Lecture 16
Part 2

Chapter 22. Phylogenetic Reconstruction

Chen Siang “Gene” Ng

Institute of Molecular and Cellular Biology

Department of Life Science

gcsng@life.nthu.edu.tw

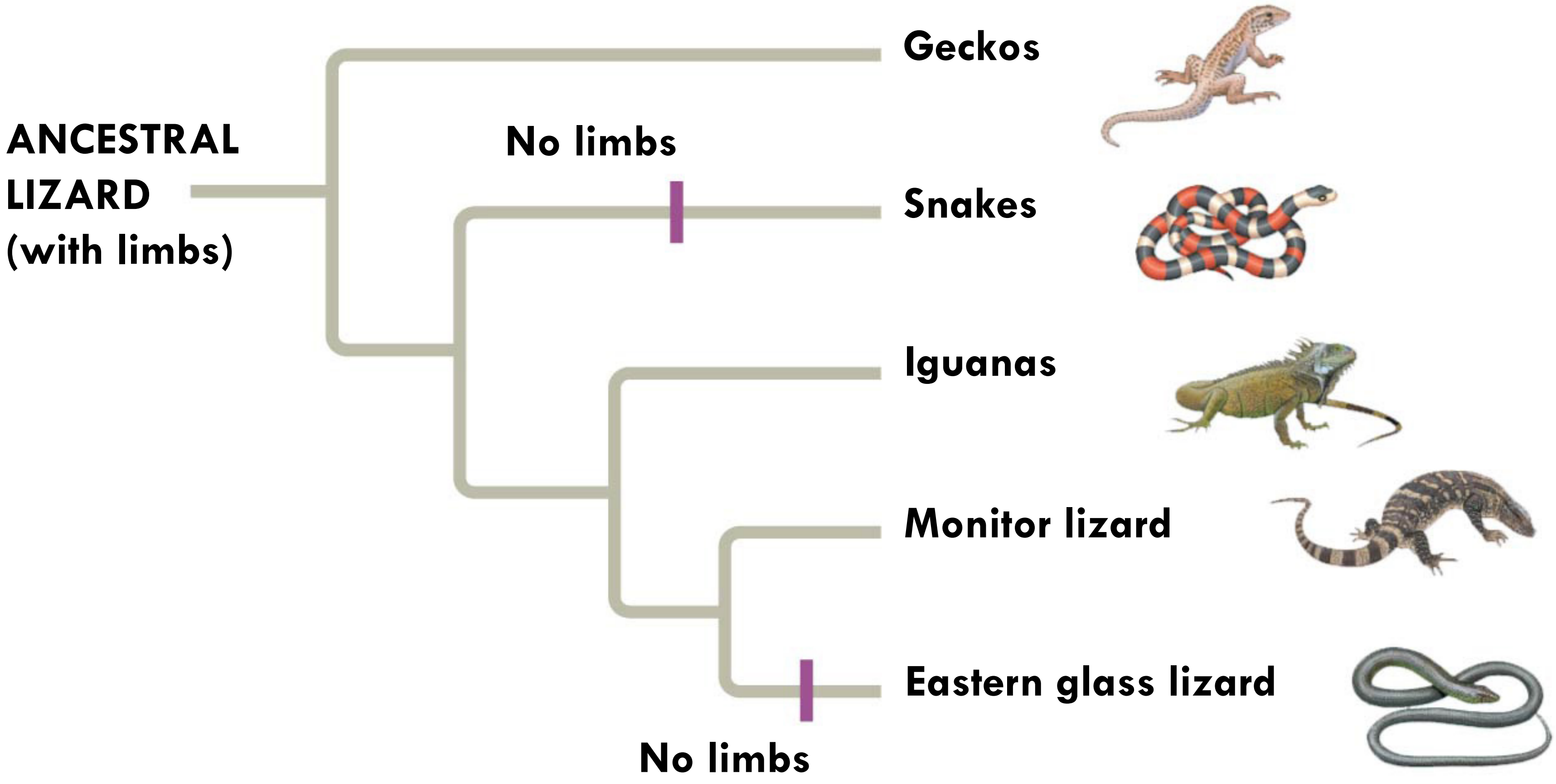
12/28/2017

Investigating the Tree of Life

- Legless lizards have evolved independently in several different groups

- **Phylogeny (譜系)** is the evolutionary history of a species or group of related species
  - For example, a phylogeny shows that legless lizards and snakes evolved from different lineages of legged lizards
- The discipline of **systematics (系統分類學)** classifies organisms and determines their evolutionary relationships

Figure 22.2


ANCESTRAL
LIZARD
(with limbs)
No limbs
Geckos
Snakes
Iguanas
Monitor lizard
Eastern glass lizard
No limbs

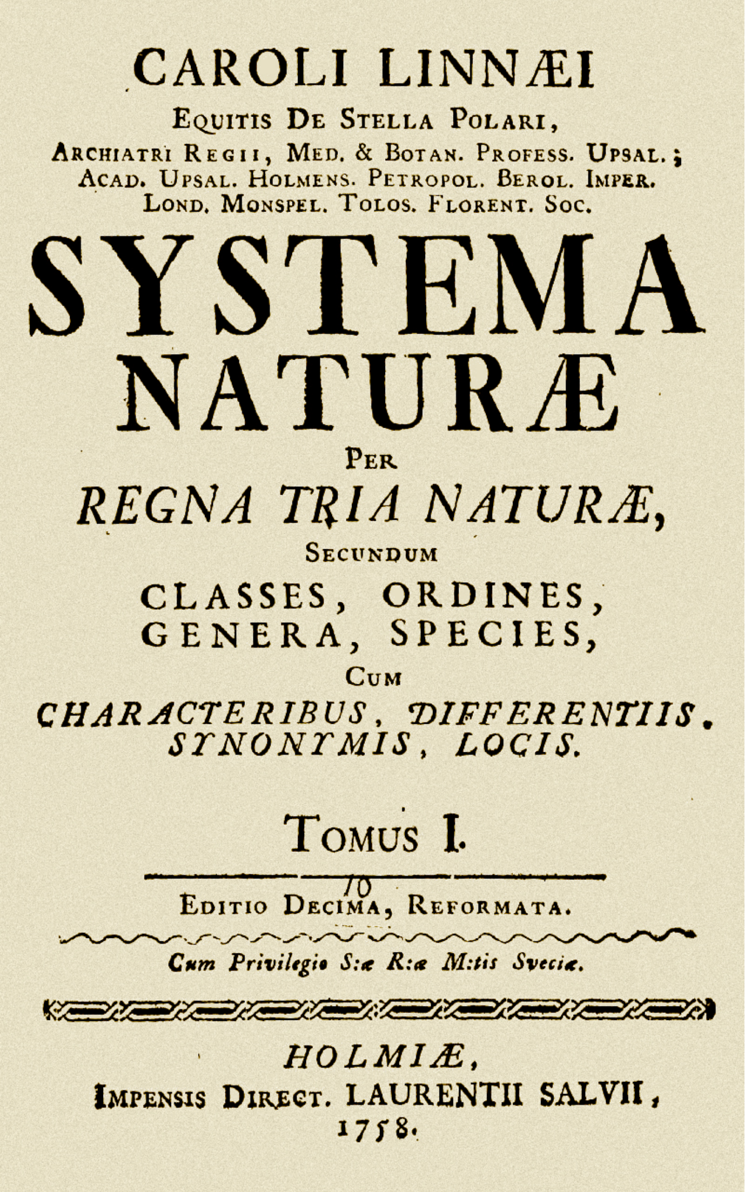
Concept 22.1: Phylogenies show evolutionary relationships

- **Taxonomy (分類學)** is the scientific discipline concerned with classifying and naming organisms

Binomial Nomenclature

- In the 18th century, Carolus Linnaeus (1707–1778) published a system of taxonomy based on resemblances

- Two key features of his system remain useful today: two-part names for species and hierarchical classification



CAROLI LINNÆI
Equitis De Stella Polari,
Archiatri Regii, Med. & Botan. Profess. Upsal.;
Acad. Upsal. Holmens. Petropol. Berol. Imper.
Lond. Monspel. Tolos. Florent. Soc.

SYSTEMA
NATURÆ
Per
REGNA TRIA NATURÆ,
Secundum
CLASSES, ORDINES,
GENERA, SPECIES,
Cum
CHARACTERIBUS, DIFFERENTIIS.
SYNONYMIS, LOCIS.

Tomus I.

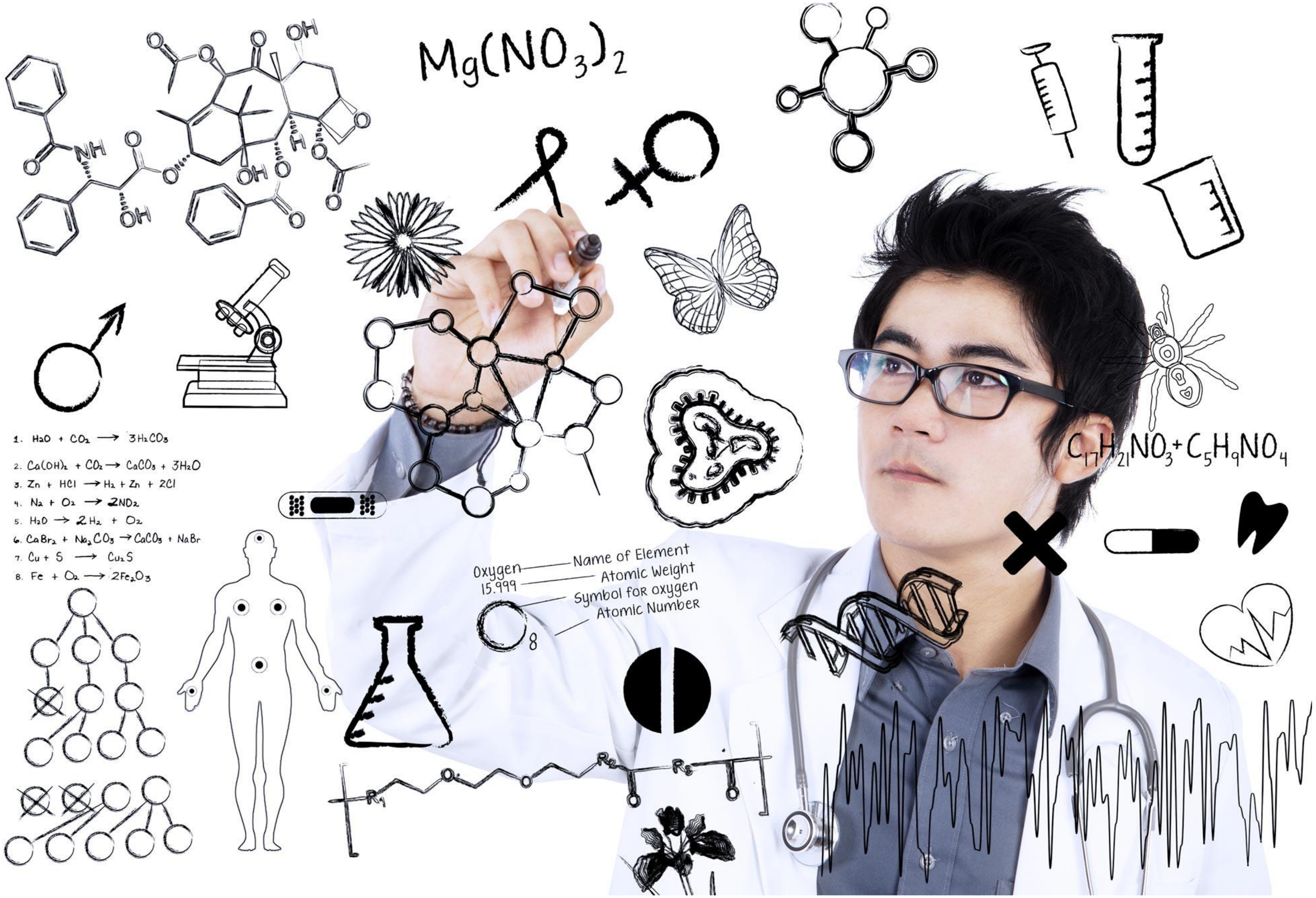
10
Editio Decima, Reformata.

Cum Privilegio S:æ R:æ M:tis Sveciæ.

HOLMIÆ,
Impensis Direct. LAURENTII SALVII,
1758.

- The two-part scientific name of a species is called a **binomial**
- The first part of the name is the **genus**
- The second part, called the specific epithet, is unique for each species within the genus
- The first letter of the genus is capitalized, and the entire species name is italicized
- Both parts together name the species (not the specific epithet alone)

Be Professional!!!



Hierarchical Classification

- Linnaeus introduced a system for grouping species in increasingly inclusive categories

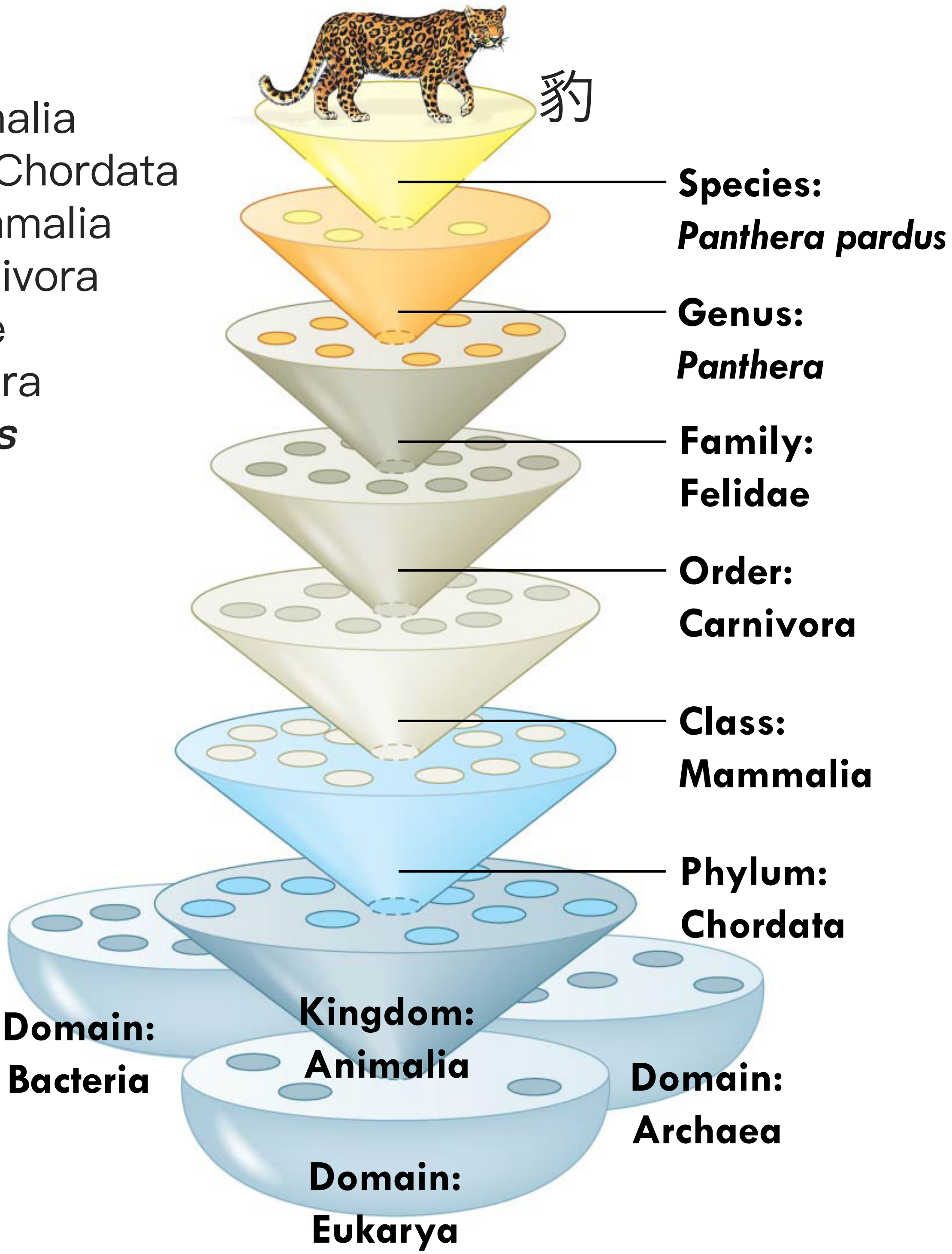
- The taxonomic groups from broad to narrow are **domain, kingdom, phylum, class, order, family**, genus, and species **域、界、門、綱、目、科、屬、種**

- A taxonomic unit at any level of hierarchy is called a **taxon (分類群)**

- The broader taxa are not comparable between lineages
  - For example, an order of snails has less genetic diversity than an order of mammals

Figure 22.3

界：動物界 Animalia
門：脊索動物門 Chordata
綱：哺乳綱 Mammalia
目：食肉目 Carnivora
科：貓科 Felidae
屬：豹屬 Panthera
種：豹 *P. pardus*



Linking Classification and Phylogeny

- The evolutionary history of a group of organisms can be represented in a branching **phylogenetic tree**

Phylogenetic Tree of Life

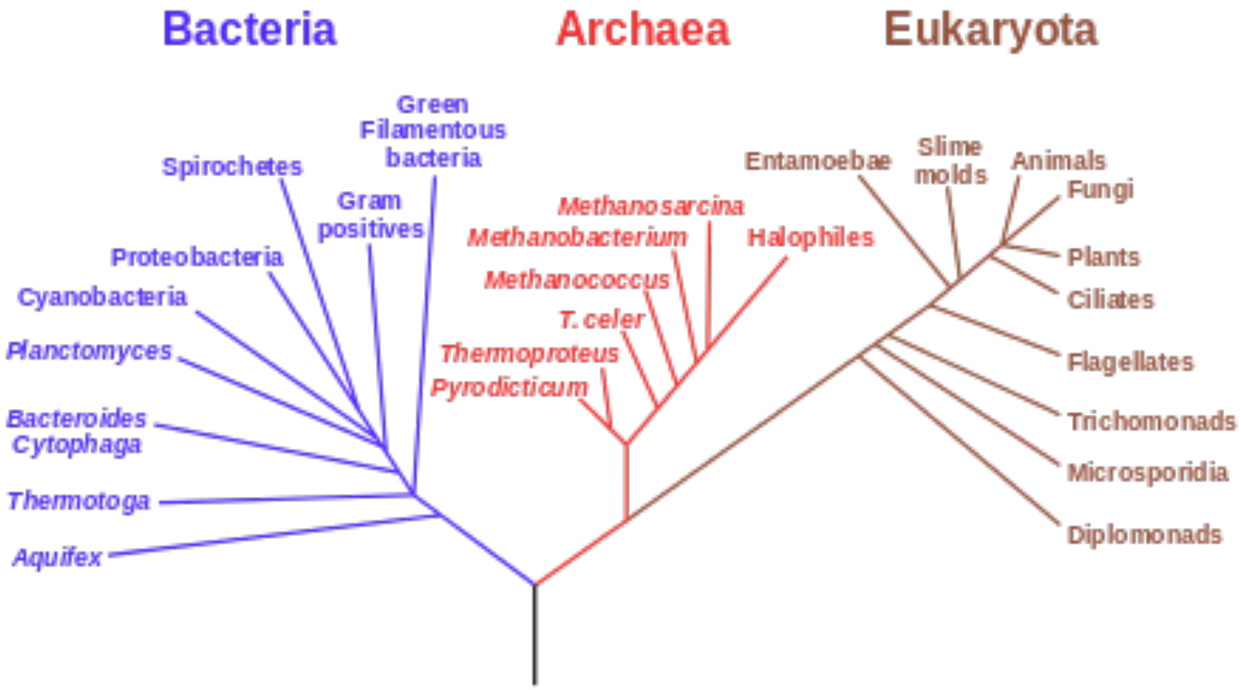
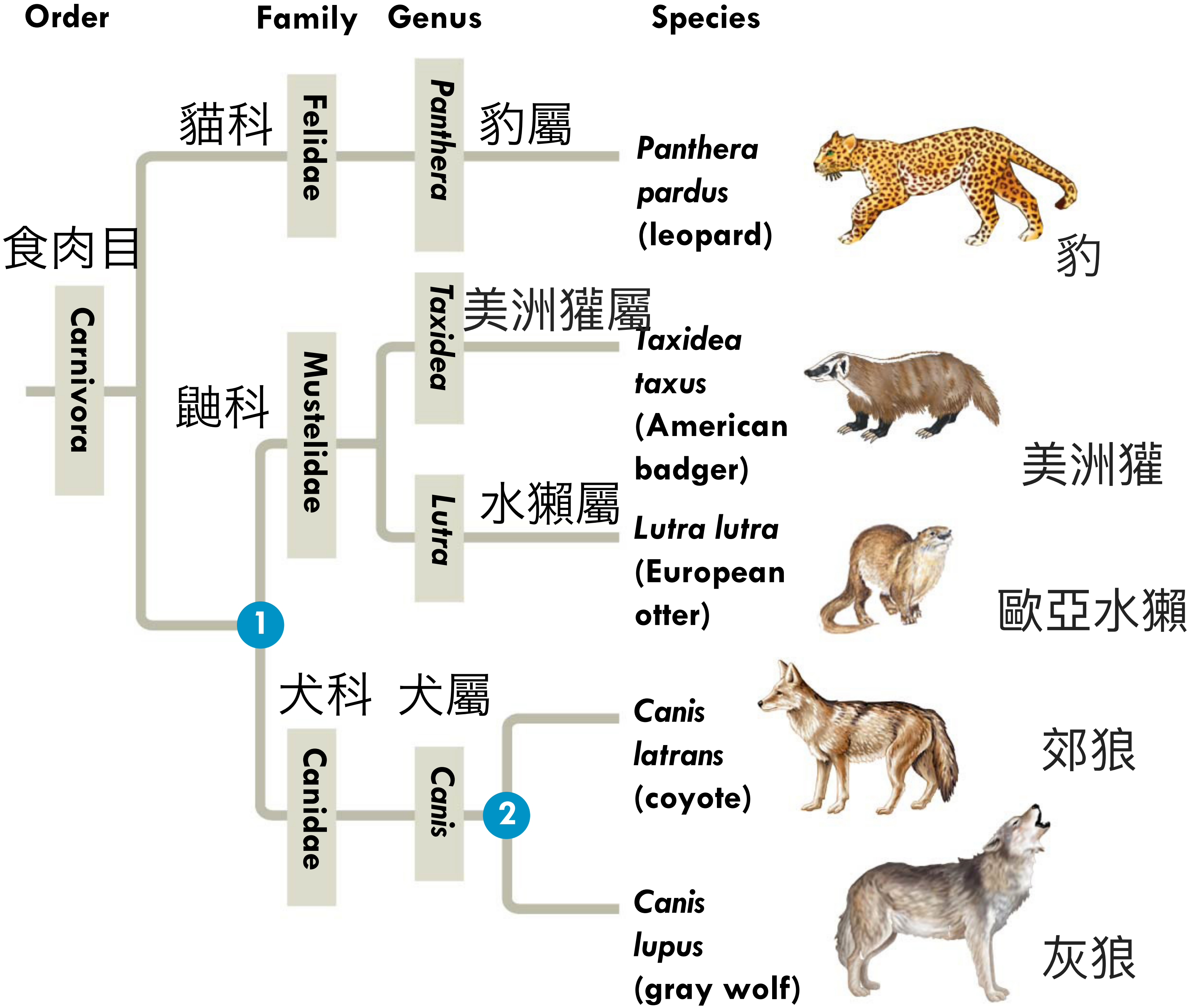
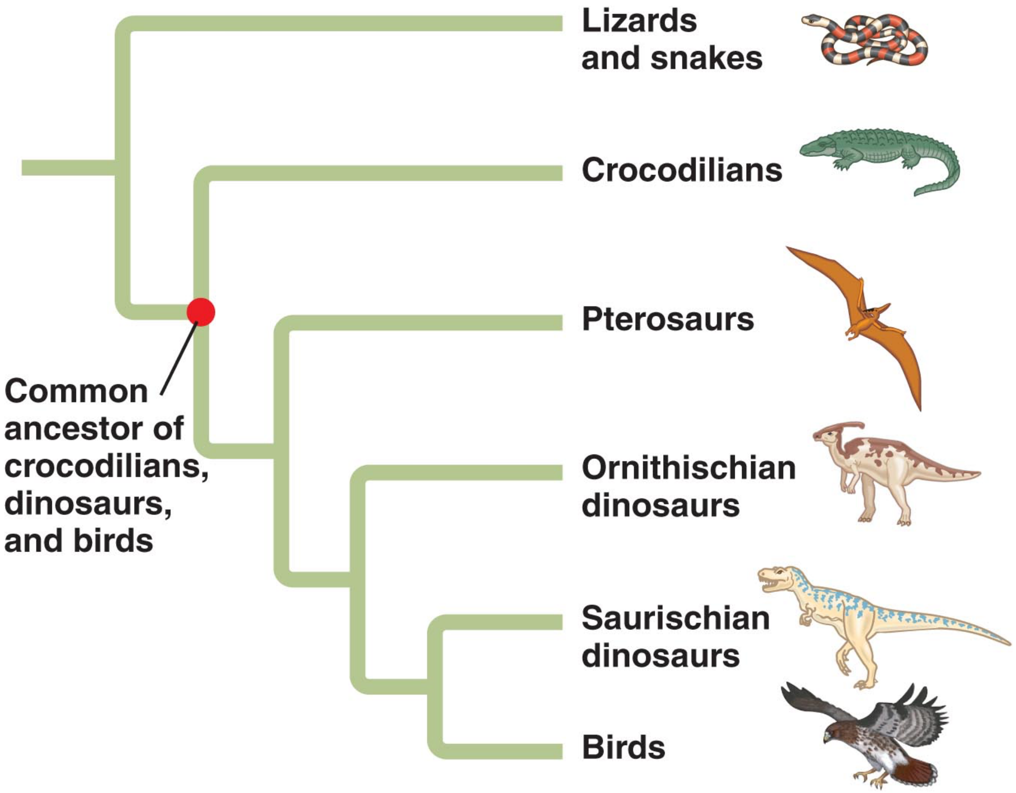


Figure 22.4



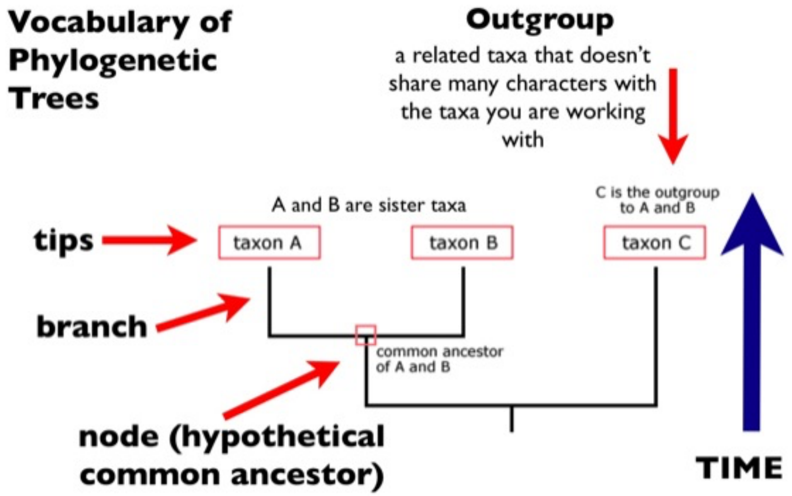
Order
Family
Genus
Species
食肉目
Carnivora
貓科
Felidae
Panthera
豹屬
Panthera pardus (leopard)
豹
鼬科
Mustelidae
Taxidea
美洲獾屬
Taxidea taxus (American badger)
美洲獾
Lutra
水獺屬
Lutra lutra (European otter)
歐亞水獺
1
犬科
犬屬
Canidae
Canis
2
Canis latrans (coyote)
郊狼
Canis lupus (gray wolf)
灰狼

- Linnaean classification and phylogeny can differ from each other
- Systematists have proposed a classification system that would recognize only groups that include a common ancestor and all its descendants



Copyright © 2009 Pearson Education, Inc.

- A phylogenetic tree represents a hypothesis about evolutionary relationships
- Each **branch point** represents the divergence of two species
- Tree branches can be rotated around a branch point without changing the evolutionary relationships
- **Sister taxa** are groups that share an immediate common ancestor



"**ROOTED** tree" - relationships in context of time

- A **rooted** tree includes a branch to represent the last common ancestor of all taxa in the tree
- A **basal taxon** diverges early in the history of a group and originates near the common ancestor of the group
- A **polytomy (多分支)** is a branch from which more than two groups emerge

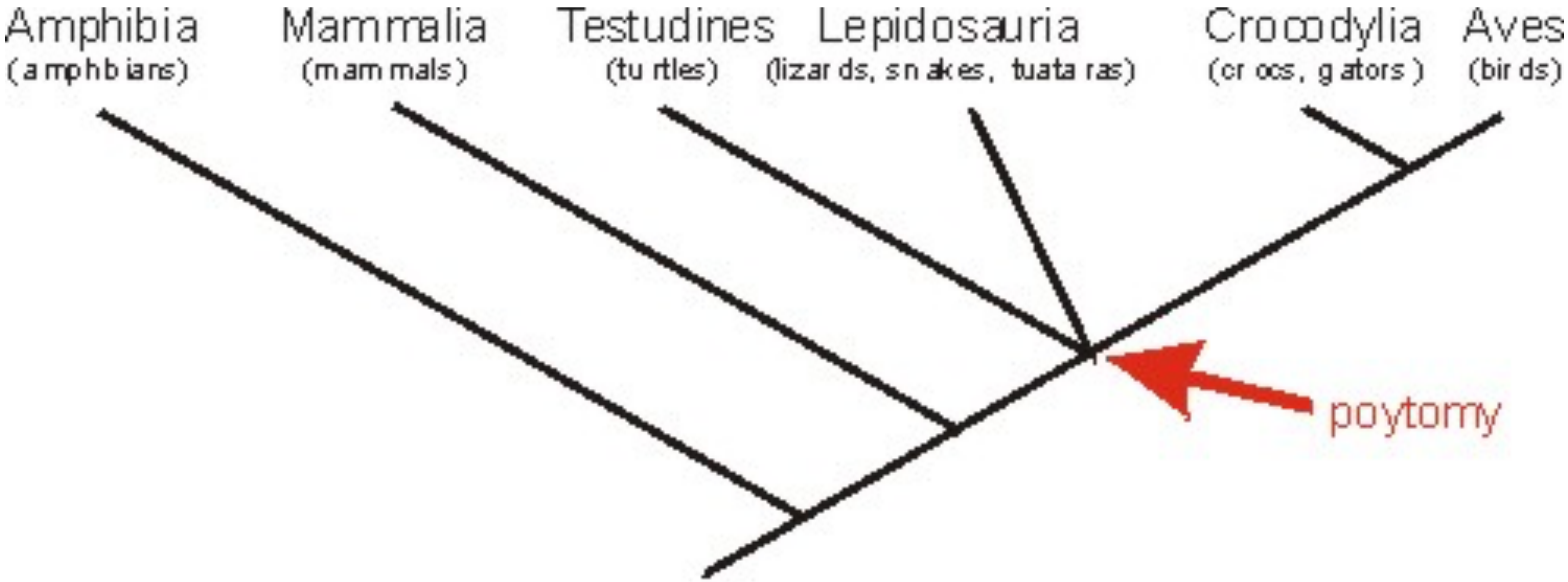
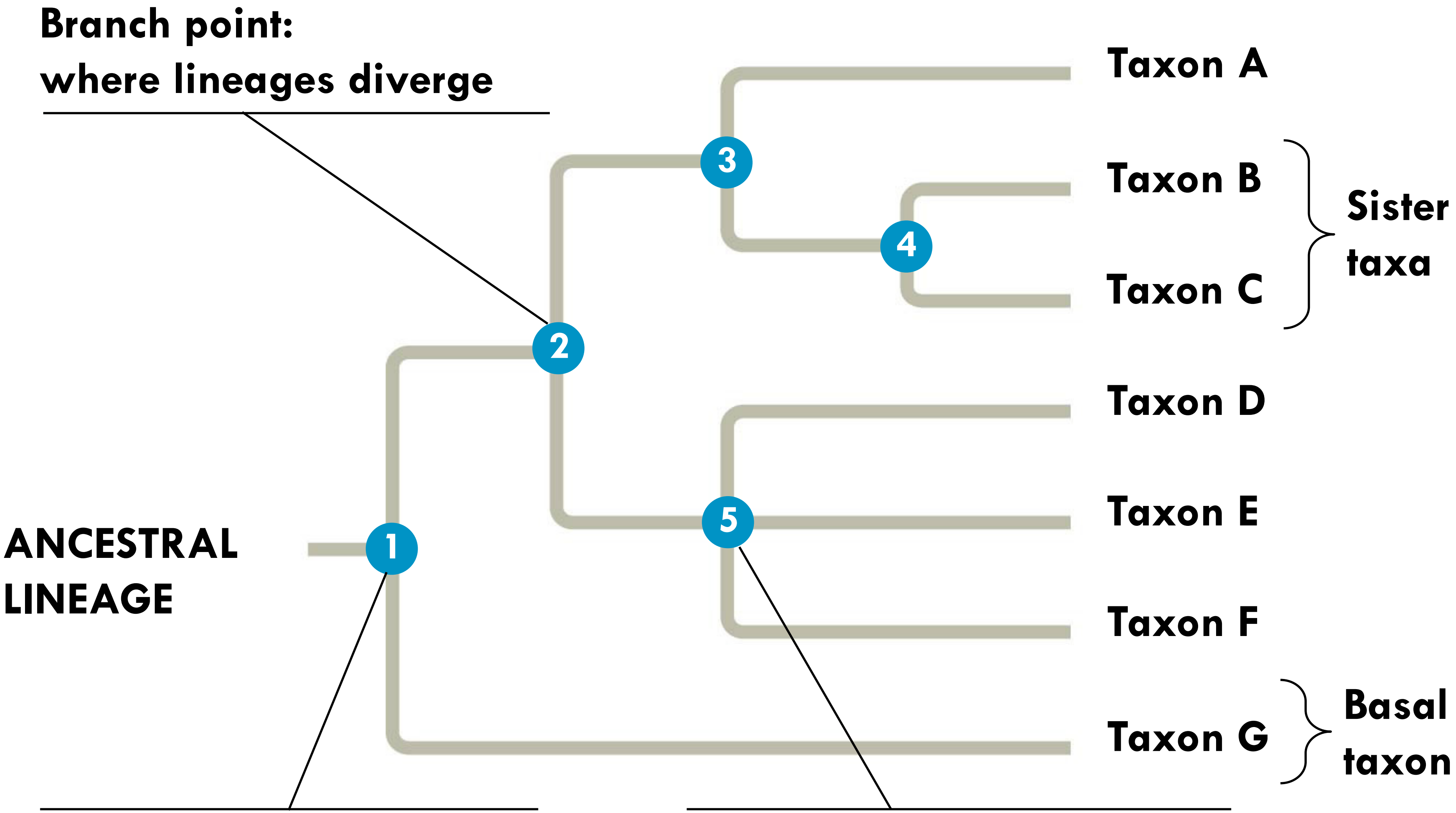


Figure 22.5


Branch point: where lineages diverge

This branch point represents the common ancestor of taxa A–G.

This branch point forms a polytomy: an unresolved pattern of divergence.

What We Can and Cannot Learn from Phylogenetic Trees

- Phylogenetic trees show patterns of descent, not phenotypic similarity
- Phylogenetic trees do not indicate when species evolved or how much change occurred in a lineage
- It should not be assumed that a taxon evolved from the taxon next to it

Figure 22.5a

Parts of a Tree

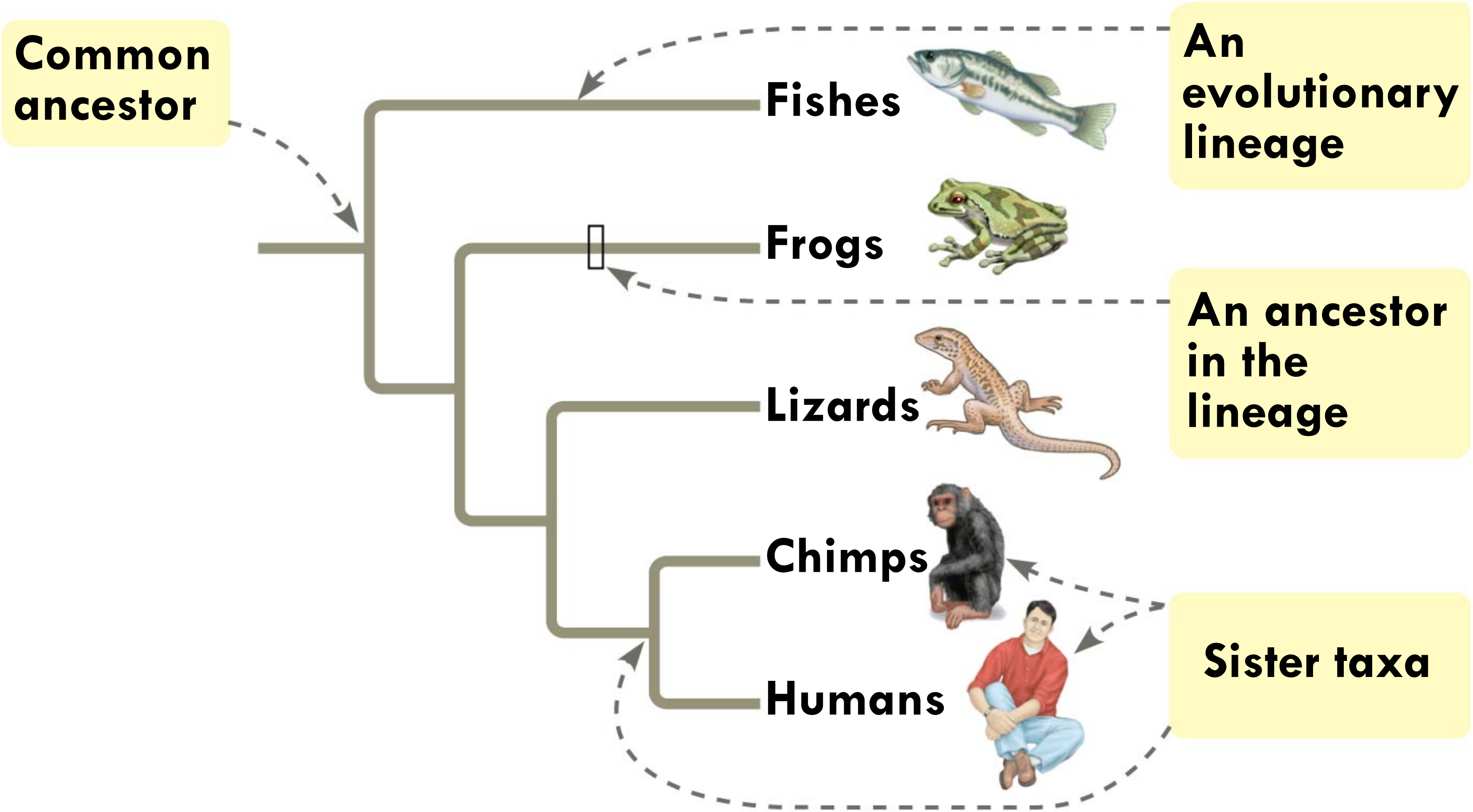


Figure 22.5b

Alternative Forms of Tree Diagrams

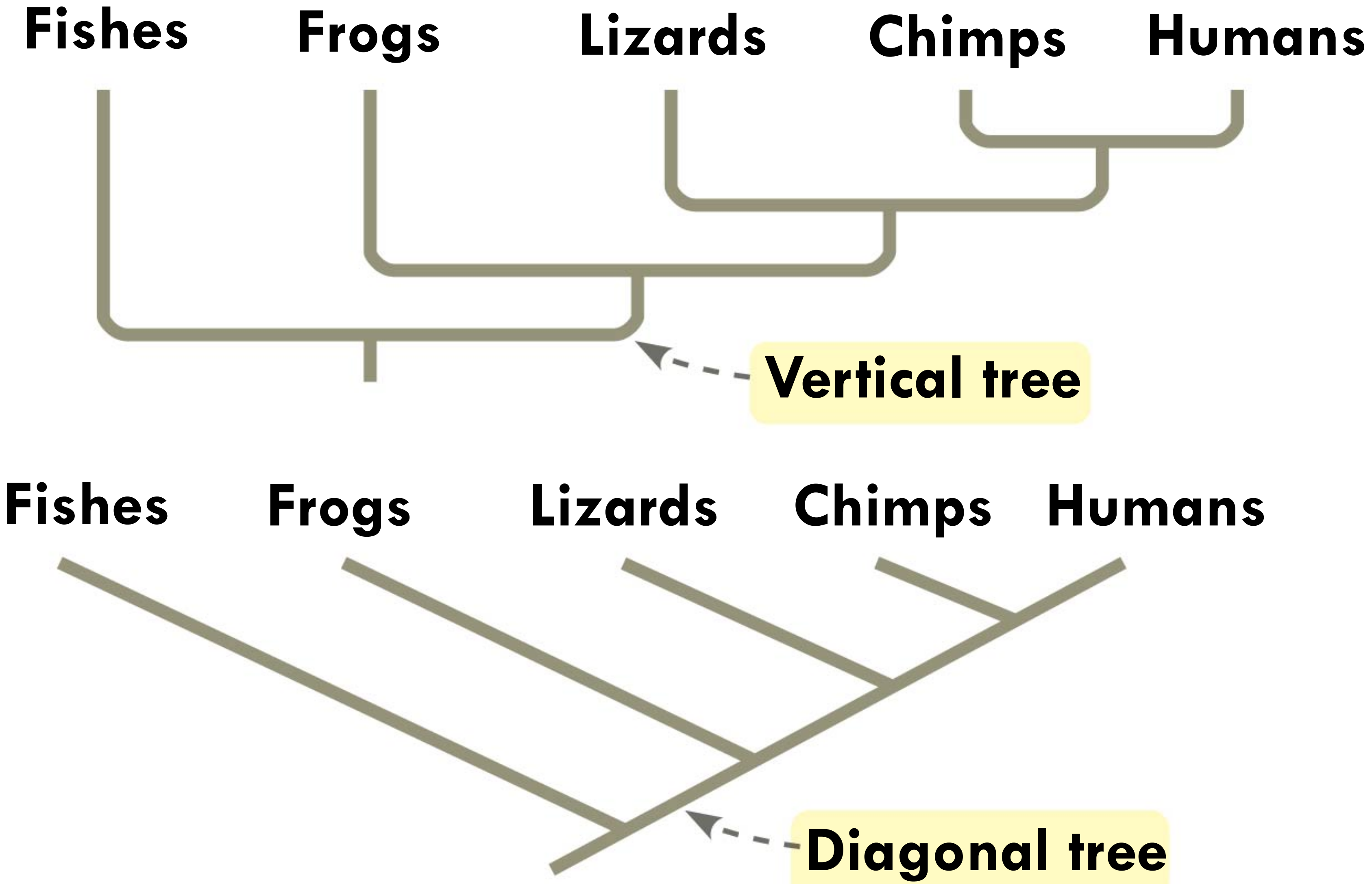
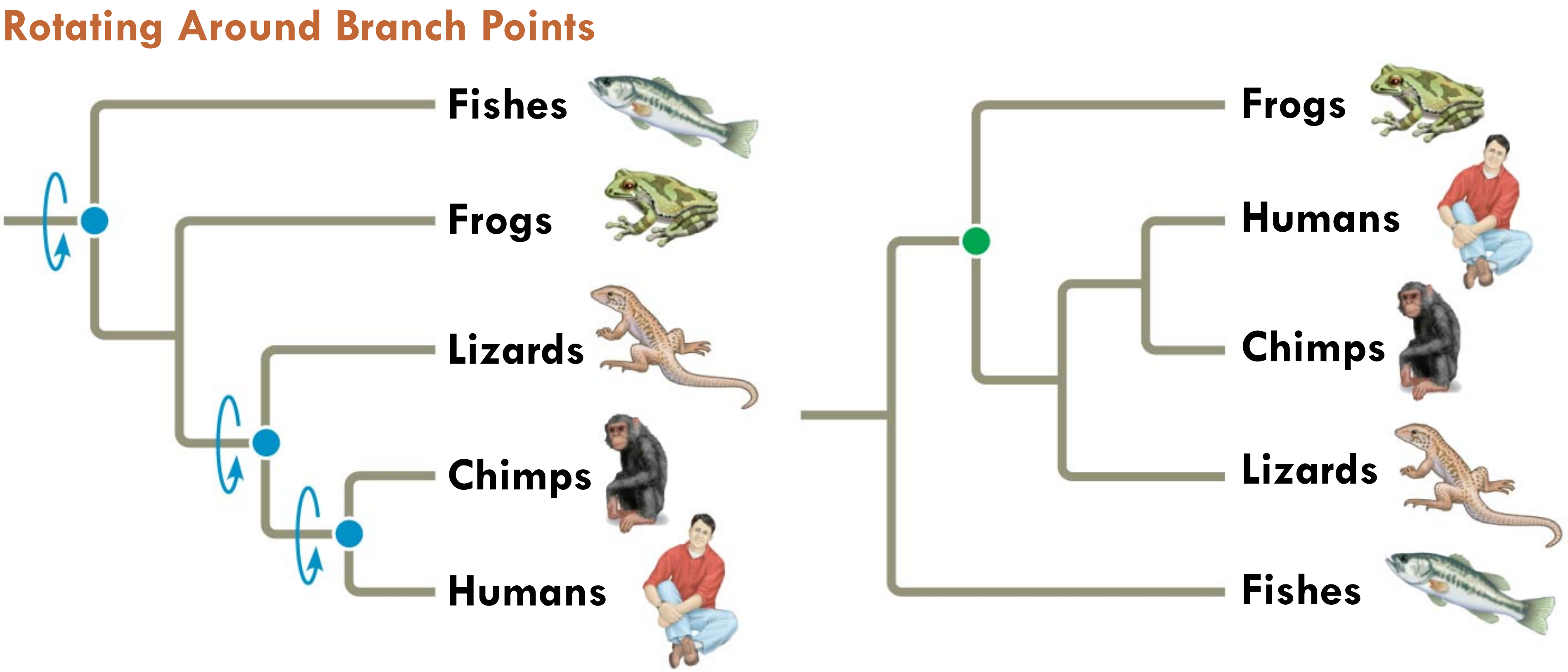


Figure 22.5c


Rotating Around Branch Points
Fishes
Frogs
Lizards
Chimps
Humans
Frogs
Humans
Chimps
Lizards
Fishes

- Phylogenetic trees can be drawn horizontally, vertically, or diagonally without changing the relationships between groups

- Tree branches can be rotated around a branch point without changing the evolutionary relationships

- A **rooted** tree includes a branch to represent the most recent common ancestor of all taxa in the tree

- A **basal taxon** diverges early in the history of a group and originates near the common ancestor of the group

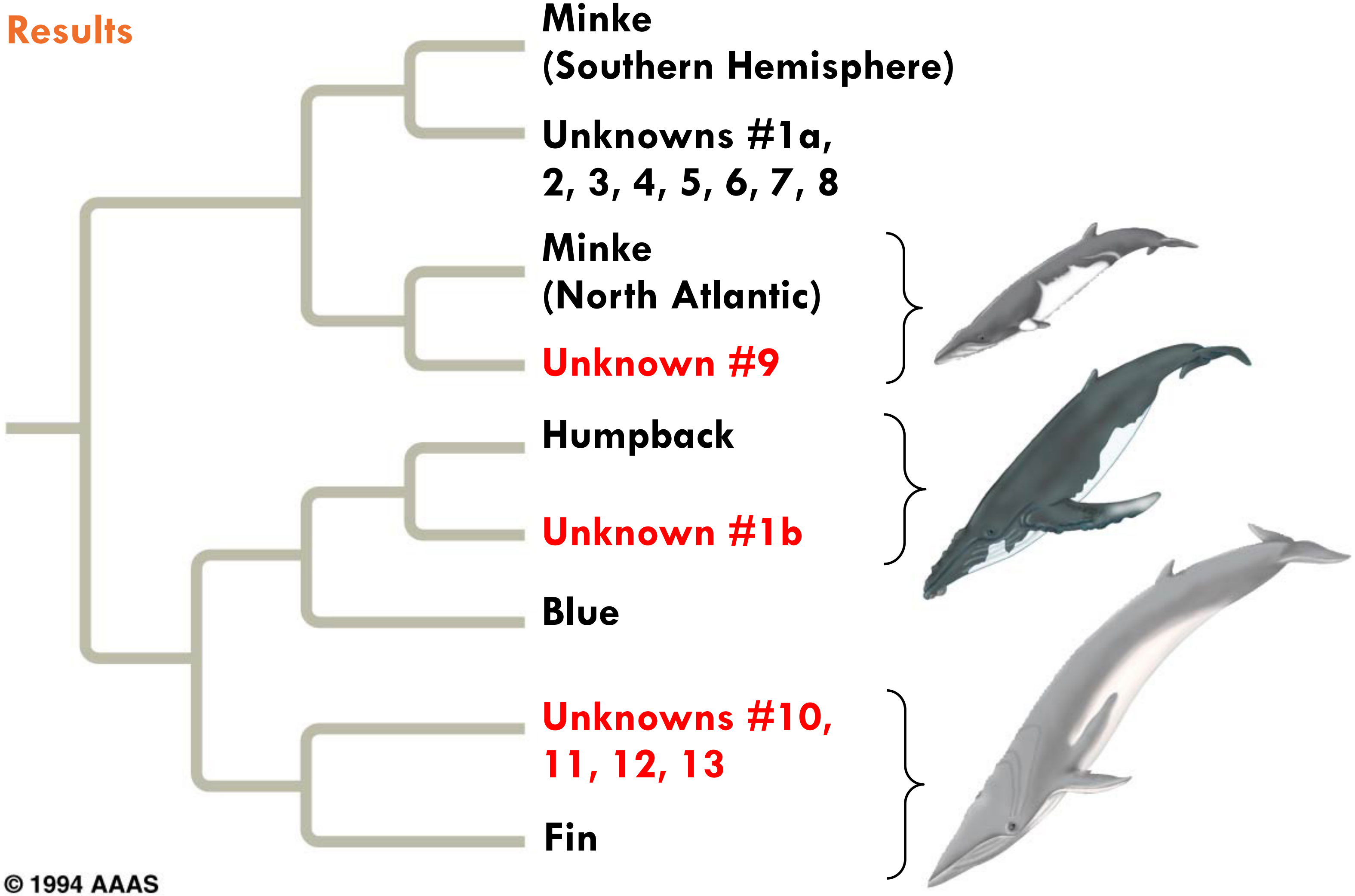
© 2014 Pearson Education, Inc.

Applying Phylogenies

- Phylogeny provides important information about similar characteristics in closely related species
- A phylogeny was used to identify the species of whale from which "whale meat" originated

Figure 22.6

Results



Minke (Southern Hemisphere)

Unknowns #1a, 2, 3, 4, 5, 6, 7, 8

Minke (North Atlantic)

Unknown #9

Humpback

Unknown #1b

Blue

Unknowns #10, 11, 12, 13

Fin

Concept 22.2: Phylogenies are inferred from morphological and molecular data

- To infer phylogenies, systematists gather information about morphologies, genes, and biochemistry of living organisms

Morphological and Molecular Homologies

- Phenotypic and genetic similarities due to shared ancestry are called homologies
- Organisms with similar morphologies or DNA sequences are likely to be more closely related than organisms with different structures or sequences

Sorting Homology from Analogy

- When constructing a phylogeny, systematists need to distinguish whether a similarity is the result of homology or **analogy**
- Homology is similarity due to shared ancestry
- Analogy is similarity due to convergent evolution

- Convergent evolution occurs when similar environmental pressures and natural selection produce similar (analogous) adaptations in organisms from different evolutionary lineages

Figure 22.7

Australian marsupial “mole”

North American eutherian mole 鼴鼠

- Bat and bird wings are homologous as forelimbs, but analogous as functional wings
- Analogous structures or molecular sequences that evolved independently are also called **homoplasies (趨同性)**
- Homology can be distinguished from analogy by comparing fossil evidence and the degree of complexity
- The more elements that are similar in two complex structures, the more likely it is that they are homologous

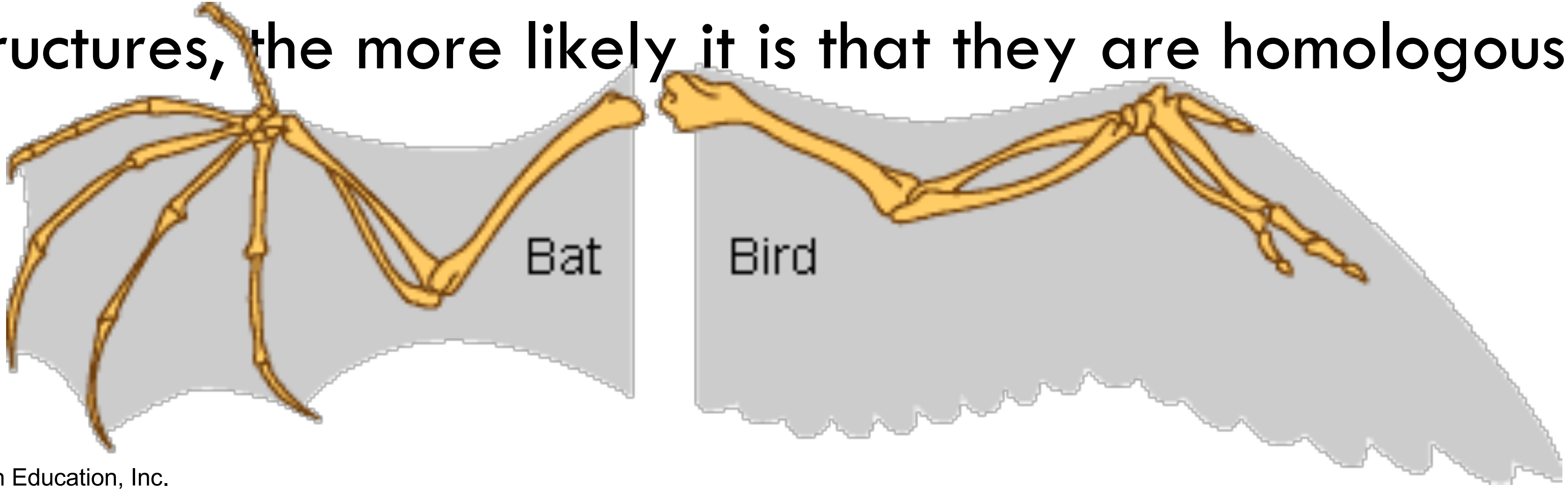


Figure 22.8-1

1 CCATCAGAGTCC

2 CCATCAGAGTCC

Figure 22.8-2

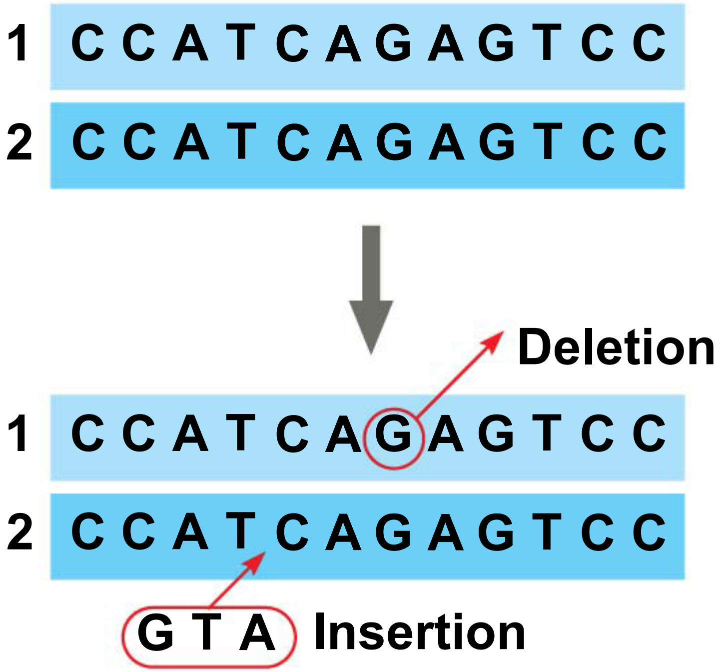
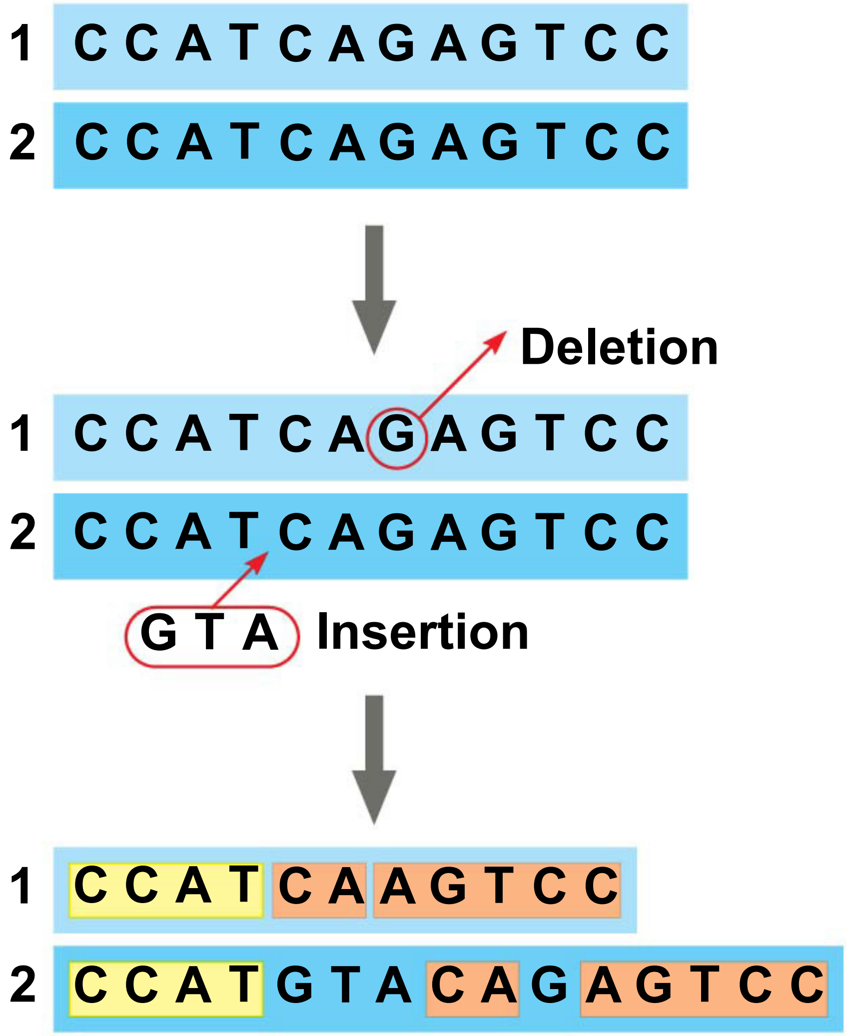

1 CCATCAGAGTCC
2 CCATCAGAGTCC
Deletion
1 CCATCAGAGTCC
2 CCATCAGAGTCC
GTA
Insertion

Figure 22.8-3



1 CCATCAGAGTCC

2 CCATCAGAGTCC

Deletion

1 CCATCAGAGTCC

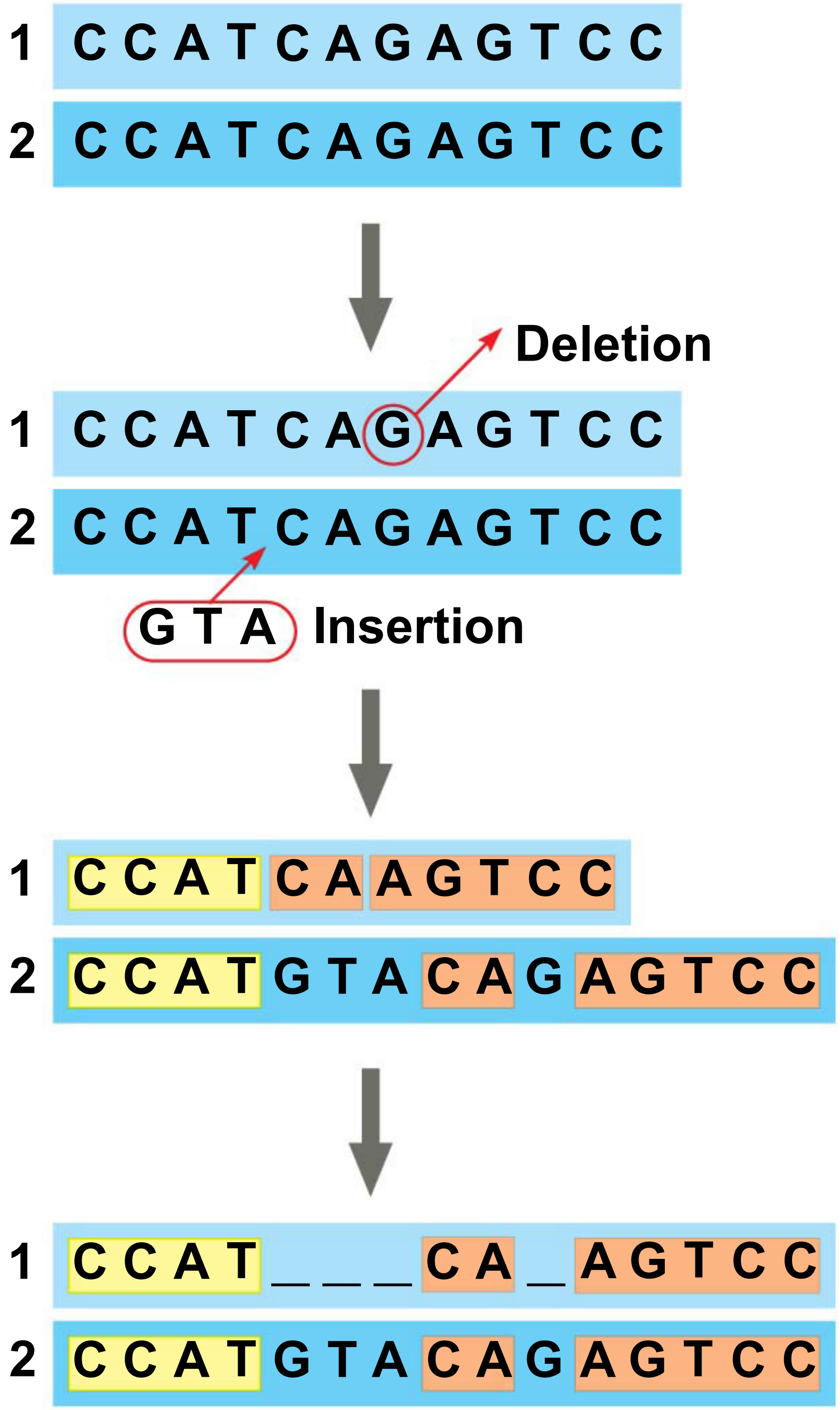
2 CCATCAGAGTCC

GTA Insertion

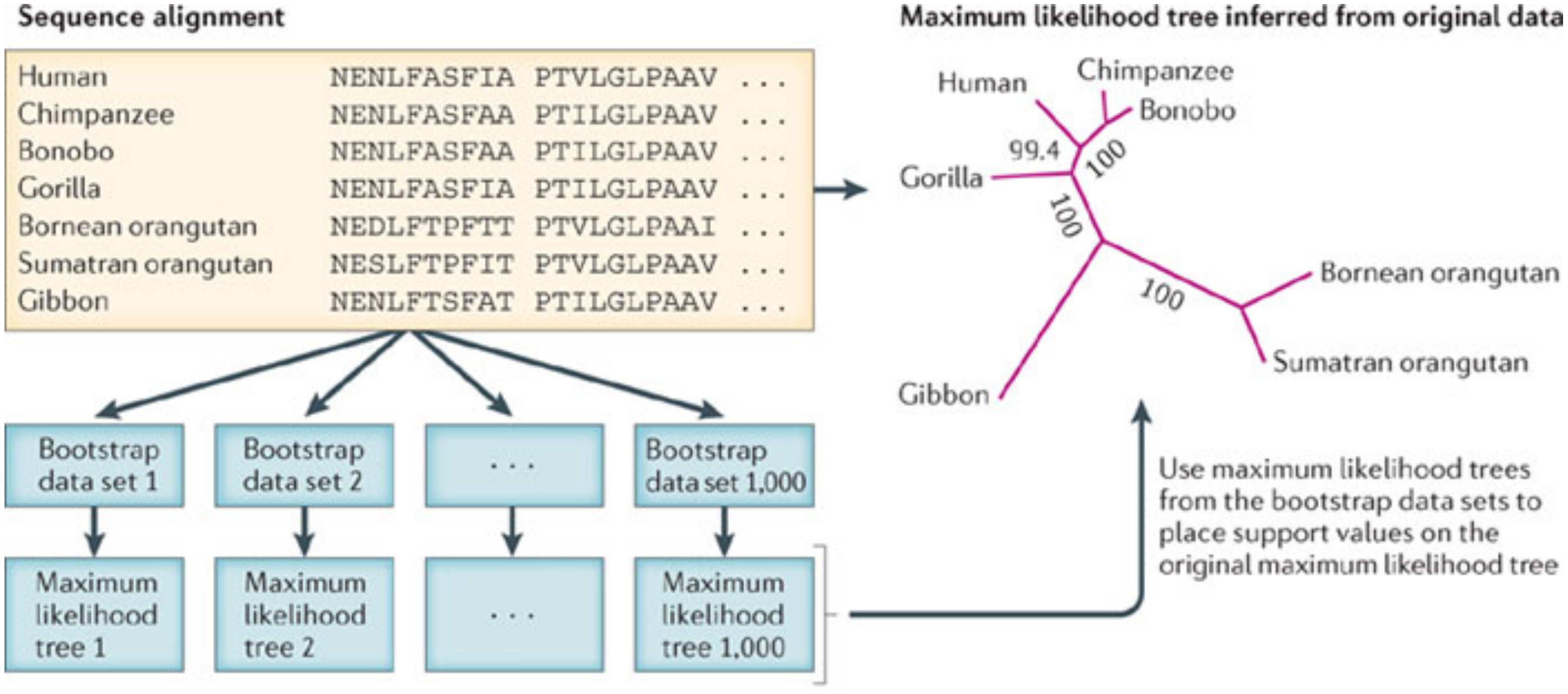
1 CCATCAAGTCC

2 CCATGTACAGAGTCC

Figure 22.8-4


1 CCATCAGAGTCC
2 CCATCAGAGTCC
Deletion
1 CCATCAGAGTCC
2 CCATCAGAGTCC
GTA
Insertion
1 CCATCAAGTCC
2 CCATGTACAGAGTCC
1 CCAT___CA_AGTCC
2 CCATGTACAGAGTCC

- It is also important to distinguish homology from analogy in molecular similarities
- Mathematical tools help to identify molecular homoplasies, or coincidences



Sequence alignment

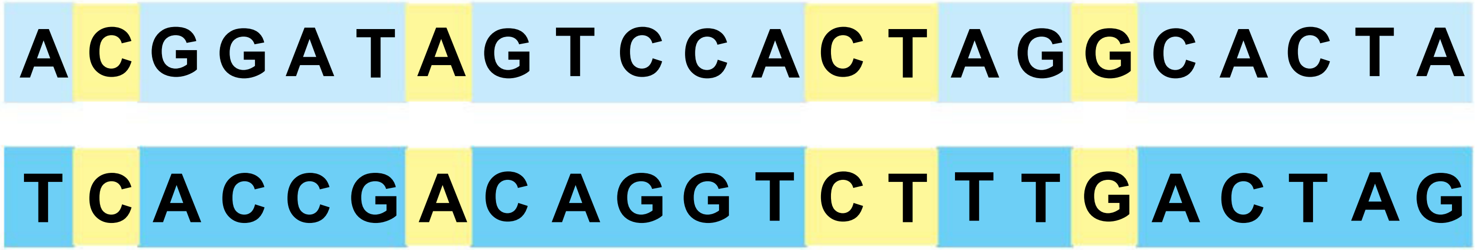
Human	NENLFASFIA	PTVLGLPAAV	...
Chimpanzee	NENLFASFAA	PTILGLPAAV	...
Bonobo	NENLFASFAA	PTILGLPAAV	...
Gorilla	NENLFASFIA	PTILGLPAAV	...
Bornean orangutan	NEDLFTPFTT	PTVLGLPAAI	...
Sumatran orangutan	NESLFTPFIT	PTVLGLPAAV	...
Gibbon	NENLFTSFAT	PTILGLPAAV	...

Maximum likelihood tree inferred from original data

Use maximum likelihood trees from the bootstrap data sets to place support values on the original maximum likelihood tree

Nature Reviews | Genetics

Figure 22.9

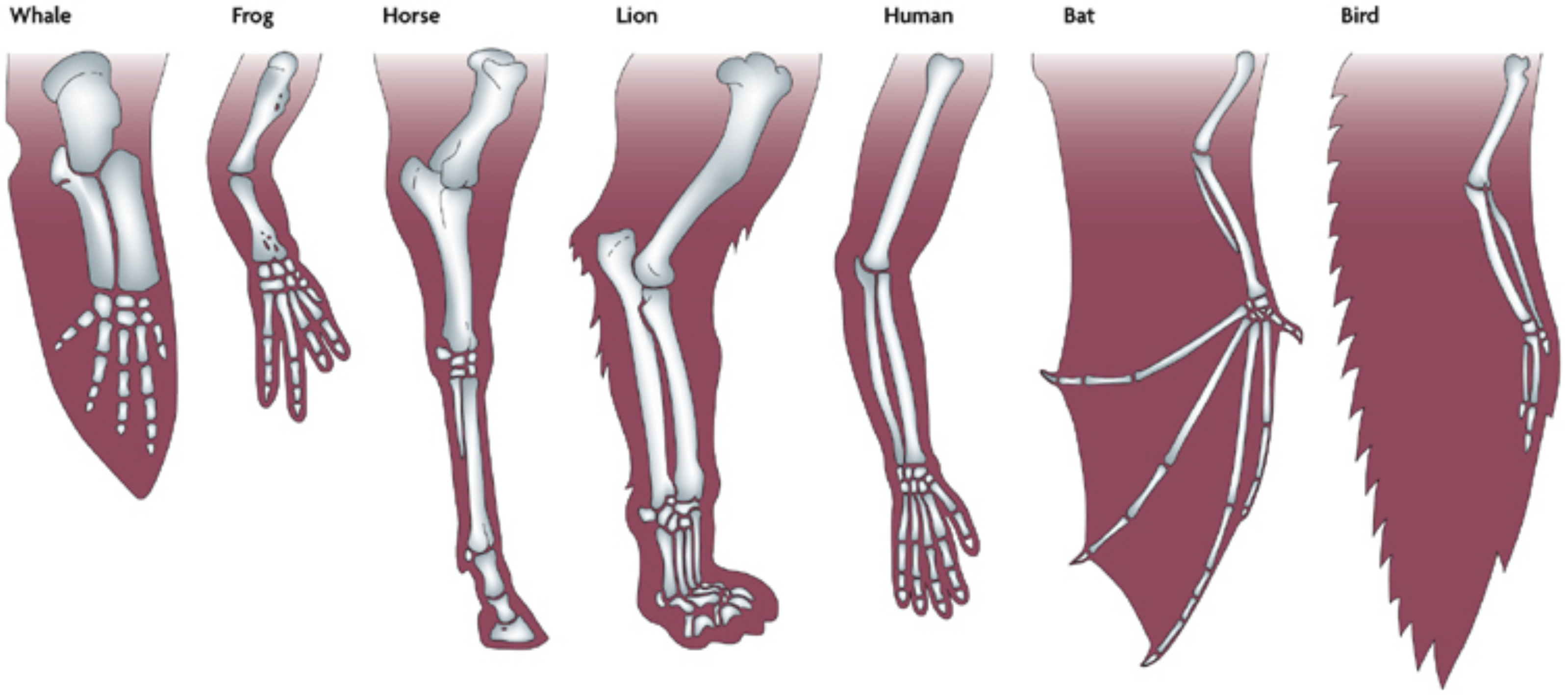


ACGGATAGTCCACTAGGCACTA

TCACCGACAGGTCTTTGACTAG

Concept 22.3: Shared characters are used to construct phylogenetic trees

- Once homologous characters have been identified, they can be used to infer a phylogeny



Nature Reviews | Genetics

Cladistics

- **Cladistics** (支序分類學) groups organisms by common descent
- A **clade** (分支) is a group of species that includes an ancestral species and all its descendants
- Clades can be nested in larger clades, but not all groupings of organisms qualify as clades

- A valid clade is **monophyletic (單系群)**, signifying that it consists of the ancestor species and all its descendants

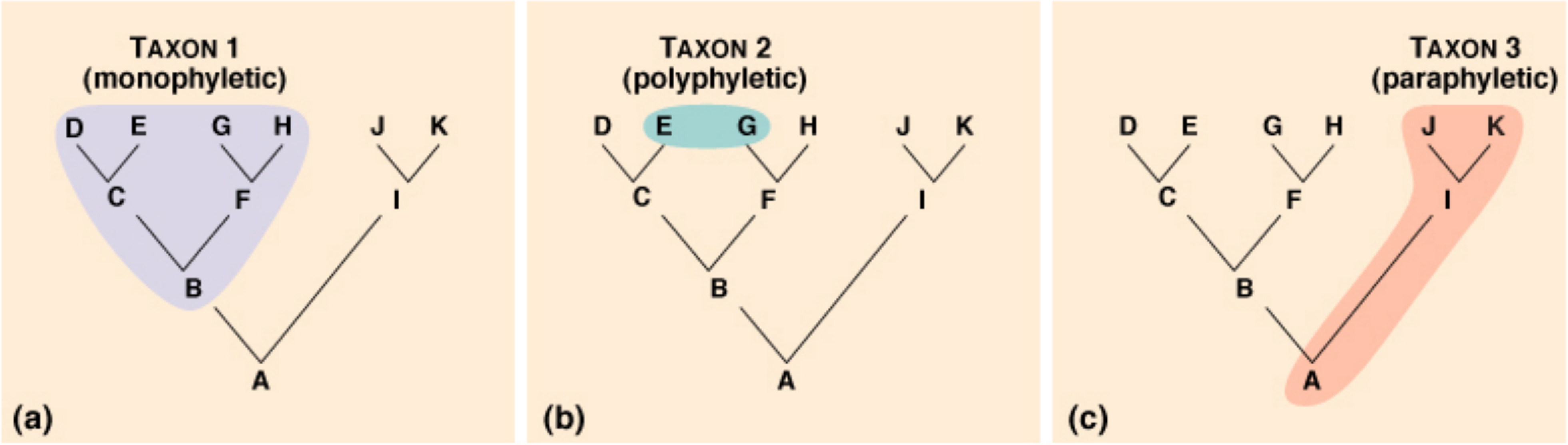
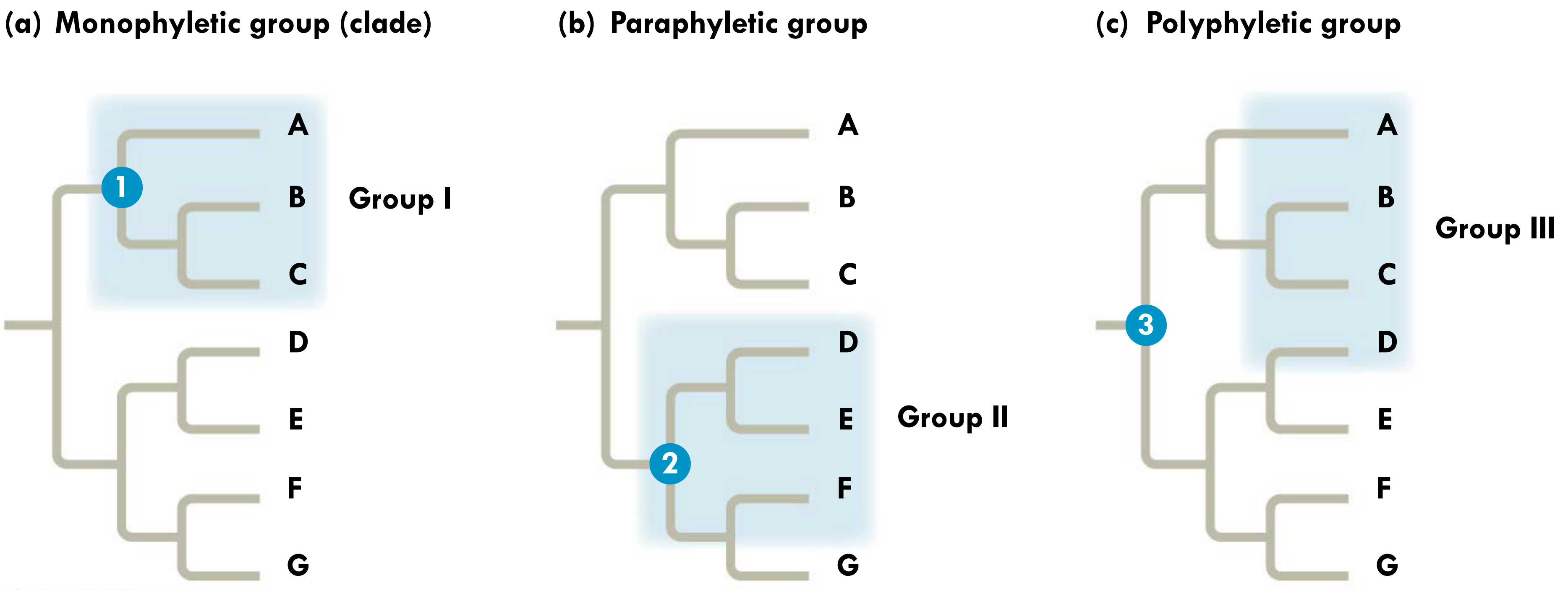
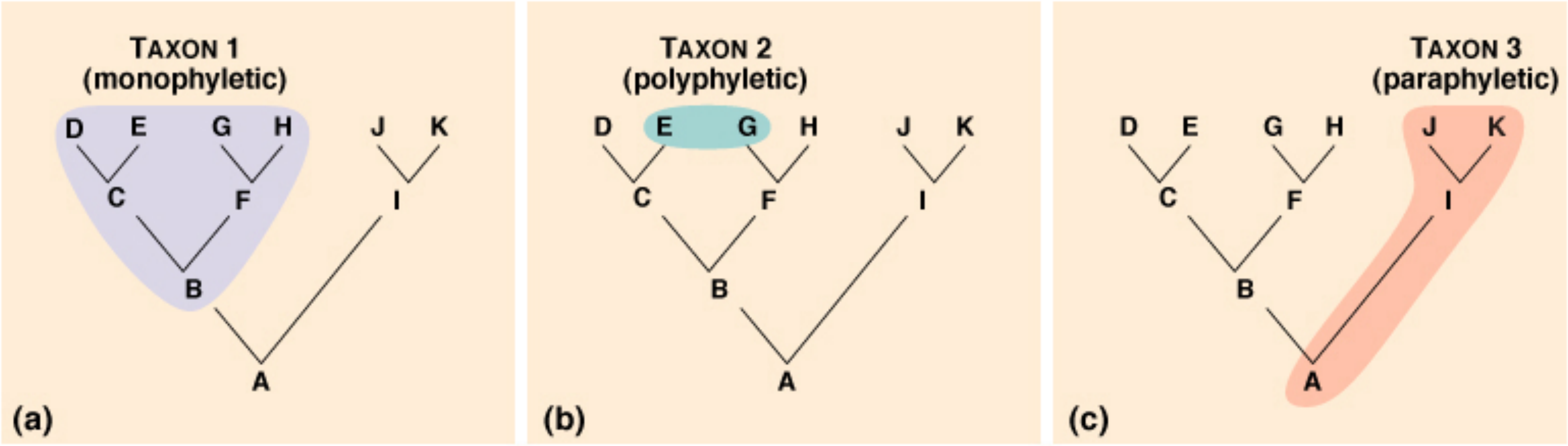


Figure 22.10

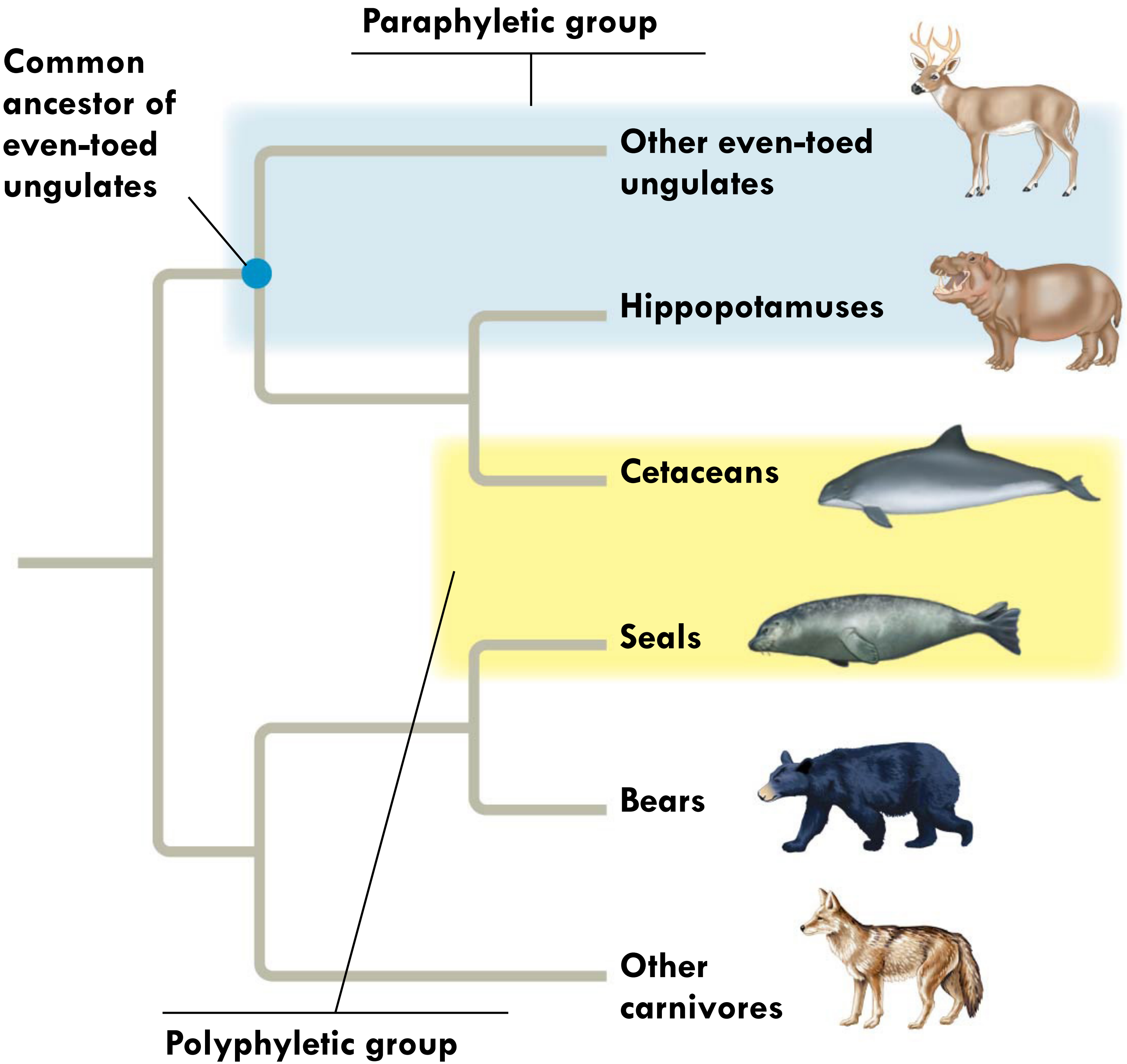

(a) Monophyletic group (clade)
(b) Paraphyletic group
(c) Polyphyletic group
1
2
3
A
B
C
D
E
F
G
Group I
Group II
Group III

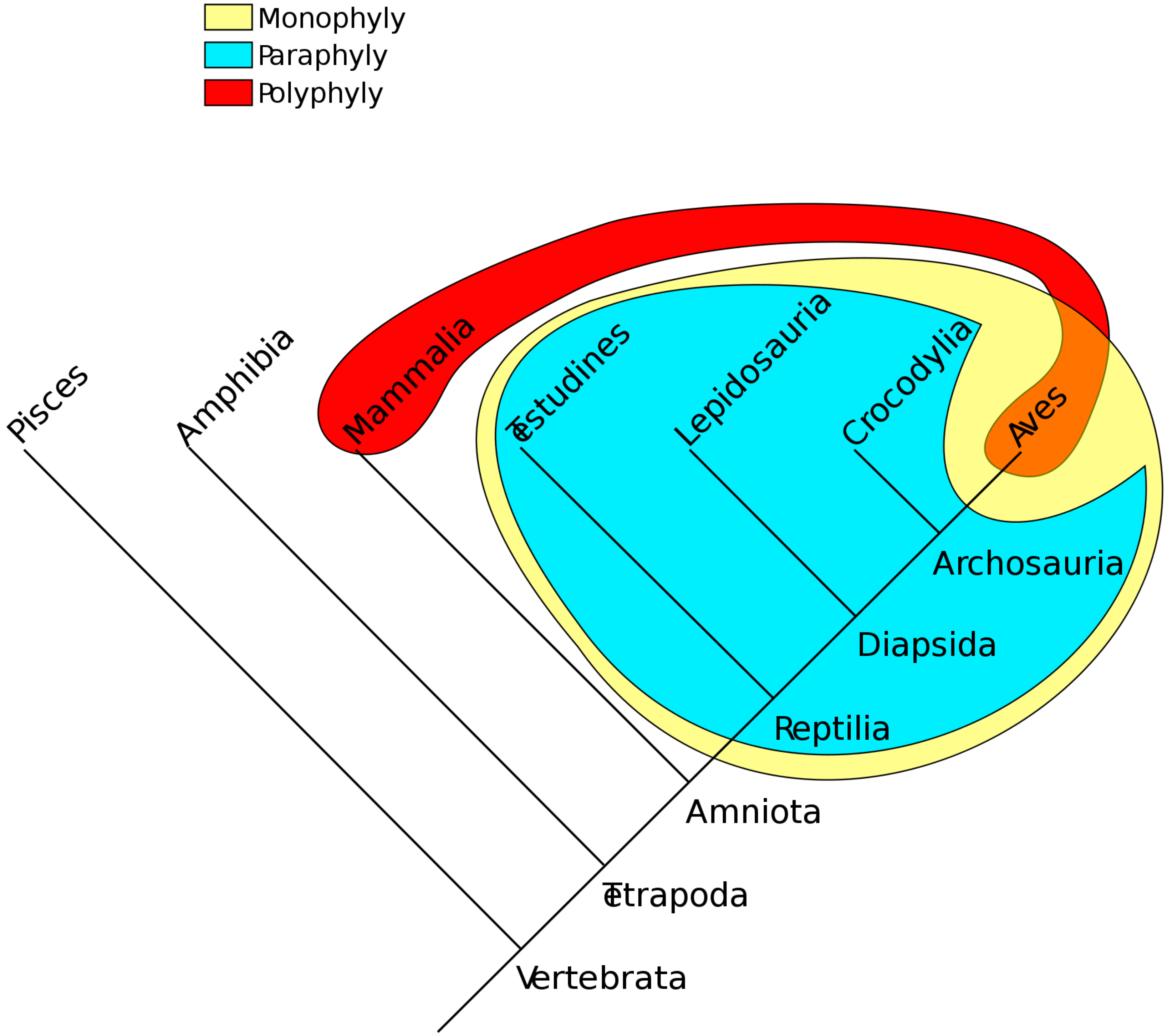
- A **paraphyletic (并系群)** grouping consists of an ancestral species and some, but not all, of the descendants
- A **polyphyletic (多系群)** grouping includes distantly related species but does not include their most recent common ancestor



- Polyphyletic groups are distinguished from paraphyletic groups by the fact that they do not include the most recent common ancestor
- Biologists avoid defining polyphyletic groups and instead reclassify organisms if evidence suggests they are polyphyletic

Figure 22.11


Paraphyletic group
Common ancestor of even-toed ungulates
Other even-toed ungulates
Hippopotamuses
Cetaceans
Seals
Bears
Other carnivores
Polyphyletic group


Monophyly
Paraphyly
Polyphyly
Pisces
Amphibia
Mammalia
Testudines
Lepidosauria
Crocodylia
Aves
Archosauria
Diapsida
Reptilia
Amniota
Tetrapoda
Vertebrata

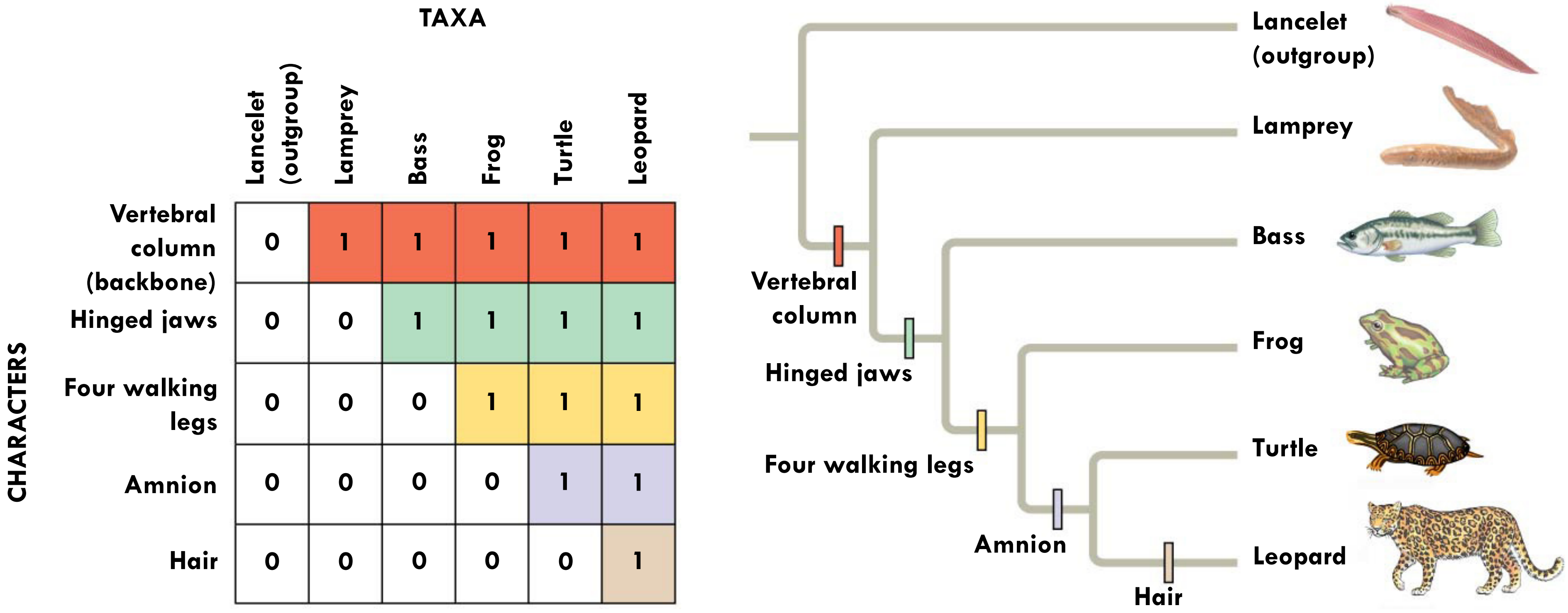
Shared Ancestral and Shared Derived Characters

- In comparison with its ancestor, an organism has both shared and different characteristics
- A **shared ancestral character** (共祖徵) is a character that originated in an ancestor of the taxon
- A **shared derived character** (共衍徵) is an evolutionary novelty unique to a particular clade
- A character can be both ancestral and derived, depending on the context

Inferring Phylogenies Using Derived Characters

- When inferring evolutionary relationships, it is useful to know in which clade a shared derived character first appeared

Figure 22.12



(a) Character table

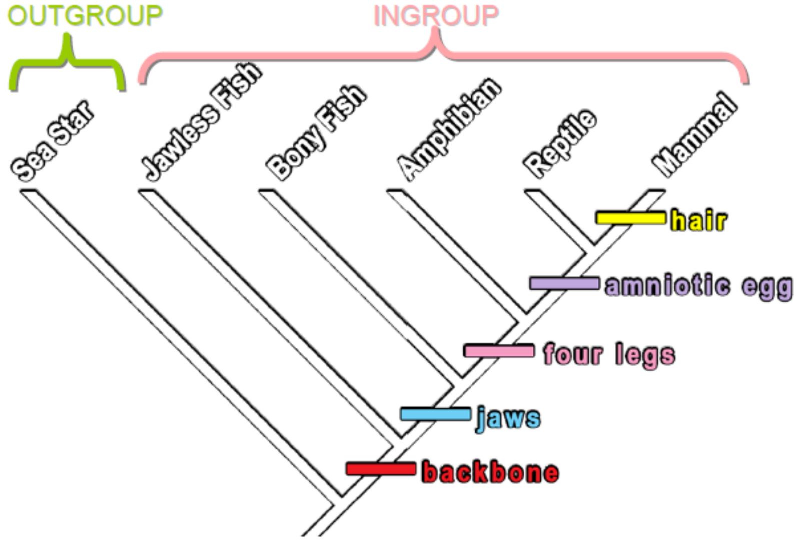
CHARACTERS	TAXA					
	Lancelet (outgroup)	Lamprey	Bass	Frog	Turtle	Leopard
Vertebral column (backbone)	0	1	1	1	1	1
Hinged jaws	0	0	1	1	1	1
Four walking legs	0	0	0	1	1	1
Amnion	0	0	0	0	1	1
Hair	0	0	0	0	0	1

(b) Phylogenetic tree

- An **outgroup (外群)** is a species or group of species that is closely related to the **ingroup (内群)**, the various species being studied

- The outgroup is a group that has diverged before the ingroup

- Systematists compare each ingroup species with the outgroup to differentiate between shared derived and shared ancestral characteristics



- Characters shared by the outgroup and ingroup are ancestral characters that predate the divergence of both groups from a common ancestor

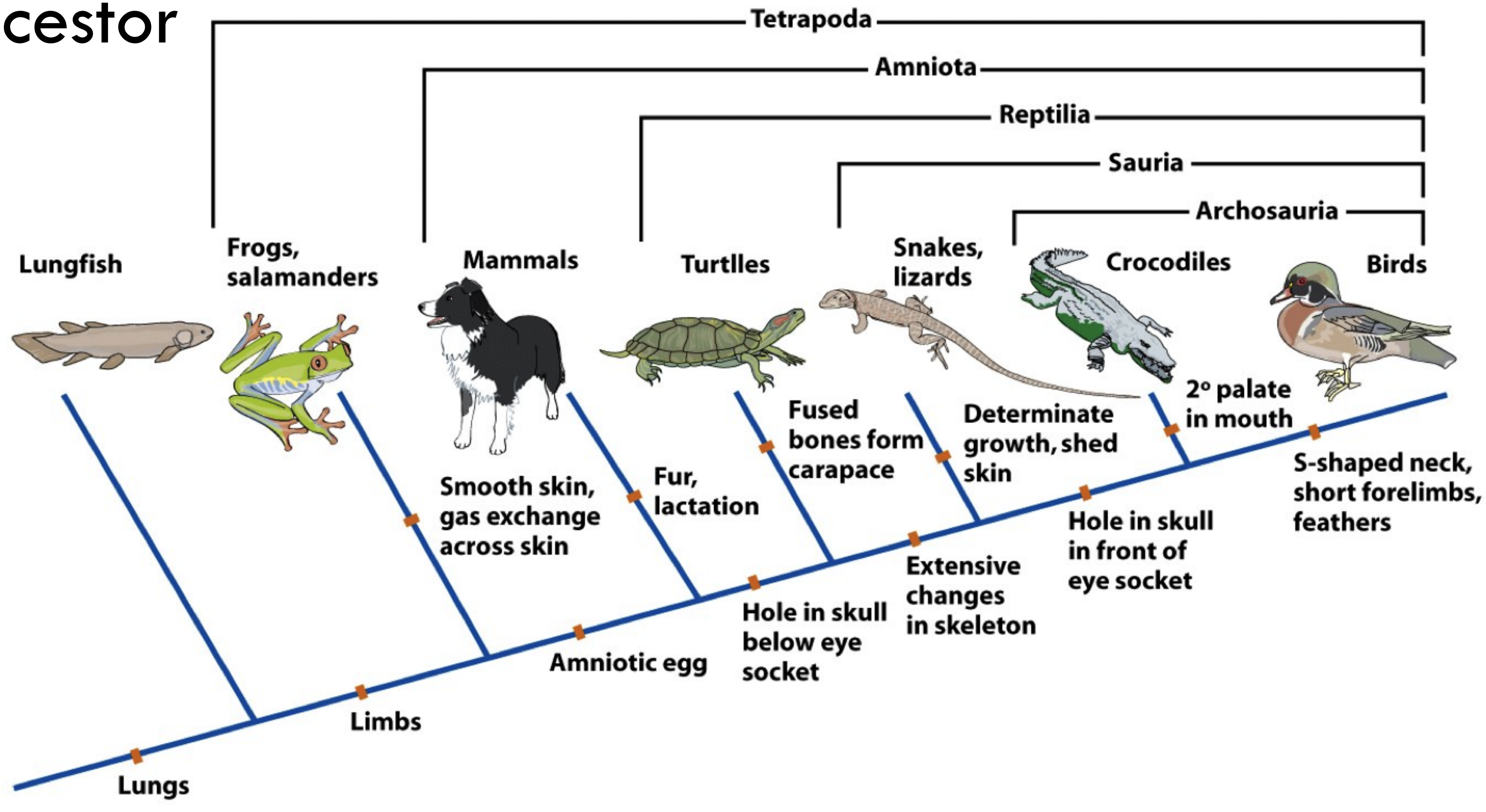


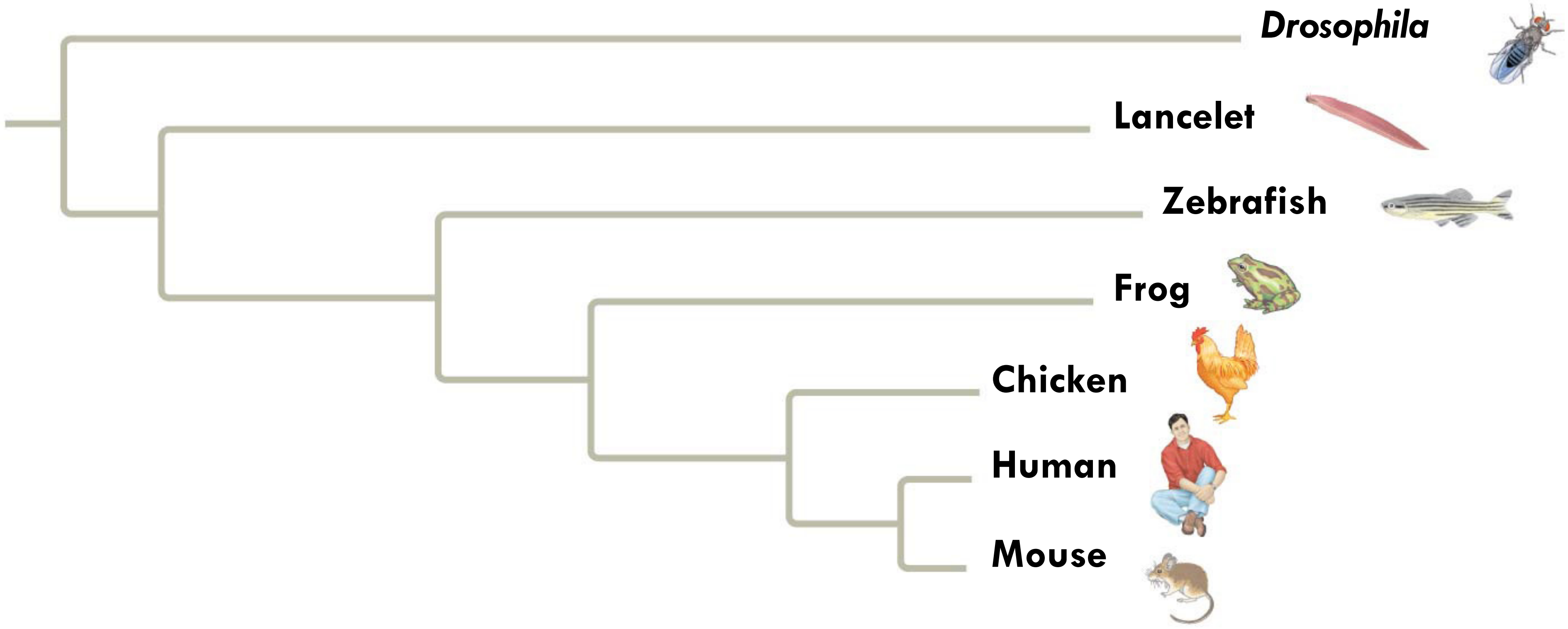
Figure 4-3 Evolutionary Analysis, 4/e

Phylogenetic Trees with Proportional Branch Lengths

- In some trees, the length of a branch can reflect the number of genetic changes that have taken place in a particular DNA sequence in that lineage

Figure 22.13

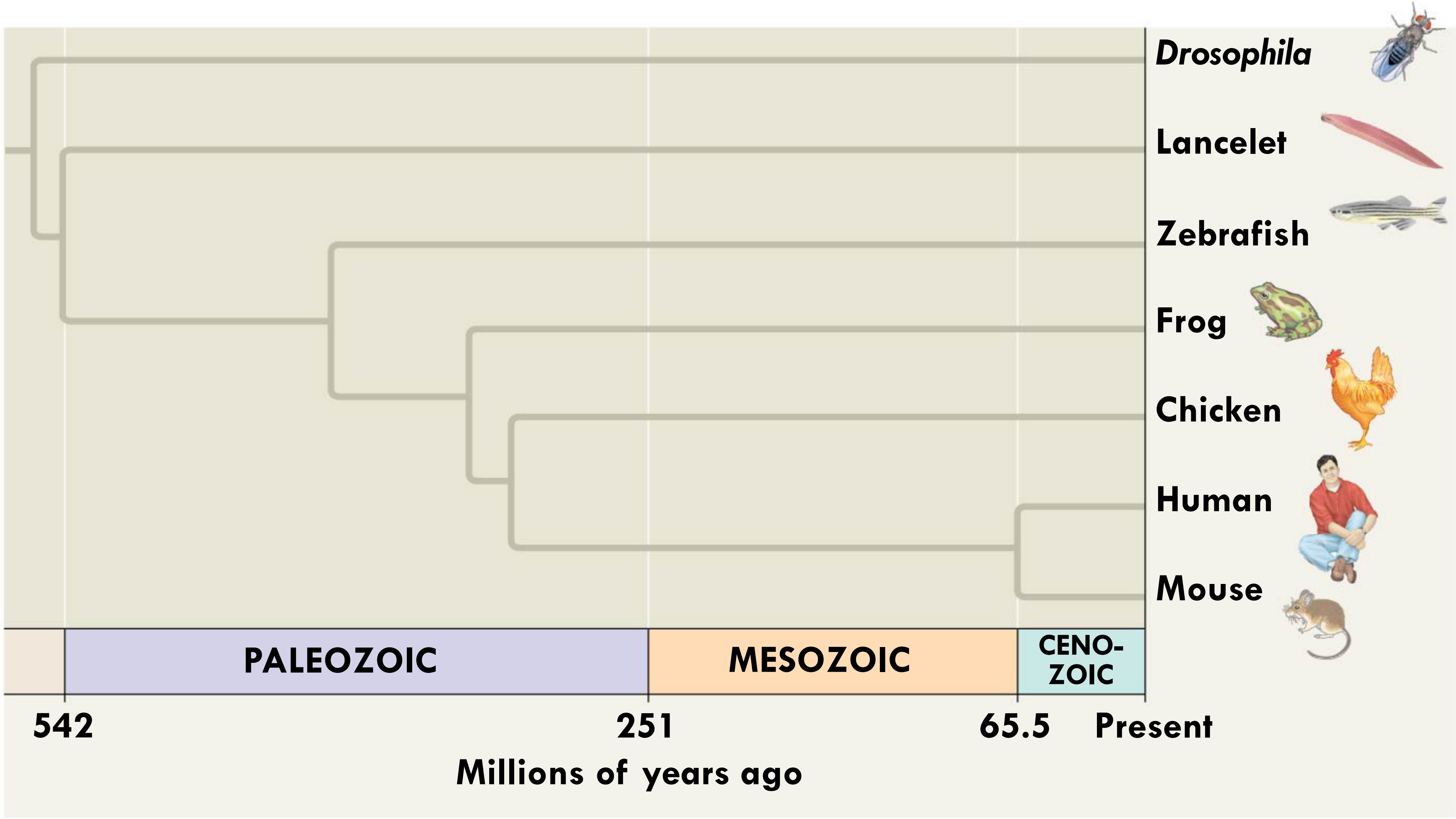
Branch length indicates *genetic changes*



- In other trees, branch length can represent chronological time, and branching points can be determined from the fossil record

Figure 22.14

Branch length indicates *time*



Maximum Parsimony and Maximum Likelihood

- Systematists can never be sure of finding the best tree in a large data set
- They narrow possibilities by applying the principles of maximum parsimony and maximum likelihood

- **Maximum parsimony (最大簡約法)** assumes that the tree that requires the fewest evolutionary events (appearances of shared derived characters) is the most likely

- The principle of **maximum likelihood (最大似然法)** states that, given certain rules about how DNA changes over time, a tree can be found that reflects the most likely sequence of evolutionary events

- Computer programs are used to search for trees that are parsimonious and likely

Evaluating Molecular Homologies

- Systematists use computer programs and mathematical tools when analyzing comparable DNA segments from different organisms

PAUP*

MEGA Molecular Evolutionary Genetics Analysis

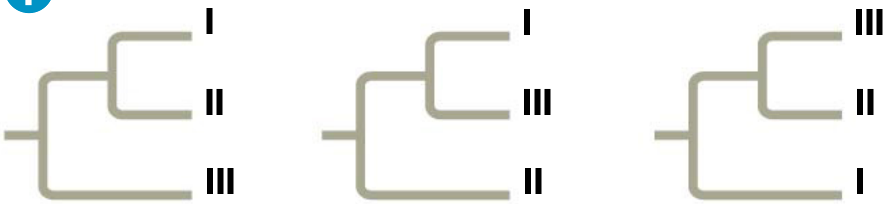
Mr Bayes
Bayesian Inference of Phylogeny

BEAST

Figure 22.15

Technique

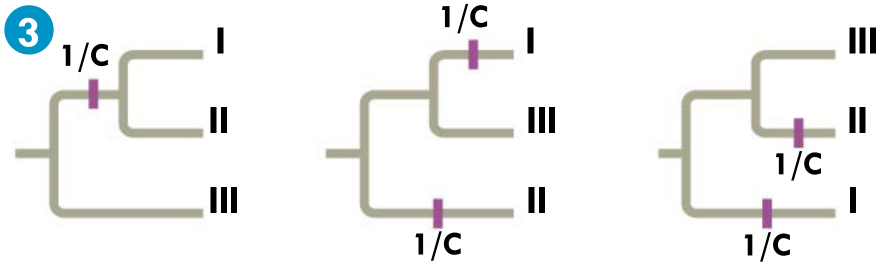
① Three phylogenetic hypotheses:



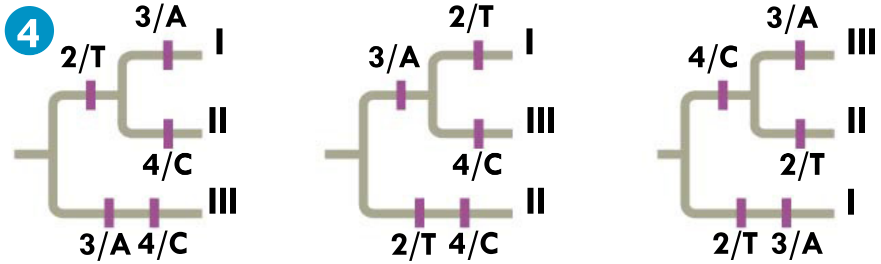
②

	Site 1	2	3	4
Species I	C	T	A	T
Species II	C	T	T	C
Species III	A	G	A	C
Ancestral sequence	A	G	T	T

③



④



Results

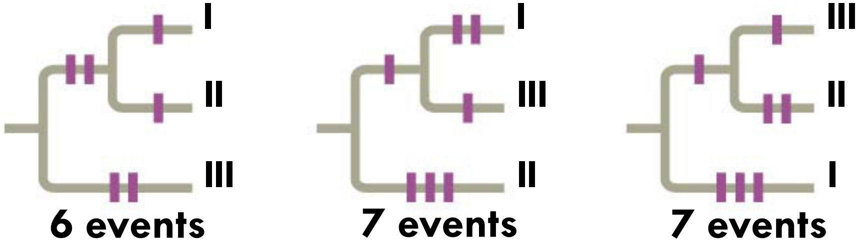
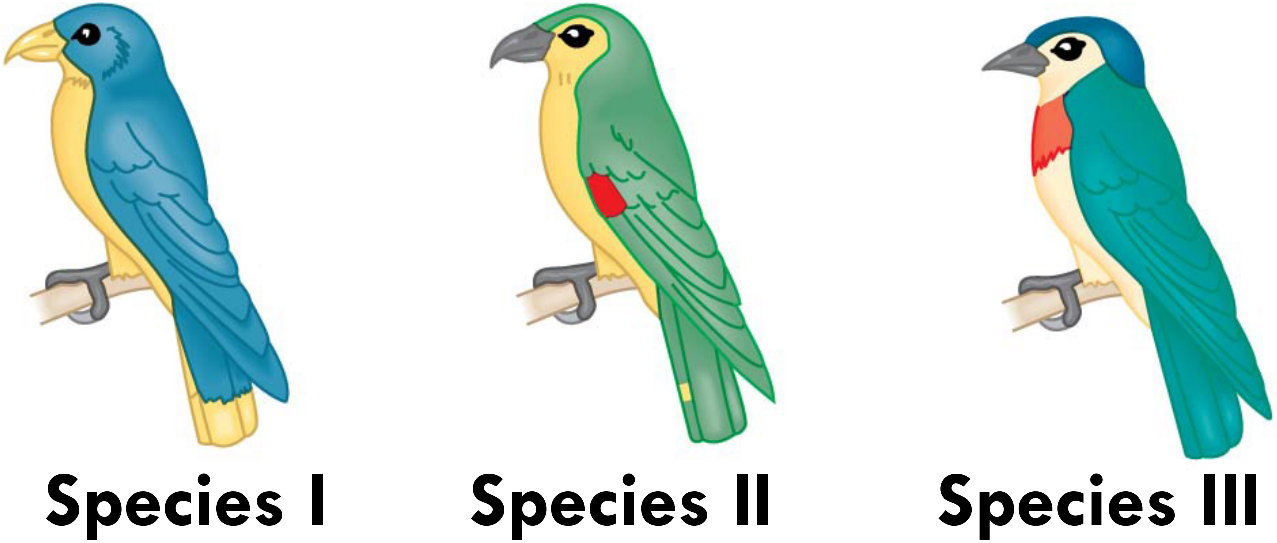


Figure 22.15a

Technique



① **Three phylogenetic hypotheses:**

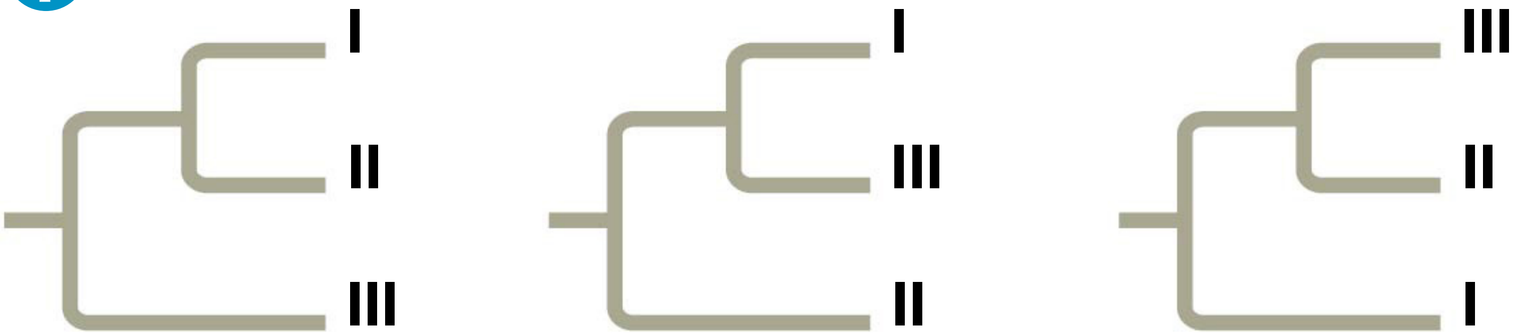


Figure 22.15b

Technique

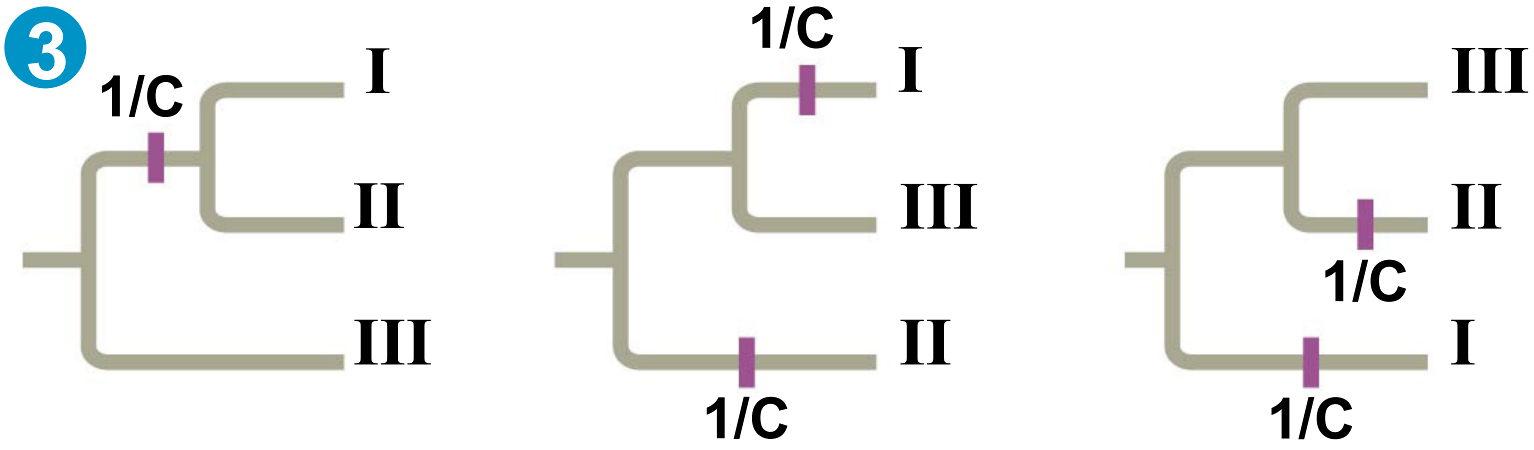
②

	Site			
	1	2	3	4
Species I	C	T	A	T
Species II	C	T	T	C
Species III	A	G	A	C
Ancestral sequence	A	G	T	T

Figure 22.15c

Technique

③



④

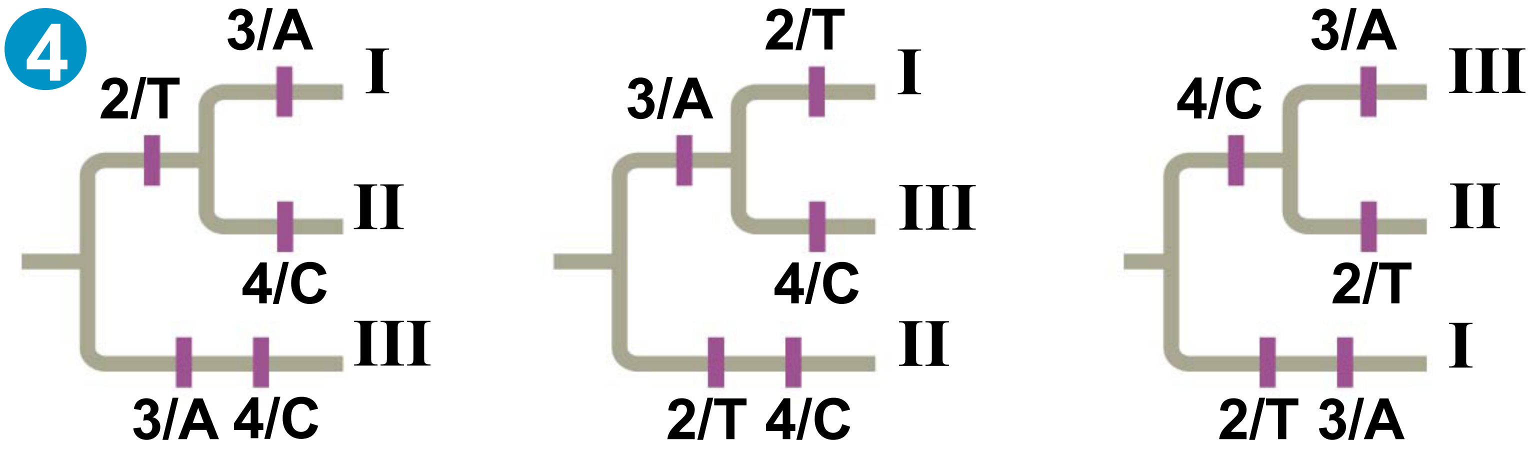
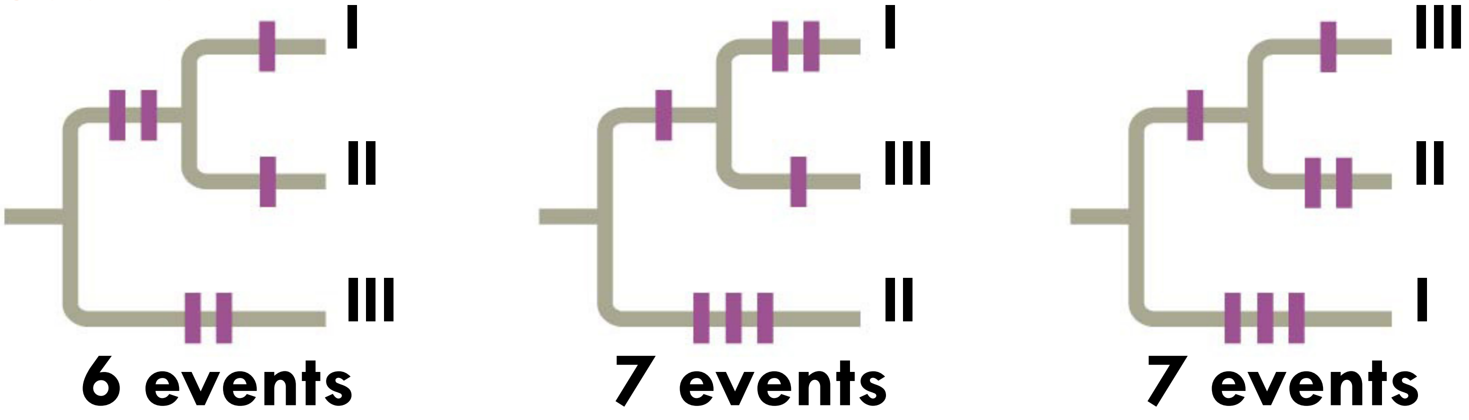


Figure 22.15d

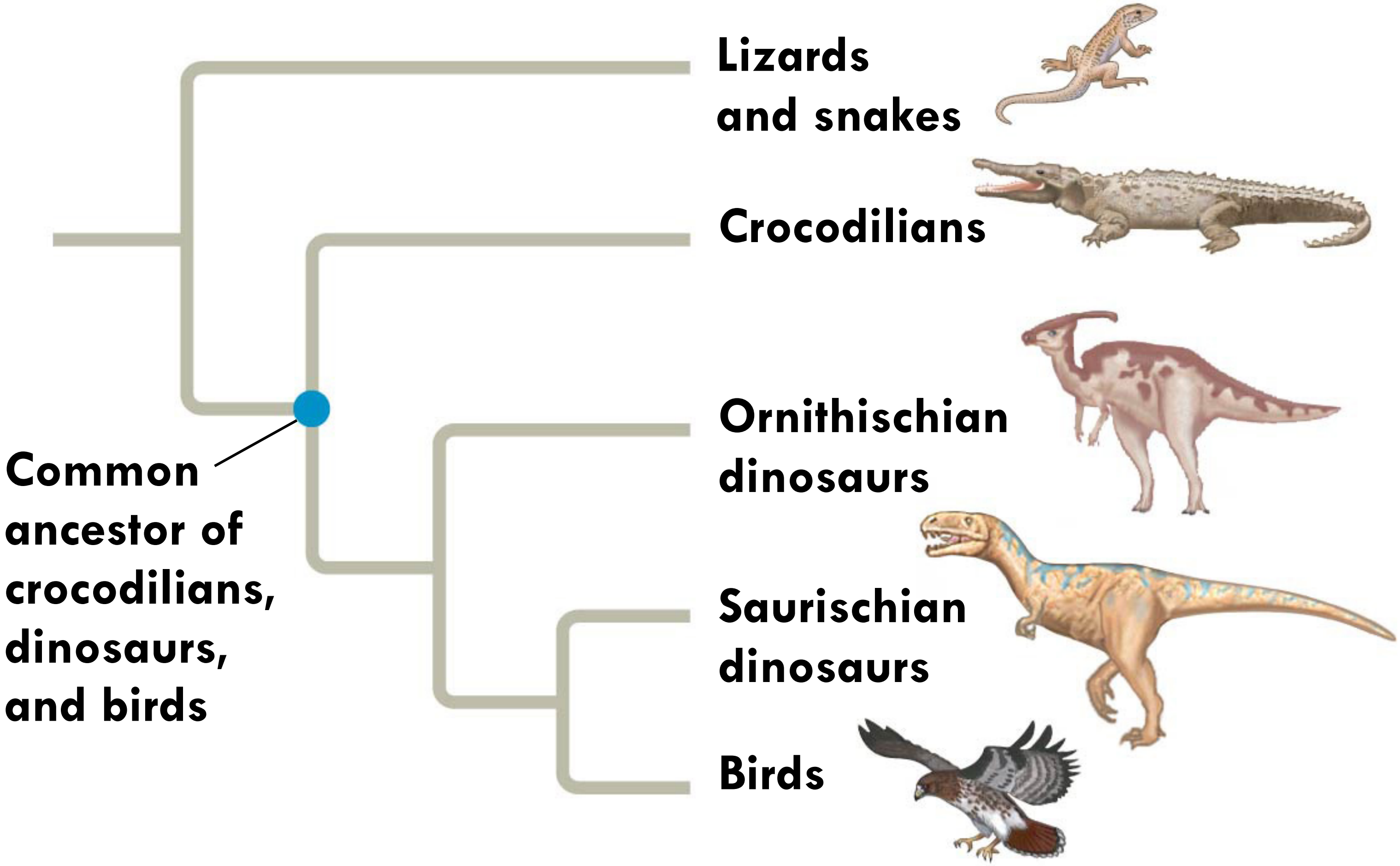
Results



Phylogenetic Trees as Hypotheses

- The best hypotheses for phylogenetic trees fit the most data: morphological, molecular, and fossil
- Phylogenetic bracketing allows us to predict features of an ancestor from features of its descendants
  - For example, phylogenetic bracketing allows us to infer characteristics of dinosaurs

Figure 22.16

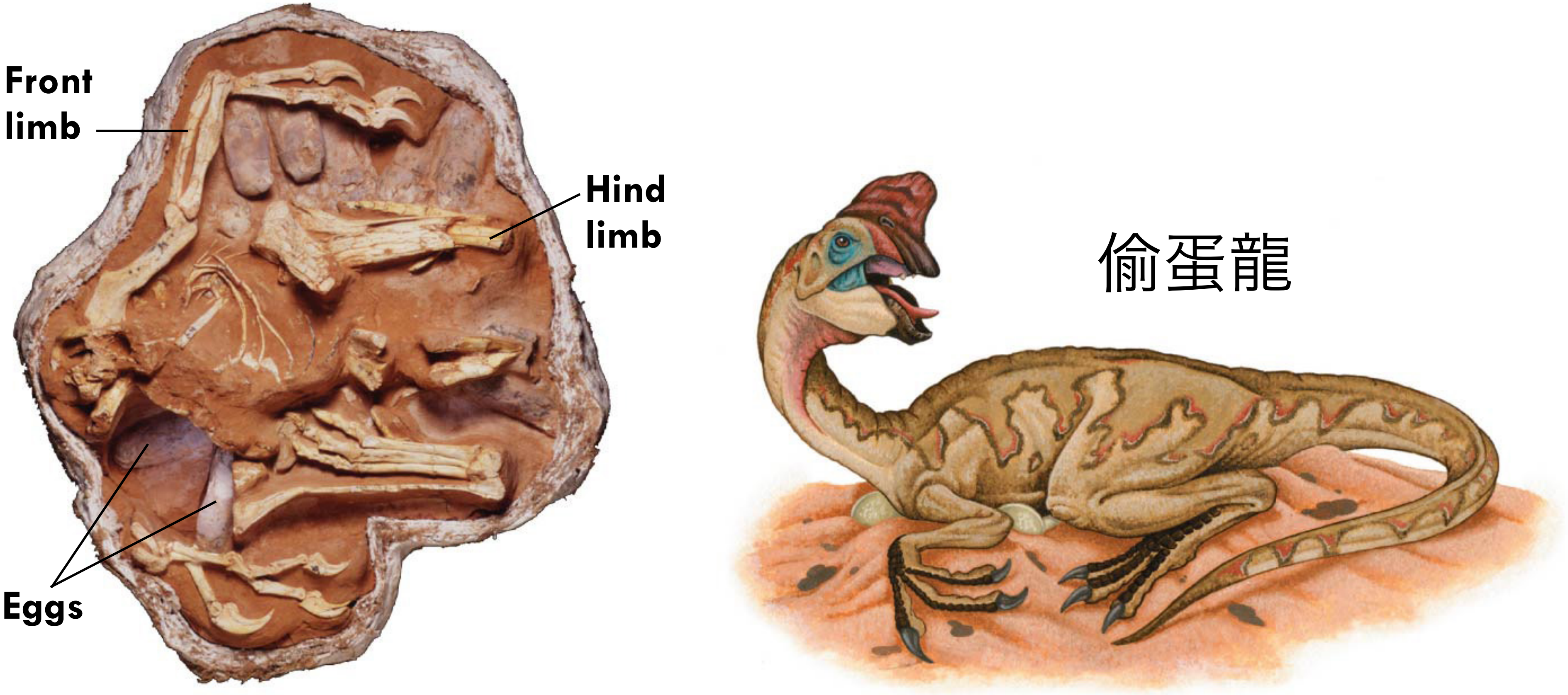

Lizards and snakes
Crocodilians
Ornithischian dinosaurs
Saurischian dinosaurs
Birds
Common ancestor of crocodilians, dinosaurs, and birds

- Birds and crocodiles share several features: four-chambered hearts, song, nest building, and brooding

- These characteristics likely evolved in a common ancestor and were shared by all of its descendants, including dinosaurs

- The fossil record supports nest building and brooding in dinosaurs

Figure 22.17



(a) **Fossil remains of *Oviraptor* and eggs**

(b) **Artist's reconstruction of the dinosaur's posture based on the fossil findings**

Concept 22.4: An organism's evolutionary history is documented in its genome

- Comparing nucleic acids or other molecules to infer relatedness is a valuable approach for tracing organisms' evolutionary history

- DNA that codes for rRNA changes relatively slowly and is useful for investigating branching points hundreds of millions of years ago

- mtDNA evolves rapidly and can be used to explore recent evolutionary events

Gene Duplications and Gene Families

- Gene duplication increases the number of genes in the genome, providing more opportunities for evolutionary changes
- Repeated gene duplications result in gene families
- Like homologous genes, duplicated genes can be traced to a common ancestor

- **Orthologous genes** are found in a single copy in the genome and are homologous between species
- They can diverge only after speciation occurs

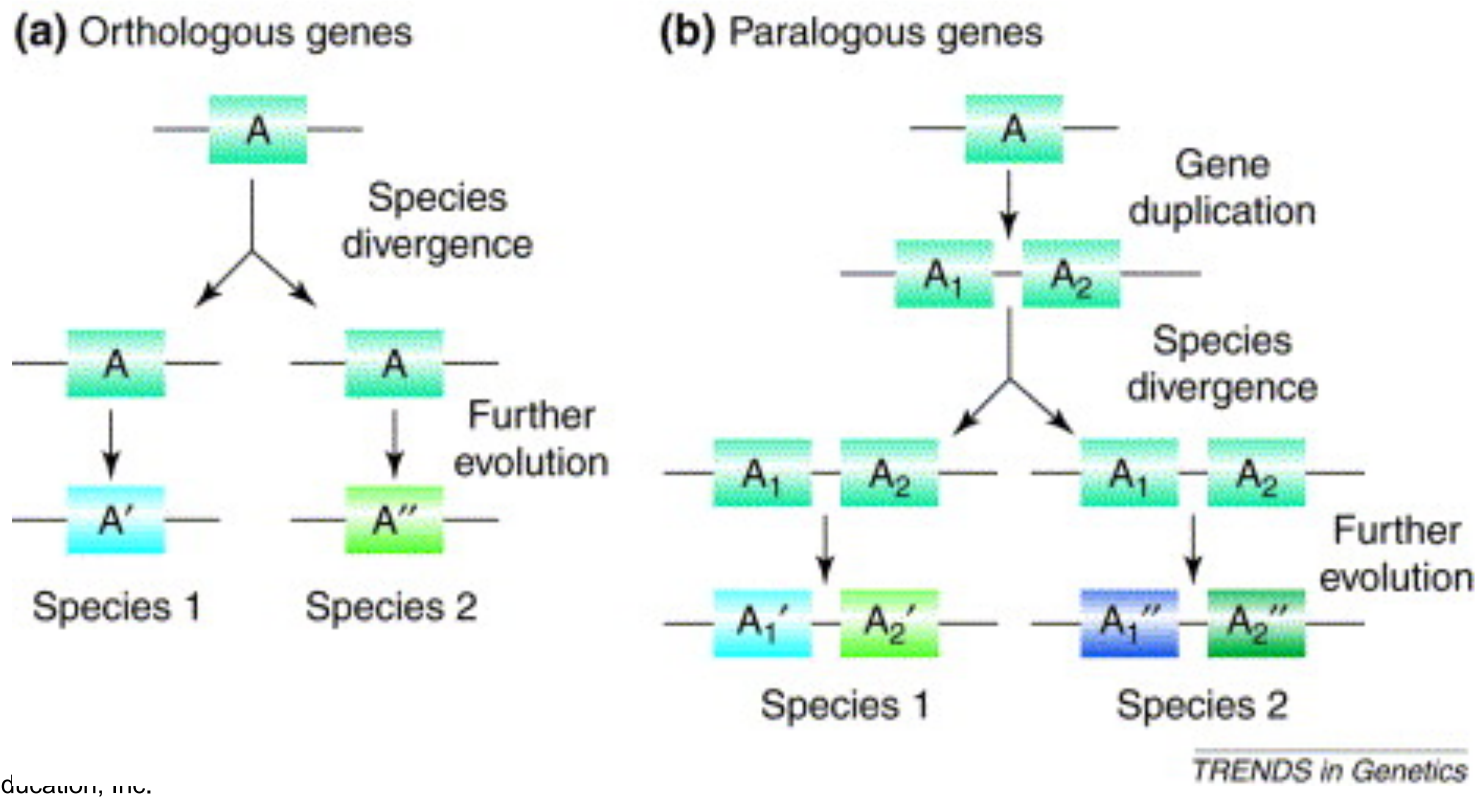
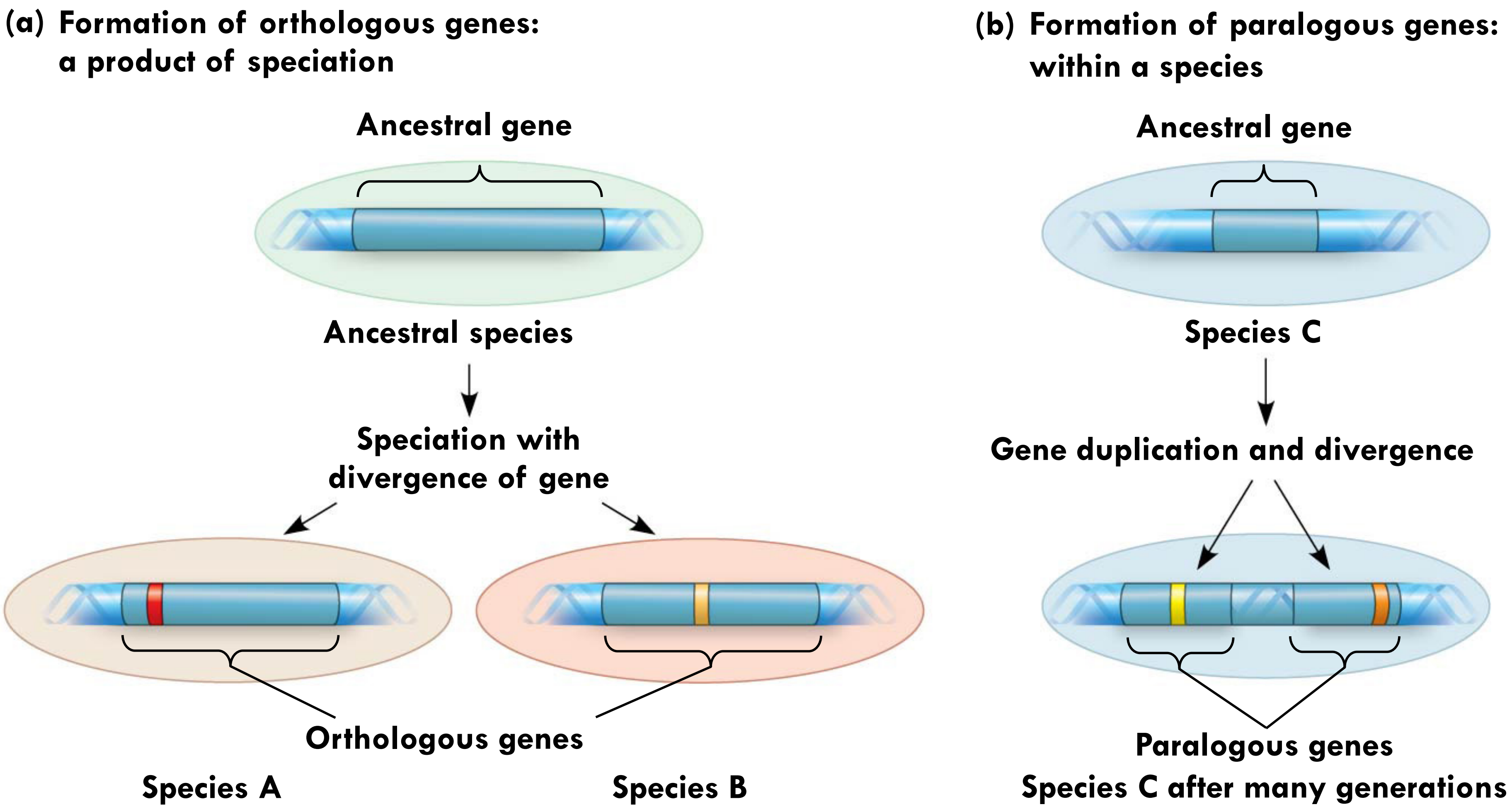


Figure 22.18



(a) Formation of orthologous genes: a product of speciation

Ancestral gene

Ancestral species

Speciation with divergence of gene

Orthologous genes

Species A

Species B

(b) Formation of paralogous genes: within a species

Ancestral gene

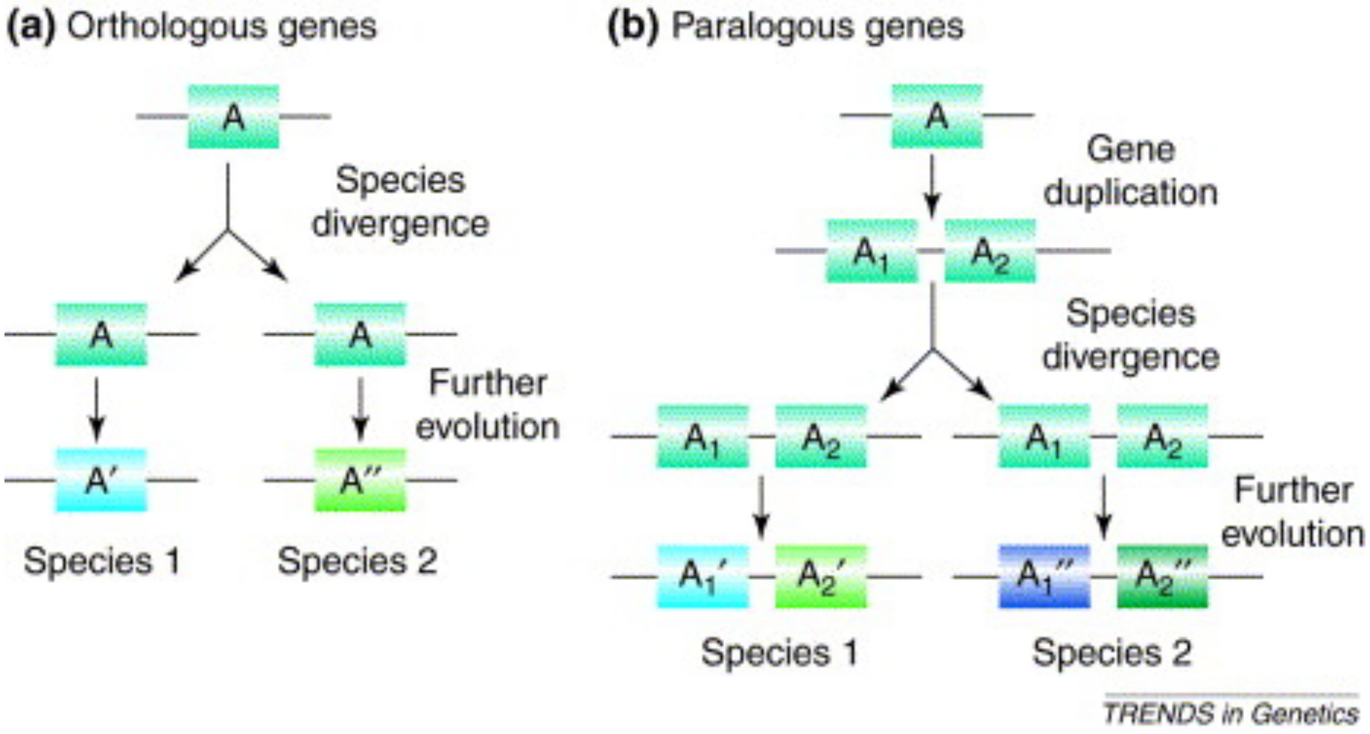
Species C

Gene duplication and divergence

Paralogous genes

Species C after many generations

- **Paralogous genes** result from gene duplication, so are found in more than one copy in the genome
- They can diverge within the clade that carries them and often evolve new functions



TRENDS in Genetics

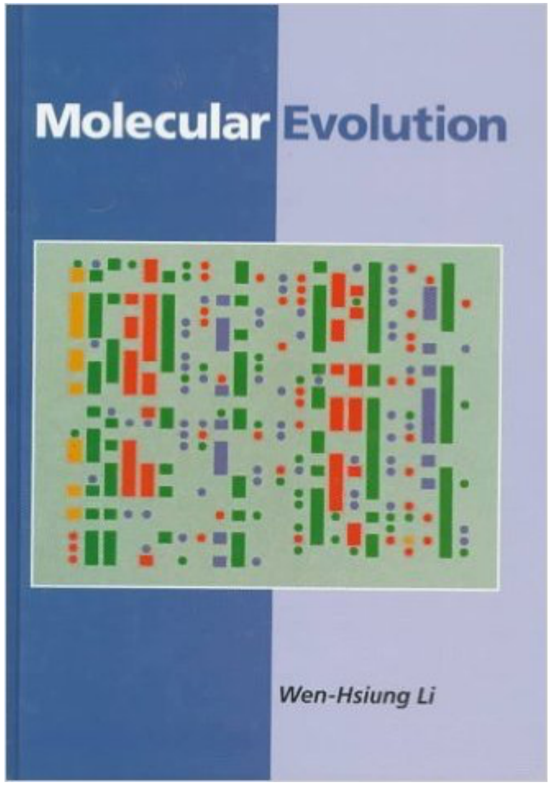
Genome Evolution

- Orthologous genes are widespread and extend across many widely varied species
  - For example, humans and mice diverged about 65 million years ago, and 99% of our genes are orthologous

- Gene number and the complexity of an organism are not strongly linked
    - For example, humans have only four times as many genes as yeast, a single-celled eukaryote
- Genes in complex organisms appear to be very versatile, and each gene can encode multiple proteins that perform many different functions

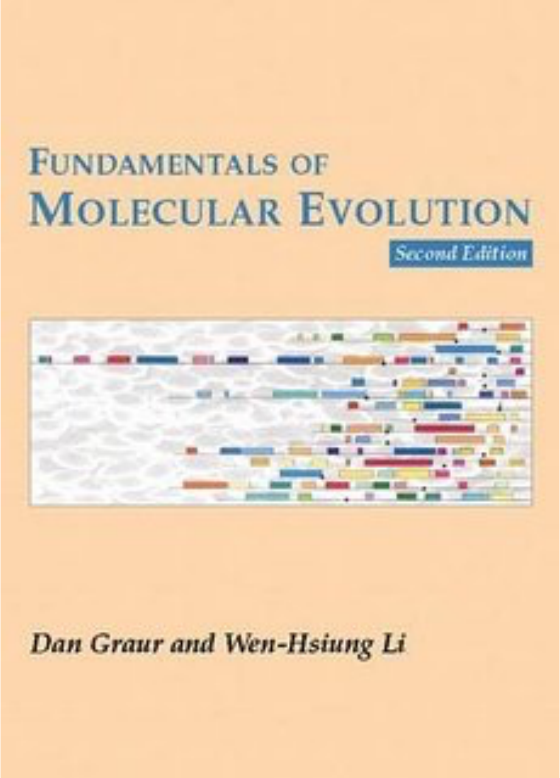
Concept 22.5: Molecular clocks help track evolutionary time

- To extend phylogenies beyond the fossil record, we must make an assumption about how molecular change occurs over time



Molecular Evolution

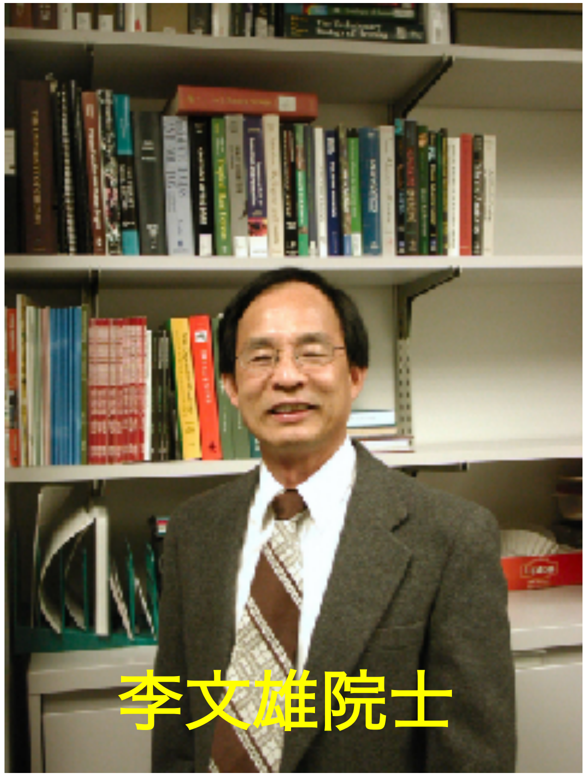
Wen-Hsiung Li



FUNDAMENTALS OF MOLECULAR EVOLUTION

Second Edition

Dan Graur and Wen-Hsiung Li



李文雄院士

Molecular Clocks

- A **molecular clock** uses constant rates of evolution in some genes to estimate the absolute time of evolutionary change
- In orthologous genes, nucleotide substitutions are assumed to be proportional to the time since they last shared a common ancestor
- In paralogous genes, nucleotide substitutions are proportional to the time since the genes became duplicated

- Molecular clocks are calibrated against branches whose dates are known from the fossil record
- Individual genes vary in how clocklike they are

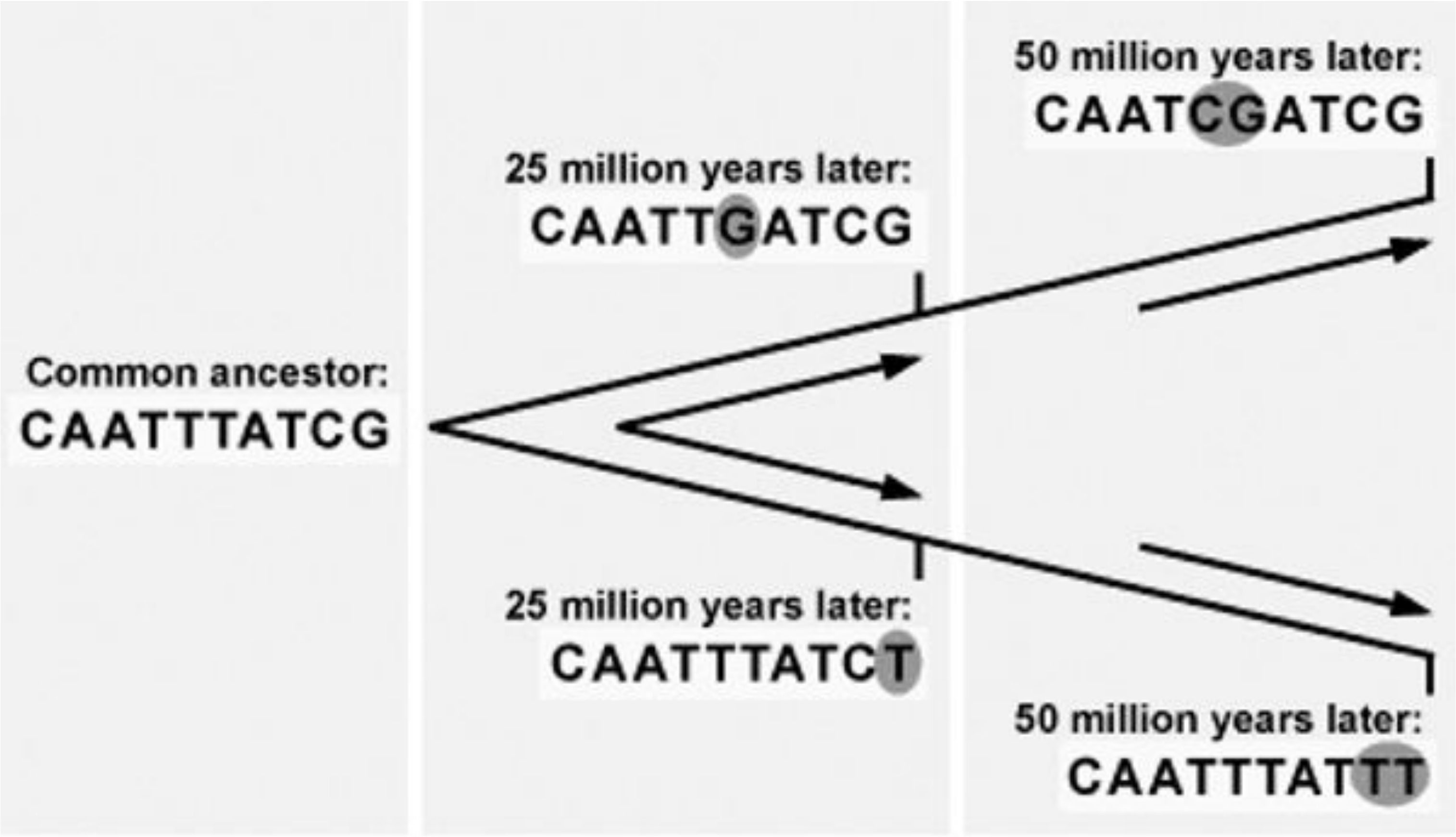
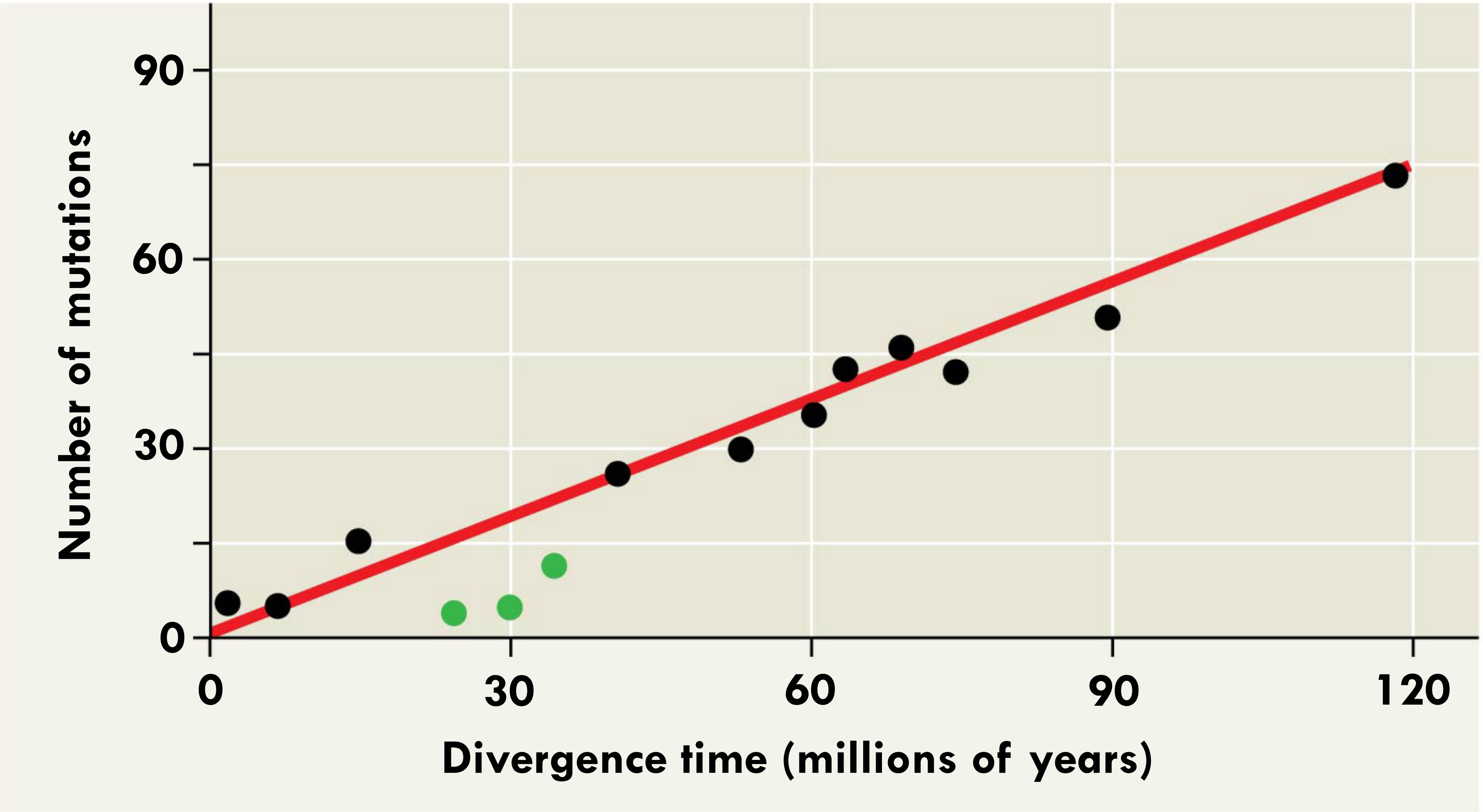


Figure 22.19


90
60
30
0
Number of mutations
0
30
60
90
120
Divergence time (millions of years)

Differences in Clock Speed

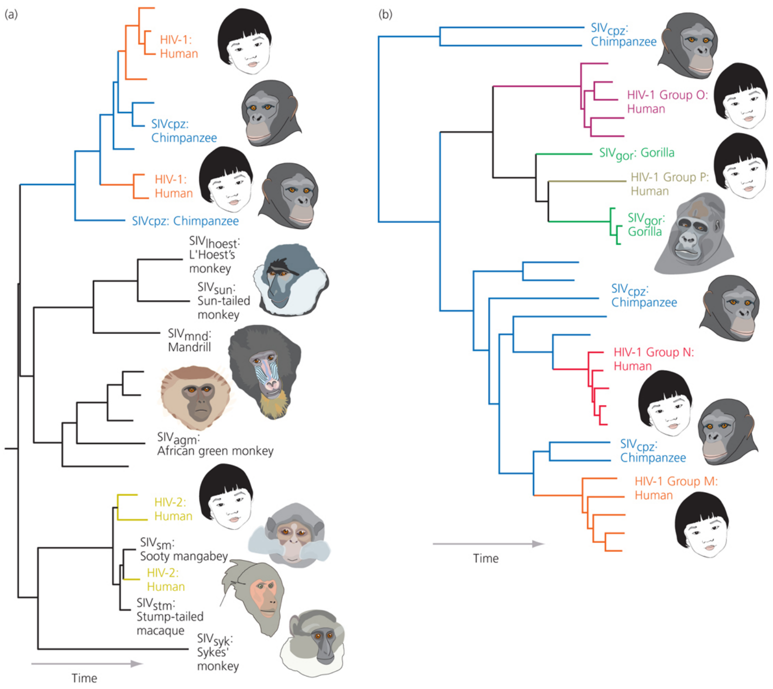
- If most of the evolutionary change in genes and proteins has no effect on fitness then the rate of molecular change should be regular like a clock
- Differences in clock rate for different genes are a function of the importance of the gene and how critical the specific amino acid is to protein function

Potential Problems with Molecular Clocks

- The molecular clock does not run as smoothly as expected if mutations were neutral
- Irregularities result from natural selection in which some DNA changes are favored over others
- Estimates of evolutionary divergences older than the fossil record have a high degree of uncertainty
- The use of multiple genes or genes that evolved in different taxa may improve estimates

Applying a Molecular Clock: Dating the Origin of HIV

- Phylogenetic analysis shows that HIV is descended from viruses that infect chimpanzees and other primates
- HIV spread to humans more than once
- Comparison of HIV samples shows that the virus evolved in a very clocklike way



- Application of a molecular clock to one strain of HIV suggests that that strain spread to humans during the 1930s

- A more advanced molecular clock approach estimated the first spread to humans around 1910

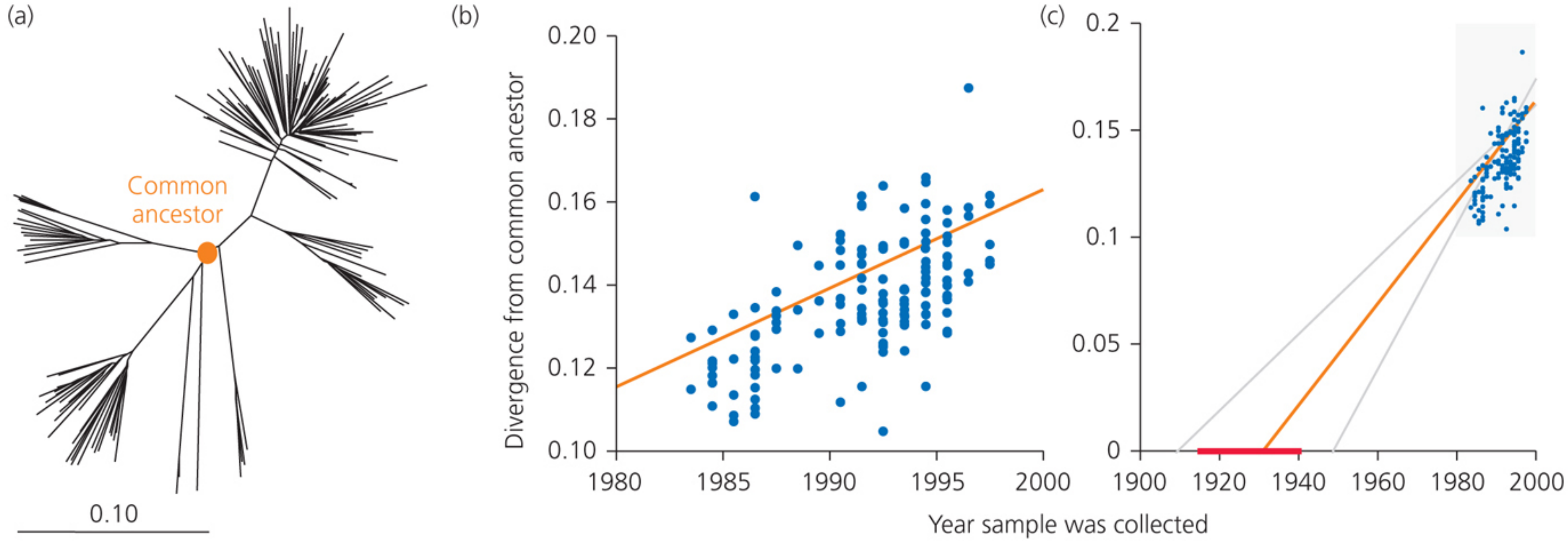
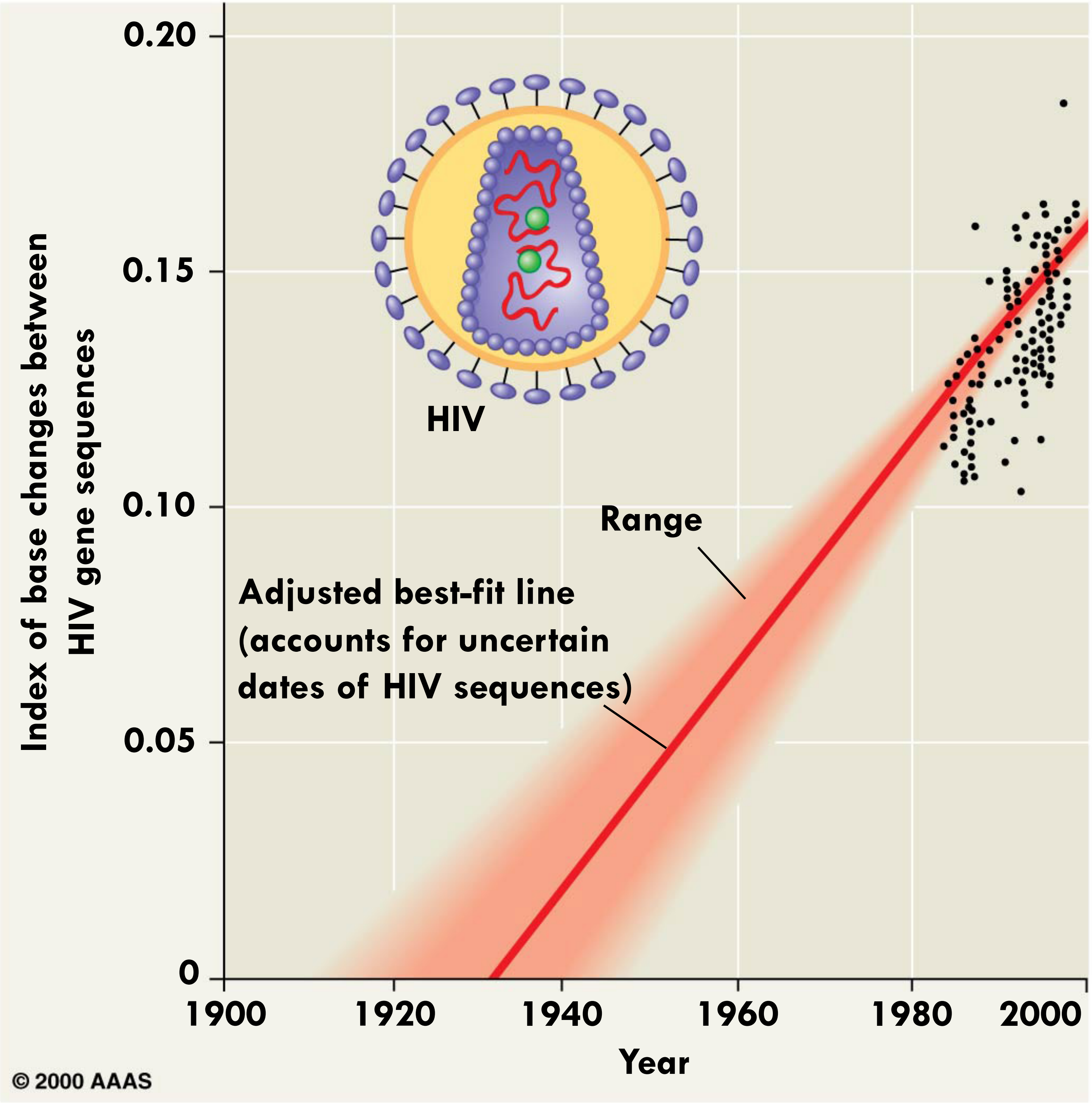


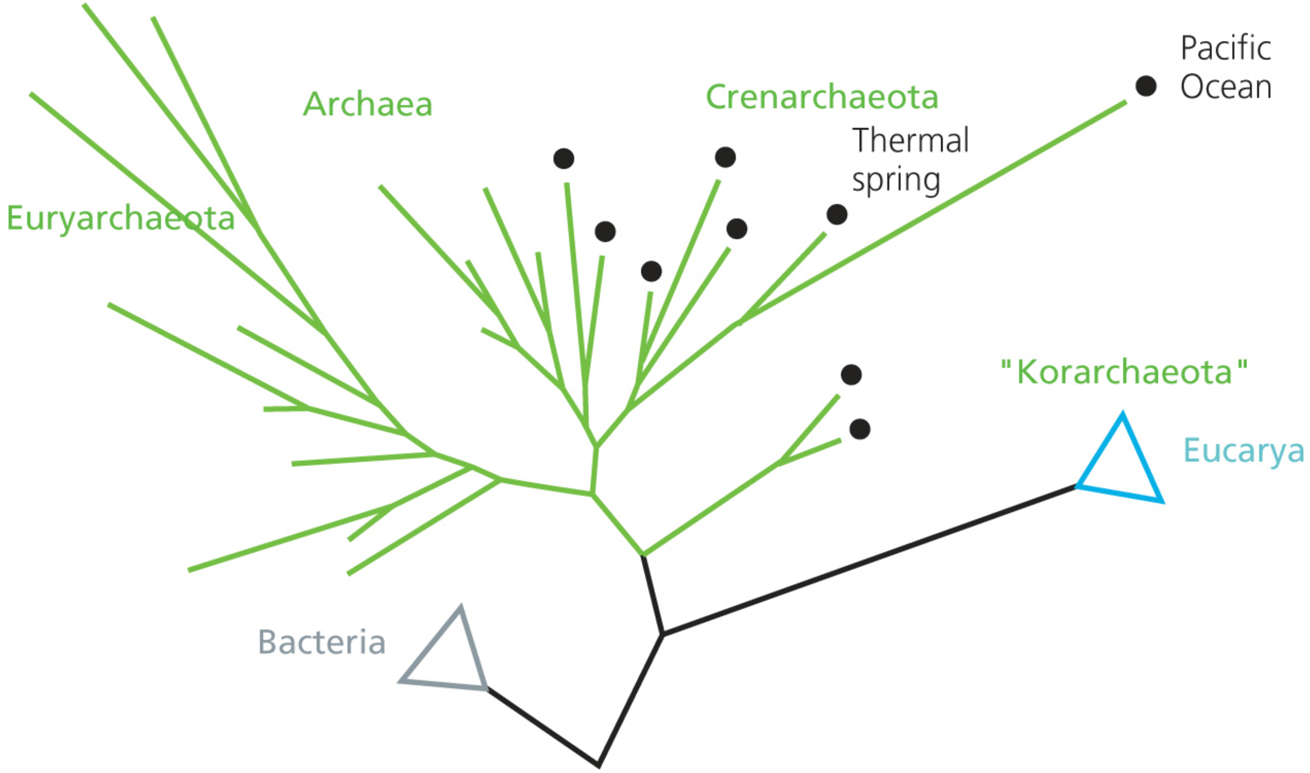
Figure 22.20



0.20
0.15
0.10
0.05
0
Index of base changes between HIV gene sequences
HIV
Range
Adjusted best-fit line (accounts for uncertain dates of HIV sequences)
1900
1920
1940
1960
1980
2000
Year
© 2000 AAAS

Concept 22.6: Our understanding of the tree of life continues to change based on new data

- Recently, we have gained insight into the very deepest branches of the tree of life through molecular systematics

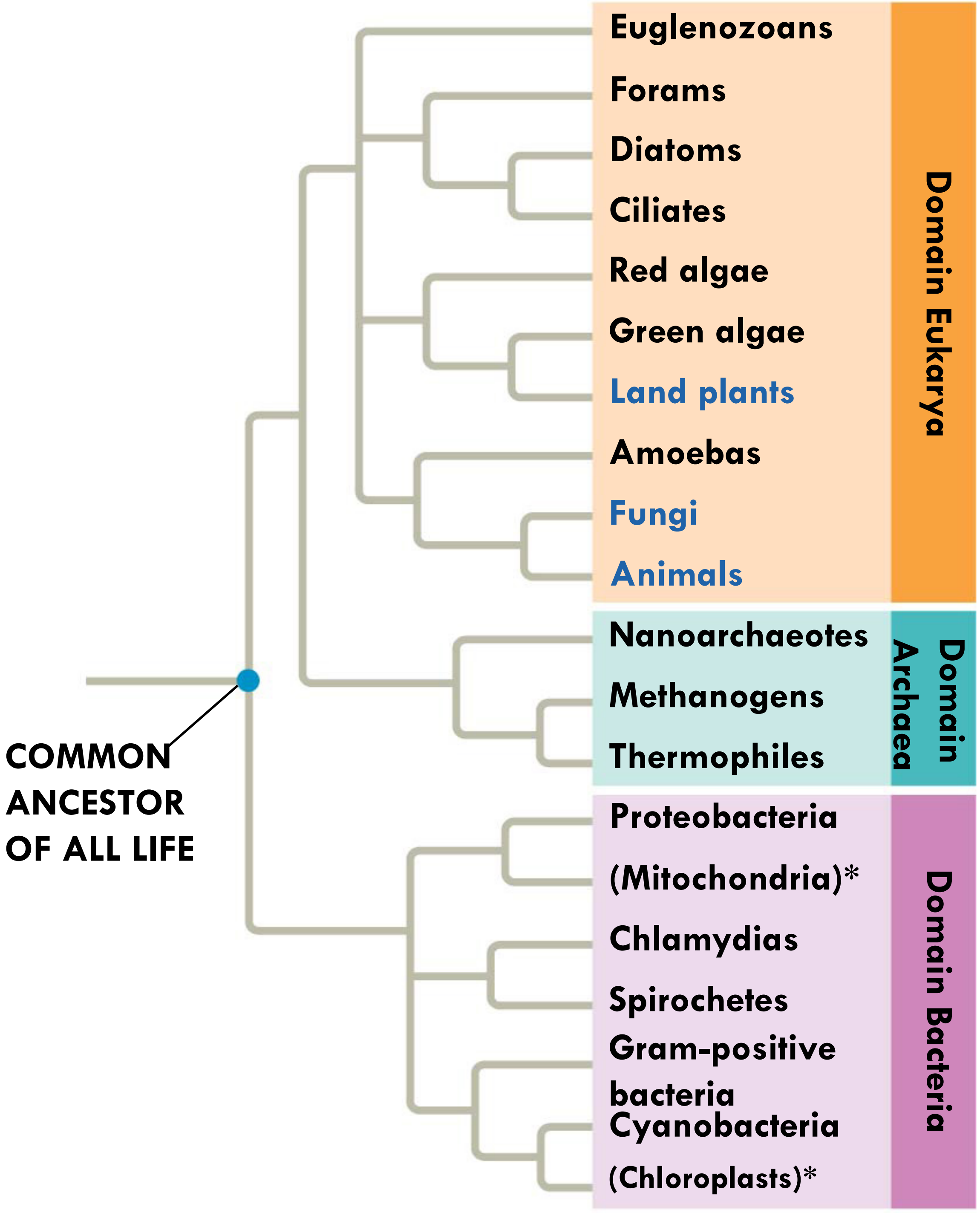


An estimate of the phylogeny of all extant organism based on rRNA

From Two Kingdoms to Three Domains

- Early taxonomists classified all species as either plants or animals
- Later, five kingdoms were recognized: Monera (prokaryotes), Protista, Plantae, Fungi, and Animalia
- More recently, the three-domain system has been adopted: Bacteria, Archaea, and Eukarya
- The three-domain system is supported by data from many sequenced genomes

Figure 22.21



Euglenozoans
Forams
Diatoms
Ciliates
Red algae
Green algae
Land plants
Amoebas
Fungi
Animals

Domain Eukarya

Nanoarchaeotes
Methanogens
Thermophiles

Domain Archaea

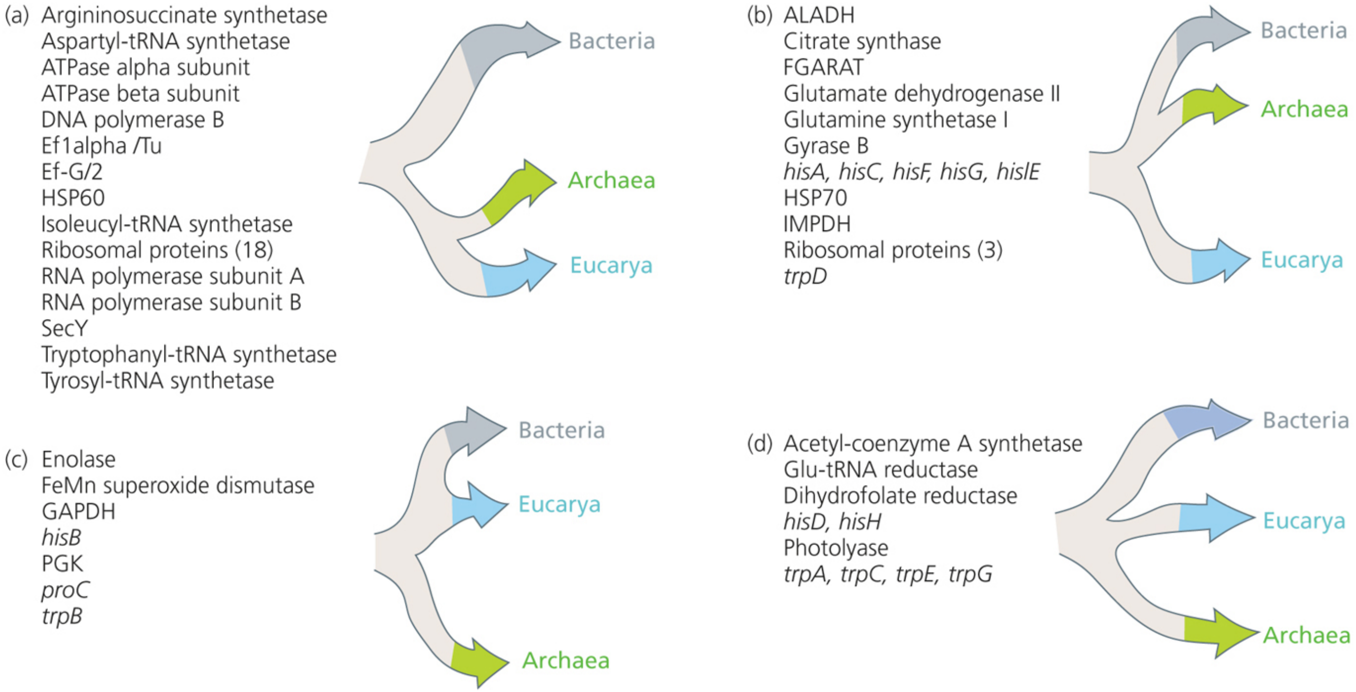
Proteobacteria
(Mitochondria)*
Chlamydias
Spirochetes
Gram-positive bacteria
Cyanobacteria
(Chloroplasts)*

Domain Bacteria

COMMON ANCESTOR OF ALL LIFE

The Important Role of Horizontal Gene Transfer

- The tree of life suggests that eukaryotes and archaea are more closely related to each other than to bacteria
- The tree of life is based largely on rRNA genes; however, some other genes reveal different relationships



(a) Argininosuccinate synthetase
Aspartyl-tRNA synthetase
ATPase alpha subunit
ATPase beta subunit
DNA polymerase B
Ef1alpha /Tu
Ef-G/2
HSP60
Isoleucyl-tRNA synthetase
Ribosomal proteins (18)
RNA polymerase subunit A
RNA polymerase subunit B
SecY
Tryptophanyl-tRNA synthetase
Tyrosyl-tRNA synthetase

(b) ALADH
Citrate synthase
FGARAT
Glutamate dehydrogenase II
Glutamine synthetase I
Gyrase B
hisA, hisC, hisF, hisG, hisIE
HSP70
IMPDH
Ribosomal proteins (3)
trpD

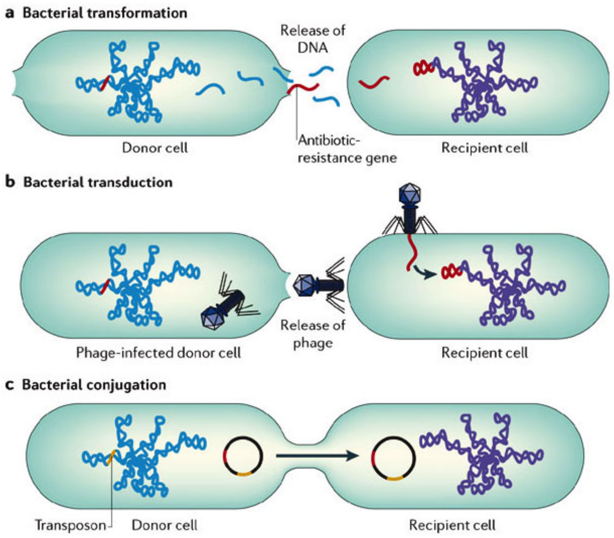
(c) Enolase
FeMn superoxide dismutase
GAPDH
hisB
PGK
proC
trpB

(d) Acetyl-coenzyme A synthetase
Glu-tRNA reductase
Dihydrofolate reductase
hisD, hisH
Photolyase
trpA, trpC, trpE, trpG

- There have been substantial interchanges of genes between organisms in different domains

- **Horizontal gene transfer** is the movement of genes from one genome to another

- Horizontal gene transfer occurs by exchange of transposable elements and plasmids, viral infection, and fusion of organisms



Copyright © 2006 Nature Publishing Group
Nature Reviews | Microbiology

- Disparities between gene trees can be explained by the occurrence of horizontal gene transfer
- Horizontal gene transfer has played a key role in the evolution of both prokaryotes and eukaryotes

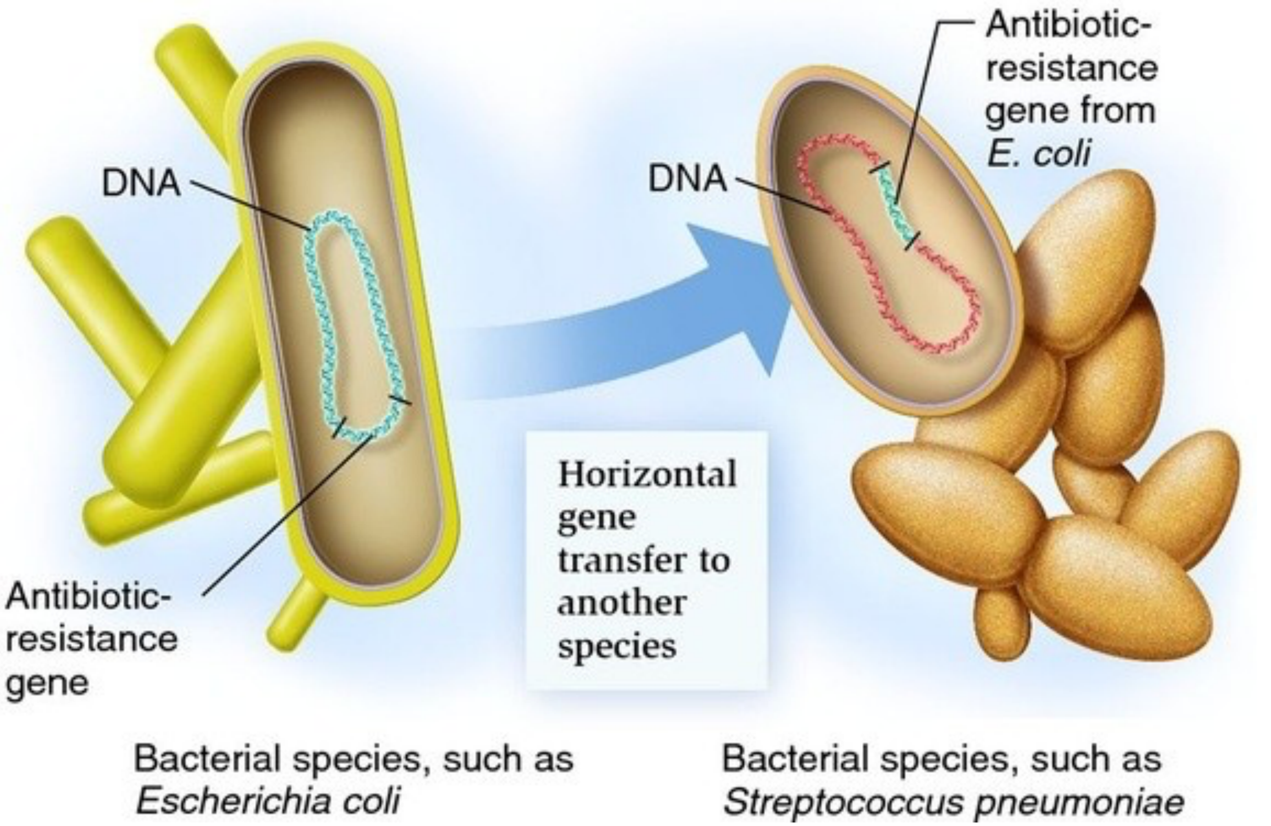


Figure 22.22

Galdieria sulphuraria

- Some biologists argue that horizontal gene transfer was so common that the early history of life should be represented as a tangled network of connected branches

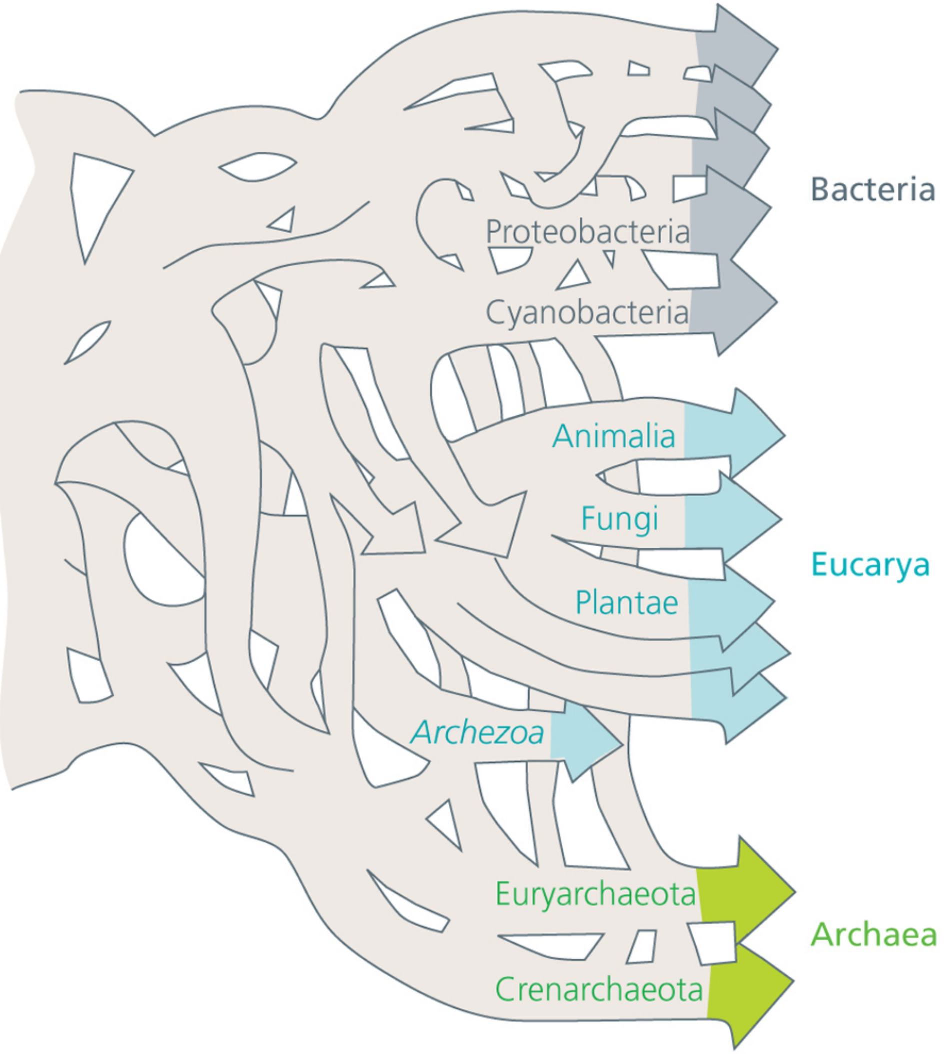
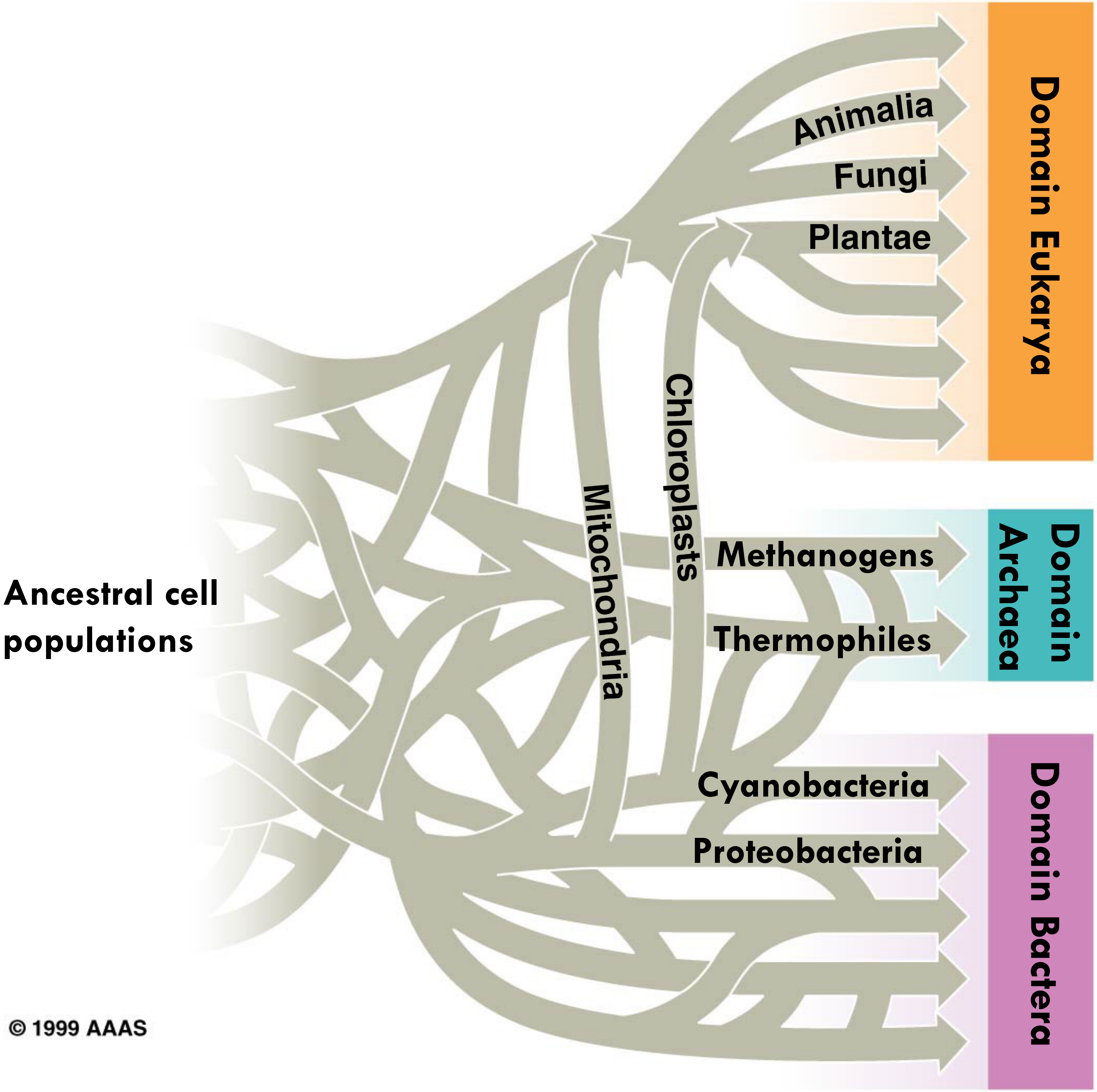


Figure 22.23


Domain Eukarya
Animalia
Fungi
Plantae
Mitochondria
Chloroplasts
Methanogens
Thermophiles
Domain Archaea
Ancestral cell populations
Cyanobacteria
Proteobacteria
Domain Bactera
© 1999 AAAS