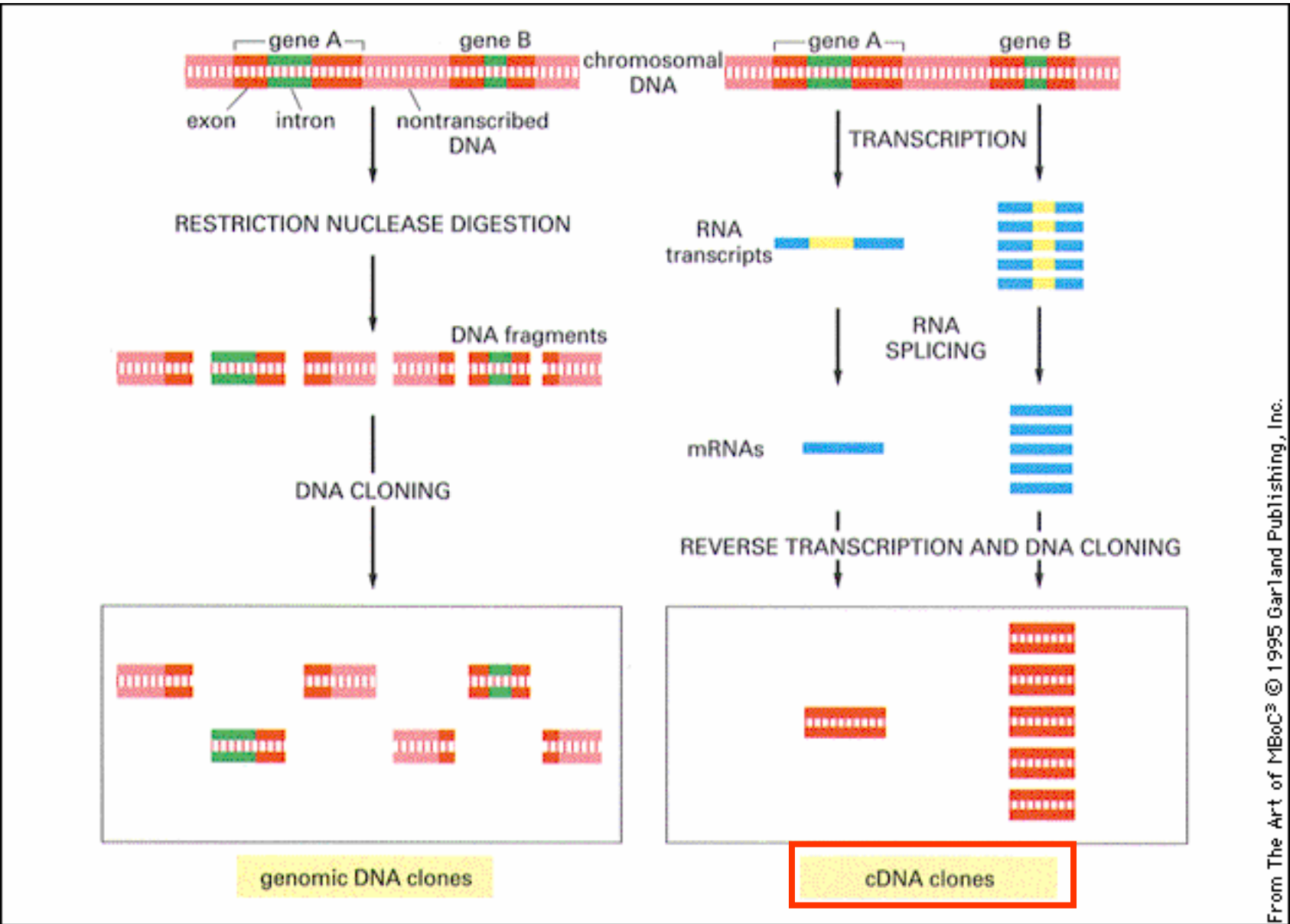


Partially ~ overlapping



# Creation of Libraries





# cDNA synthesis

**mRNA**

5' A A A A A A A A

↓ mRNA template is annealed to synthetic oligonucleotide (oligo dT) primer.

5' A A A A A A A A  
3' T T T T T T T T

↓ Reverse transcriptase and dNTPs yield a complementary DNA strand.

**mRNA-DNA hybrid**

5' A A A A A A A A  
3' T T T T T T T T

↓ mRNA is degraded with alkali.

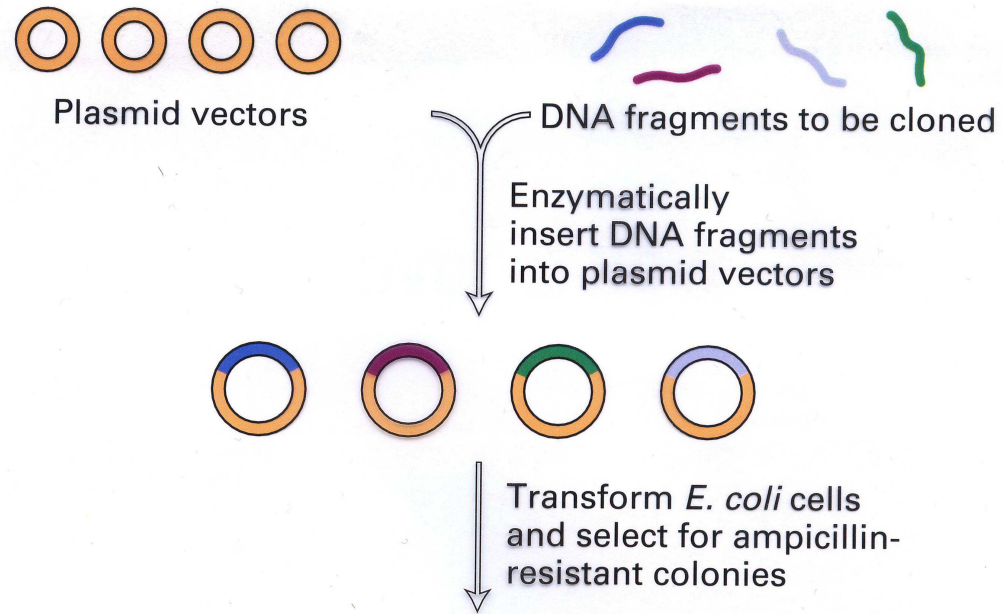
3' T T T T T T T T

↓ DNA polymerase I and dNTPs yield double-stranded DNA.

**Duplex DNA**

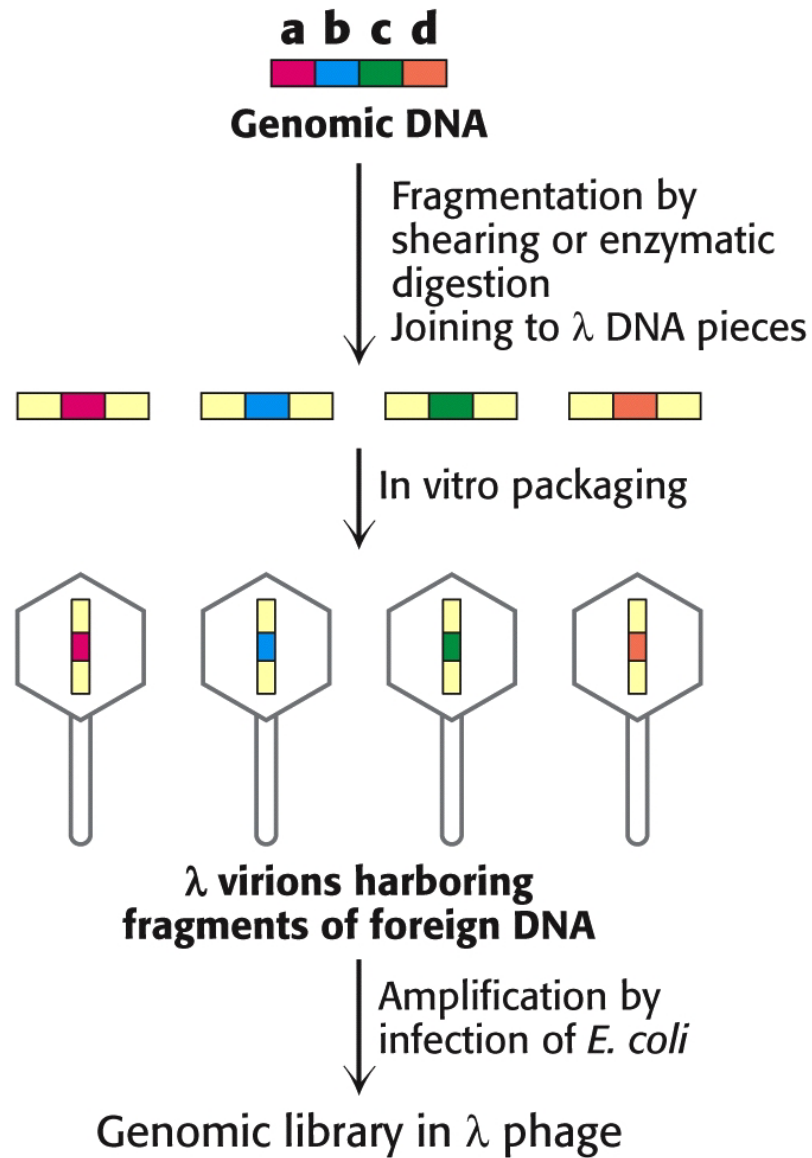
5' A A A A A A A A  
3' T T T T T T T T

# DNA Library

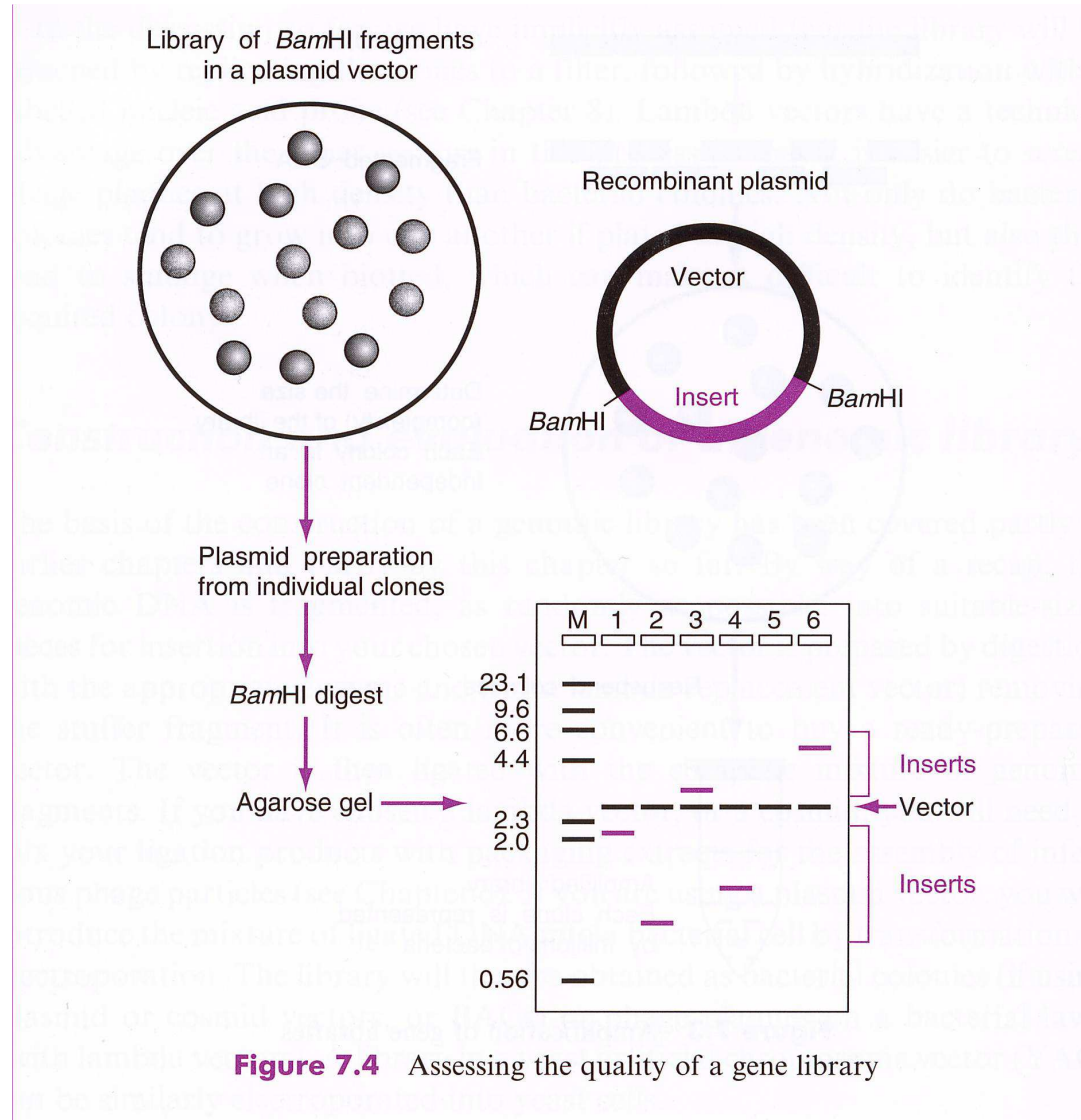


Clones →  
genetically  
identical

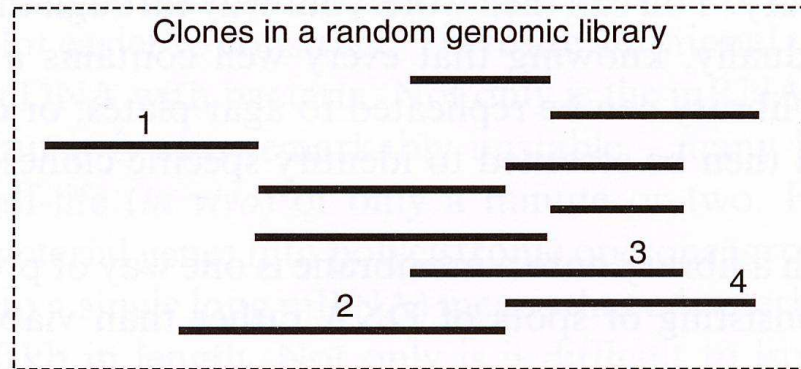
# Genomic phage library



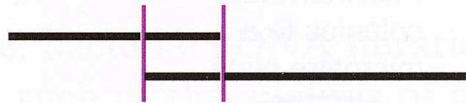
# Evaluation of library



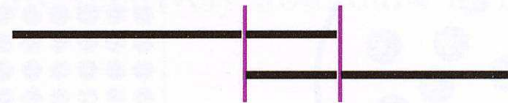
# Ordered library



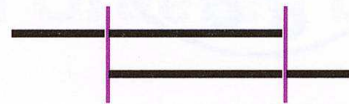
(a) Clone 1 hybridizes to clone 2



(b) Clone 2 hybridizes to clone 3



(c) Clone 3 hybridizes to clone 4



→ "Chromosome Walking"  
→ also used in "Human  
Genome Project"

Screening of the  
library - later

**Figure 7.11** Production of an ordered library

# Evaluation of library

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

Organism	Genome size	Vector type	Insert size	<i>P</i>	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) <sup>a</sup>	Contour length (μm)
<b>Viruses</b>		
Polyoma, SV40	5.2	1.7
λ Bacteriophage	48.6	17
T2, T4, T6 bacteriophage	166	55
Fowlpox	280	193
<b>Bacteria</b>		
<i>Mycoplasma hominis</i>	760	260
<i>Escherichia coli</i>	4,600	1,600
<b>Eukaryotes</b>		
Yeast (in 17 haploid chromosomes)	12,000	4,100
<i>Drosophila</i> (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

<sup>a</sup>kb = kilobase pair = 1000 base pairs (bp).

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

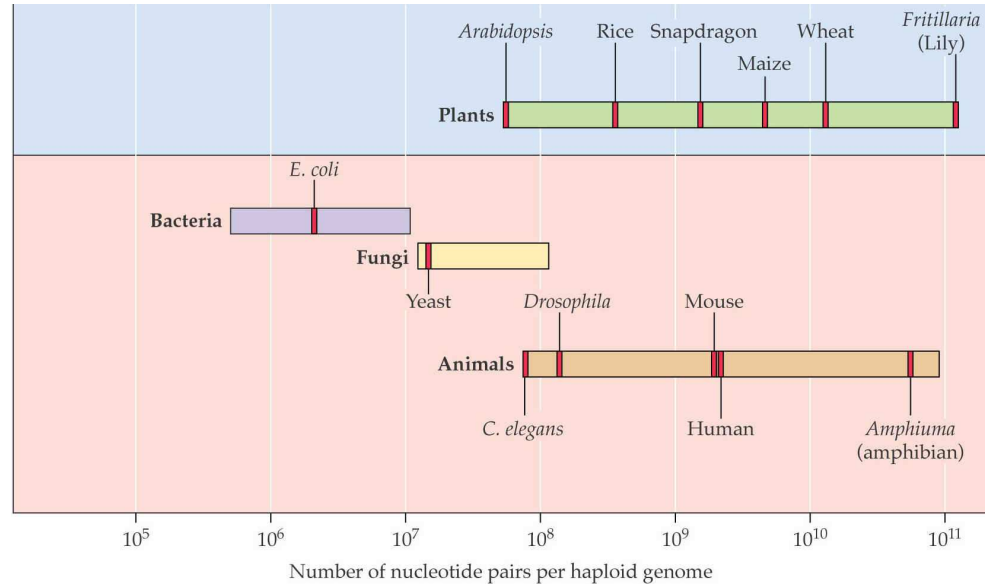
# How big genomes are?



*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***



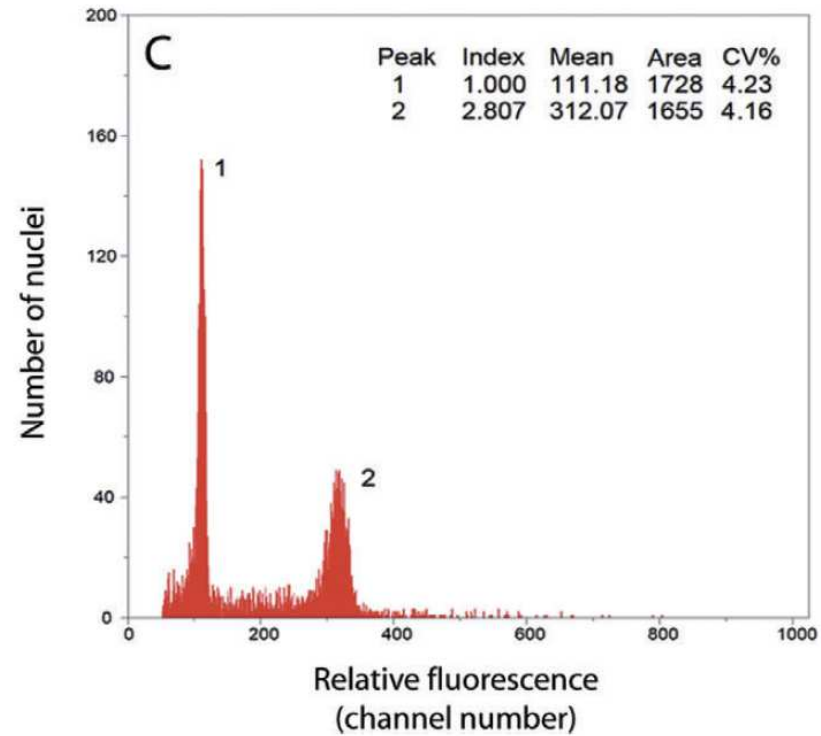
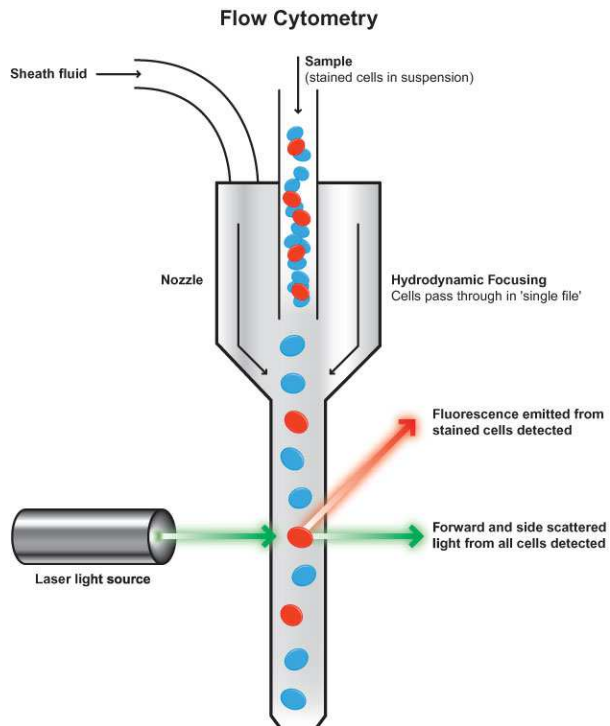
**149 000 Mbp - *Paris japonica***



**- *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

$$\text{genome size (bp)} = (0.910 \times 10^9) \times \text{DNA content (pg)}$$

$$\text{DNA content (pg)} = \text{genome size (bp)} / (0.910 \times 10^9)$$

$$1 \text{ pg} = \text{cca } 910 \text{ Mbp}; \text{ MW (1 bp)} = \text{cca } 660 \text{ Da}$$

## 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

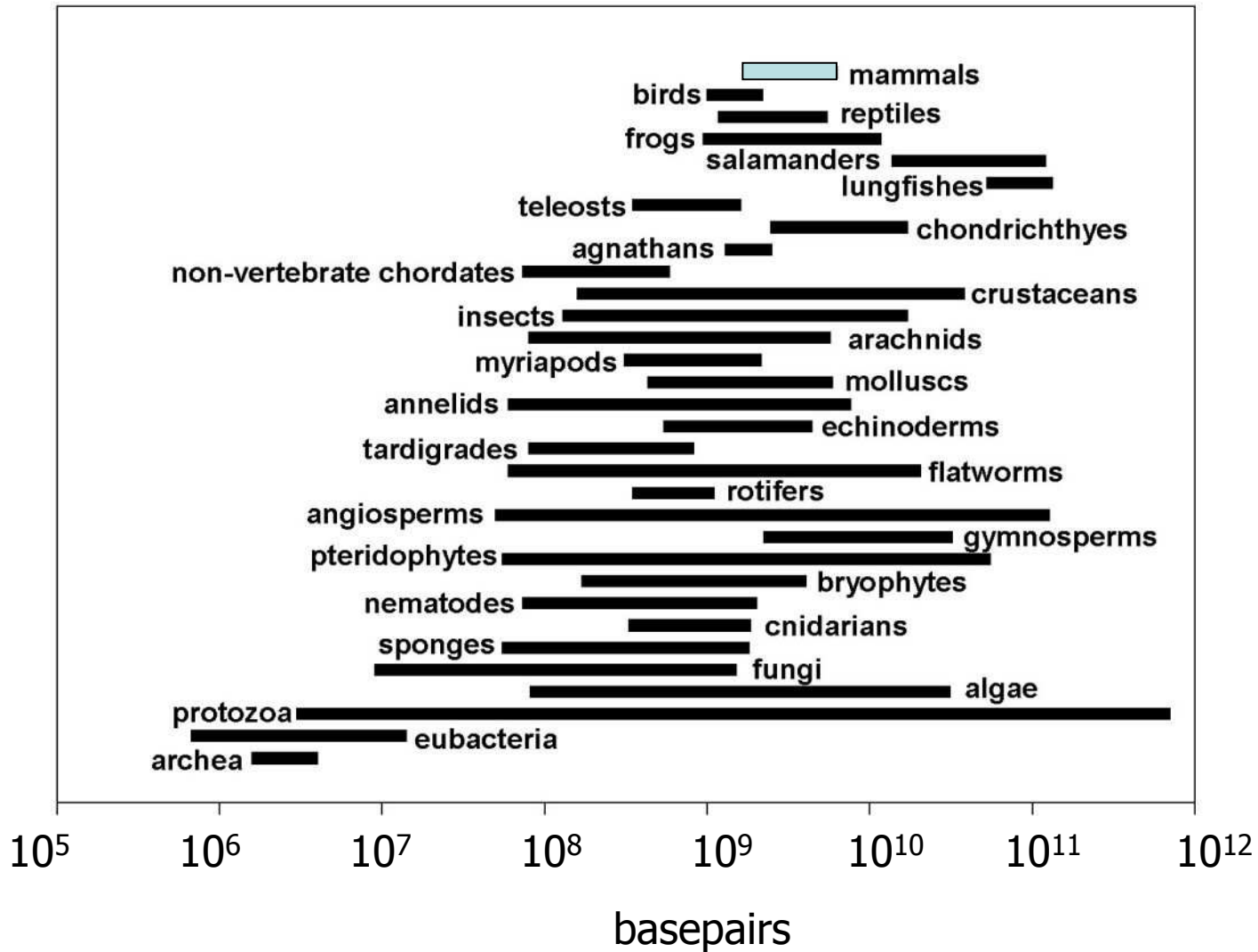


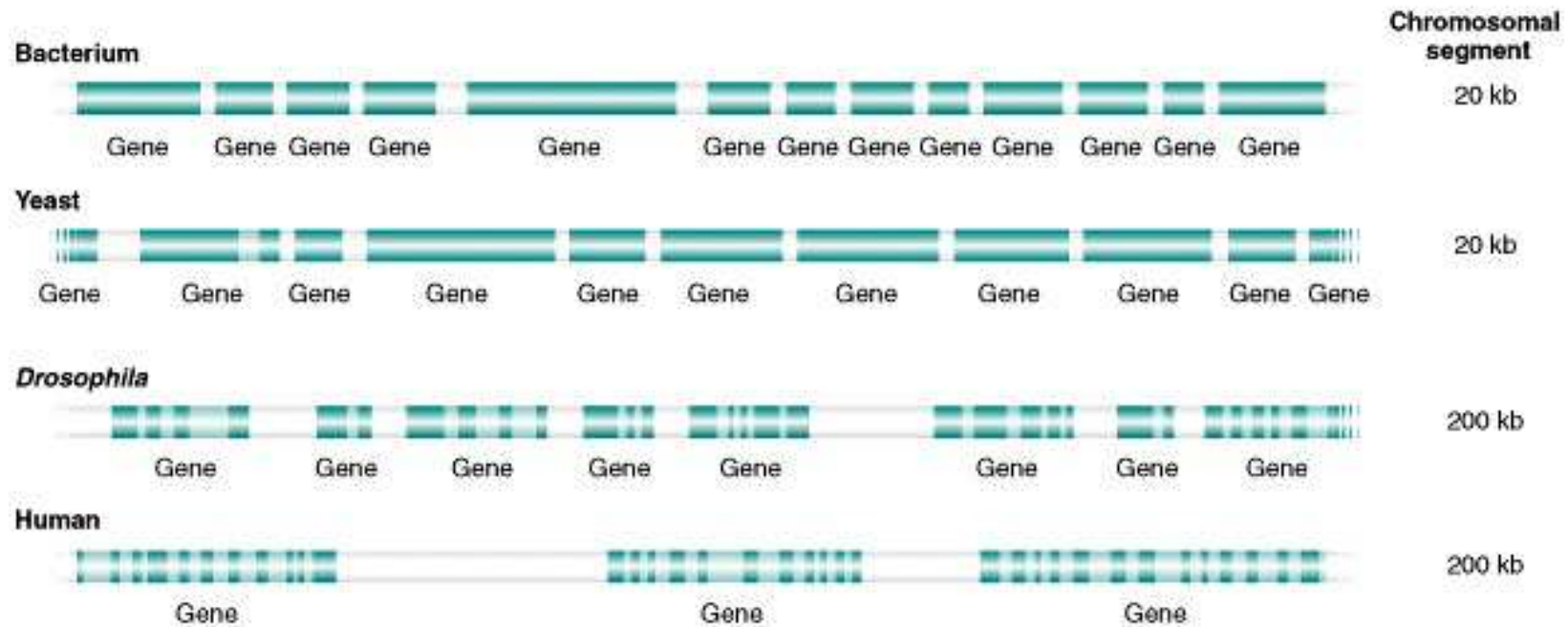
TABLE 17.1

## Representative Sequenced Genomes

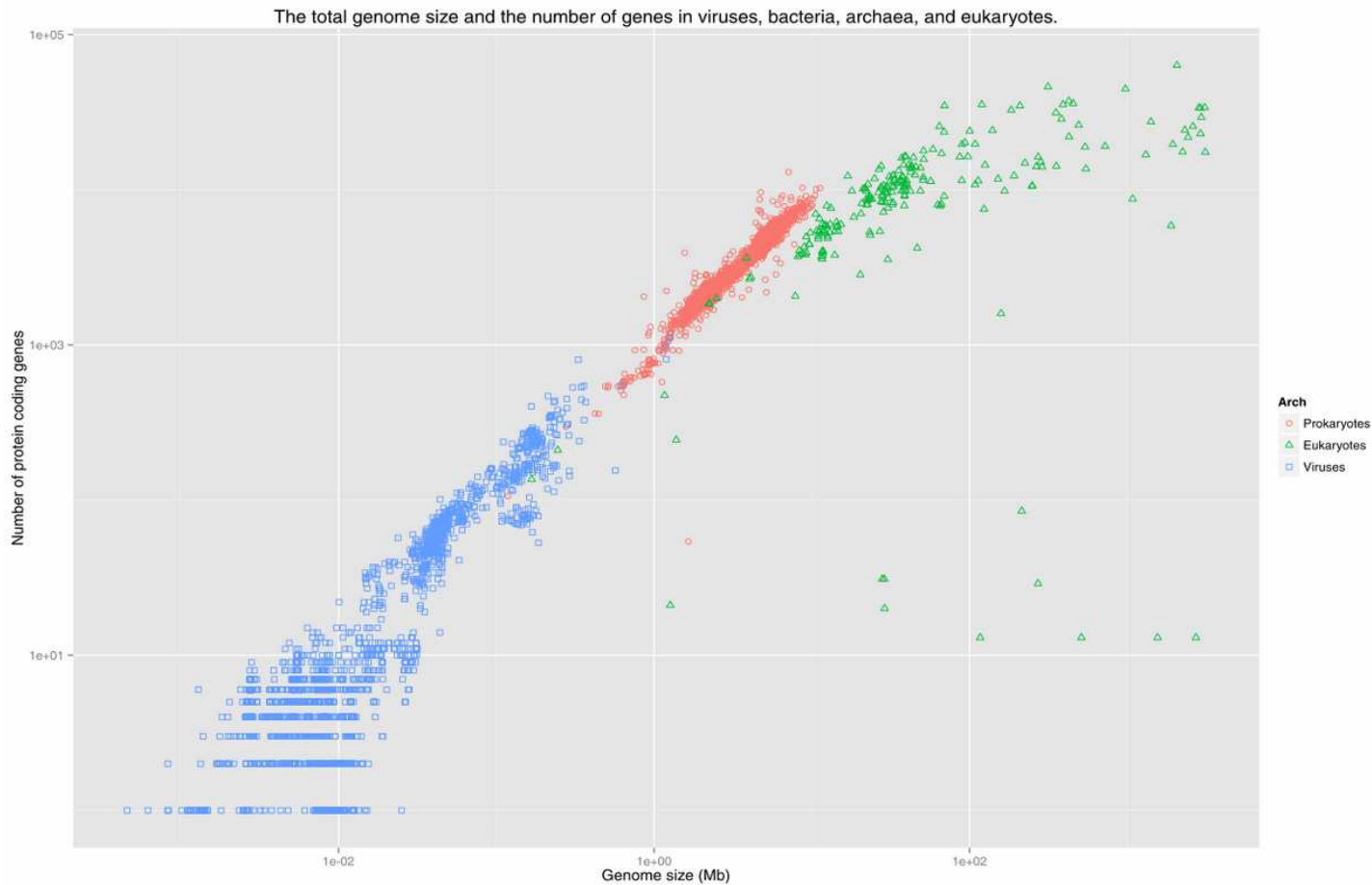
ORGANISM	HAPLOID GENOME SIZE (Mb)	NUMBER OF GENES	PROTEIN- CODING SEQUENCE
Bacteria			
<i>M. genitalium</i>	0.58	485	88%
<i>H. influenzae</i>	1.8	1,738	89%
<i>E. coli</i>	4.6	4,377	88%
Yeasts			
<i>S. cerevisiae</i>	12.5	5,770	70%
<i>S. pombe</i>	12.5	4,929	60%
Plants			
<i>A. thaliana</i>	115	28,000	25%
Rice	390	37,544	12%
Animals			
<i>C. elegans</i>	100	19,427	25%
<i>D. melanogaster</i>	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



# 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

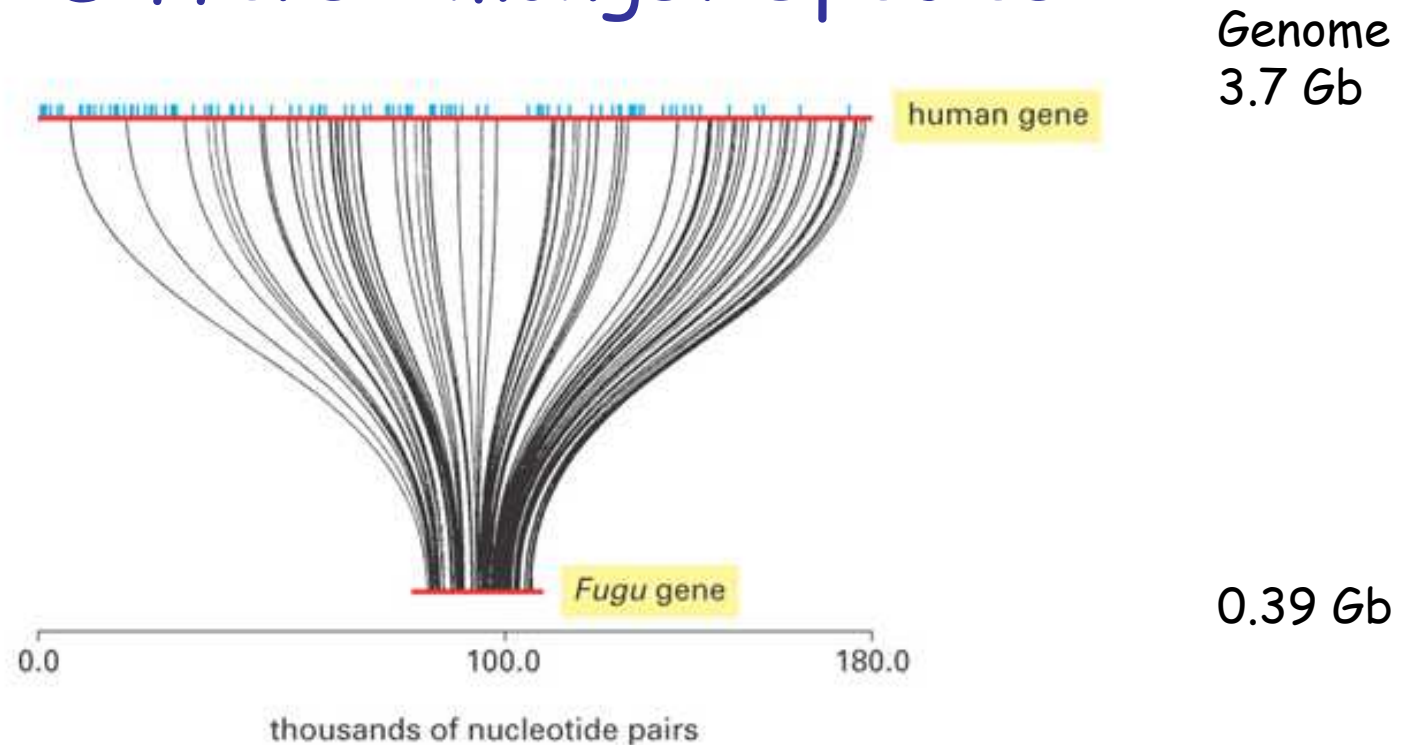


Figure 9-21 Essential Cell Biology, 2/e. (© 2004 Garland Science)

# 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  
Arizona and New Mexico: It  
cloudy in New Mexico, thunder  
in the mountains. Partly sunny  
where. Highs 90 mountains, ove  
deserts. Weather map is on Page

No. 51,432 Copyright © 2000 The New York Times

TUESDAY, JUNE 27, 2000

Printed in Arizona ONE DOLL

## Genetic Code of Human Life Is Cracked by Scientist

become part that Congress was entitled to the last  
word because Miranda's presump-  
tion that a confession was not valid.

### The Book of Life

The 3 billion  
base pairs ...

**BASE PAIRS**  
Rungs between  
the strands of  
the double helix

**BASES**  
A adenine  
C cytosine  
G guanine  
T thymine

... of the intertwining  
double helix of DNA ...

... that make up the set of  
chromosomes in our cells,  
have been sequenced.

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

### A SHARED SUCCESS

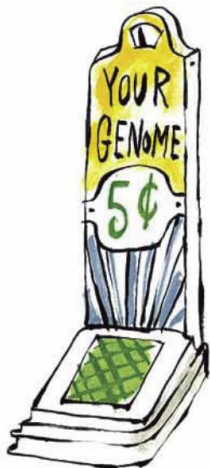
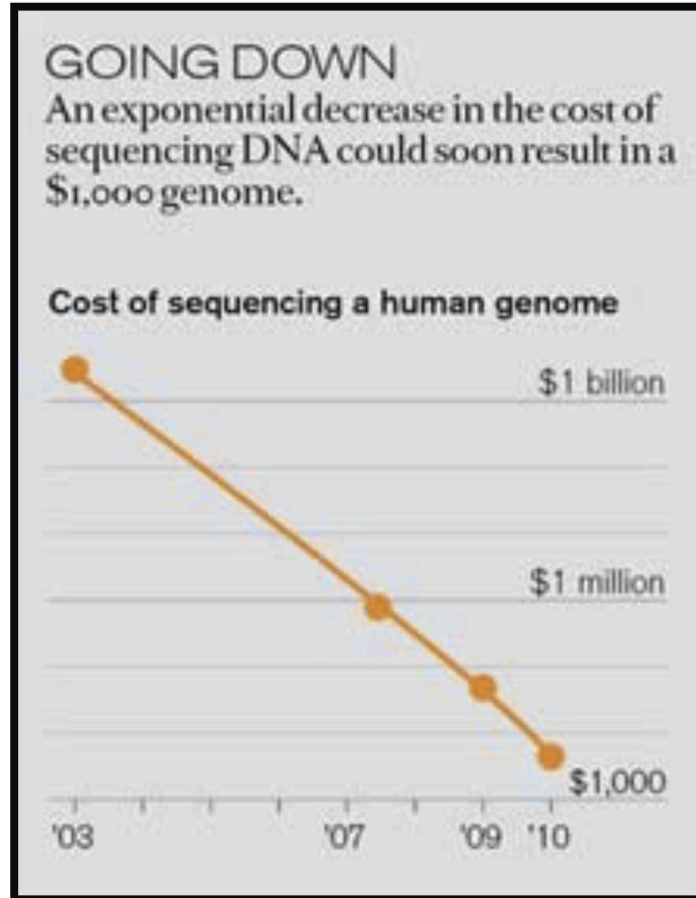
### 2 Rivals' Announcements Marks New Medical Era, Risks and All

By NICHOLAS WADE  
WASHINGTON, June 26 —  
An achievement that represents a  
mile of human self-knowledge  
rival groups of scientists said  
that they had deciphered the he  
terary script, the set of instructions  
that defines the human organism

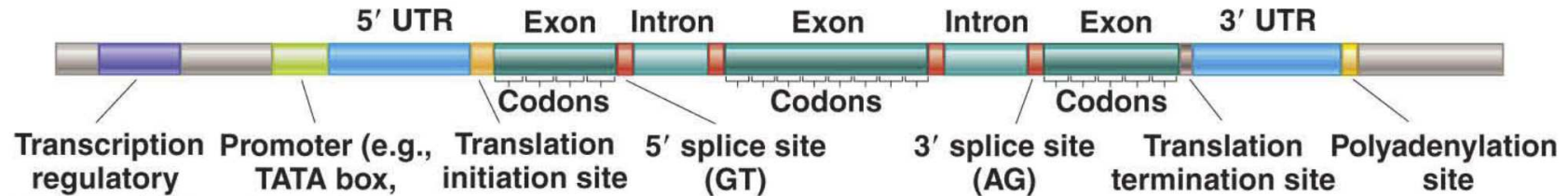
Public & Private Effort Using Different Strategies - A Race!

3 Billion Dollars & Took 15 Years

# Cost of Sequencing is Going Down Precipitously!



# Finding and Annotating Genes Within Genome Sequences



Unannotated

(a)

```

gagccacacc ctagggttgg ccaatctact cccaggagca gggaggcag gagccagggc
tgggcataaa agtcagggca gagccatcta ttgcttacct ttgctctga cacaccctgt
ttcacttaga acctcaacaa gaccacatga tgcacttacc tccctggagag aagtttgcag
ttactgccct gtgggccaag gtgaacctgg atgaagtgg ttgttaggccc ctgggcaagt
tggttacaag gttacaagac aggtttaagg agaccaatag aaactgggca tgtggagaca
gagaagactc ttgggtttct gataggcaact gactctctct gcoatttggc ctattttccc
acccttaggc tgotgttggc ctacccttgg acccagaggt tctttgagtc ctttggggat
ctgtccactc ctgatgctgt tatgggcaac cctaaagtga aggtctatgg caagaagatg
ctcgttgcct ttatgtatgg cctggtctac ctggkacacc ttaaggggcc ctttgcacaa
ctgagtgagc tgcactgtga caagctgcaac gtggatctct agaacttccg ggtgagteta
tgggaccctt gatgtttctc ttcccctctc ttctctatgt taagtctatg tcataggaag
ggggaagta acaggttaca gtttagaatg gaaacacagc gaatgattgc atcagtgtgg
aagtctcagg atcgttttag tttcttttat ttgctgtcca taacaattgt tttcttttgg
tcaattcttg ctctcttttt ttctctctcc cgaatttttc actattatcc ttaagtcctt
aacatttggc atacaacaaq gaatatcttc ttgactatct taagttaact aaaaasaaac
ttacacagt ctgctagta cattaactat tggatatac gtgtgcttat tgcataatc
ataatctccc tactttatct tcttttattt ttaattgata cataatcatt atacatattt
atgggttaaa gtttaattgt ttaatatgtg tacacatatt gaccanaoca gggtaatttt
gcattgttaa ttttaaaaa ttgcttcttc ttttaataa ctbttttgg ttctctattt
ctaatacttt cctaactctc ttctctctcc ttgactatct atacactgta tcatgctctt
tggacccttt ctaagaata acagtataa tttctgggtt aagcnaatag caatatttct
gcataaaat atttctgat ataaatgta actgatgaa gaggttctat attgctaata
gcagctacaa tccagctacc atctgtctt tattttatgg ttgggataag gctggattat
totgagtcca agotaggccc ttttgctaat catgttcaata cctcttatct tctcccaca
gtctctgggc aactgtctg totgtgtctt gggccatcac ttggcnaag aatccccc
accagtgcaag gctgcttata agaaatggt ggtggtgtg gctaaatgccc tggccacaa
gtataactaa gctgcttctc ttgctgtcca atttctata aagttctct tgttccotaa
gtccaactac taaactggg gatattatga agggcttga gcatctgat ttctgctaat
aaaaaacatt tatttctatt gcaatgatgt atttaaata tttctgata tttactaaa
    
```

A HUGE Job!  
Especially if Genes  
Have Unknown Functions

Annotated

(b)

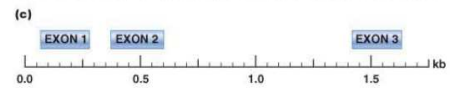
```

gagccacacc ctagggttgg ccaatctact cccaggagca gggaggcag gagccagggc
tgggcataaa agtcagggca gagccatcta ttgcttacct ttgctctga cacaccctgt
ttcacttaga acctcaacaa gaccacatga tgcacttacc tccctggagag aagtttgcag
ttactgccct gtgggccaag gtgaacctgg atgaagtgg ttgttaggccc ctgggcaagt
tggttacaag gttacaagac aggtttaagg agaccaatag aaactgggca tgtggagaca
gagaagactc ttgggtttct gataggcaact gactctctct gcoatttggc ctattttccc
acccttaggc tgotgttggc ctacccttgg acccagaggt tctttgagtc ctttggggat
ctgtccactc ctgatgctgt tatgggcaac cctaaagtga aggtctatgg caagaagatg
ctcgttgcct ttatgtatgg cctggtctac ctggkacacc ttaaggggcc ctttgcacaa
ctgagtgagc tgcactgtga caagctgcaac gtggatctct agaacttccg ggtgagteta
tgggaccctt gatgtttctc ttcccctctc ttctctatgt taagtctatg tcataggaag
ggggaagta acaggttaca gtttagaatg gaaacacagc gaatgattgc atcagtgtgg
aagtctcagg atcgttttag tttcttttat ttgctgtcca taacaattgt tttcttttgg
tcaattcttg ctctcttttt ttctctctcc cgaatttttc actattatcc ttaagtcctt
aacatttggc atacaacaaq gaatatcttc ttgactatct taagttaact aaaaasaaac
ttacacagt ctgctagta cattaactat tggatatac gtgtgcttat tgcataatc
ataatctccc tactttatct tcttttattt ttaattgata cataatcatt atacatattt
atgggttaaa gtttaattgt ttaatatgtg tacacatatt gaccanaoca gggtaatttt
gcattgttaa ttttaaaaa ttgcttcttc ttttaataa ctbttttgg ttctctattt
ctaatacttt cctaactctc ttctctctcc ttgactatct atacactgta tcatgctctt
tggacccttt ctaagaata acagtataa tttctgggtt aagcnaatag caatatttct
gcataaaat atttctgat ataaatgta actgatgaa gaggttctat attgctaata
gcagctacaa tccagctacc atctgtctt tattttatgg ttgggataag gctggattat
totgagtcca agotaggccc ttttgctaat catgttcaata cctcttatct tctcccaca
gtctctgggc aactgtctg totgtgtctt gggccatcac ttggcnaag aatccccc
accagtgcaag gctgcttata agaaatggt ggtggtgtg gctaaatgccc tggccacaa
gtataactaa gctgcttctc ttgctgtcca atttctata aagttctct tgttccotaa
gtccaactac taaactggg gatattatga agggcttga gcatctgat ttctgctaat
aaaaaacatt tatttctatt gcaatgatgt atttaaata tttctgata tttactaaa
    
```

Exon 1

Exon 2

Exon 3



PHASE TWO: INTERPRETATION

SEIDMAN *for the Ledger*

