Chapter 21

Genomes and Their Evolution

PowerPoint® Lecture Presentations for

Biology

Eighth Edition

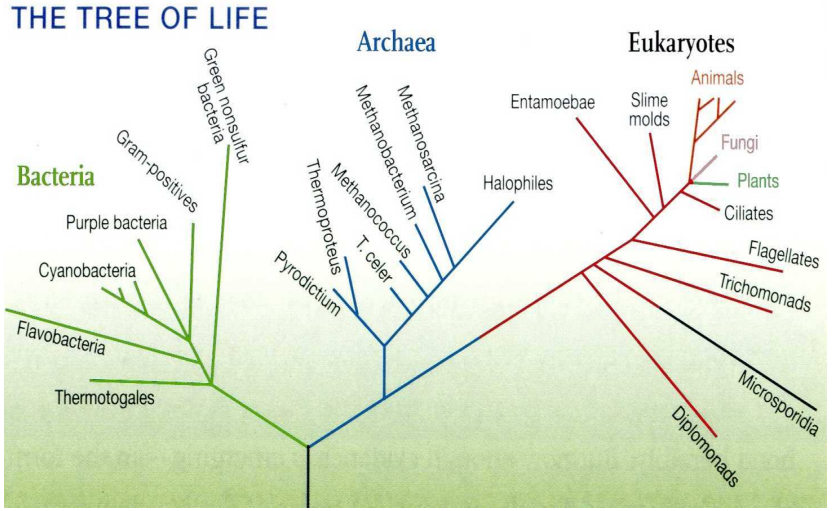
Neil Campbell and Jane Reece

Lectures by Chris Romero, updated by Erin Barley with contributions from Joan Sharp

Overview: Reading the Leaves from the Tree of Life

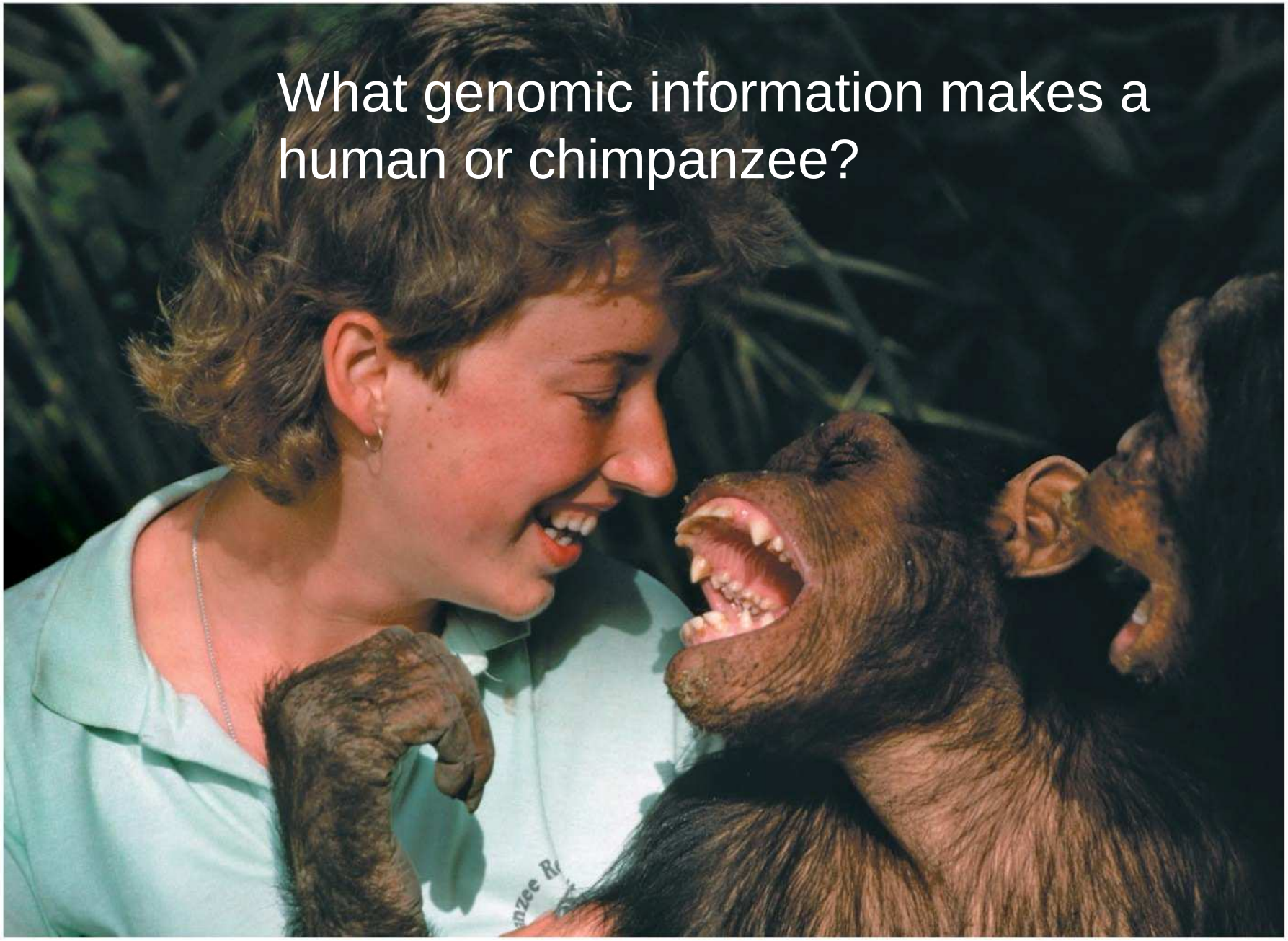
- Complete genome sequences exist for a
  - Human, chimpanzee, rhesus macaque
  - House mouse
  - Fruit fly,
  - Nematode,
  - Zebrafish
  - *E. coli*, brewer's yeast

  And many other organisms



- Comparisons of genomes among organisms provide information about the evolutionary history of genes and taxonomic groups

Fig. 21-1



Genomics 基因體學

- **Genomics** is the study of whole sets of genes and their interactions
- **Bioinformatics** is the application of computational methods to the storage and analysis of biological data

Concept 21.1: New approaches have accelerated the pace of genome sequencing

- The most ambitious mapping project to date has been the sequencing of the human genome
- Officially begun as the **Human Genome Project** in 1990, the sequencing was largely completed by 2003
- The project had three stages:
  - **Genetic (or linkage) mapping**
  - **Physical mapping**
  - **DNA sequencing**

Approach 1:

Three-Stage Approach to Genome Sequencing

Stage 1: Linkage mapping

- A **linkage map** (genetic map) maps the location of several thousand genetic markers on each chromosome
- A genetic marker is a gene or other identifiable DNA sequence
- **Recombination frequencies** are used to determine the order and relative distances between genetic markers

Fig. 21-2-1

Cytogenetic map

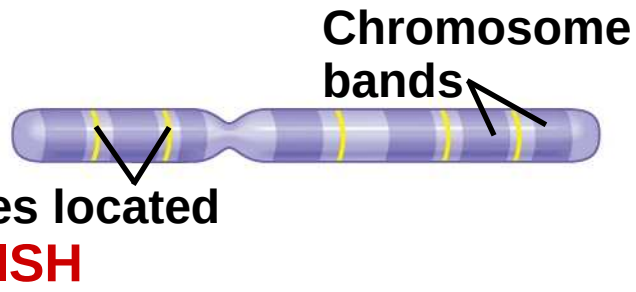


Fig. 21-2-2

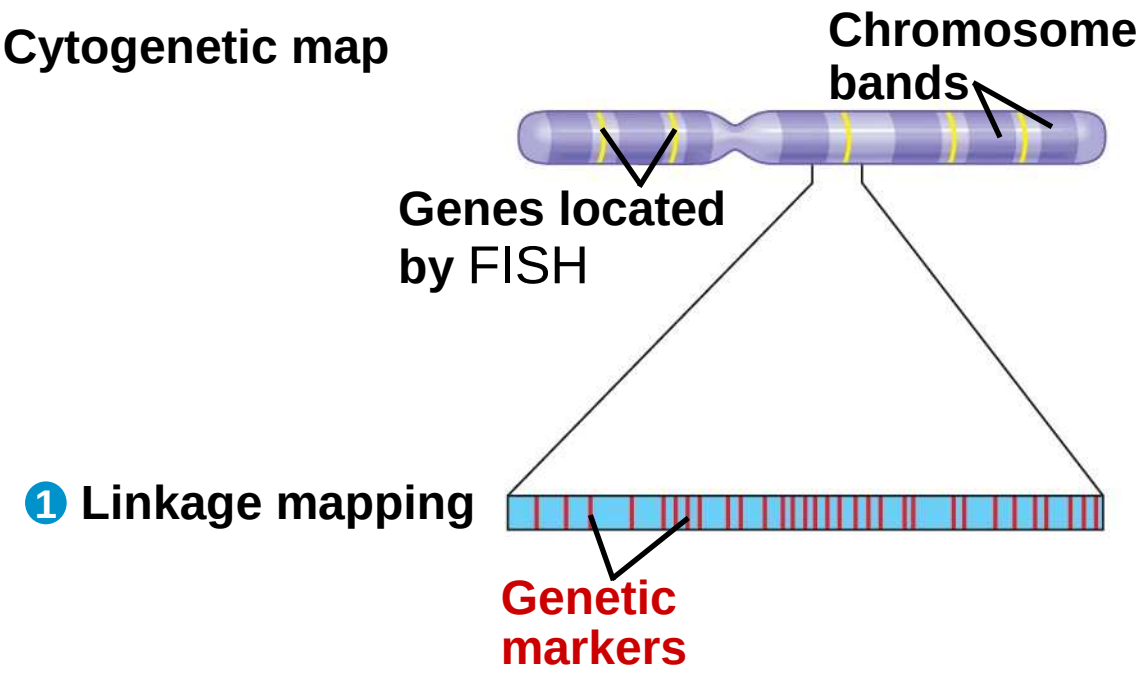


Fig. 21-2-3

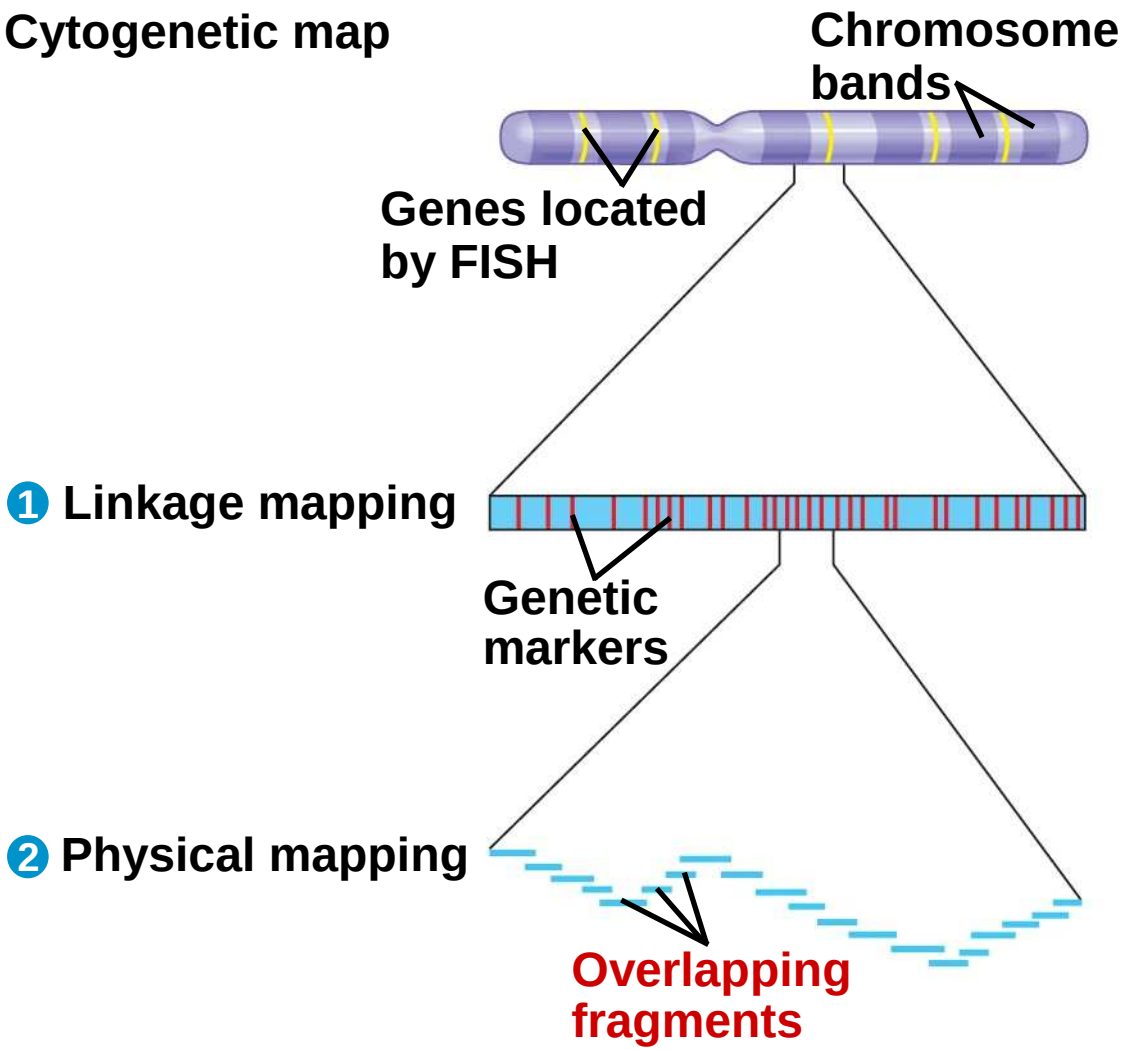
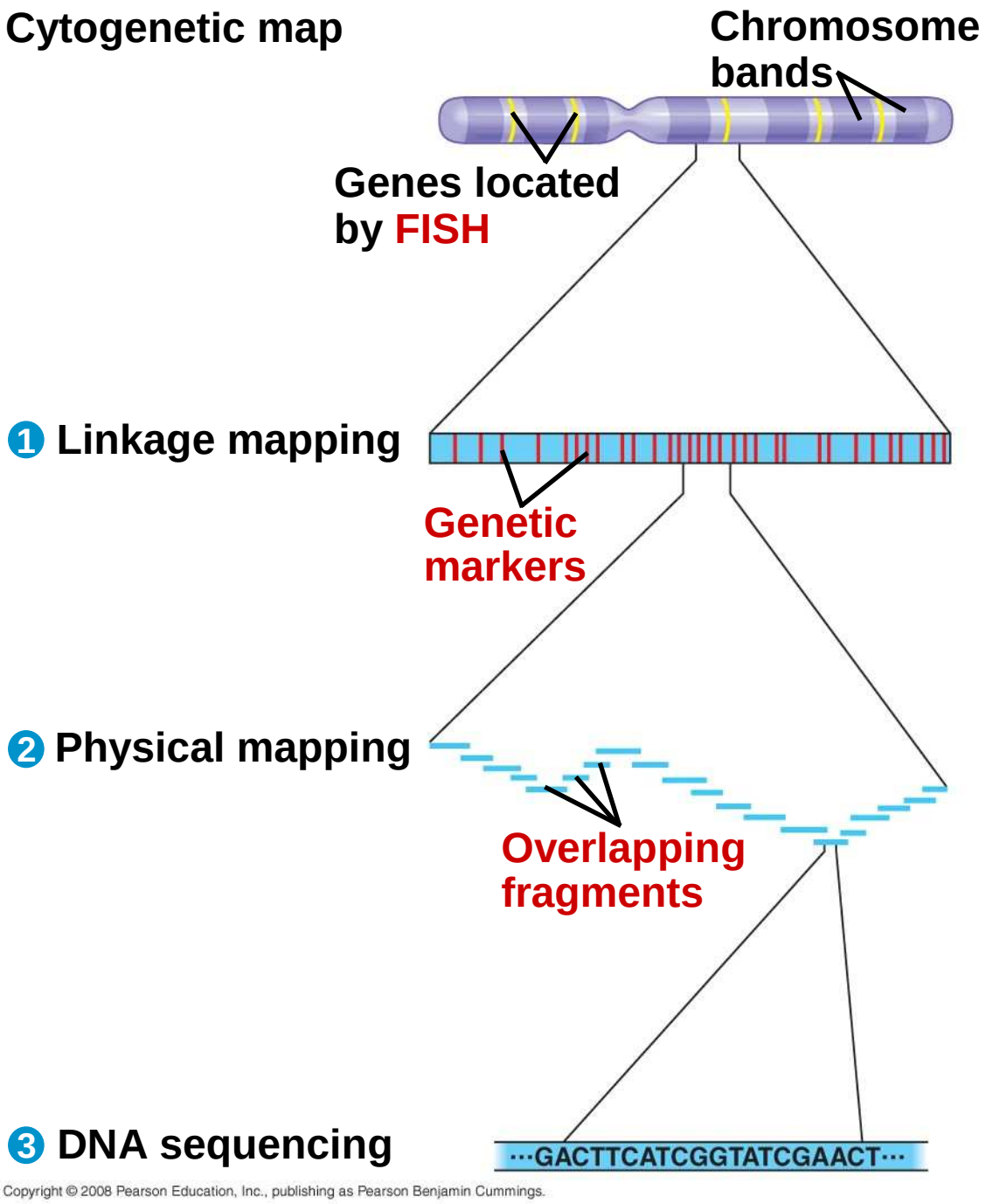
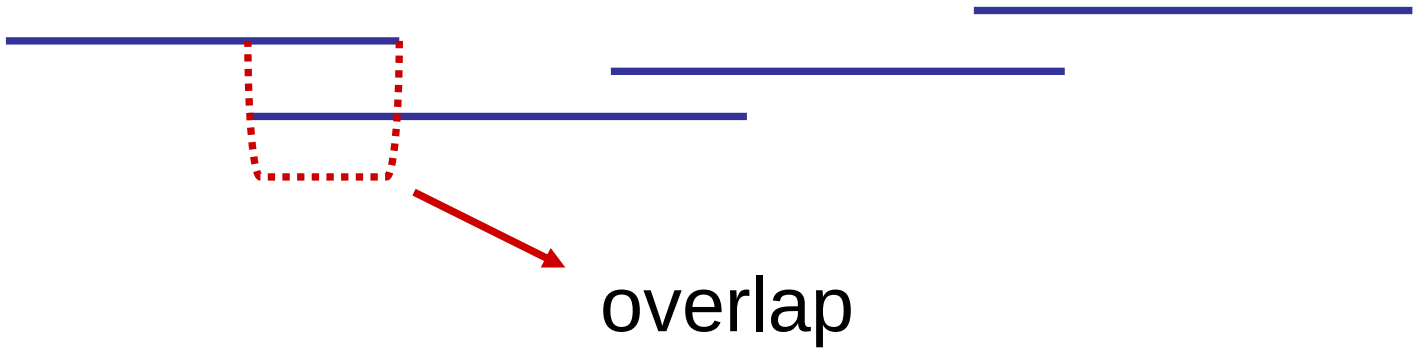


Fig. 21-2-4



Stage 2: Physical mapping

- A **physical map** expresses the distance between genetic markers, usually as the number of base pairs along the DNA
- It is constructed by cutting a DNA molecule into many short fragments and arranging them in order by identifying overlaps

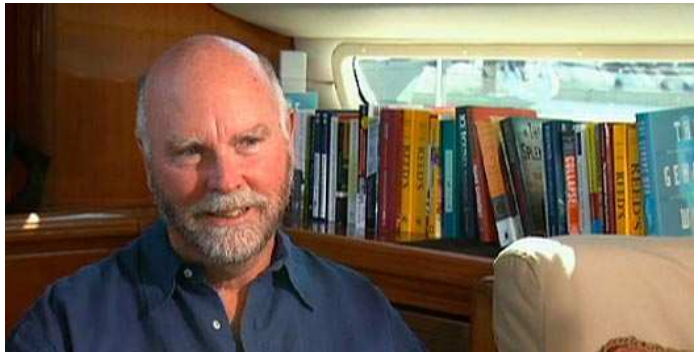


Stage 3: DNA sequencing

- Sequencing machines are used to determine the complete nucleotide sequence of each chromosome
- A complete haploid set of human chromosomes consists of 3.2 billion base pairs

Approach 2: Whole-Genome Shotgun Approach to Genome Sequencing

- The whole-genome shotgun approach was developed by J. Craig Venter in 1992



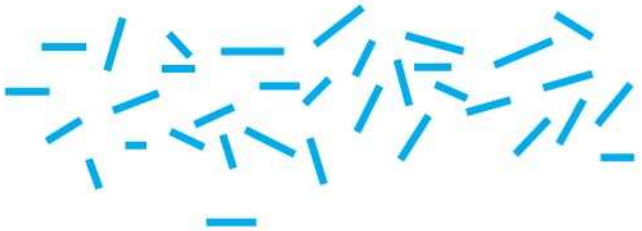
- This approach skips genetic and physical mapping and sequences random DNA fragments directly
- Powerful computer programs are used to order fragments into a continuous sequence

Fig. 21-3-1

Whole-genome shotgun approach to sequencing

1 **Cut the DNA into overlapping fragments short enough for sequencing**

Whole-genome shotgun approach to sequencing



2 **Clone the fragments in plasmid or phage vectors.**

Fig. 21-3-2

Whole-genome shotgun approach to sequencing

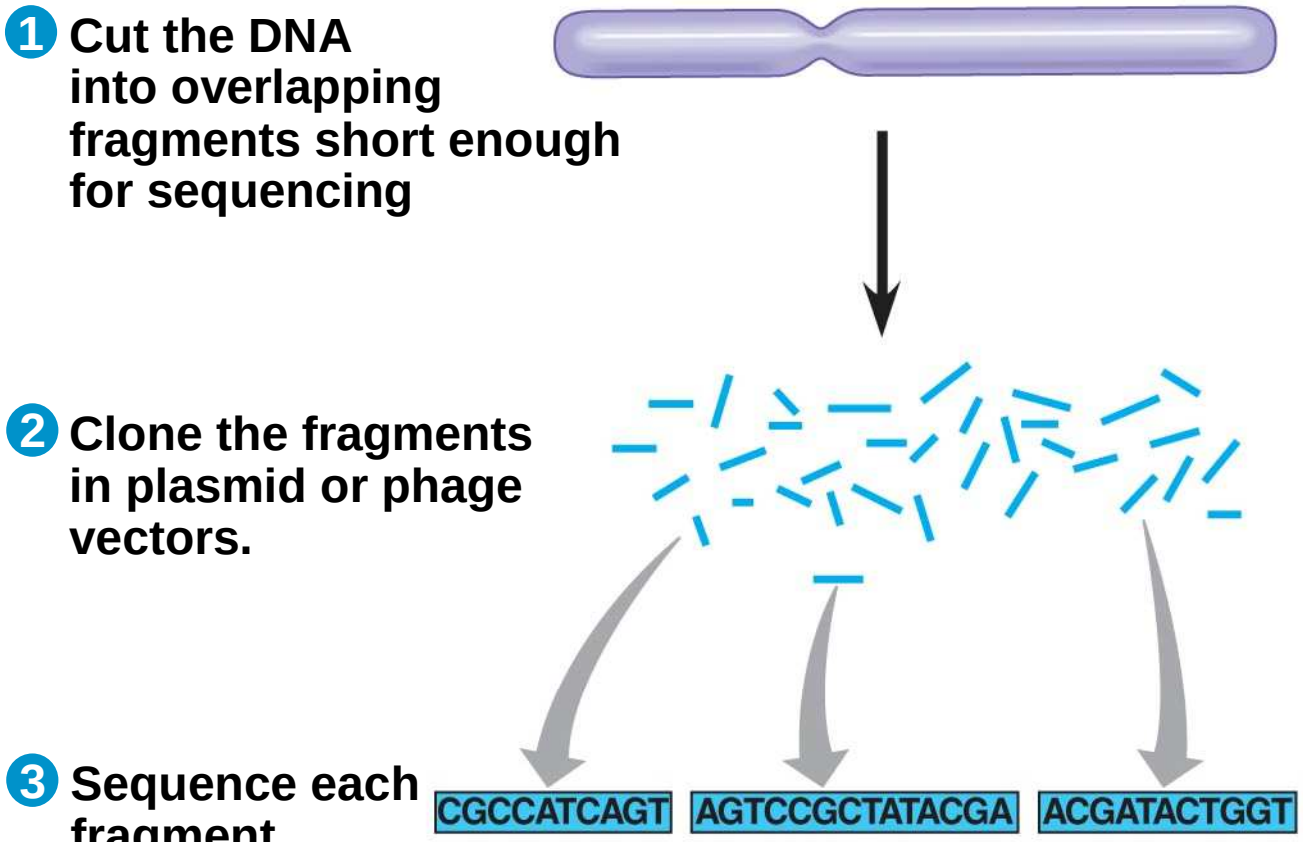
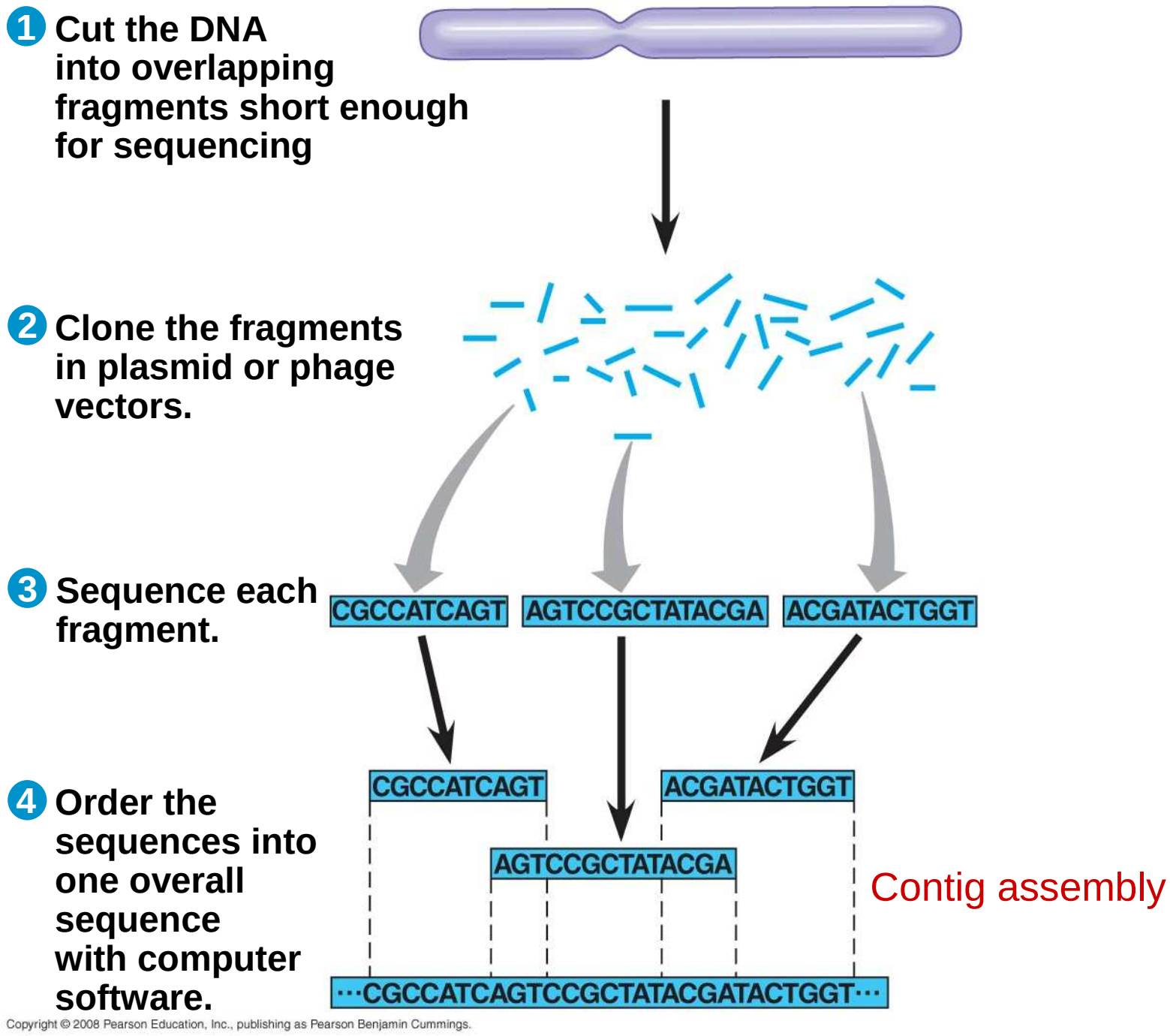


Fig. 21-3-3

Whole-genome shotgun approach to sequencing



Whole-genome shotgun approach

- Both the three-stage process and the whole-genome shotgun approach were used for the Human Genome Project and for genome sequencing of other organisms
- At first many scientists were skeptical about the whole-genome shotgun approach, but it is now widely used as the sequencing method of choice
- A hybrid of the two approaches may be the most useful in the long run

Next Generation DNA Sequencing

1964 - 77 bp, 4 researchers, three years

2003 - US$3 billion, ~10 years

2009 - US$5000, weeks

2012 – US$1000, weeks

before 2020 – US$100, days!

Table 1. Second-generation DNA sequencing technologies

▲ Figures and tables index

	Feature generation	Sequencing by synthesis	Cost per megabase	Cost per instrument	Paired ends?	1° error modality	Read-length
454	Emulsion PCR	Polymerase (pyrosequencing)	~$60	$500,000	Yes	Indel	250 bp
Solexa	Bridge PCR	Polymerase (reversible terminators)	~$2	$430,000	Yes	Subst.	36 bp
SOLiD	Emulsion PCR	Ligase (octamers with two-base encoding)	~$2	$591,000	Yes	Subst.	35 bp
Polonator	Emulsion PCR	Ligase (nonamers)	~$1	$155,000	Yes	Subst.	13 bp
HeliScope	Single molecule	Polymerase (asynchronous extensions)	~$1	$1,350,000	Yes	Del	30 bp

By Next Generation Sequencing

NATURE | Vol 462 |17 December 2009

SNAPSHOT

China celebrates panda genome

With just 1,600 giant pandas estimated to remain in the wild, Chinese scientists have led the task of immortalizing the charismatic critter's 2.25 billion base pairs of DNA, reporting their findings online in *Nature* last week. Although it is unlikely to have a significant effect on conservation, the work is a proof-of-principle for next-generation sequencing technologies, and allows China to trumpet work involving a national animal. Indeed, one tactic for researchers hoping to win funding may be to sequence similarly patriotic symbols. "Australia has the most interesting animals in the world," says Jenny Graves, a geneticist at the Australian National University in Canberra and deputy director of the Australian Research Council's Centre for Kangaroo Genomics, who analysed sequences from the first marsupial (a South American opossum, ironically) and the duck-billed platypus. Graves says that such efforts are not just gimmicks; the kangaroo genomics project has helped researchers to work out that the *SRY* gene determines sex in humans and other mammals (J. W. Foster *et al. Nature* 359, 531–533; 1992). Other patriotic sequencing projects are detailed in the table.

Brendan Borrell

Country	Organism	Status
China	Giant panda	Draft assembly in 2009
Australia	Tammar wallaby	Whole-genome map in 2008
United States (Hawaii)	Transgenic papaya	Draft assembly in 2008
France and Italy	Wine grape (Pinot Noir strain)	Draft assembly in 2007
China and United States	Rice	Draft assembly in 2002
Sweden	Norway (European) spruce	Recently announced

Concept 21.2 Scientists use bioinformatics to analyze genomes and their functions

- The Human Genome Project established databases and refined analytical software to make data available on the Internet
- These databases and software (Bioinformatics) have accelerated progress in DNA sequence analysis

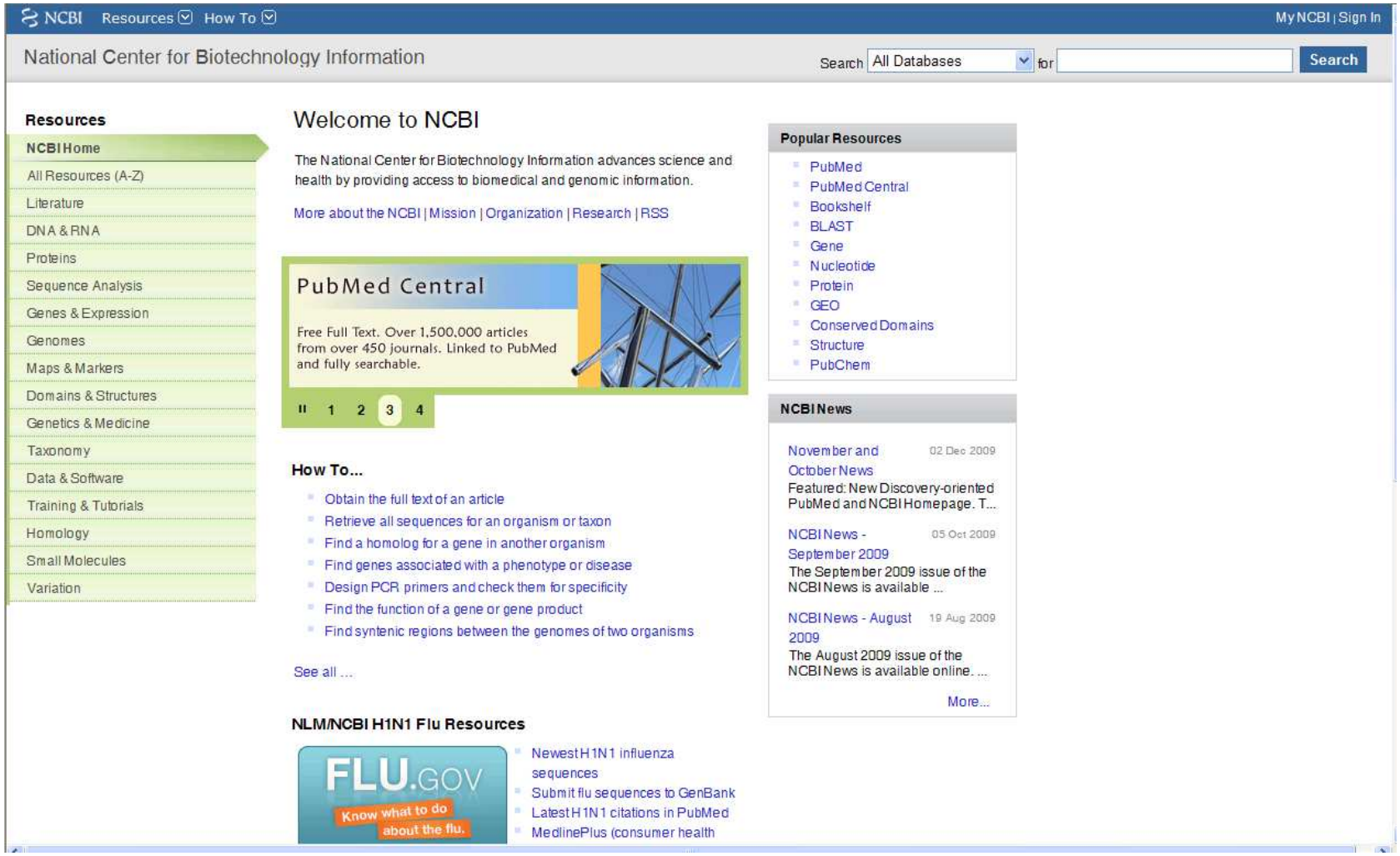
Centralized Resources for Analyzing Genome Sequences

- Bioinformatics resources are provided by a number of sources:
  - National Library of Medicine and the National Institutes of Health (NIH) created the National Center for Biotechnology Information (NCBI)
  - European Molecular Biology Laboratory
  - DNA Data Bank of Japan

Bioinformatics on internet – NCBI as an example

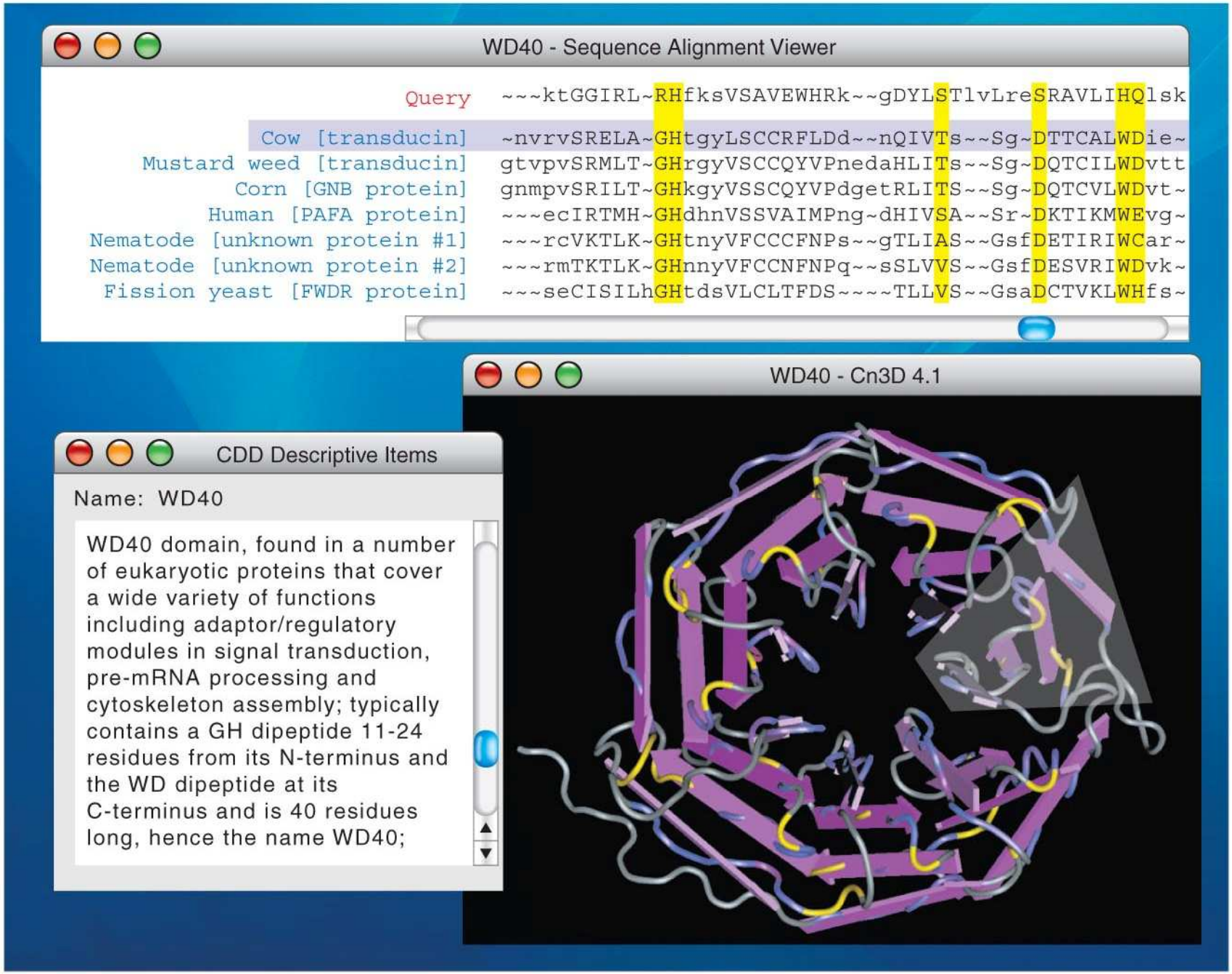
- Genbank, the NCBI database of sequences, doubles its data approximately every 18 months
- Software is available that allows online visitors to search Genbank for matches to:
  - A specific DNA sequence
  - A predicted protein sequence
  - Common stretches of amino acids in a protein
- The NCBI website also provides 3-D views of all protein structures that have been determined

NCBI – National Center for Biotechnology Infomraiotn



http://www.ncbi.nlm.nih.gov/

Fig. 21-4

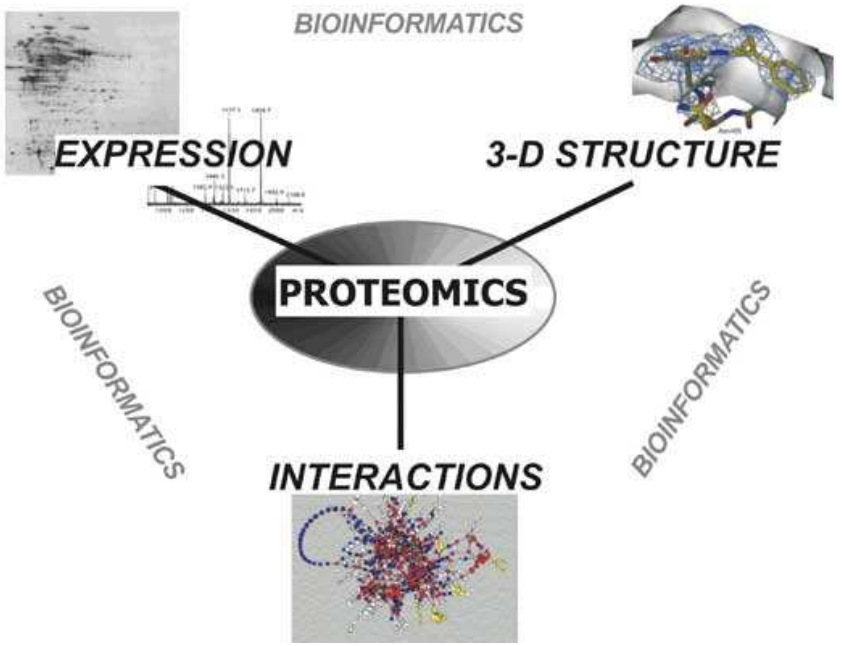


Identifying Protein-Coding Genes Within DNA Sequences

- Computer analysis of genome sequences helps identify sequences likely to encode proteins
- Comparison of sequences of "new" genes with those of known genes in other species may help identify new genes
  - 比對基因或蛋白質序列而獲得新資訊

Understanding Genes and Their Products at the Systems Level

- **Proteomics** is the systematic study of all proteins encoded by a genome
- Proteins, not genes, carry out most of the activities of the cell

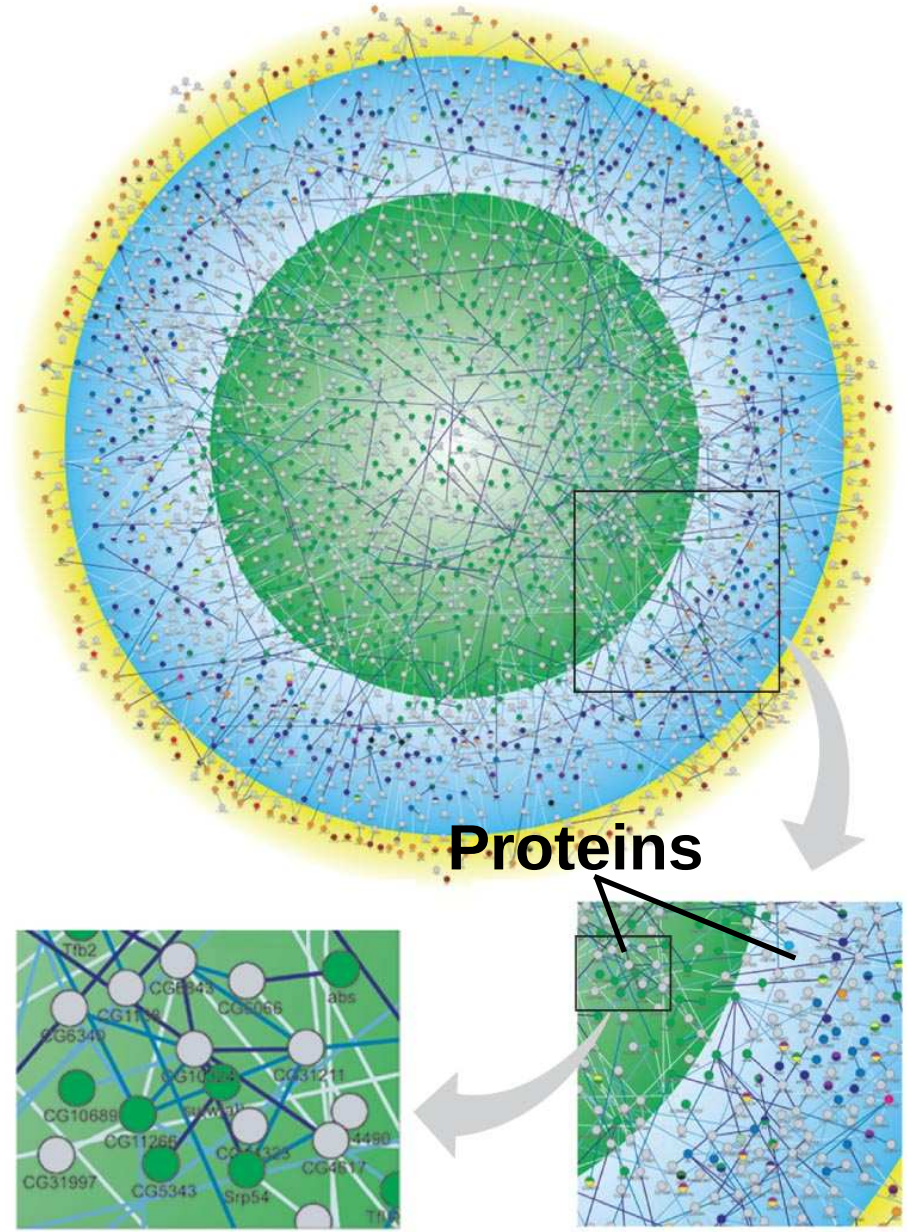


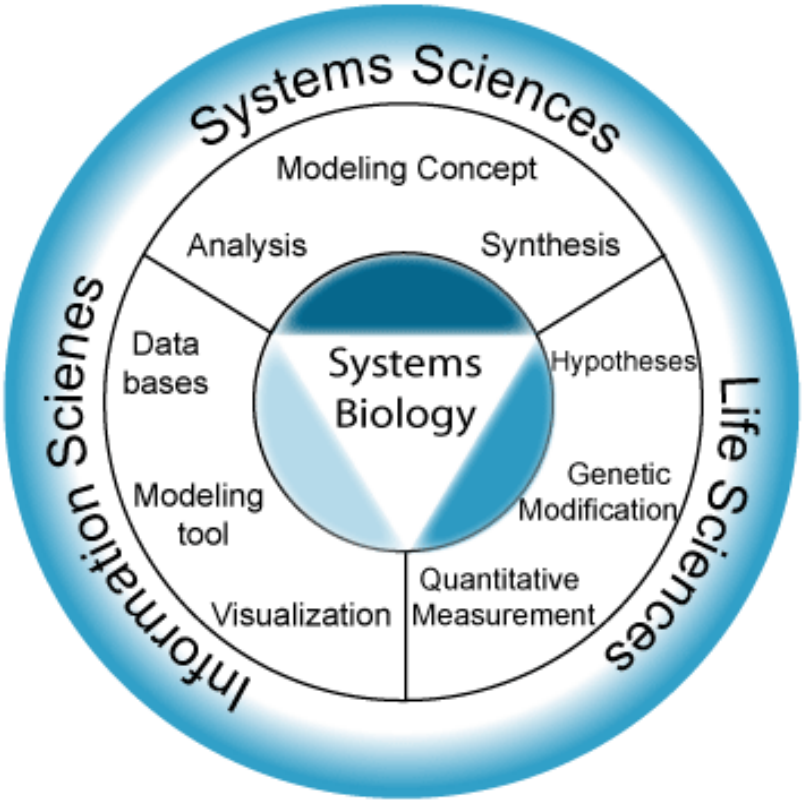
How Systems Are Studied: An Example

- A systems biology (系統生物學）approach can be applied to define gene circuits and protein interaction networks
- Researchers working on *Drosophila* used powerful computers and software to predict 4,700 protein products that participated in 4,000 interactions
- The systems biology approach is possible because of advances in bioinformatics

Fig. 21-5

Systems Biology





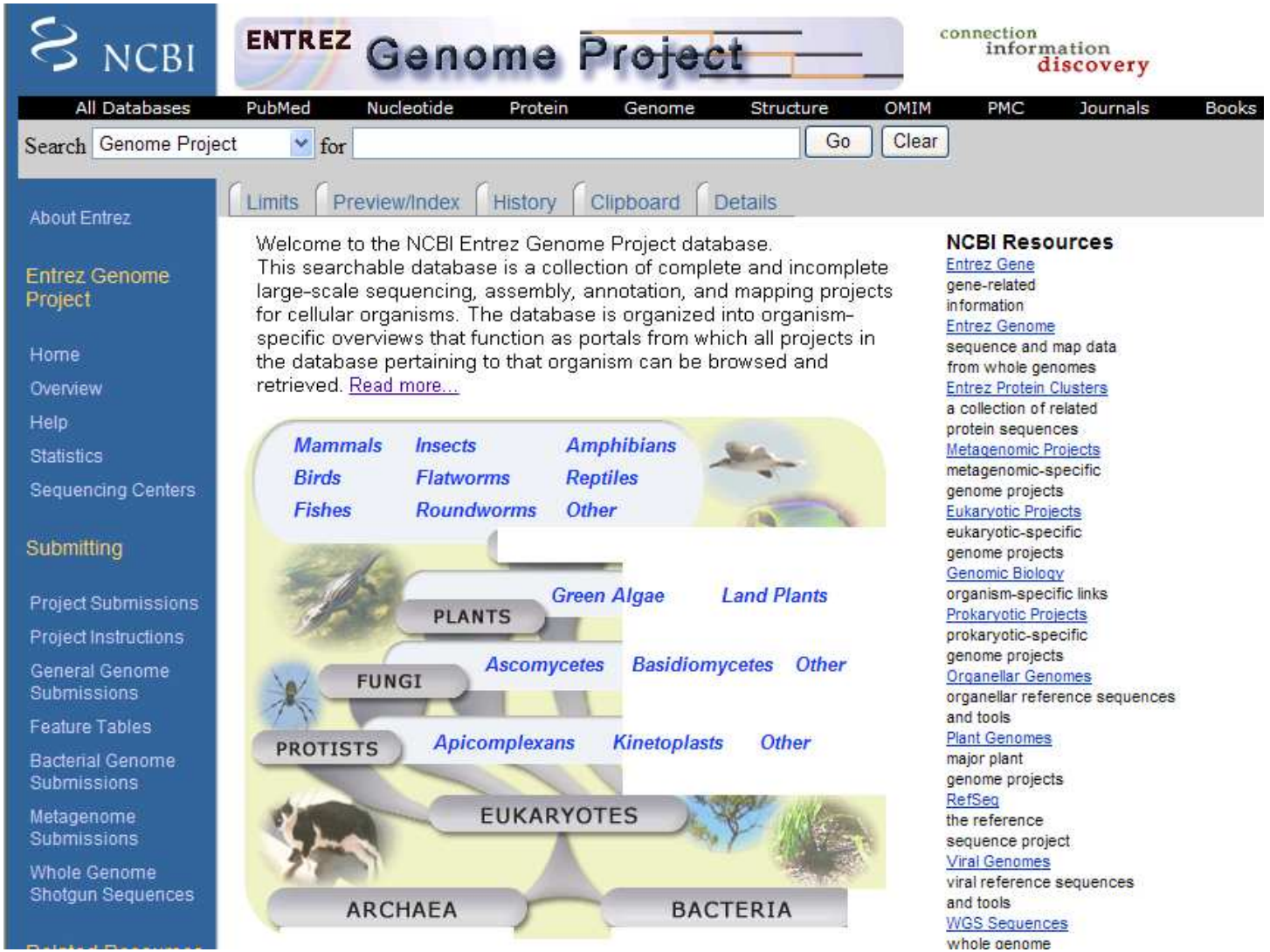
Protein-Protein Interaction Network

Application of Systems Biology to Medicine

- A systems biology approach has several medical applications:
  - The Cancer Genome Atlas project is currently monitoring 2,000 genes in cancer cells for changes due to mutations and rearrangements
  - Treatment of cancers and other diseases can be individually tailored following analysis of gene expression patterns in a patient
  - In future, DNA sequencing may highlight diseases to which an individual is predisposed

Concept 21.3 Genomes vary in size, number of genes, and gene density

http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj



Genome Size

- Genomes of most bacteria and archaea range from 1 to 6 million base pairs (Mb); genomes of eukaryotes are usually larger
- Most plants and animals have genomes greater than 100 Mb; humans have **3,200 Mb**
- Within each domain there is no systematic relationship between genome size and phenotype

Table 21-1

Table 21.1 Genome Sizes and Estimated Numbers of Genes*

Organism	Haploid Genome Size (Mb)	Number of Genes	Genes per Mb
Bacteria			
Haemophilus influenzae	1.8	1,700	940
Escherichia coli	4.6	4,400	950
Archaea			
Archaeoglobus fulgidus	2.2	2,500	1,130
Methanosarcina barkeri	4.8	3,600	750
Eukaryotes			
Saccharomyces cerevisiae (yeast)	13	6,200	480
Caenorhabditis elegans (nematode)	100	20,000	200
Arabidopsis thaliana (plant)	118	25,500	215
Drosophila melanogaster (fruit fly)	180	13,700	76
Oryza sativa (rice)	390	40,000	140
Danio rerio (zebrafish)	1,700	23,000	13
Mus musculus (house mouse)	2,600	22,000	11
Homo sapiens (human)	3,200	20,500	7
Fritillaria assyriaca (plant)	120,000	ND	ND

*Some values given here are likely to be revised as genome analysis continues. Mb = million base pairs. ND = not determined.

Number of Genes

- Free-living bacteria and archaea
  - have 1,500 to 7,500 genes
- Unicellular fungi
  - have ~ 5,000 genes
- Multicellular eukaryotes (animals and plants)
  - Have ~ 14,000~40,000 genes

Genome size ≠ Gene number

- Number of genes is not correlated to genome size
- For example, it is estimated that the nematode *C. elegans* has 100 Mb and 20,000 genes, while humans have 3,200 Mb and 20,488 genes
- Vertebrate genomes can produce more than one polypeptide per gene because of **alternative splicing** of RNA transcripts

Gene Density and Noncoding DNA

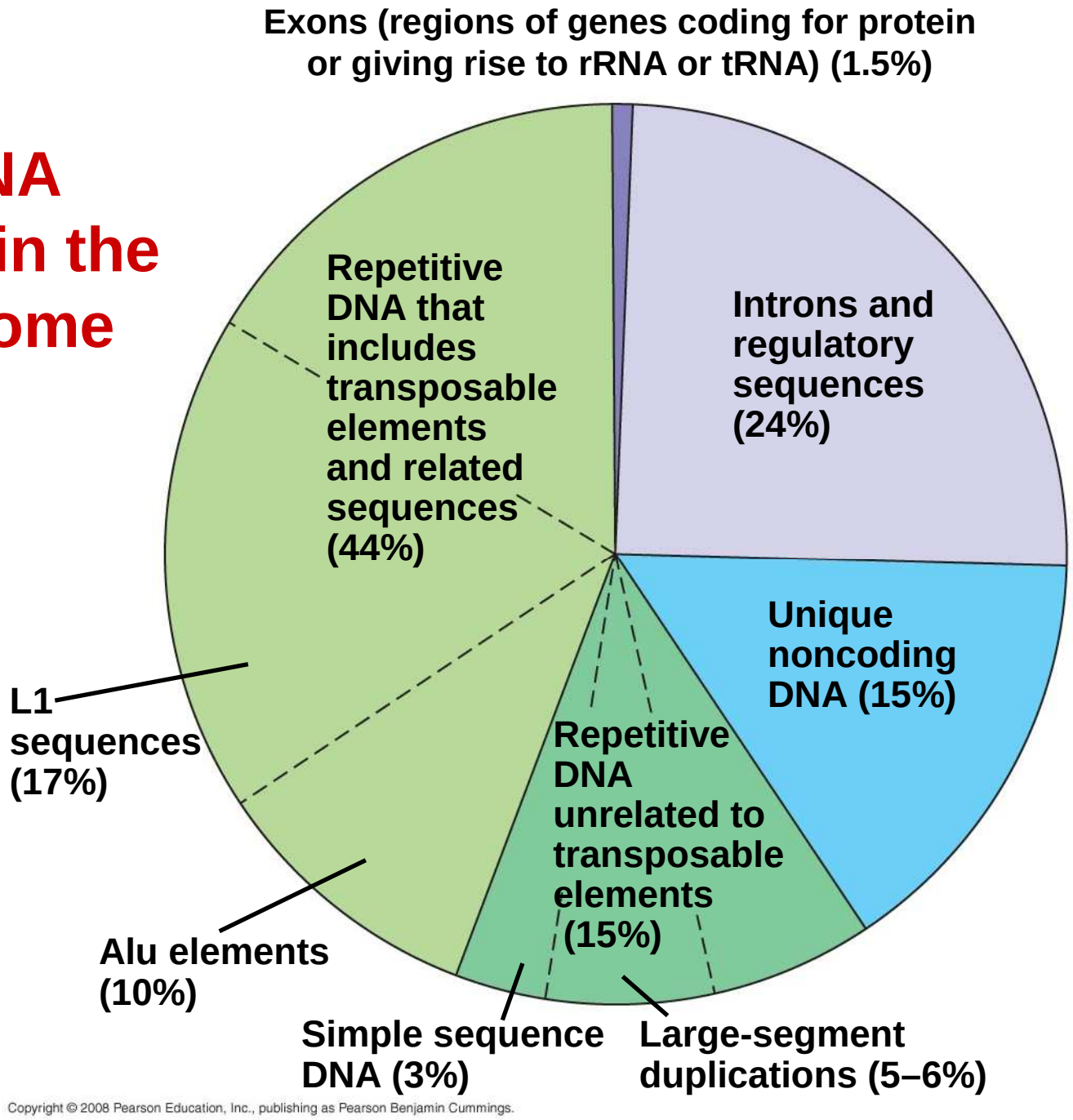
- Humans and other mammals have the lowest gene density, or number of genes, in a given length of DNA
- Multicellular eukaryotes have many introns within genes and noncoding DNA between genes

Concept 21.4: Multicellular eukaryotes have much noncoding DNA and many multigene families

- The bulk of most eukaryotic genomes consists of noncoding DNA sequences, often described in the past as "junk DNA"
- Sequencing of the human genome reveals that 98.5% does not code for proteins, rRNAs, or tRNAs
- Much evidence indicates that noncoding DNA plays important roles in the cell
  - For example, genomes of humans, rats, and mice show high sequence conservation for about **500 noncoding regions**

Fig. 21-7

Types of DNA sequences in the human genome



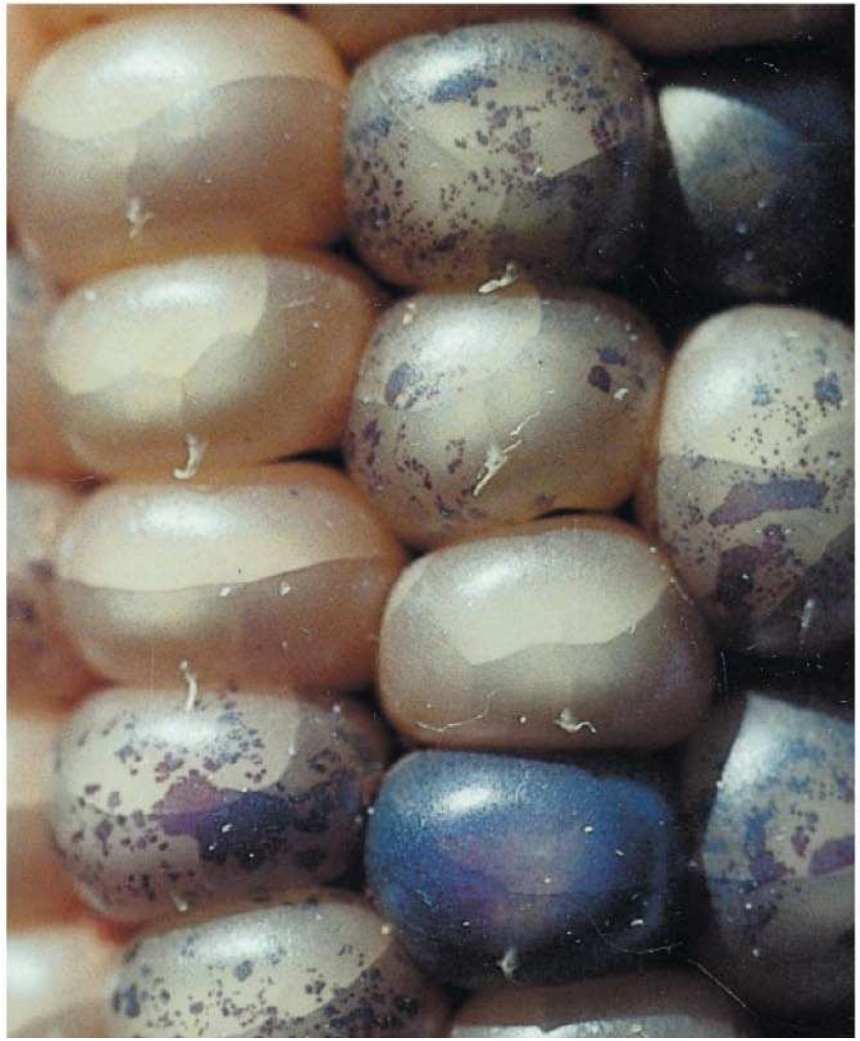
Non-coding region in the genome

- About **24%** of the human genome codes for introns and gene-related regulatory sequences
- Intergenic DNA is noncoding DNA found between genes
  - **Pseudogenes** are former genes that have accumulated mutations and are nonfunctional
  - **Repetitive DNA** is present in multiple copies in the genome
- About three-fourths of repetitive DNA is made up of transposable elements and sequences related to them

Transposable Elements and Related Sequences

- The first evidence for wandering DNA segments came from geneticist Barbara McClintock's breeding experiments with Indian corn
- McClintock identified changes in the color of corn kernels that made sense only by postulating that some genetic elements move from other genome locations into the genes for kernel color
- These **transposable elements** move from one site to another in a cell's DNA; they are present in both prokaryotes and eukaryotes

Fig. 21-8



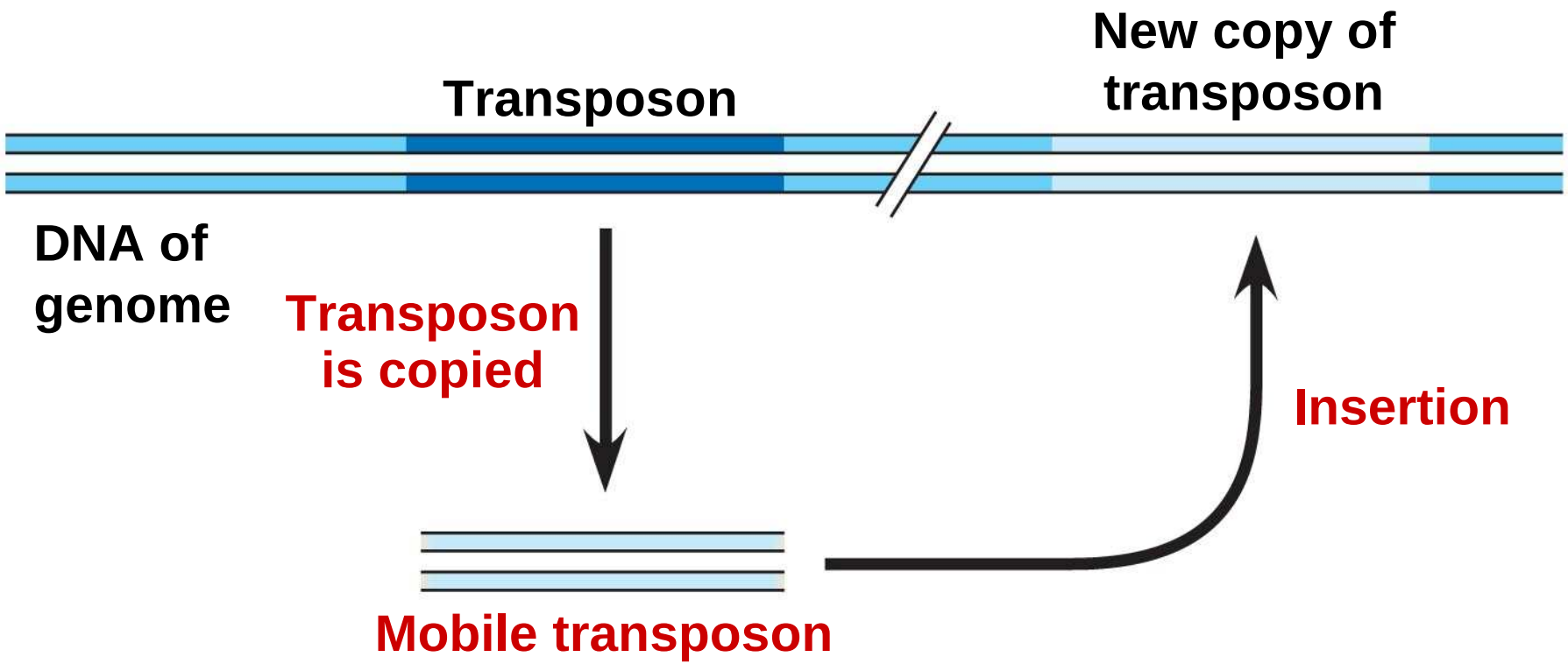
McClintock identified changes in the color of corn kernels are caused by transportable DNA elements.

Movement of Transposons and Retrotransposons

- Eukaryotic transposable elements are of two types:
  - **Transposons (轉位子)** , which move within a genome by means of a DNA intermediate
  - **Retrotransposons**, which move by means of a RNA intermediate

Fig. 21-9a

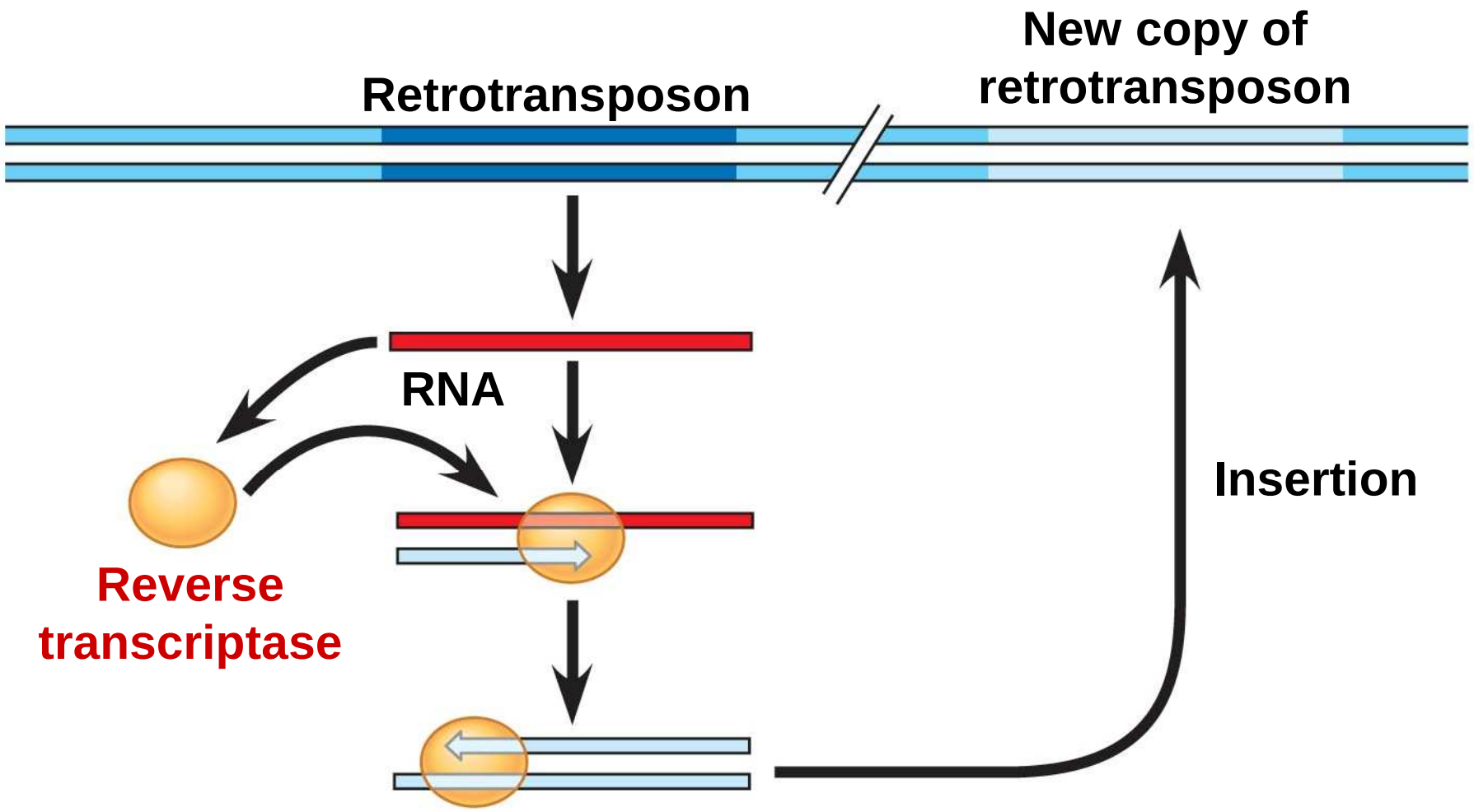
Movement of eukaryotic transposable elements



(a) Transposon movement ("copy-and-paste" mechanism)

Fig. 21-9b

Movement of eukaryotic transposable elements



(b) Retrotransposon movement

Sequences Related to Transposable Elements

- Multiple copies of transposable elements and related sequences are scattered throughout the eukaryotic genome
- In primates, a large portion of transposable element–related DNA consists of a family of similar sequences called ***Alu elements***
- Many *Alu* elements are transcribed into RNA molecules; however, their function is unknown

- The human genome also contains many sequences of a type of retrotransposon called *LINE-1* (*L1*)
- L1 sequences have a low rate of transposition and may help regulate gene expression

Other Repetitive DNA, Including Simple Sequence DNA

- About **15%** of the human genome consists of duplication of long sequences of DNA from one location to another
- In contrast, **simple sequence DNA** contains many copies of tandemly repeated short sequences

$$---(GTTAC)_n---$$

Short tandem repeat

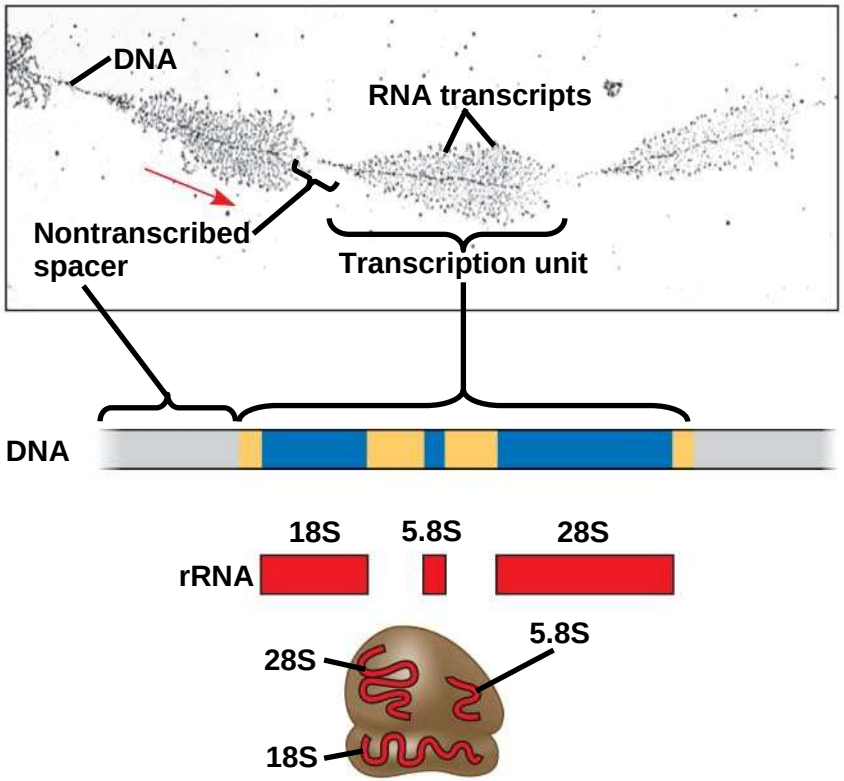
- A series of repeating units of 2 to 5 (or 2~16) nucleotides is called a **short tandem repeat** (**STR**)
  - The repeat number for STRs can vary among sites (within a genome) or individuals
  - Simple sequence DNA is common in centromeres and telomeres, where it probably plays structural roles in the chromosome

Genes and Multi-gene Families

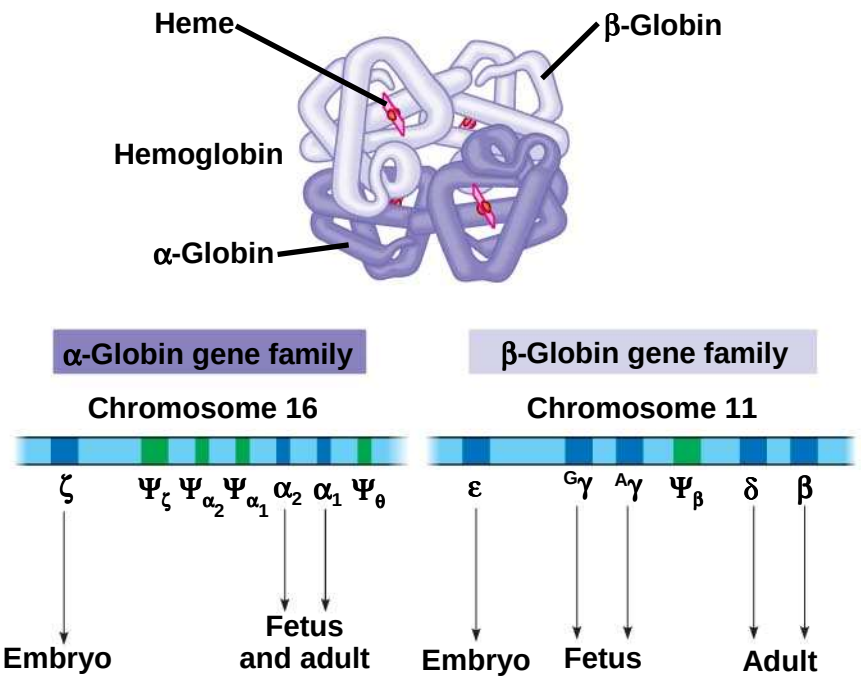
- Many eukaryotic genes are present in one copy per haploid set of chromosomes
- The rest of the genome occurs in **multigene families**, collections of identical or very similar genes
- Some multigene families consist of identical DNA sequences, usually clustered tandemly, such as those that code for RNA products

Fig. 21-10

Gene families

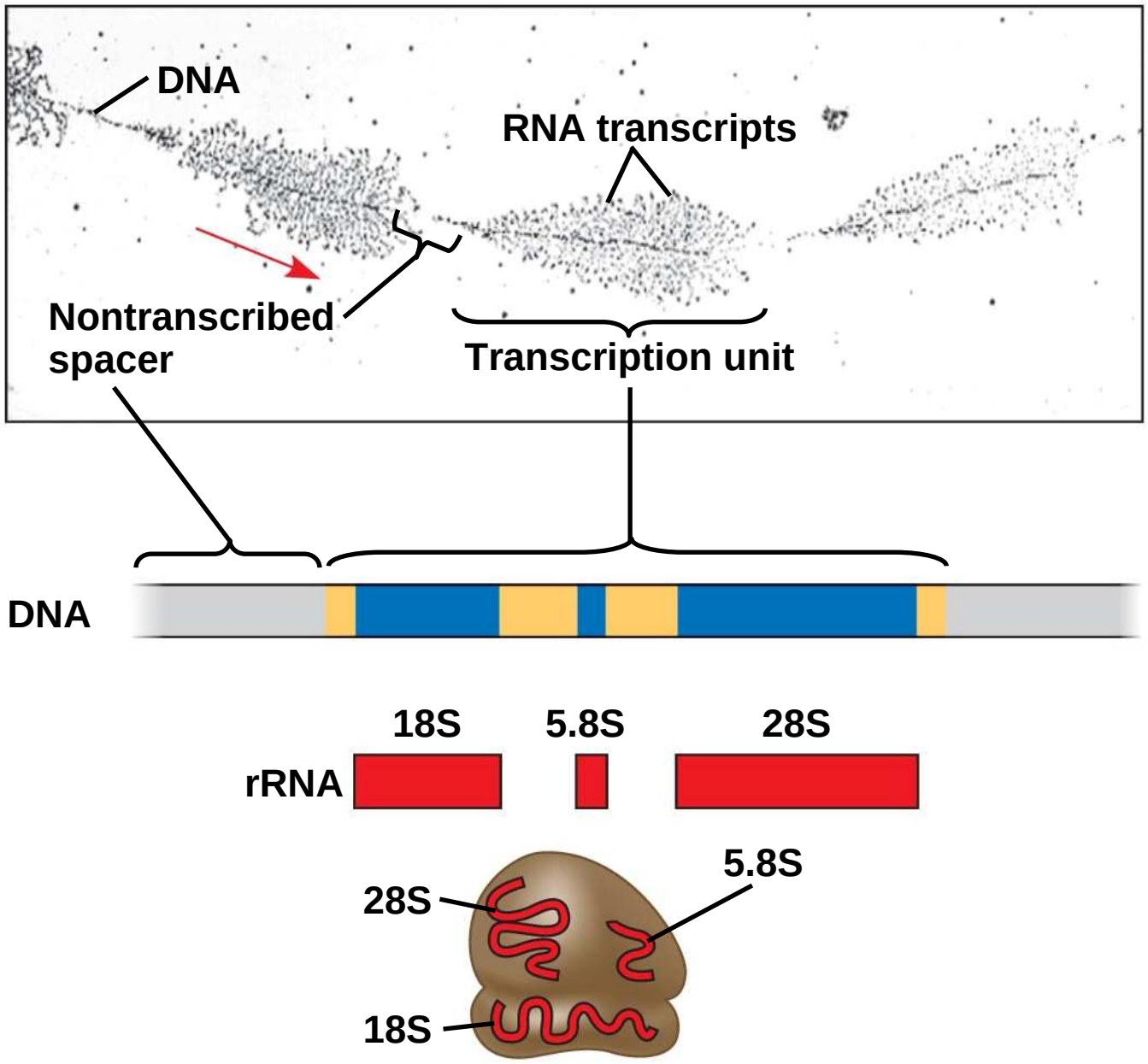


(a) Part of the ribosomal RNA gene family



(b) The human α-globin and β-globin gene families

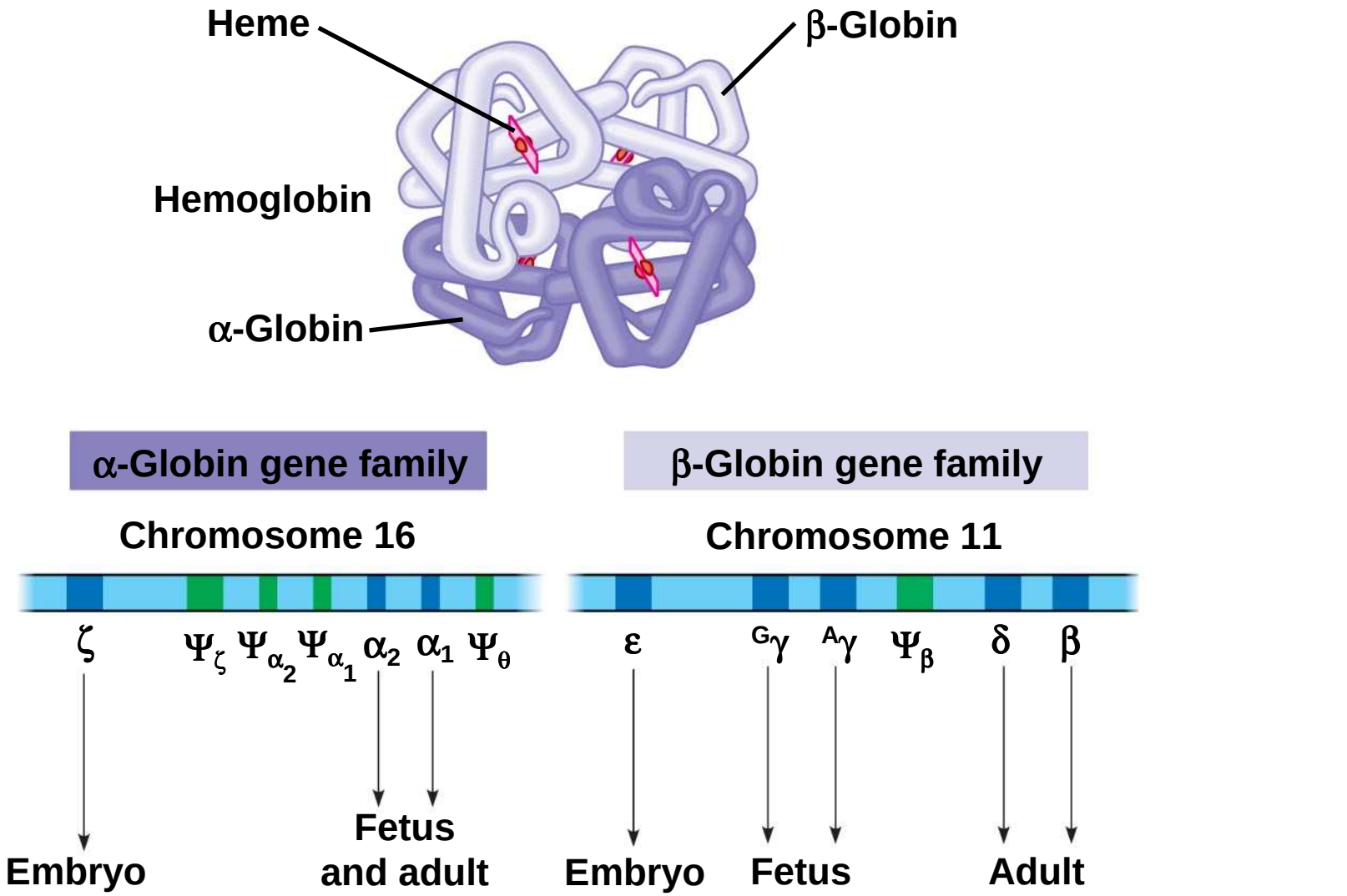
Fig. 21-10a



(a) Part of the ribosomal RNA gene family

- The classic examples of multigene families of nonidentical genes are two related families of genes that encode globins
- α-globins and β-globins are polypeptides of hemoglobin and are coded by genes on different human chromosomes

Fig. 21-10b



(b) The human α-globin and β-globin gene families

Concept 21.5: Duplication, rearrangement, and mutation of DNA contribute to genome evolution

- The basis of change at the genomic level is mutation, which underlies much of genome evolution
- The earliest forms of life likely had a minimal number of genes, including only those necessary for survival and reproduction
- The size of genomes has increased over evolutionary time, with the extra genetic material providing raw material for gene diversification

Duplication of Entire Chromosome Sets

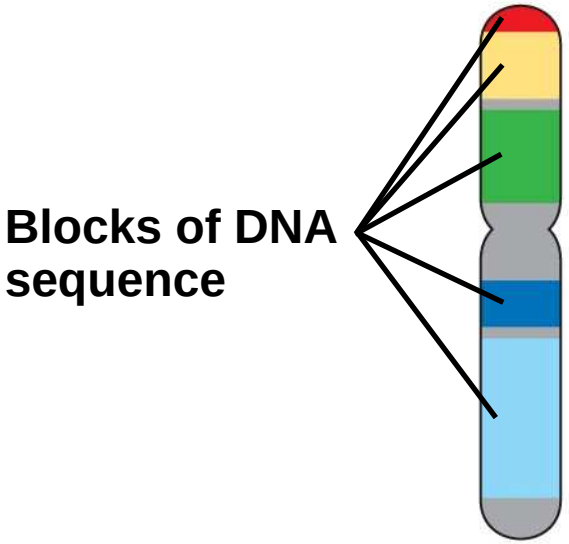
- Accidents in meiosis can lead to one or more extra sets of chromosomes, a condition known as polyploidy
- The genes in one or more of the extra sets can diverge by accumulating mutations; these variations may persist if the organism carrying them survives and reproduces

Alterations of Chromosome Structure

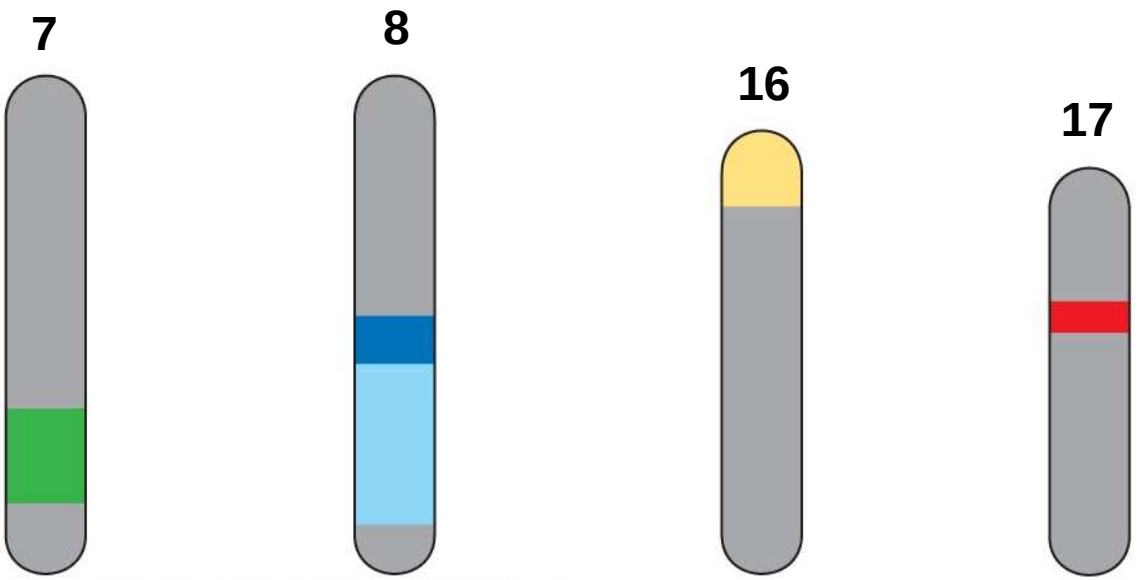
- Humans have 23 pairs of chromosomes, while chimpanzees have 24 pairs
- Following the divergence of humans and chimpanzees from a common ancestor, two ancestral chromosomes fused in the human line
- Duplications and inversions result from mistakes during meiotic recombination
- Comparative analysis between chromosomes of humans and 7 mammalian species paints a hypothetical chromosomal evolutionary history

Fig. 21-11

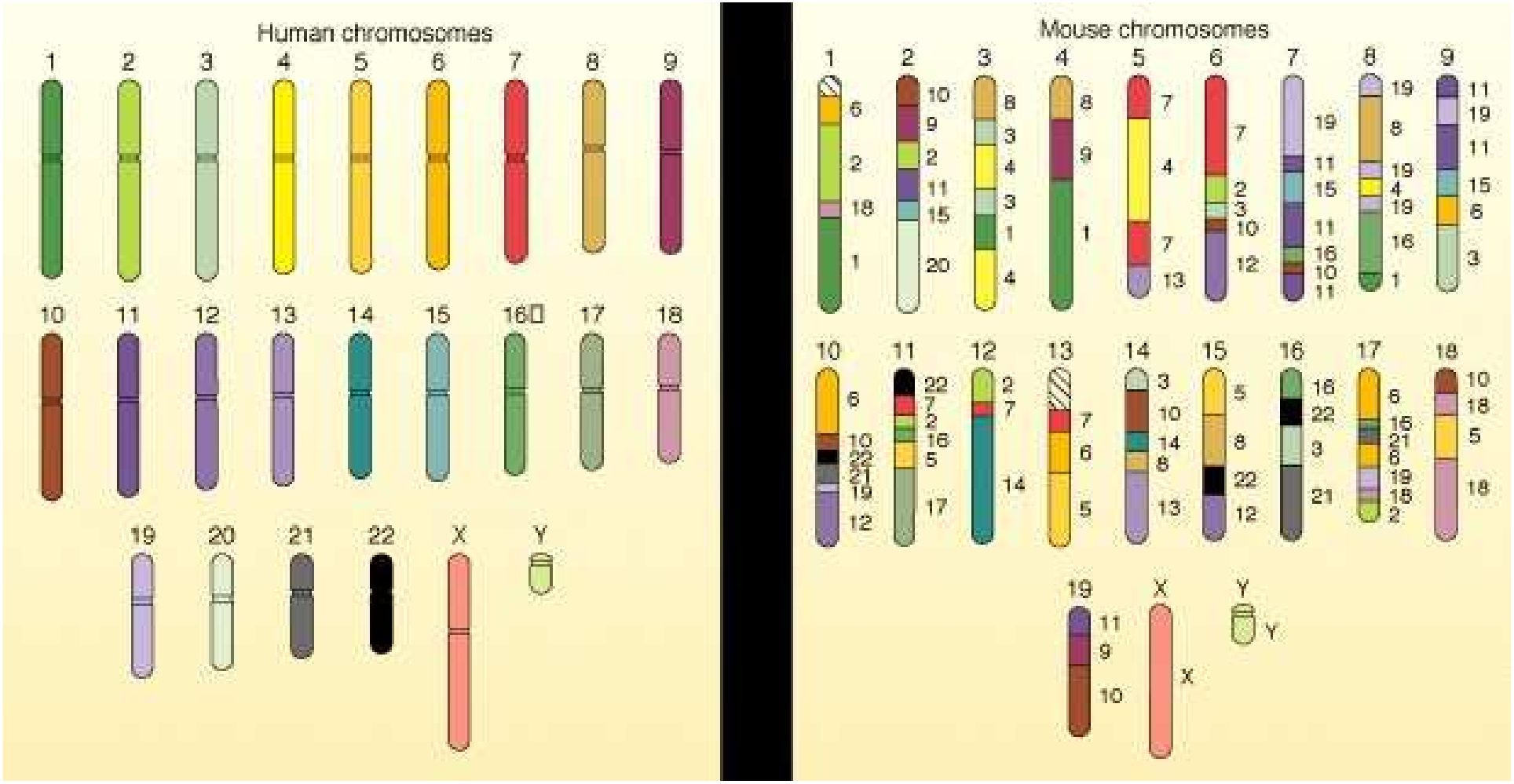
Human chromosome 16



Blocks of similar sequences in four mouse chromosomes:



Human vs. Mouse genome



http://fig.cox.miami.edu/Faculty/Dana/synteny.jpg

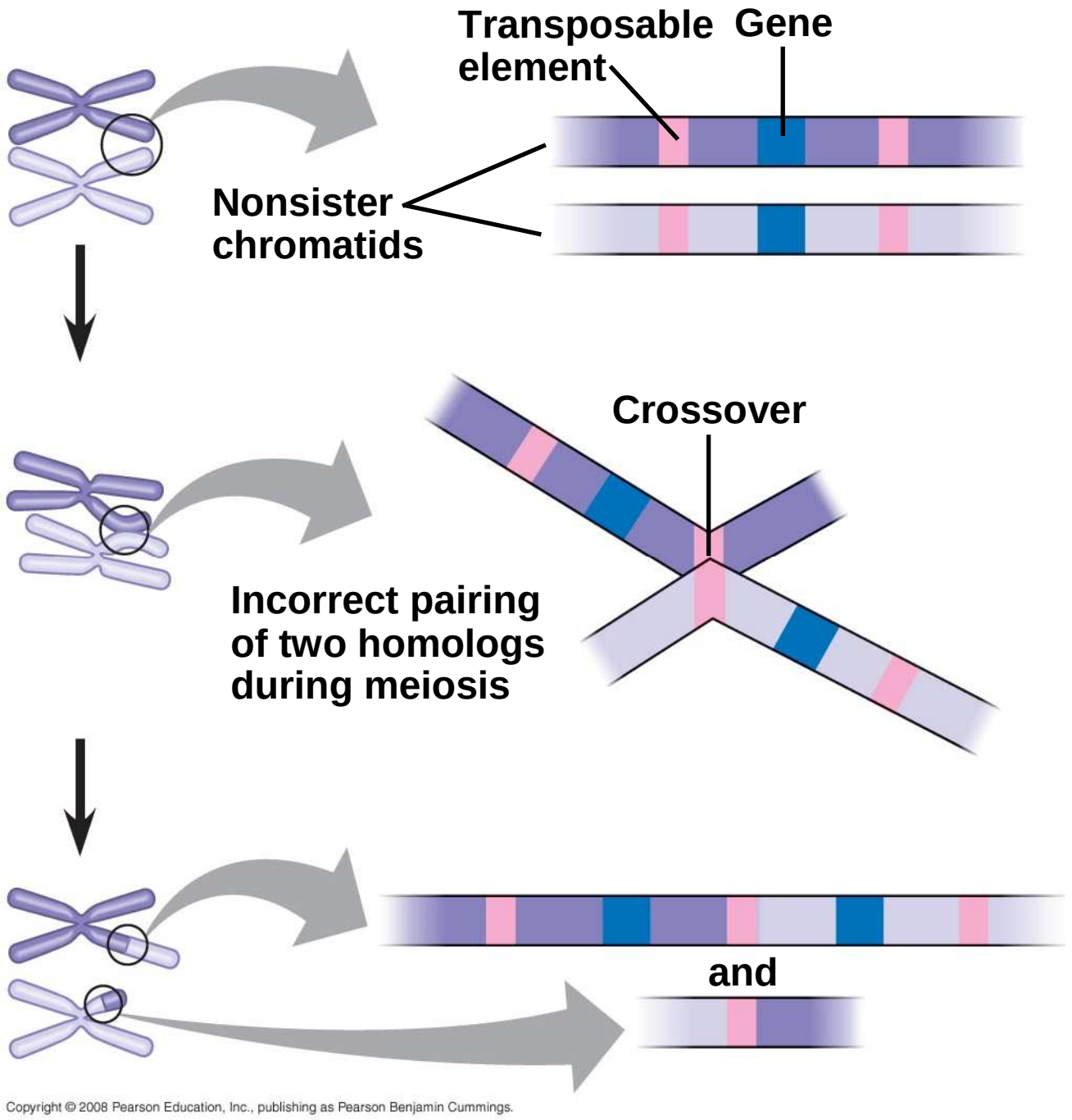
Chromosomal rearrangement

- The rate of duplications and inversions seems to have accelerated about 100 million years ago
- This coincides with when large dinosaurs went extinct and mammals diversified
- Chromosomal rearrangements are thought to contribute to the generation of new species
- Some of the recombination "hot spots" associated with chromosomal rearrangement are also locations that are associated with diseases

Duplication and Divergence of Gene-Sized Regions of DNA

- Unequal crossing over during prophase I of meiosis can result in one chromosome with a deletion and another with a duplication of a particular region
- Transposable elements can provide sites for crossover between nonsister chromatids

Fig. 21-12



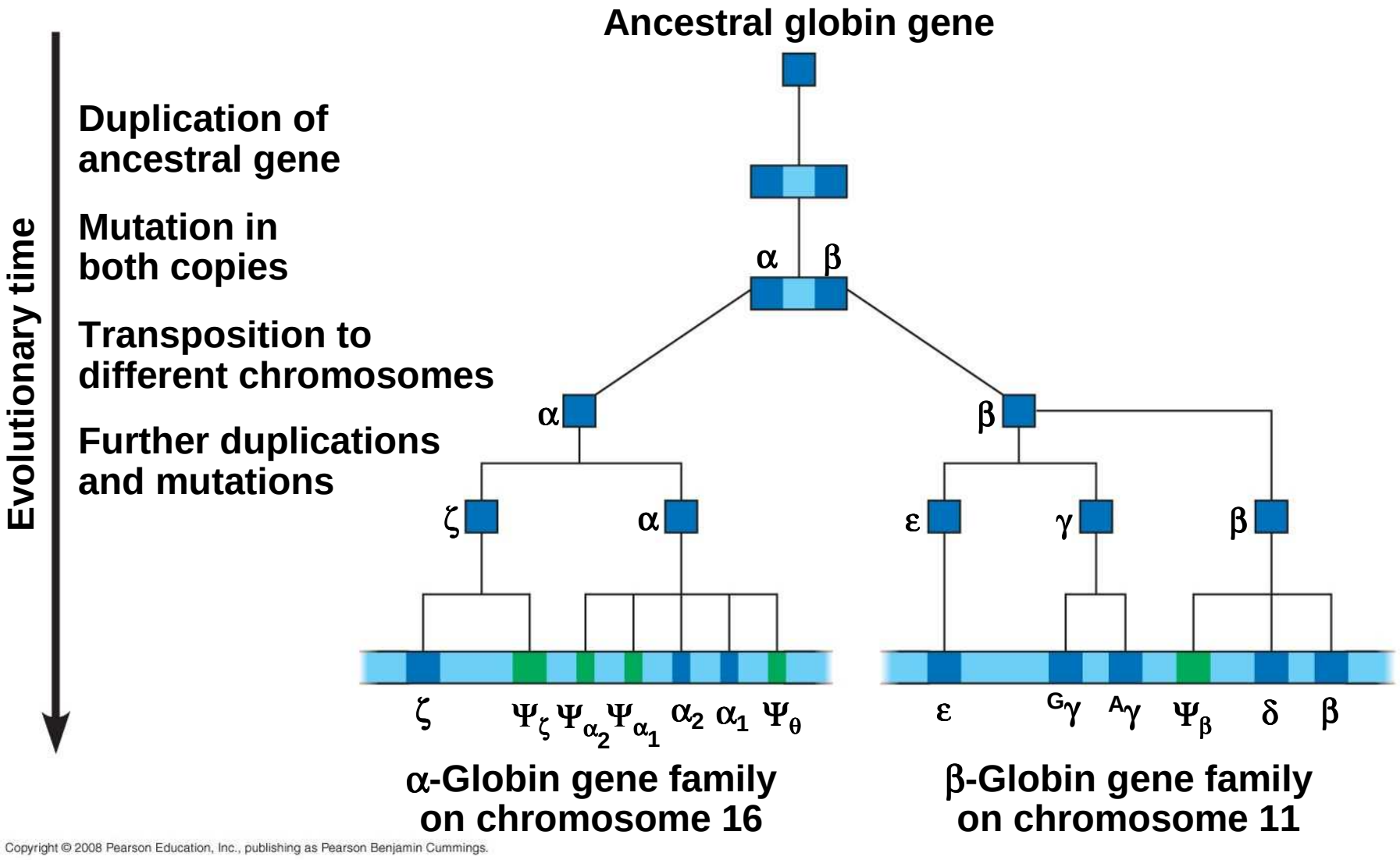
Gene duplication due to unequal crossing over

Evolution of Genes with Related Functions: The Human Globin Genes

- The genes encoding the various globin proteins evolved from one common ancestral globin gene, which duplicated and diverged about 450–500 million years ago
- After the duplication events, differences between the genes in the globin family arose from the accumulation of mutations

Fig. 21-13

A model for the evolution of the human α-globin and β-globin gene families from a single ancestral globin gene



- Subsequent duplications of these genes and random mutations gave rise to the present globin genes, which code for oxygen-binding proteins
- The similarity in the amino acid sequences of the various globin proteins supports this model of gene duplication and mutation

Table 21-2

Table 21.2 Percentage of Similarity in Amino Acid Sequence Between Human Globin Proteins

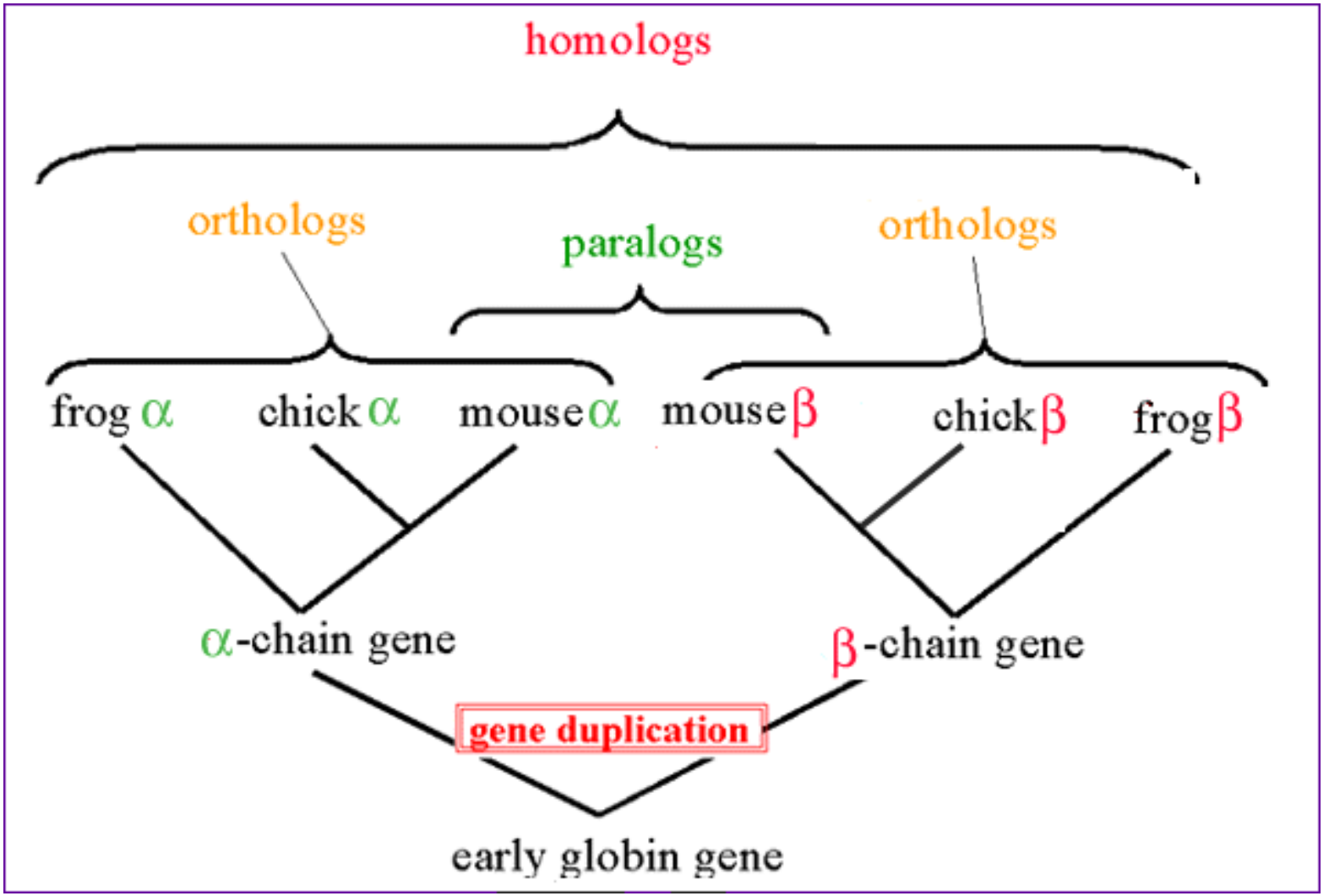
		α-Globins		β-Globins		
		α	ζ	β	γ	ϵ
α-Globins	α	100	58	42	39	37
	ζ	58	100	34	38	37
β-Globins	β	42	34	100	73	75
	γ	39	38	73	100	80
	ϵ	37	37	75	80	100

Greek alphabet

Α α	alpha	Ν ν	nu
Β β	beta	Ξ ξ	ksi
Γ γ	gamma	Ο ο	omicron
Δ δ	delta	Π π	pi
Ε ε	epsilon	Ρ ρ	rho
Ζ ζ	zeta	Σ σς	sigma
Η η	eta	Τ τ	tau
Θ θ	theta	Υ υ	upsilon
Ι ι	iota	Φ φ	phi
Κ κ	kappa	Χ χ	chi
Λ λ	lambda	Ψ ψ	psi
Μ μ	mu	Ω ω	omega

http://en.wikipedia.org/wiki/Alpha_(letter)

Homolog, Ortholog, Paralog
同源基因,直向同源基因,横向同源基因



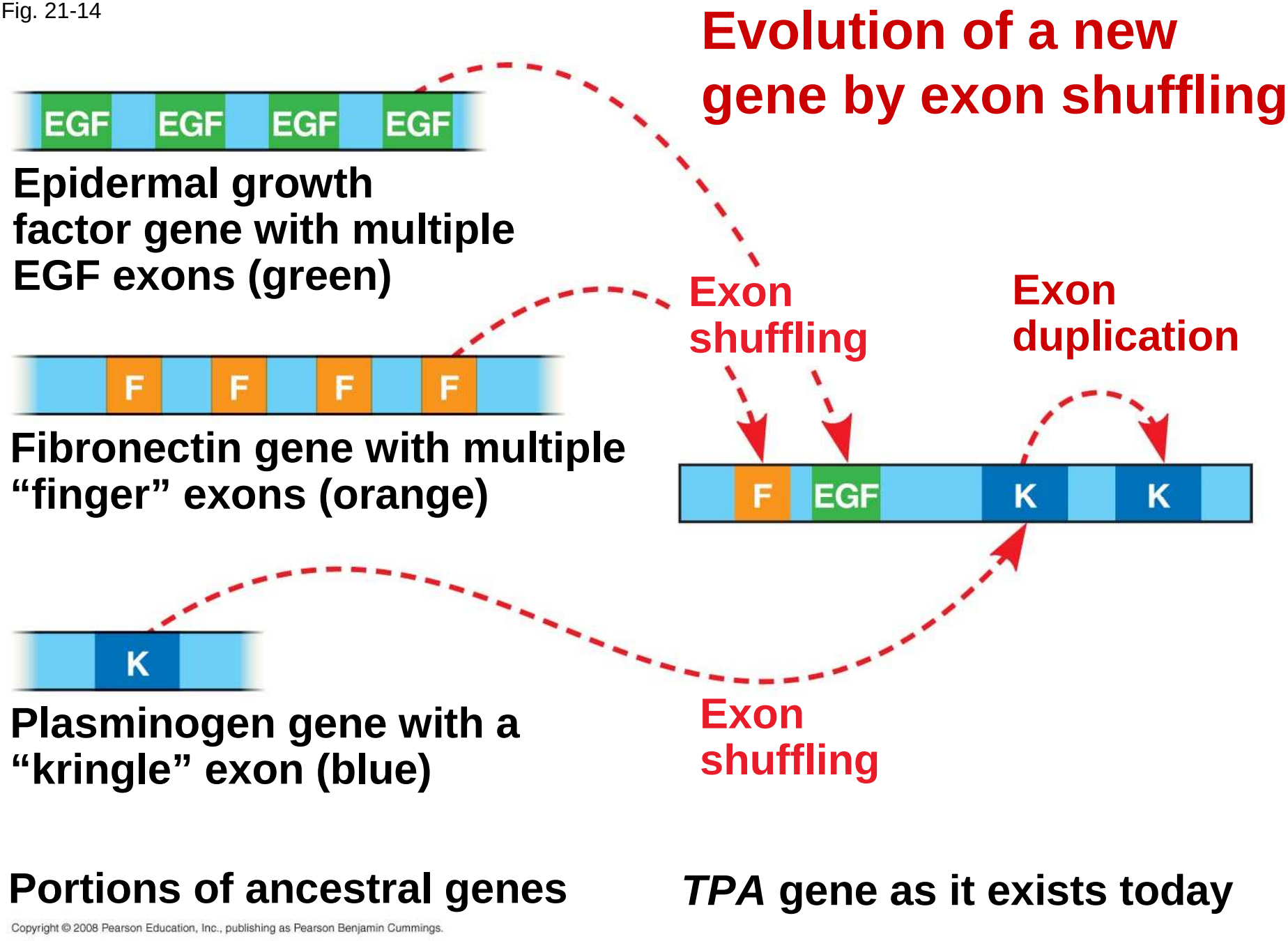
Evolution of Genes with Novel Functions

- The copies of some duplicated genes have diverged so much in evolution that the functions of their encoded proteins are now very different
- For example the lysozyme gene was duplicated and evolved into the α-lactalbumin gene in mammals
  - Lysozyme is an enzyme that helps protect animals against bacterial infection
  - α-lactalbumin is a nonenzymatic protein that plays a role in milk production in mammals

Rearrangements of Parts of Genes: Exon Duplication and Exon Shuffling

- The duplication or repositioning of exons has contributed to genome evolution
- Errors in meiosis can result in an exon being duplicated on one chromosome and deleted from the homologous chromosome
- In exon shuffling, errors in meiotic recombination lead to some mixing and matching of exons, either within a gene or between two nonallelic genes

Fig. 21-14



How Transposable Elements Contribute to Genome Evolution

- Multiple copies of similar transposable elements may facilitate recombination, or crossing over, between different chromosomes
- Insertion of transposable elements within a protein-coding sequence may block protein production
- Insertion of transposable elements within a regulatory sequence may increase or decrease protein production

Bad moves vs. good moves

- Transposable elements may carry a gene or groups of genes to a new location
- Transposable elements may also create new sites for alternative splicing in an RNA transcript
- In all cases, changes are usually detrimental but may on occasion prove advantageous to an organism

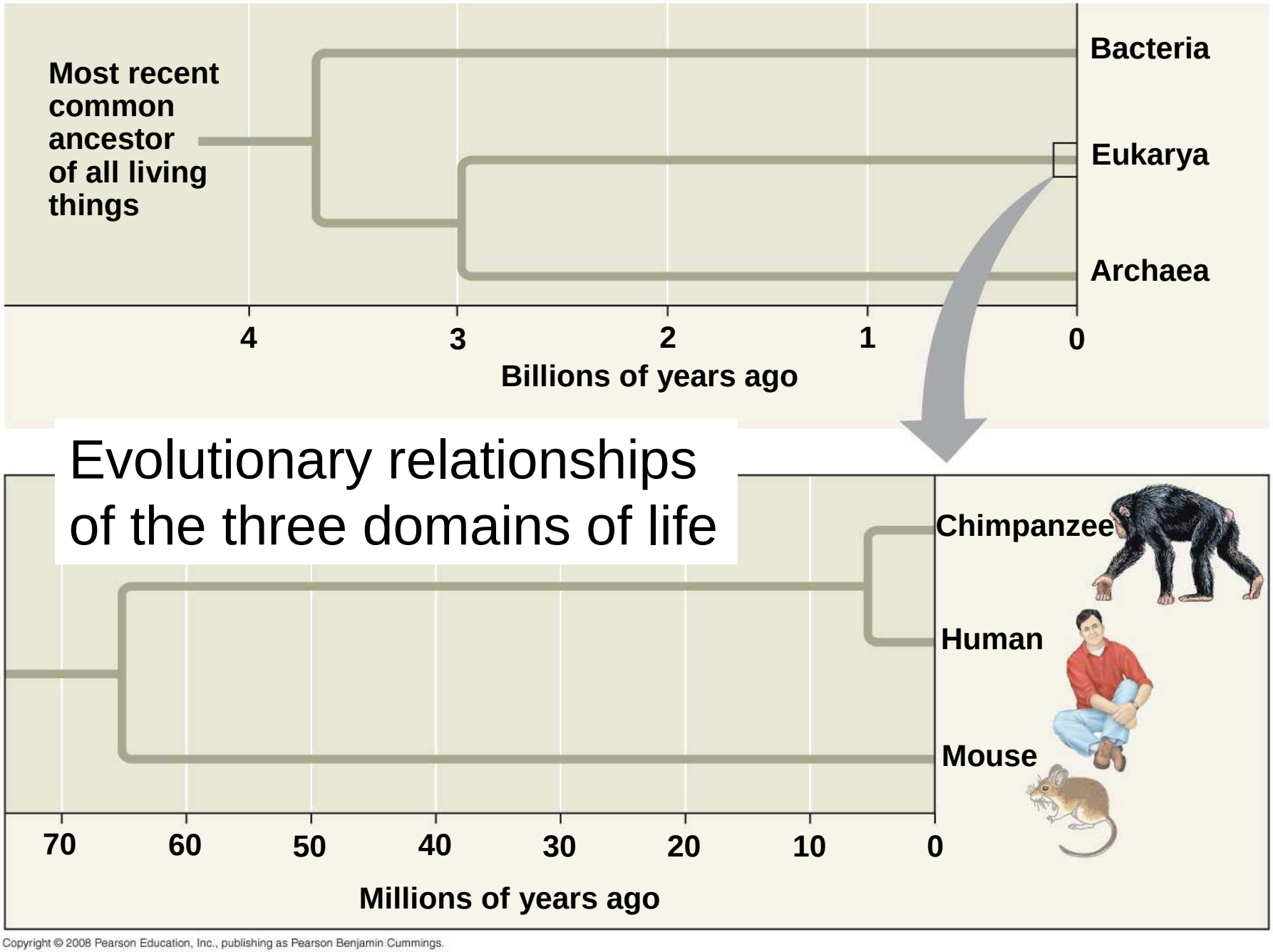
Concept 21.6: Comparing genome sequences provides clues to evolution and development

- Genome sequencing has advanced rapidly in the last 20 years
- Comparative studies of genomes
  - Advance our understanding of the evolutionary history of life
  - Help explain how the evolution of development leads to morphological diversity

Comparing Genomes

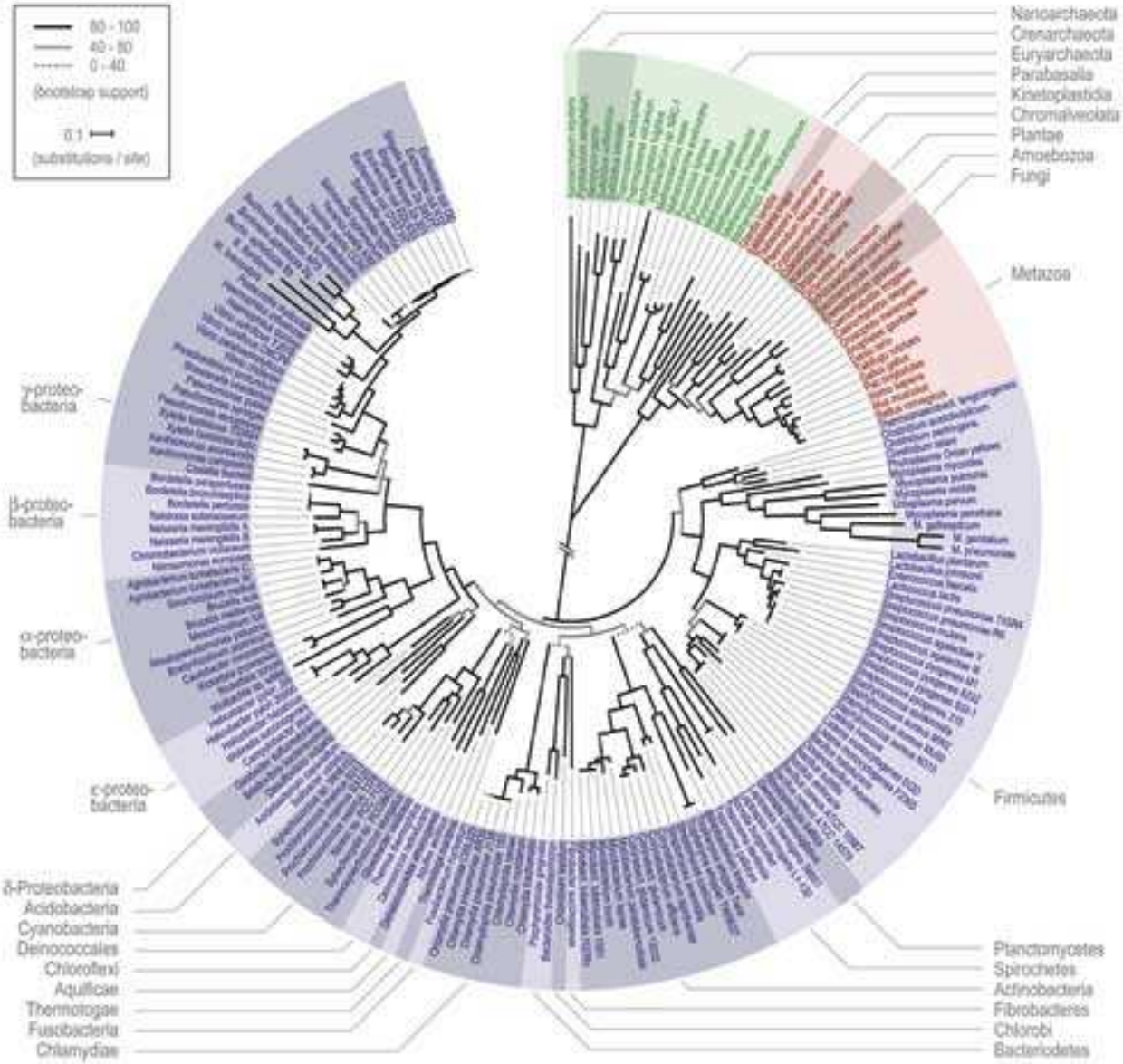
- Genome comparisons of closely related species help us understand recent evolutionary events
- Genome comparisons of distantly related species help us understand ancient evolutionary events
- Relationships among species can be represented by a tree-shaped diagram (Tree of Life)

Fig. 21-15



Evolutionary relationships of the three domains of life

Tree of Life : http://tolweb.org/tree/



Comparing Distantly Related Species

- Highly conserved genes are genes that have changed very little over time
  - These inform us about relationships among species that diverged from each other a long time ago
- Bacteria, archaea, and eukaryotes diverged from each other between 2 and 4 billion years ago
- Highly conserved genes can be studied in one model organism, and the results applied to other organisms

Comparing Closely Related Species

- Genetic differences between closely related species can be correlated with phenotypic differences
  - For example, genetic comparison of several mammals with nonmammals helps identify what it takes to make a mammal

- Human and chimpanzee genomes differ by 1.2%, at single base-pairs, and by 2.7% because of insertions and deletions
  - Several genes are evolving faster in humans than chimpanzees
  - These include genes involved in defense against malaria and tuberculosis, regulation of brain size, and genes that code for transcription factors

“Speech” gene – *FOXP2*

- Humans and chimpanzees differ in the expression of the *FOXP2* gene whose product turns on genes involved in vocalization
- Differences in the *FOXP2* gene may explain why humans but not chimpanzees communicate by speech

Fig. 21-16

What is the function of a gene (*FOXP2*) that is rapidly evolving in the human lineage?

EXPERIMENT

Wild type: two normal copies of *FOXP2*

Heterozygote: one copy of *FOXP2* disrupted

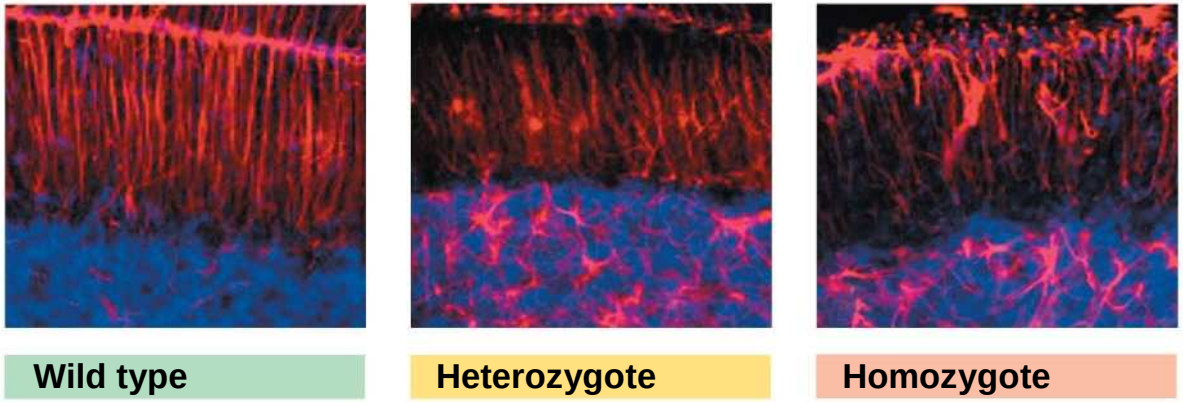
Homozygote: both copies of *FOXP2* disrupted

Experiment 1: Researchers cut thin sections of brain and stained them with reagents, allowing visualization of brain anatomy in a UV fluorescence microscope.

Experiment 2: Researchers separated each newborn pup from its mother and recorded the number of ultrasonic whistles produced by the pup.

RESULTS

Experiment 1



Wild type | **Heterozygote** | **Homozygote**

Experiment 2

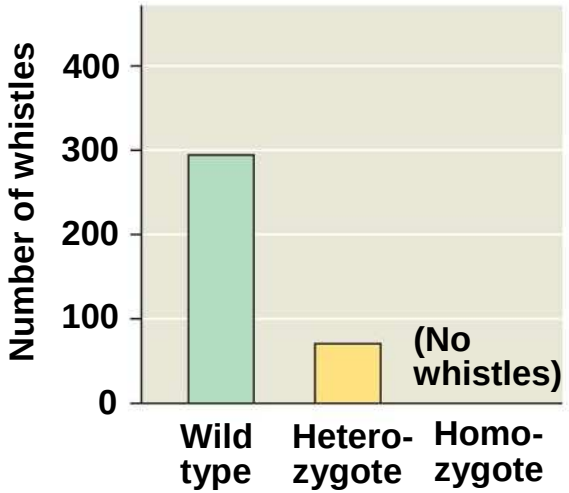


Fig. 21-16a

EXPERIMENT

Wild type: two normal copies of *FOXP2*

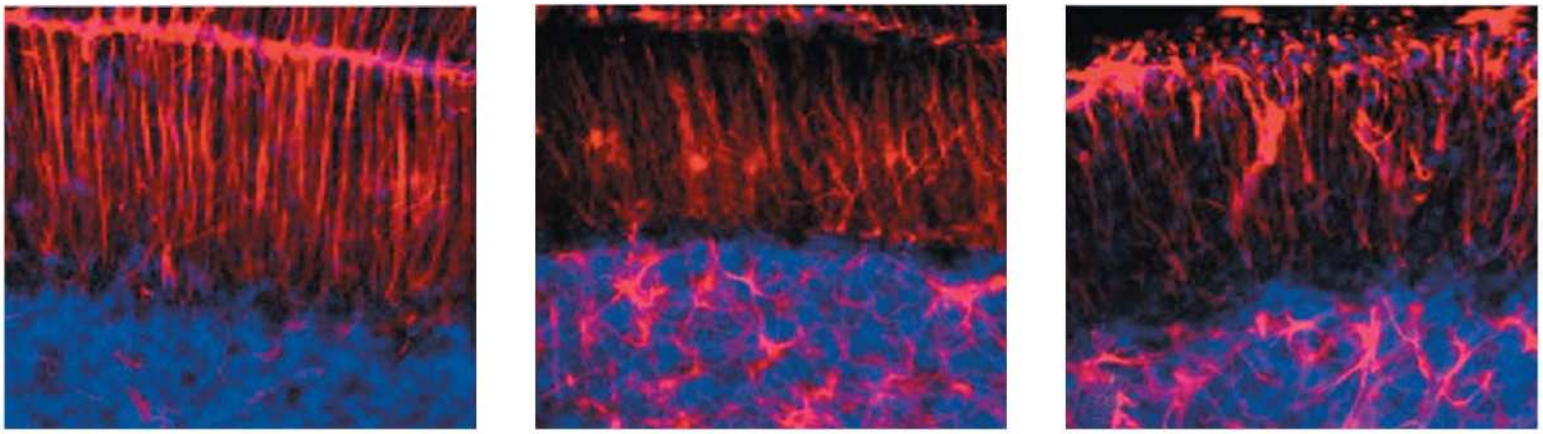
Heterozygote: one copy of *FOXP2* disrupted

Homozygote: both copies of *FOXP2* disrupted

Experiment 1: Researchers cut thin sections of brain and stained them with reagents, allowing visualization of brain anatomy in a UV fluorescence microscope.

RESULTS

Experiment 1



Wild type | **Heterozygote** | **Homozygote**

Fig. 21-16b

EXPERIMENT

Wild type: two normal copies of *FOXP2*

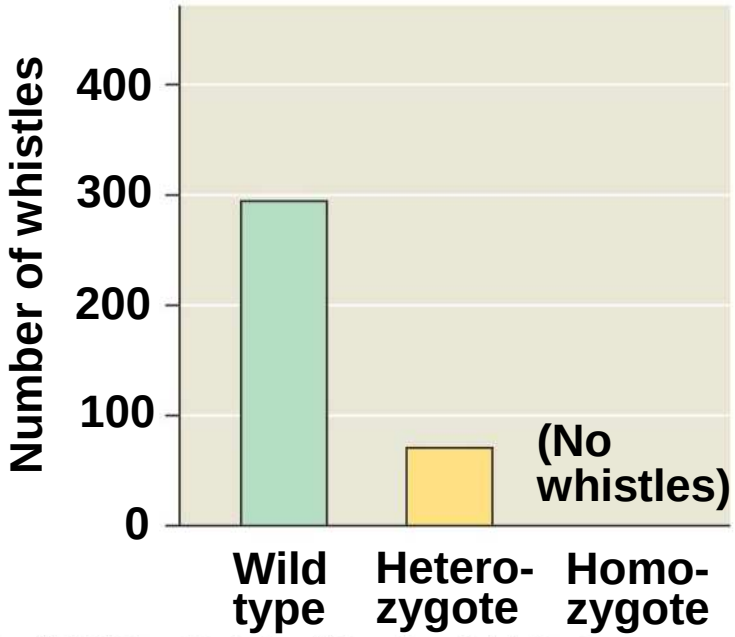
Heterozygote: one copy of *FOXP2* disrupted

Homozygote: both copies of *FOXP2* disrupted

Experiment 2: Researchers separated each newborn pup from its mother and recorded the number of ultrasonic whistles produced by the pup.

RESULTS

Experiment 2



What if:
Replace chimpanzee FOXP2 with human FOXP2?

Comparing Genomes Within a Species

- As a species, humans have only been around about 200,000 years and have low within-species genetic variation
- Variation within humans is due to single nucleotide polymorphisms, inversions, deletions, and duplications
- These variations are useful for studying human evolution and human health

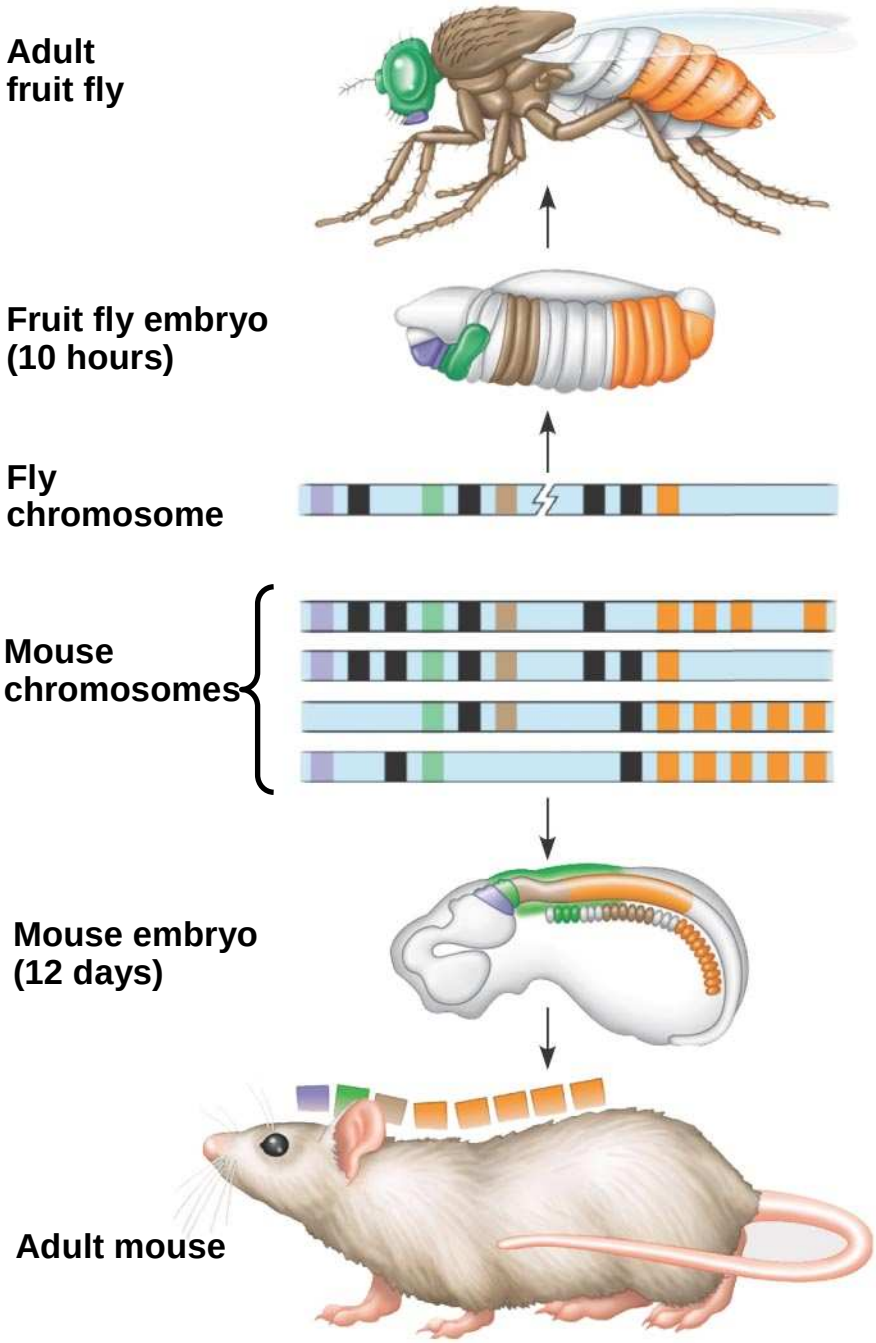
Comparing Developmental Processes

- Evolutionary developmental biology, or **evo-devo**, is the study of the evolution of developmental processes in multicellular organisms
- Genomic information shows that minor differences in gene sequence or regulation can result in major differences in form

Widespread Conservation of Developmental Genes Among Animals

- Molecular analysis of the homeotic genes (同源基因) in *Drosophila* has shown that they all include a sequence called a **homeobox**
- An identical or very similar nucleotide sequence has been discovered in the homeotic genes of both vertebrates and invertebrates
- Homeobox genes code for a domain that allows a protein to bind to DNA and to function as a transcription regulator
- Homeotic genes in animals are called *Hox* genes

Fig. 21-17



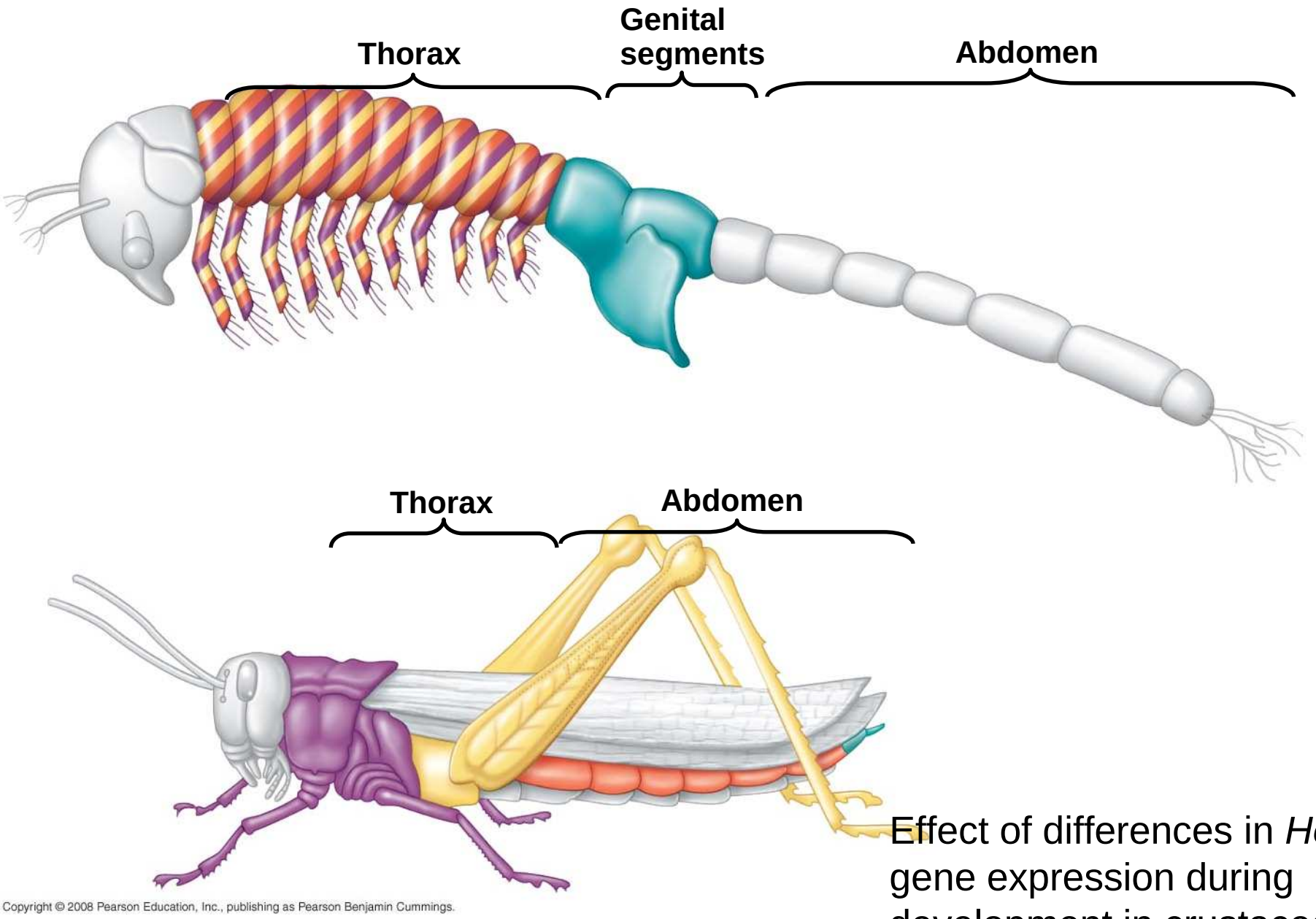
Conservation of homeotic genes in a fruit fly and a mouse

Hox genes, specify the anterior-posterior axis and segment identity during early development of metazoan (animal) organisms.

They are critical for the proper placement and number of embryonic segment structures (such as legs, antennae, and eyes).

- Related homeobox sequences have been found in regulatory genes of yeasts, plants, and even prokaryotes
- In addition to homeotic genes, many other developmental genes are highly conserved from species to species

Fig. 21-18



Effect of differences in *Hox* gene expression during development in crustaceans and insects

- Sometimes small changes in regulatory sequences of certain genes lead to major changes in body form
- For example, variation in *Hox* gene expression controls variation in leg-bearing segments of crustaceans and insects
- In other cases, genes with conserved sequences play different roles in different species

Comparison of Animal and Plant Development

- In both plants and animals, development relies on a cascade of transcriptional regulators turning genes on or off in a finely tuned series
- Molecular evidence supports the separate evolution of developmental programs in plants and animals
- ***Mads-box* genes** in plants are the regulatory equivalent of ***Hox* genes** in animals

You should now be able to:

1. Explain how linkage mapping, physical mapping, and DNA sequencing each contributed to the Human Genome Project
2. Define and compare the fields of proteomics and genomics
3. Describe the surprising findings of the Human Genome Project with respect to the size of the human genome
4. Distinguish between transposons and retrotransposons

5. Explain how polyploidy may facilitate gene evolution

6. Describe in general terms the events that may have led to evolution of the globin superfamily

7. Explain the significance of the rapid evolution of the *FOXP2* gene in the human lineage

8. Provide evidence that suggests that the homeobox DNA sequence evolved very early in the history of life

補充資料： Next Generation Sequencing (NGS)

	454 GS FLX*	AB SOLiD	Illumina GAII
Chemistry	Pyrosequencing	Ligation based	Reversible terminators
	Standard	**Fragment**	**Fragment**
Run Time	7 hours	3-6.5 days	3 days
Read Lengths (bp)	250	25, 50	35, 50
Ave. Reads per Run	400K	$150x10^6$	$85x10^6$
Data per run	100MB	up to 7GB	up to 4.3GB
Throughput	100MB	1.1GB/day	1.4GB/day
	Titanium	**Mate-Paired**	**Mate-Paired**
Run Time	10 hours	7-13 days	6.5 days
Read Lengths (bp)	400+	2x25, 2x35	2x50
Ave. Reads per Run	$1x10^6$	$250x10^6$	$90x10^6$ pairs
Data per run	400MB	up to 8.75GB	9Gb
Throughput	400MB	900MB/day	1.3GB/day

*Metrics apply to both Fragment and Mate-Paired runs.