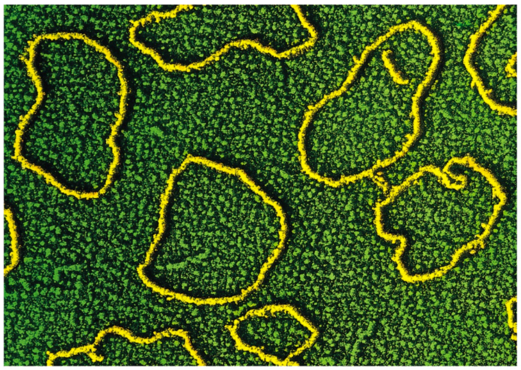
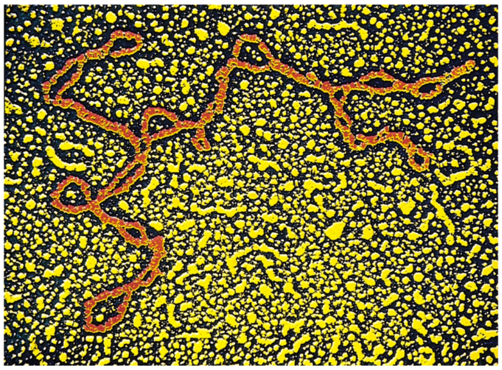
# Chromosome structure and DNA Sequence organization

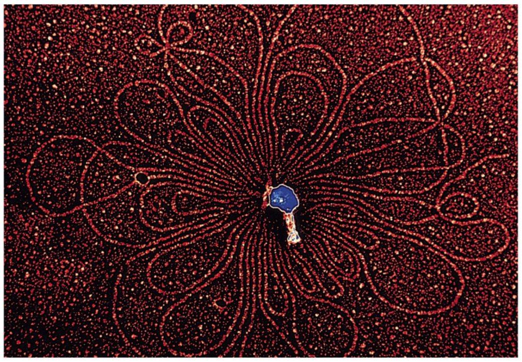
## Plasmid DNA



## Mitochondrial DNA



## T2 phage DNA



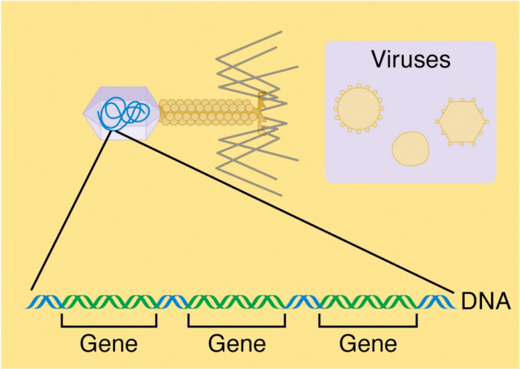
## *E. coli* DNA

## 2n=6 muntjac deer

# Viral genomes

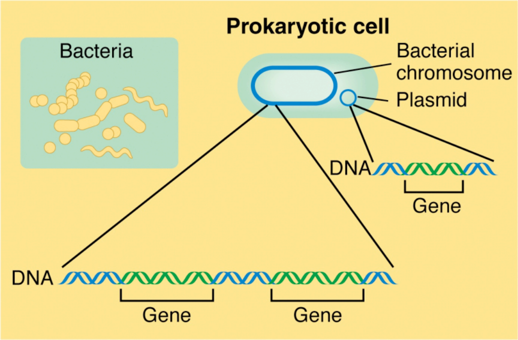
- Nonliving particle
  - nucleic acid
  - protein
- DNA or RNA
  - single-stranded or double-stranded
  - linear or circular
- Compact genomes with little spacer DNA

In prokaryotes, viruses are sometimes referred to as bacteriophages.

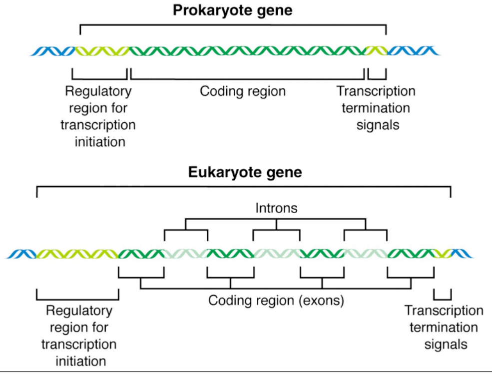


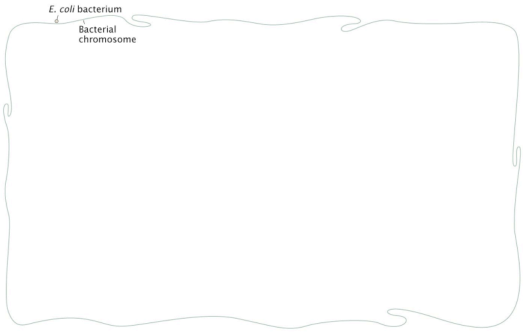
# Prokaryotic genome

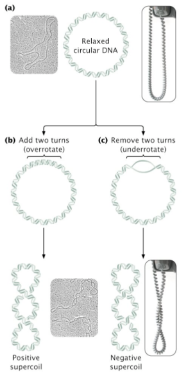
- Usually circular double helix
  - occupies nucleoid region of cell
  - attached to plasma membrane
- Genes are close together with little intergenic spacer
- Operon
  - tandem cluster of coordinately regulated genes
  - transcribed as single mRNA
- Introns very rare

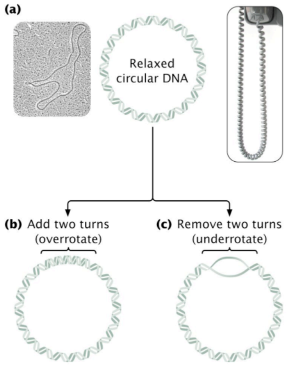


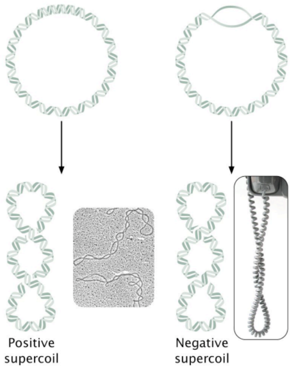
# Gene structure

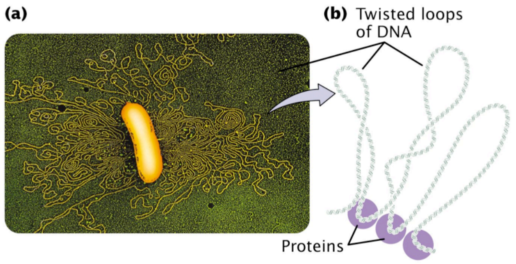












# Eukaryotic nuclear genomes

- Each species has characteristic chromosome number
- Genes are segments of nuclear chromosomes
- Ploidy refers to number of complete sets of chromosomes
  - haploid (1*n*): one complete set of genes
  - diploid (2*n*)
  - polyploid (≥3*n*)
- In diploids, chromosomes come in homologous pairs (homologs)
  - structurally similar
  - same sequence of genes
  - may contain different alleles

**In humans, somatic cells have 2*n* = 46 chromosomes.**

# Eukaryotic chromosomes (1)

- Cytogenetics: microscopic study of chromosomes
- Considerable difference in size and number of genes
- Variable centromere position
  - telocentric: centromere at end
  - acrocentric: centromere close to end
  - metacentric: centromere in middle
  - *p* arm is shortest, *q* arm is longest
- Telomere: end of chromosome
- Nucleolar organizer (rRNA)
- Chromomere

# Eukaryotic chromosomes (2)

- Heterochromatin
  - – densely stained regions of highly compact DNA
  - – mostly repetitive sequences
- Euchromatin: poorly stained, less compact, contains transcribed genes
- Banding patterns (metaphase chromosomes)
  - – differential uptake of dyes
  - – G bands, Giemsa stain (A/T rich)
  - – R bands, reverse of Giemsa (G/C rich)
- Polytene chromosomes
  - – replicated, unseparated chromosomes
  - – present in certain tissues of dipteran insects

# Nuclear DNA

- Highly organized, various degrees of coiling
- Nucleosome
  - – fundamental unit of chromatin
  - – DNA wound around histone core (octamer)
    - histones are highly conserved proteins
    - H2A, H2B, H3, H4
  - – 10 nm fiber
  - – solenoid, 30 nm fiber
- Higher order coiling
  - – solenoid loops attach to scaffold
  - – scaffold attachments contain topoisomerase II
  - – form larger diameter fibers

A haploid set of human chromosomes consists of about 1 meter of DNA.

# Gene and mRNA size

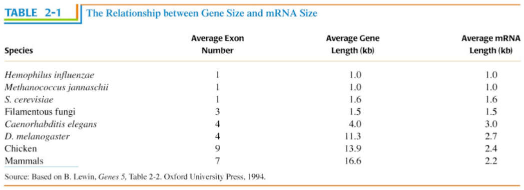
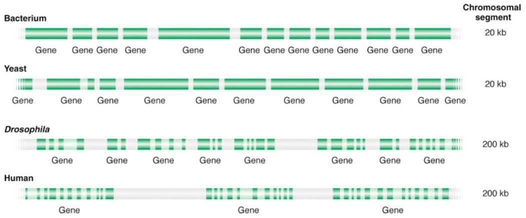


**TABLE 2-1** The Relationship between Gene Size and mRNA Size

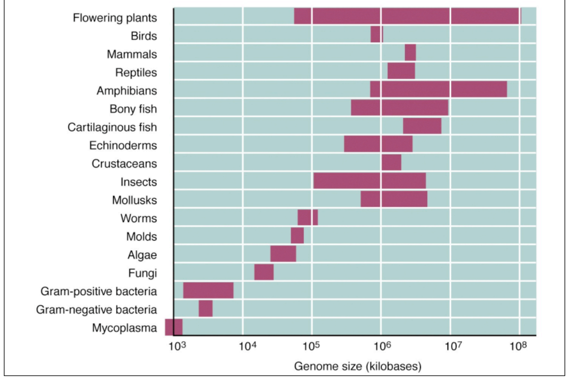
| Species | Average Exon Number | Average Gene Length (kb) | Average mRNA Length (kb) |
|---|---|---|---|
| *Hemophilus influenzae* | 1 | 1.0 | 1.0 |
| *Methanococcus jannaschii* | 1 | 1.0 | 1.0 |
| *S. cerevisiae* | 1 | 1.6 | 1.6 |
| Filamentous fungi | 3 | 1.5 | 1.5 |
| *Caenorhabditis elegans* | 4 | 4.0 | 3.0 |
| *D. melanogaster* | 4 | 11.3 | 2.7 |
| Chicken | 9 | 13.9 | 2.4 |
| Mammals | 7 | 16.6 | 2.2 |

Source: Based on B. Lewin, *Genes 5*, Table 2-2. Oxford University Press, 1994.

# Gene topography (intron, exons, regulatory and spacer regions)



# Genome sizes



# Genome size vs. number of genes

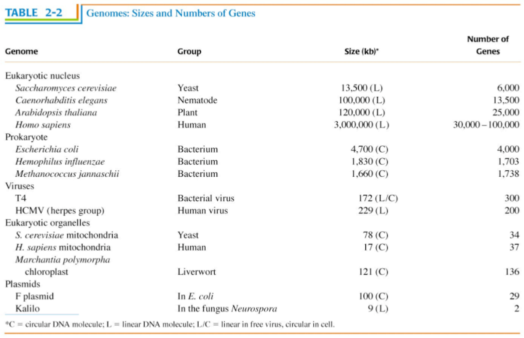


**TABLE 2-2** Genomes: Sizes and Numbers of Genes

| Genome | Group | Size (kb)* | Number of Genes |
|---|---|---|---|
| Eukaryotic nucleus | | | |
| *Saccharomyces cerevisiae* | Yeast | 13,500 (L) | 6,000 |
| *Caenorhabditis elegans* | Nematode | 100,000 (L) | 13,500 |
| *Arabidopsis thaliana* | Plant | 120,000 (L) | 25,000 |
| *Homo sapiens* | Human | 3,000,000 (L) | 30,000 – 100,000 |
| Prokaryote | | | |
| *Escherichia coli* | Bacterium | 4,700 (C) | 4,000 |
| *Hemophilus influenzae* | Bacterium | 1,830 (C) | 1,703 |
| *Methanococcus jannaschii* | Bacterium | 1,660 (C) | 1,738 |
| Viruses | | | |
| T4 | Bacterial virus | 172 (L/C) | 300 |
| HCMV (herpes group) | Human virus | 229 (L) | 200 |
| Eukaryotic organelles | | | |
| *S. cerevisiae* mitochondria | Yeast | 78 (C) | 34 |
| *H. sapiens* mitochondria | Human | 17 (C) | 37 |
| *Marchantia polymorpha* chloroplast | Liverwort | 121 (C) | 136 |
| Plasmids | | | |
| F plasmid | In *E. coli* | 100 (C) | 29 |
| Kalilo | In the fungus *Neurospora* | 9 (L) | 2 |

*C = circular DNA molecule; L = linear DNA molecule; L/C = linear in free virus, circular in cell.

# Numbers of chromosomes

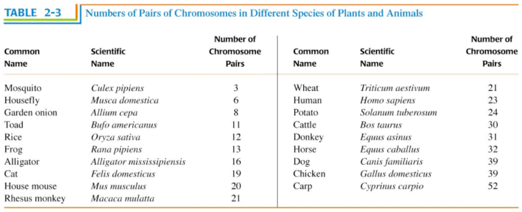
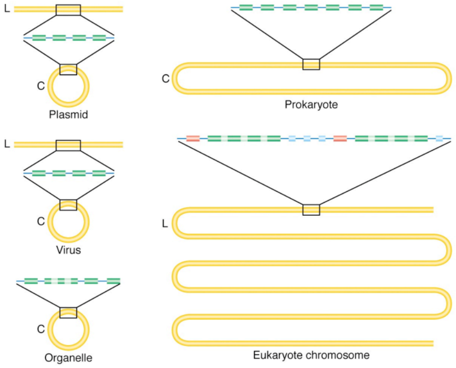


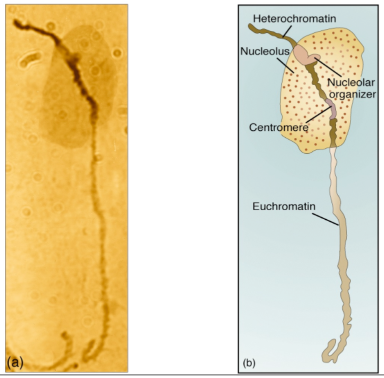
**TABLE 2-3** Numbers of Pairs of Chromosomes in Different Species of Plants and Animals

| Common Name | Scientific Name | Number of Chromosome Pairs | Common Name | Scientific Name | Number of Chromosome Pairs |
|---|---|---|---|---|---|
| Mosquito | *Culex pipiens* | 3 | Wheat | *Triticum aestivum* | 21 |
| Housefly | *Musca domestica* | 6 | Human | *Homo sapiens* | 23 |
| Garden onion | *Allium cepa* | 8 | Potato | *Solanum tuberosum* | 24 |
| Toad | *Bufo americanus* | 11 | Cattle | *Bos taurus* | 30 |
| Rice | *Oryza sativa* | 12 | Donkey | *Equus asinus* | 31 |
| Frog | *Rana pipiens* | 13 | Horse | *Equus caballus* | 32 |
| Alligator | *Alligator mississipiensis* | 16 | Dog | *Canis familiaris* | 39 |
| Cat | *Felis domesticus* | 19 | Chicken | *Gallus domesticus* | 39 |
| House mouse | *Mus musculus* | 20 | Carp | *Cyprinus carpio* | 52 |
| Rhesus monkey | *Macaca mulatta* | 21 | | | |

# Form and sizes of genomes



Chromosome 2 of tomato (nucleolus & nuclear organizer)



# Centromeres

- Repetitive DNA sequences where replicated chromosomes stay attached until mitosis or meiosis
- Kinetochores assemble on centromeres
- Microtubules assemble on kinetochores

  to help separate the chromosomes during mitosis and meiosis

# Centromeres

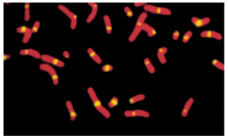


Figure 7.36b

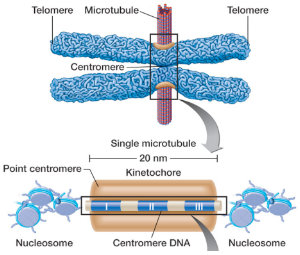
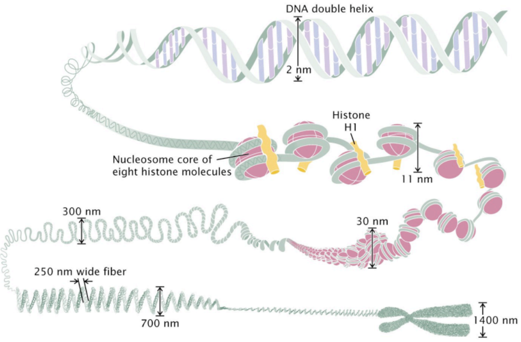
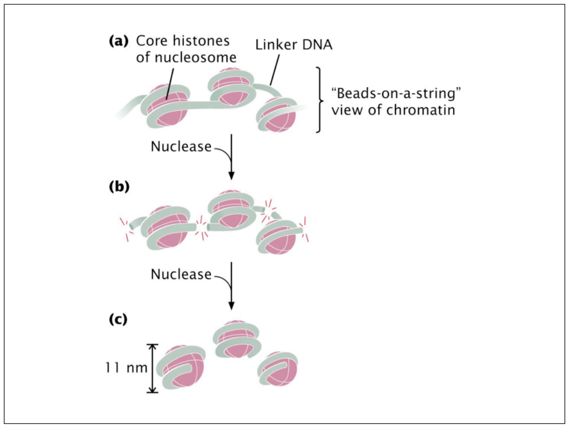
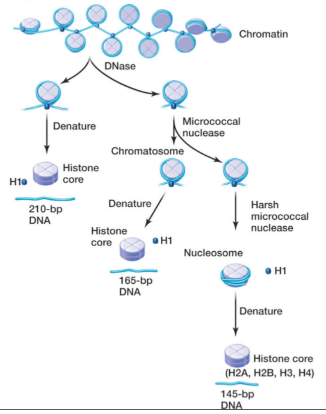


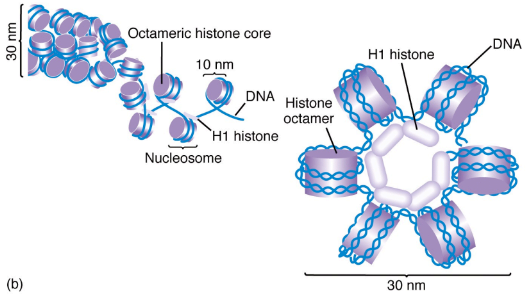
Figure 7.38







# Winding the DNA



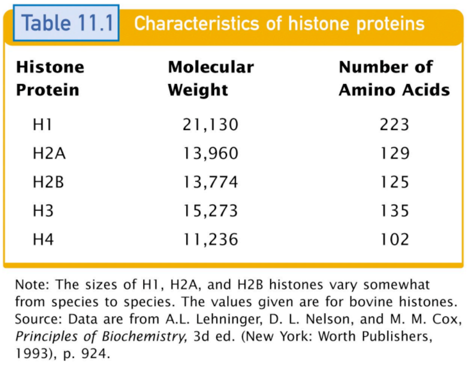
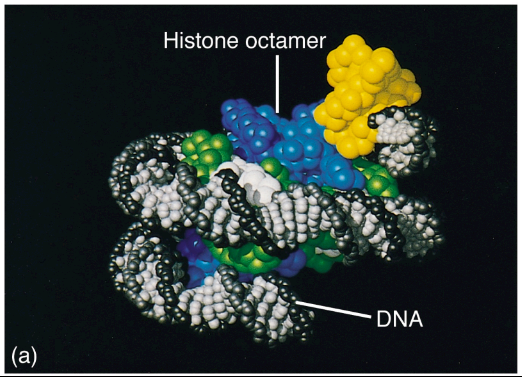


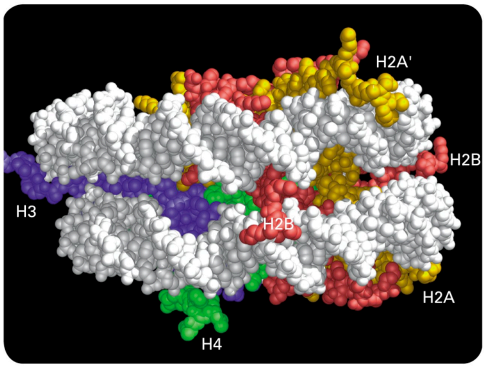
**Table 11.1** Characteristics of histone proteins

| Histone Protein | Molecular Weight | Number of Amino Acids |
|---|---|---|
| H1 | 21,130 | 223 |
| H2A | 13,960 | 129 |
| H2B | 13,774 | 125 |
| H3 | 15,273 | 135 |
| H4 | 11,236 | 102 |

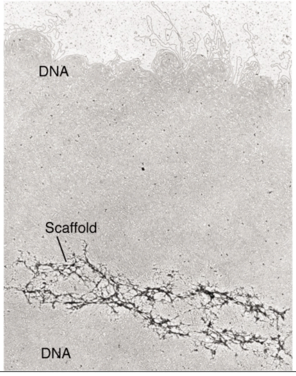
Note: The sizes of H1, H2A, and H2B histones vary somewhat from species to species. The values given are for bovine histones. Source: Data are from A.L. Lehninger, D. L. Nelson, and M. M. Cox, *Principles of Biochemistry,* 3d ed. (New York: Worth Publishers, 1993), p. 924.

# Histone octomer & DNA

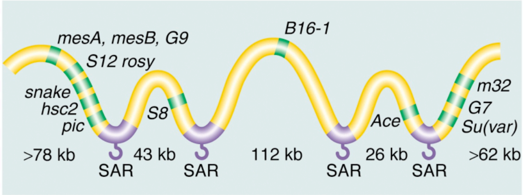




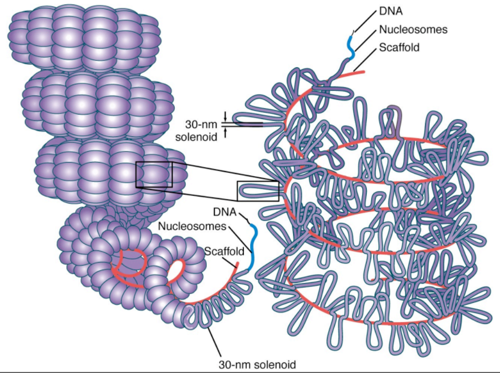
# DNA scaffolding

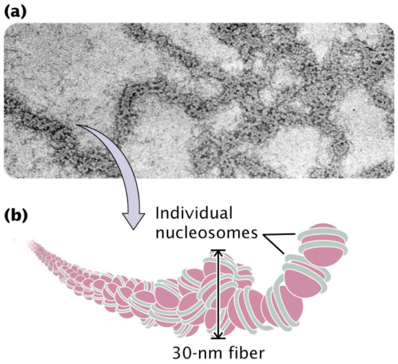


# Scaffold attachment regions

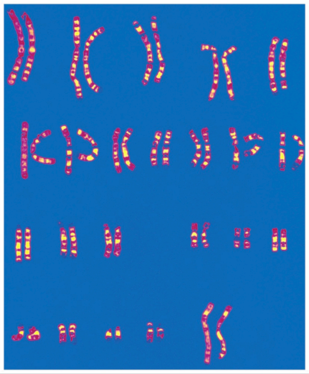


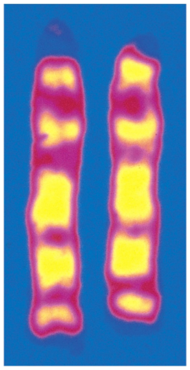
# Chromosome structure



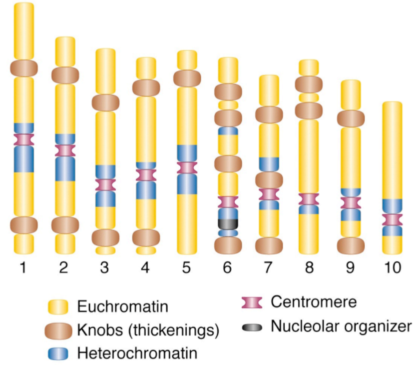


# Karyotype of human female/chromosome 13



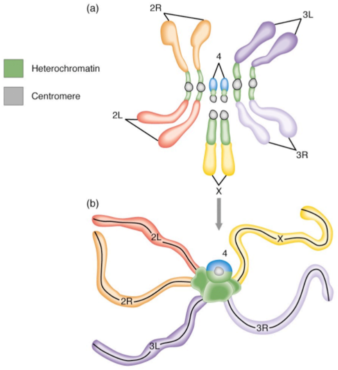


# Chromosome landmarks (corn)

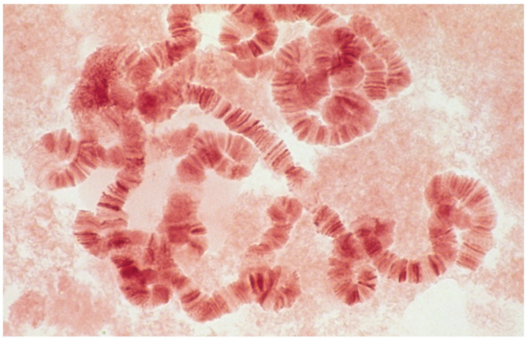


# Polytene chromosome formation

Heterochromatin forms the chromocenter



# Polytene chromosomes



# Lampbrush Chromosomes

Visible loops of DNA covered with an RNA matrix: sites of active transcription

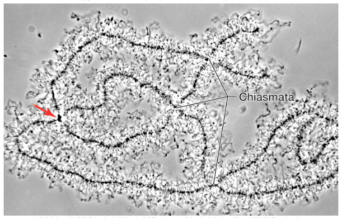


Figure 7.34

Joseph G. Gall, figure 2 in D.M. Prescott, ed., Methods in Cell Physiology, vol. 2 (New York: Academic Press, 1966), 39. Reproduced by permission

# Chromosome Banding

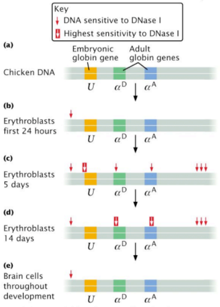
Characteristic banding pattern in stained chromosomes can be used to identify:

- Euchromatin (transcriptionally active)
- Heterochromatin (transcriptionally inactive)
- Facultative heterochromatin (can shift between transcriptionally active and inactive states)

# DNA Structure during development

**Question:** Is chromatin structure altered during transcription?

**Method:** Sensitivity to DNase I was tested on different tissues and at different times in development.



**Conclusion:** Sensitivity of DNA to digestion by DNase I is correlated with gene expression, suggesting that chromatin structure changes during transcription.

# Centromeres

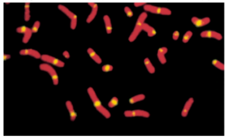


Figure 7.36b

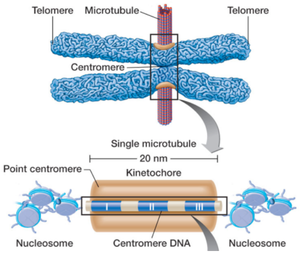


Figure 7.38

---

# Telomeres

- Repetitive DNA sequence located at the end of linear chromosomes
- Protect the ends of chromosomes from being degraded by nucleases
- Allow ends of the chromosomes to be properly replicated

---

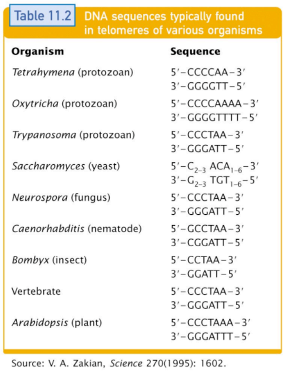
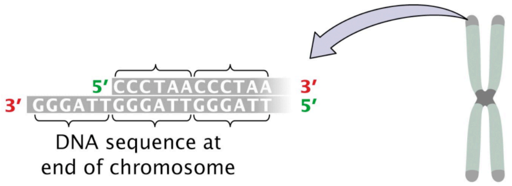


**Table 11.2** DNA sequences typically found in telomeres of various organisms

| **Organism** | **Sequence** |
|---|---|
| *Tetrahymena* (protozoan) | 5′–CCCCAA–3′<br>3′–GGGGTT–5′ |
| *Oxytricha* (protozoan) | 5′–CCCCAAAA–3′<br>3′–GGGGTTTT–5′ |
| *Trypanosoma* (protozoan) | 5′–CCCTAA–3′<br>3′–GGGATT–5′ |
| *Saccharomyces* (yeast) | $5'–C_{2-3}\ ACA_{1-6}–3'$<br>$3'–G_{2-3}\ TGT_{1-6}–5'$ |
| *Neurospora* (fungus) | 5′–CCCTAA–3′<br>3′–GGGATT–5′ |
| *Caenorhabditis* (nematode) | 5′–GCCTAA–3′<br>3′–CGGATT–5′ |
| *Bombyx* (insect) | 5′–CCTAA–3′<br>3′–GGATT–5′ |
| Vertebrate | 5′–CCCTAA–3′<br>3′–GGGATT–5′ |
| *Arabidopsis* (plant) | 5′–CCCTAAA–3′<br>3′–GGGATTT–5′ |

Source: V. A. Zakian, *Science* 270(1995): 1602.

---



DNA sequence at end of chromosome

---

# C Value Paradox

- DNA genome size (C value) does not correlate with complexity of organism
- Genome contains highly repetitive DNA:

  Transposable genetic elements

  retrotransposons-LINES and SINES
- Much eukaryotic DNA is junk DNA, lacks genes and has no known purpose

---

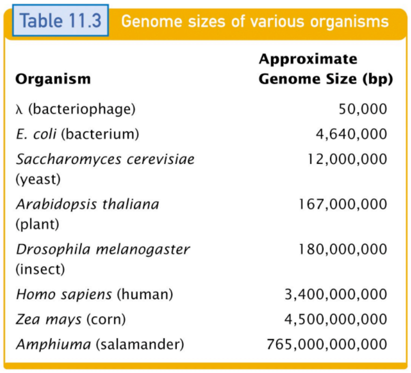
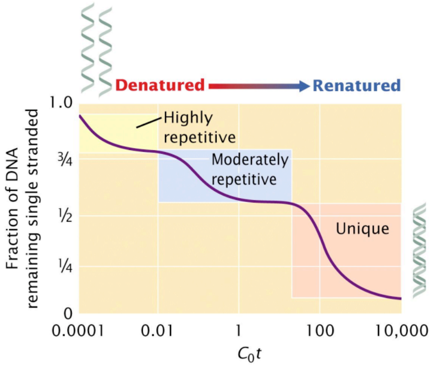
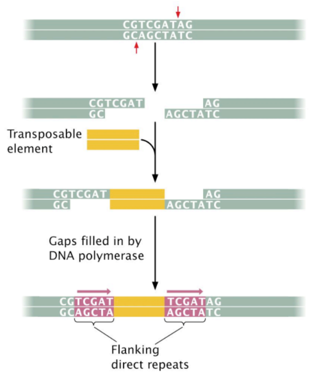


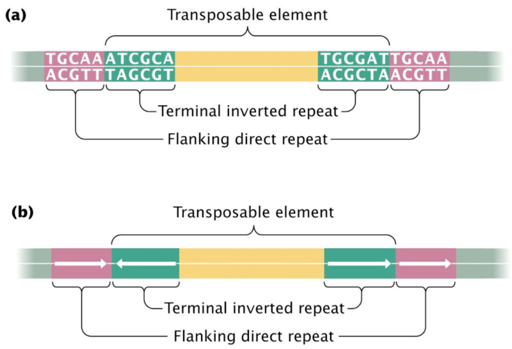
**Table 11.3** Genome sizes of various organisms

| **Organism** | **Approximate Genome Size (bp)** |
|---|---|
| λ (bacteriophage) | 50,000 |
| *E. coli* (bacterium) | 4,640,000 |
| *Saccharomyces cerevisiae* (yeast) | 12,000,000 |
| *Arabidopsis thaliana* (plant) | 167,000,000 |
| *Drosophila melanogaster* (insect) | 180,000,000 |
| *Homo sapiens* (human) | 3,400,000,000 |
| *Zea mays* (corn) | 4,500,000,000 |
| *Amphiuma* (salamander) | 765,000,000,000 |



# Transposable element

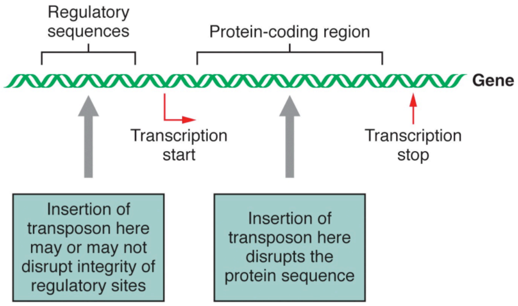




# Mobile elements and mutation

- Also known as transposable elements
- Encode transposase enzyme
- Types of prokaryotic mobile elements
  - insertion sequence (IS)
    - plasmids or chromosome
    - may move from one location to another
  - bacterial transposons (TN)
    - include genes conferring drug resistance (R factors)
    - ends consist of identical IS sequences in opposite orientation

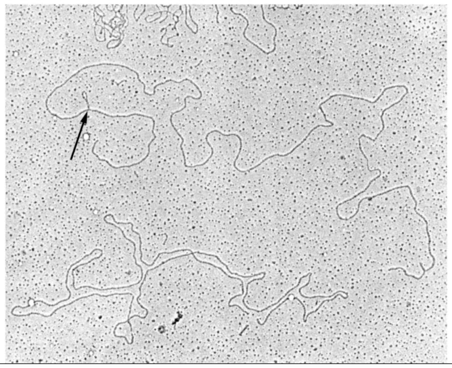
# Transposable elements



Insertion of transposon here may or may not disrupt integrity of regulatory sites

Insertion of transposon here disrupts the protein sequence

# Insertional sequence (IS) in λ*dgal*



# Prokaryotic IS elements

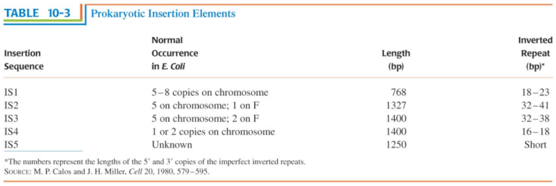
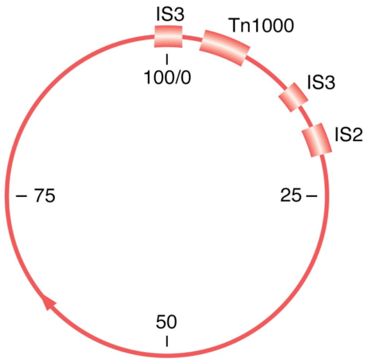


**TABLE 10-3** Prokaryotic Insertion Elements

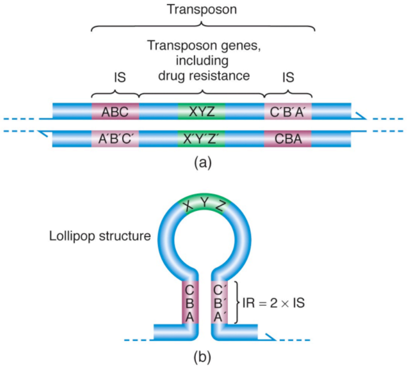
| Insertion Sequence | Normal Occurrence in *E. Coli* | Length (bp) | Inverted Repeat (bp)* |
|---|---|---|---|
| IS1 | 5–8 copies on chromosome | 768 | 18–23 |
| IS2 | 5 on chromosome; 1 on F | 1327 | 32–41 |
| IS3 | 5 on chromosome; 2 on F | 1400 | 32–38 |
| IS4 | 1 or 2 copies on chromosome | 1400 | 16–18 |
| IS5 | Unknown | 1250 | Short |

*The numbers represent the lengths of the 5′ and 3′ copies of the imperfect inverted repeats.
Source: M. P. Calos and J. H. Miller, *Cell* 20, 1980, 579–595.

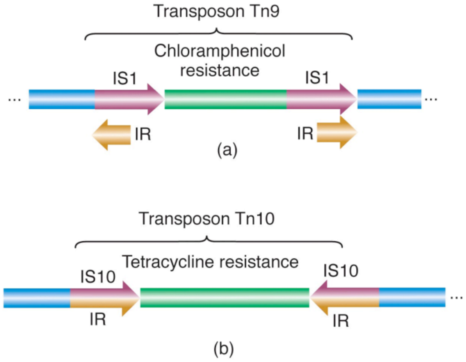
# F factor with IS elements



# Transposon with inverted repeats



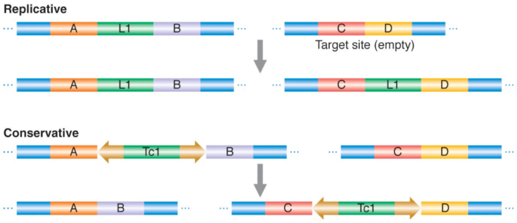
# Two different transposons



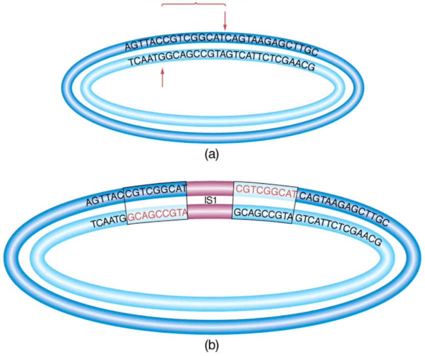
# Mechanisms of transposition

- Replicative transposition
  - – copy of transposon left behind
  - – new copy inserted elsewhere in genome
  - – mediated by transposon-specific transposase enzyme
- Conservative transposition
  - – excise from location and integrate elsewhere
  - – mediate by transposon-specific transposase
  - – transposon often flanked by duplicate repeat sequence generated by insertion

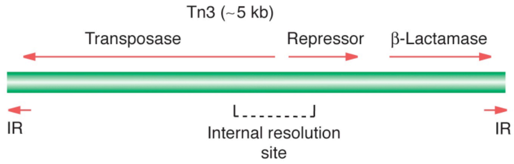
# Replicative vs conservative modes



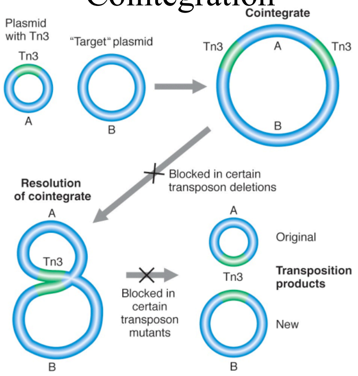
# Insertion results in repeats



# Tn3 transposon



# Cointegration



# Eukaryotic mobile elements

- Historically, mobile elements were discovered in eukaryotes by genetic analysis
  - Barbara McClintock working with maize
  - Ac (activator) element
  - Unstable mutations
- Mobile elements are common in eukaryotes
  - utilize transposase, as in prokaryotes
  - may cause mutation by insertion into gene
- Several categories
  - based on mode of replication and transposition
  - also based on types of extra genes present

# Rhoades

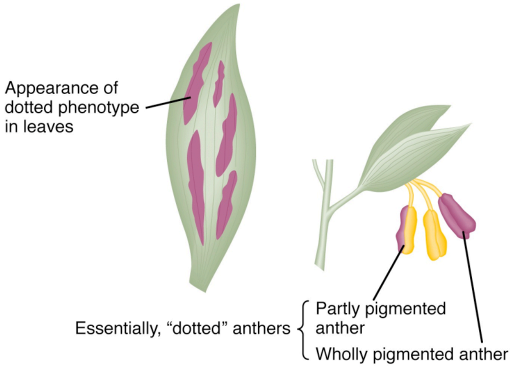
$$A_1 A_1\, dt\, dt \xrightarrow{\text{Two mutations}} A_1 a_1\, Dt\, dt \xrightarrow{\text{Selfed}} \text{Progeny}$$

$$\text{Progeny: } \left.\begin{array}{l} \frac{9}{16}\ A_1/- \;;\; Dt/- \\ \frac{3}{16}\ A_1/- \;;\; dt/dt \end{array}\right\}\ \frac{12}{16}\ \text{pigmented}$$

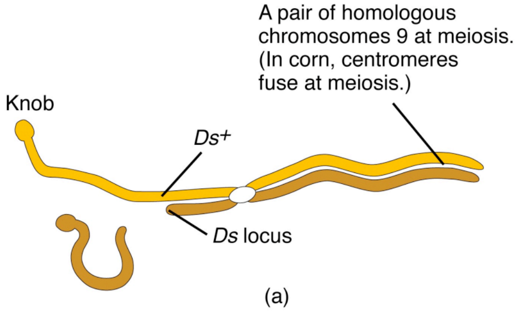
$$\left.\frac{3}{16}\ a_1/a_1 \;;\; Dt/-\right\}\ \frac{3}{16}\ \text{dotted}$$

$$\left.\frac{1}{16}\ a_1/a_1 \;;\; dt/dt\right\}\ \frac{1}{16}\ \text{colorless}$$

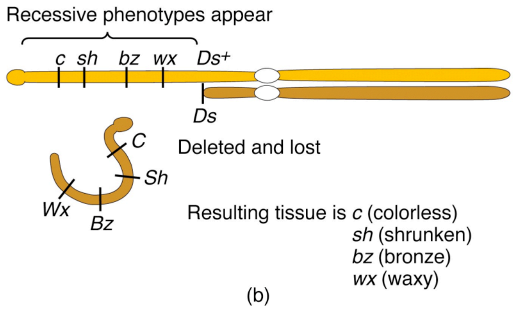
# Rhodes



# McClintock



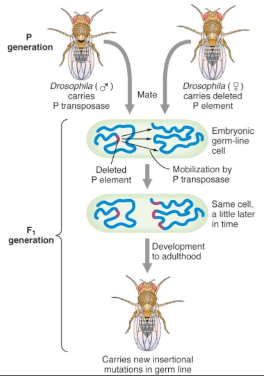
# McClintock



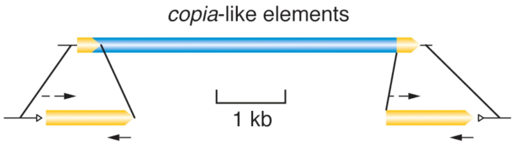
# Retroviral-like mobile elements

- Transpose through RNA intermediate
  - e.g., *copia* in *Drosophila*
    - 4-9 kb in length
    - long terminal repeat (LTR)
    - reverse transcriptase copies RNA into DNA
  - may or may not have LTR
- In mammals, includes
  - LINES (long)
    - functional elements
  - SINES (short)
    - nonfunctional elements

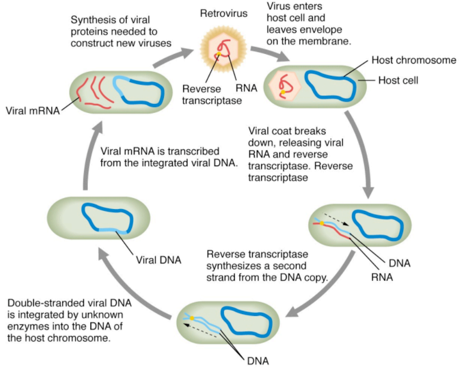
# Crosses



# Copia-like elements in *Drosophila*

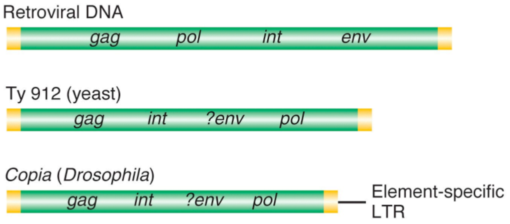


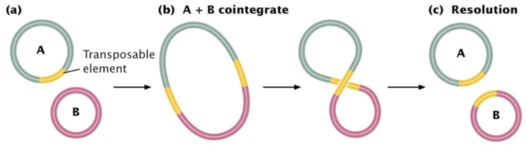
# Retroviruses

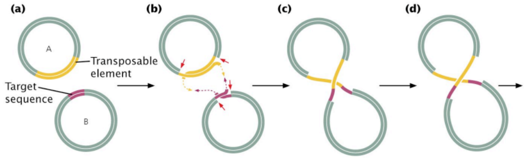


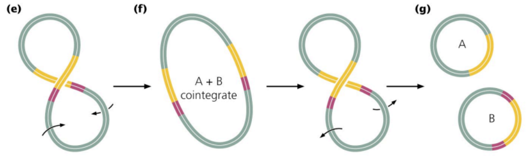
# Retrovirus genes

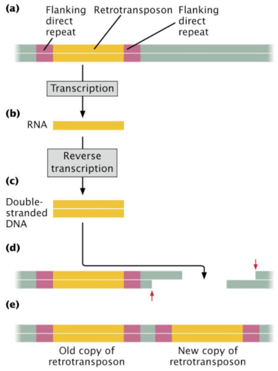
Genes in retroviral DNA and viral retrotransposons

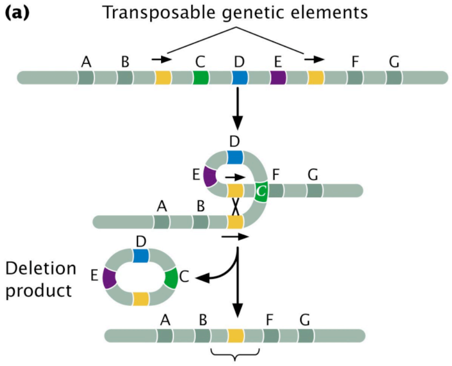


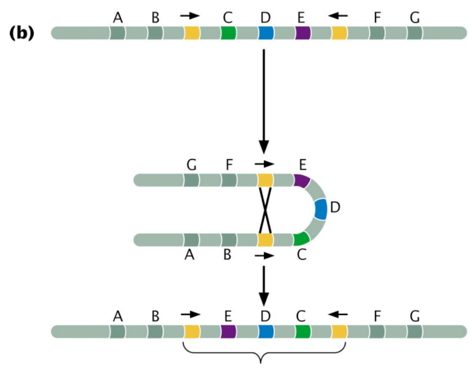


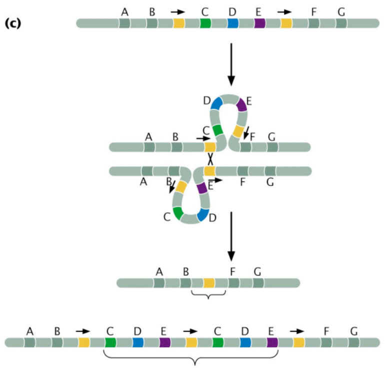


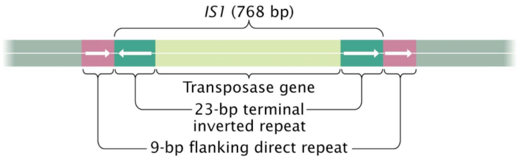


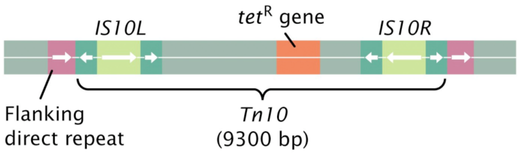












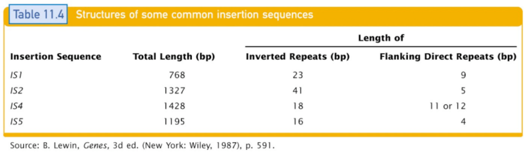


Table 11.4 Structures of some common insertion sequences

| | | Length of | |
|---|---|---|---|
| **Insertion Sequence** | **Total Length (bp)** | **Inverted Repeats (bp)** | **Flanking Direct Repeats (bp)** |
| *IS1* | 768 | 23 | 9 |
| *IS2* | 1327 | 41 | 5 |
| *IS4* | 1428 | 18 | 11 or 12 |
| *IS5* | 1195 | 16 | 4 |

Source: B. Lewin, *Genes*, 3d ed. (New York: Wiley, 1987), p. 591.

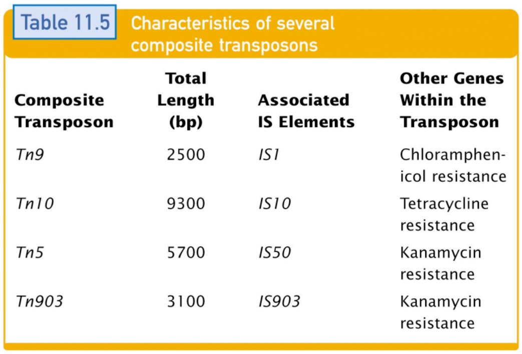
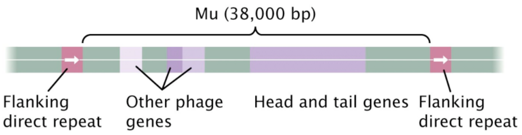
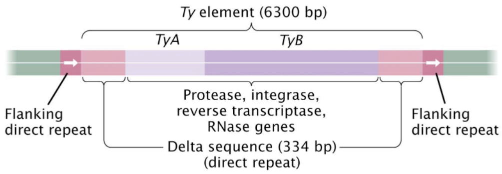
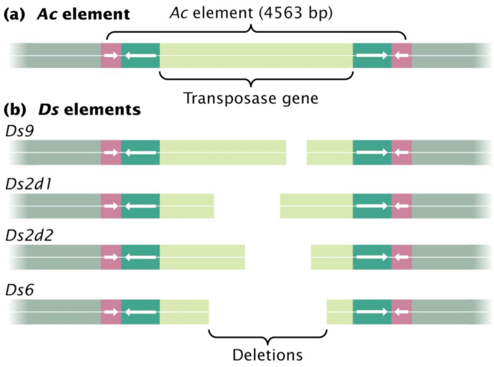


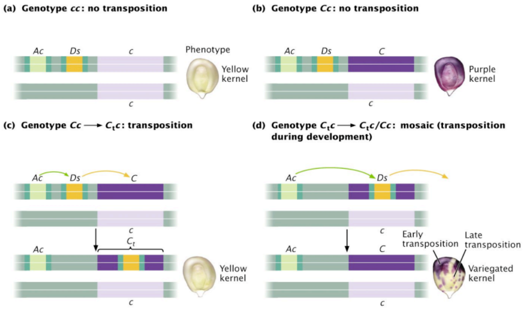
Table 11.5 Characteristics of several composite transposons

| **Composite Transposon** | **Total Length (bp)** | **Associated IS Elements** | **Other Genes Within the Transposon** |
|---|---|---|---|
| *Tn9* | 2500 | *IS1* | Chloramphenicol resistance |
| *Tn10* | 9300 | *IS10* | Tetracycline resistance |
| *Tn5* | 5700 | *IS50* | Kanamycin resistance |
| *Tn903* | 3100 | *IS903* | Kanamycin resistance |

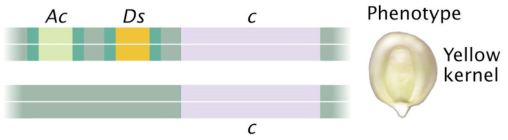

Mu (38,000 bp)
Flanking direct repeat
Other phage genes
Head and tail genes
Flanking direct repeat


*Ty* element (6300 bp)
*TyA*
*TyB*
Flanking direct repeat
Protease, integrase, reverse transcriptase, RNase genes
Delta sequence (334 bp) (direct repeat)
Flanking direct repeat

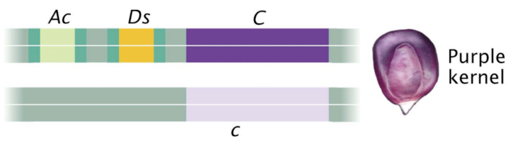

**(a) *Ac* element**
*Ac* element (4563 bp)
Transposase gene
**(b) *Ds* elements**
*Ds9*
*Ds2d1*
*Ds2d2*
*Ds6*
Deletions


(a) Genotype *cc*: no transposition
*Ac*
*Ds*
*c*
Phenotype
Yellow kernel
*c*
(b) Genotype *Cc*: no transposition
*Ac*
*Ds*
*C*
Purple kernel
*c*
(c) Genotype *Cc* ⟶ $C_t$*c*: transposition
*Ac*
*Ds*
*C*
*c*
$C_t$
*Ac*
Yellow kernel
*c*
(d) Genotype $C_t$*c* ⟶ $C_t$*c*/*Cc*: mosaic (transposition during development)
*Ac*
*Ds*
*c*
*Ac*
*C*
Early transposition
Late transposition
Variegated kernel
*c*

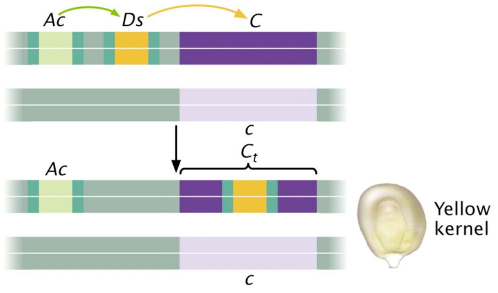
**(a) Genotype *cc*: no transposition**



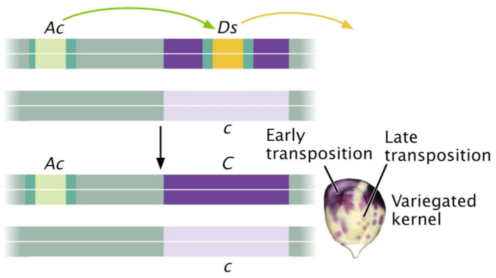
**(b) Genotype *Cc*: no transposition**

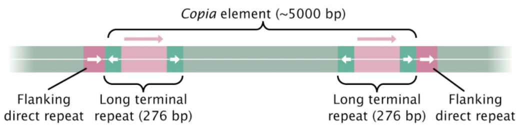


**(c) Genotype *Cc* ⟶ $C_t c$: transposition**

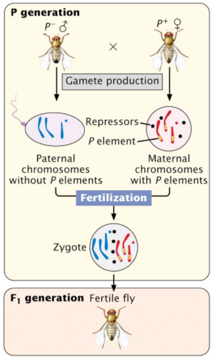


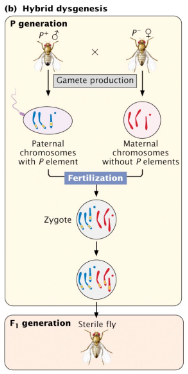
**(d) Genotype $C_t c$ ⟶ $C_t c/Cc$: mosaic (transposition during development)**





**(a) No hybrid dysgenesis**





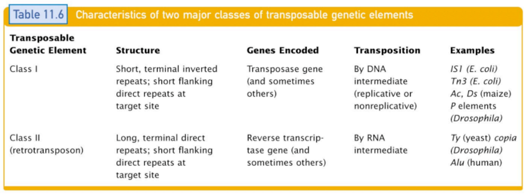


Table 11.6 Characteristics of two major classes of transposable genetic elements

| Transposable Genetic Element | Structure | Genes Encoded | Transposition | Examples |
|---|---|---|---|---|
| Class I | Short, terminal inverted repeats; short flanking direct repeats at target site | Transposase gene (and sometimes others) | By DNA intermediate (replicative or nonreplicative) | *IS1 (E. coli)*<br>*Tn3 (E. coli)*<br>*Ac*, *Ds* (maize)<br>*P* elements *(Drosophila)* |
| Class II (retrotransposon) | Long, terminal direct repeats; short flanking direct repeats at target site | Reverse transcriptase gene (and sometimes others) | By RNA intermediate | *Ty* (yeast) *copia (Drosophila)*<br>*Alu* (human) |