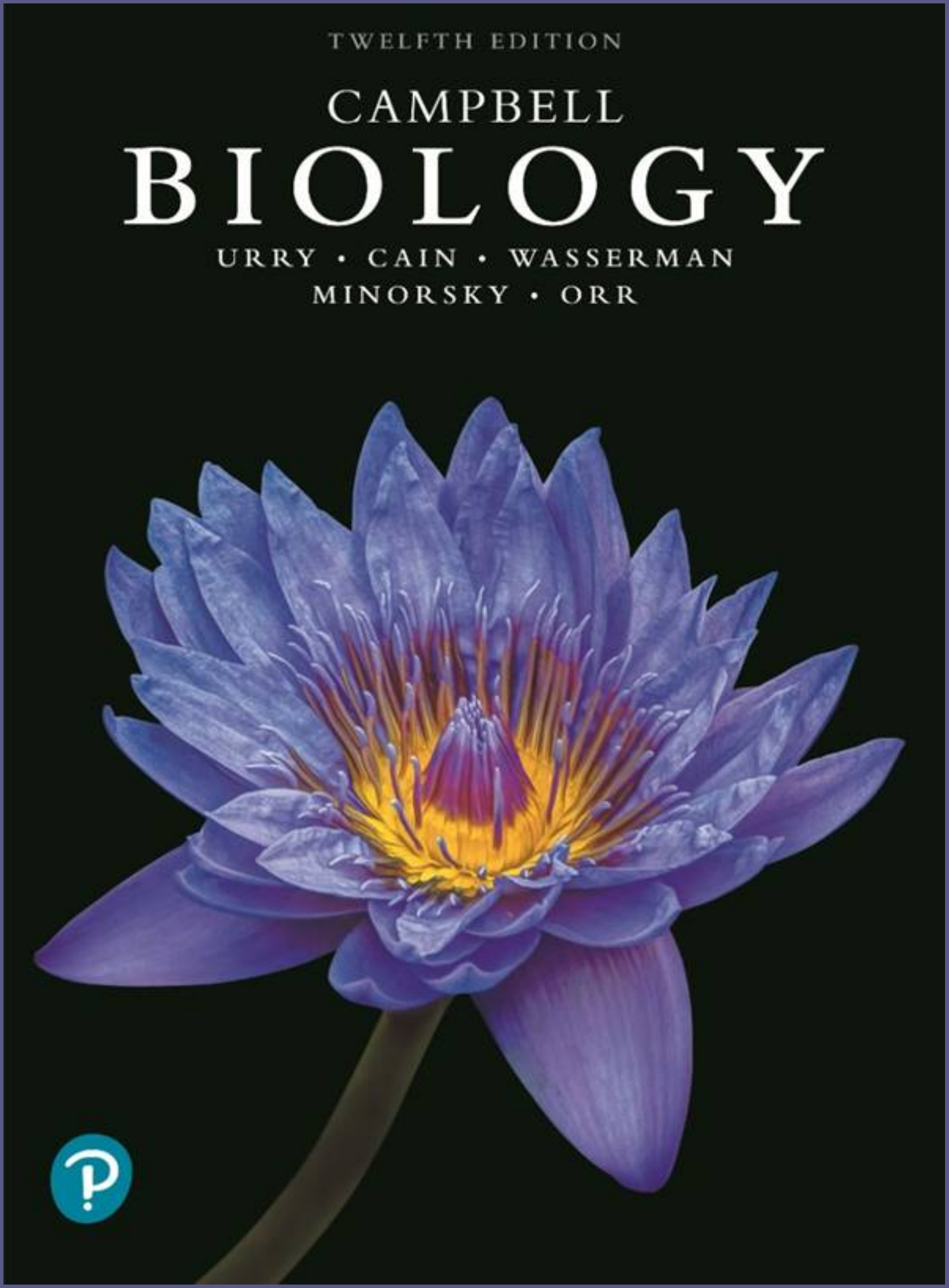


Chapter 26

Phylogeny and the Tree of Life

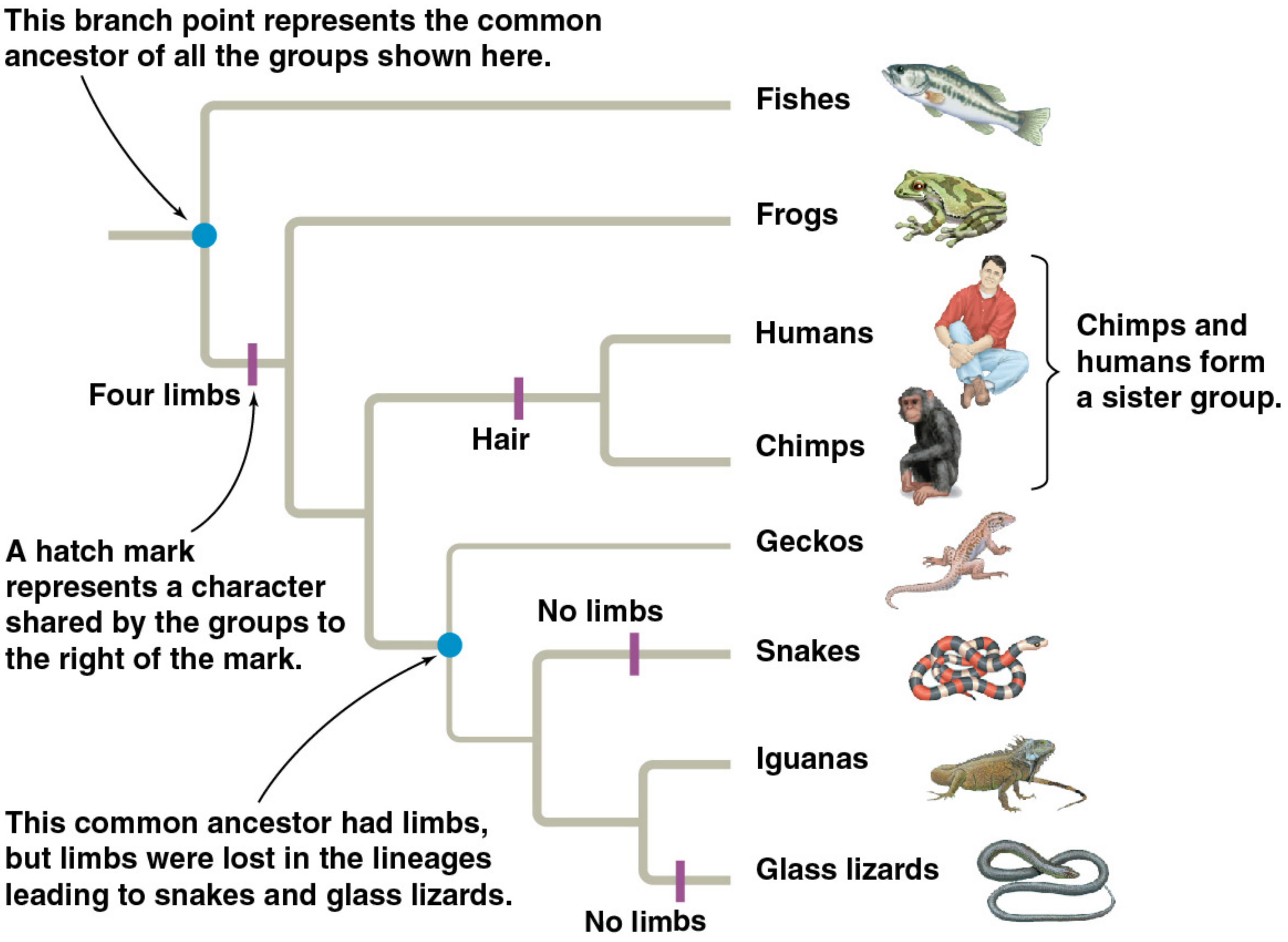
Lecture Presentations by
Nicole Tunbridge and
Kathleen Fitzpatrick

Figure 26.1a

Figure 26.1b

How do biologists distinguish and categorize the millions of species on Earth?

Traits shared due to common ancestry are used to classify organisms into groups that reflect their evolutionary history:



This branch point represents the common ancestor of all the groups shown here.

A hatch mark represents a character shared by the groups to the right of the mark.

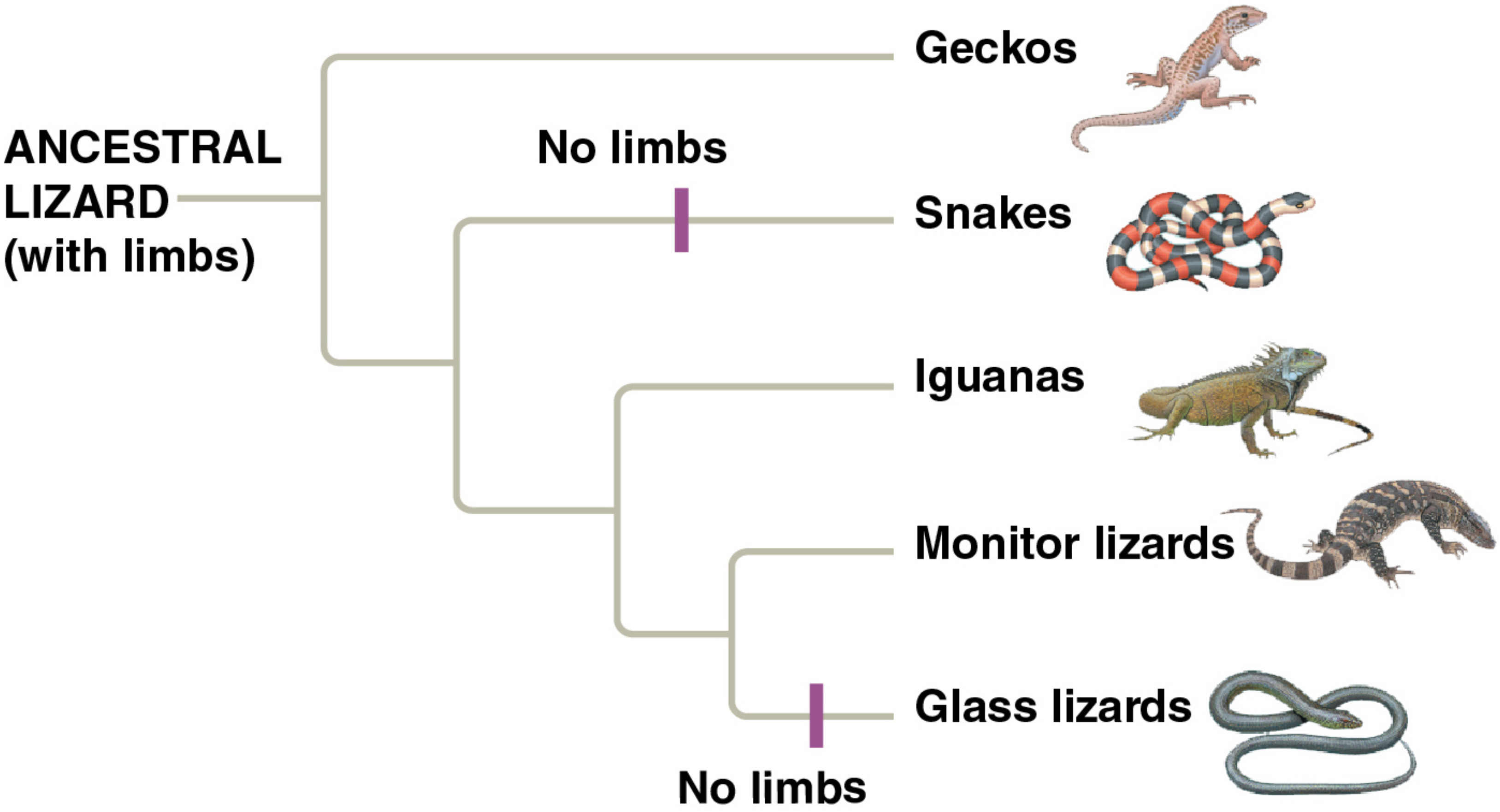
This common ancestor had limbs, but limbs were lost in the lineages leading to snakes and glass lizards.

Chimps and humans form a sister group.

CONCEPT 26.1 Phylogenies show evolutionary relationships

- **Phylogeny** is the evolutionary history of a species or group of related species
- **Systematics** is a discipline focused on classifying organisms and determining their evolutionary relationships

Figure 26.2


ANCESTRAL LIZARD (with limbs)
No limbs
Geckos
Snakes
Iguanas
Monitor lizards
Glass lizards
No limbs

Binomial Nomenclature

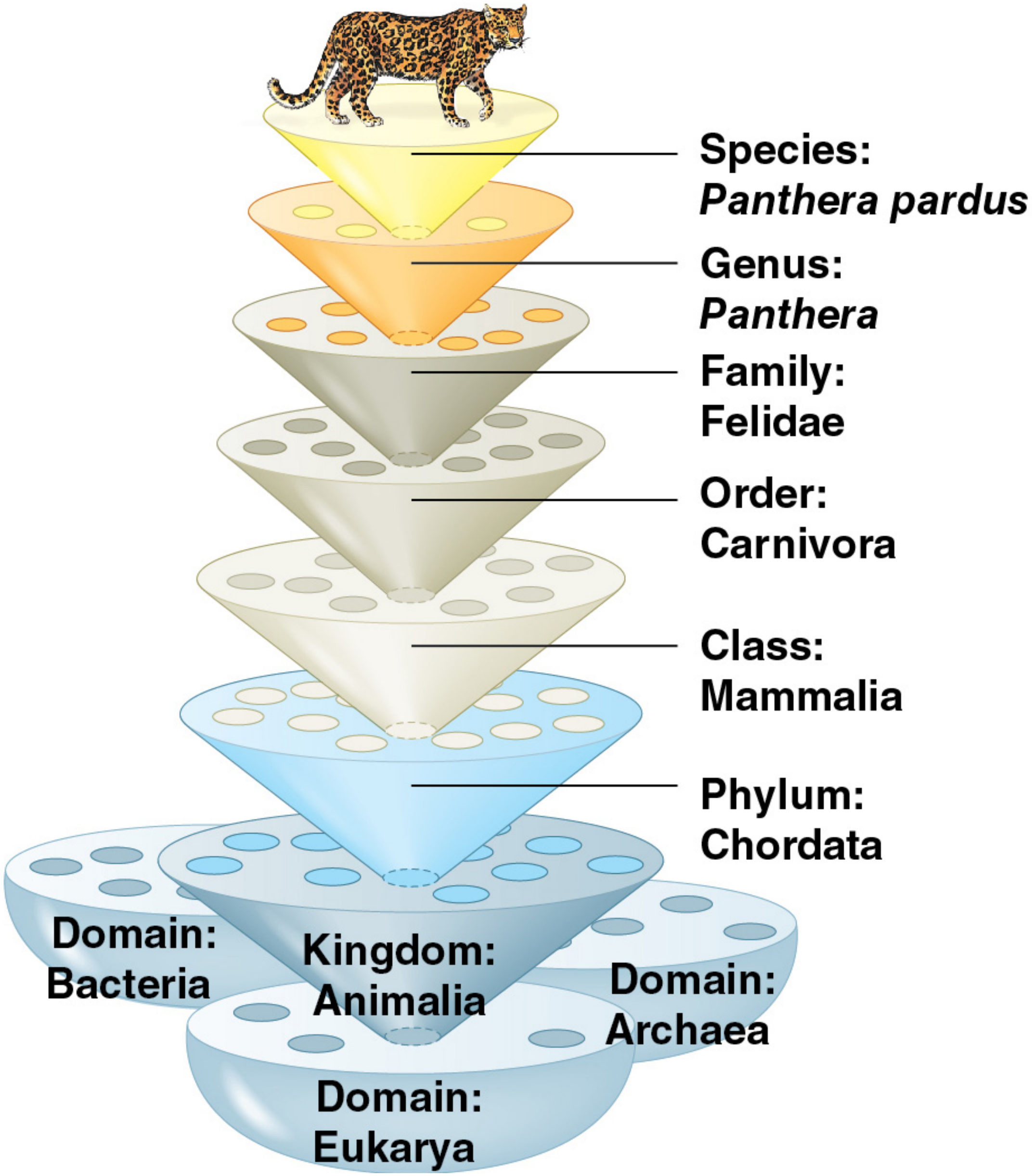
- In the 18th century, Carolus Linnaeus published a system of taxonomy based on resemblances
- Two key features of his system remain useful today: two-part names for species and hierarchical classification

- The two-part scientific name of a species is a **binomial**
- The first part of the name is the **genus**
- The second part, 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

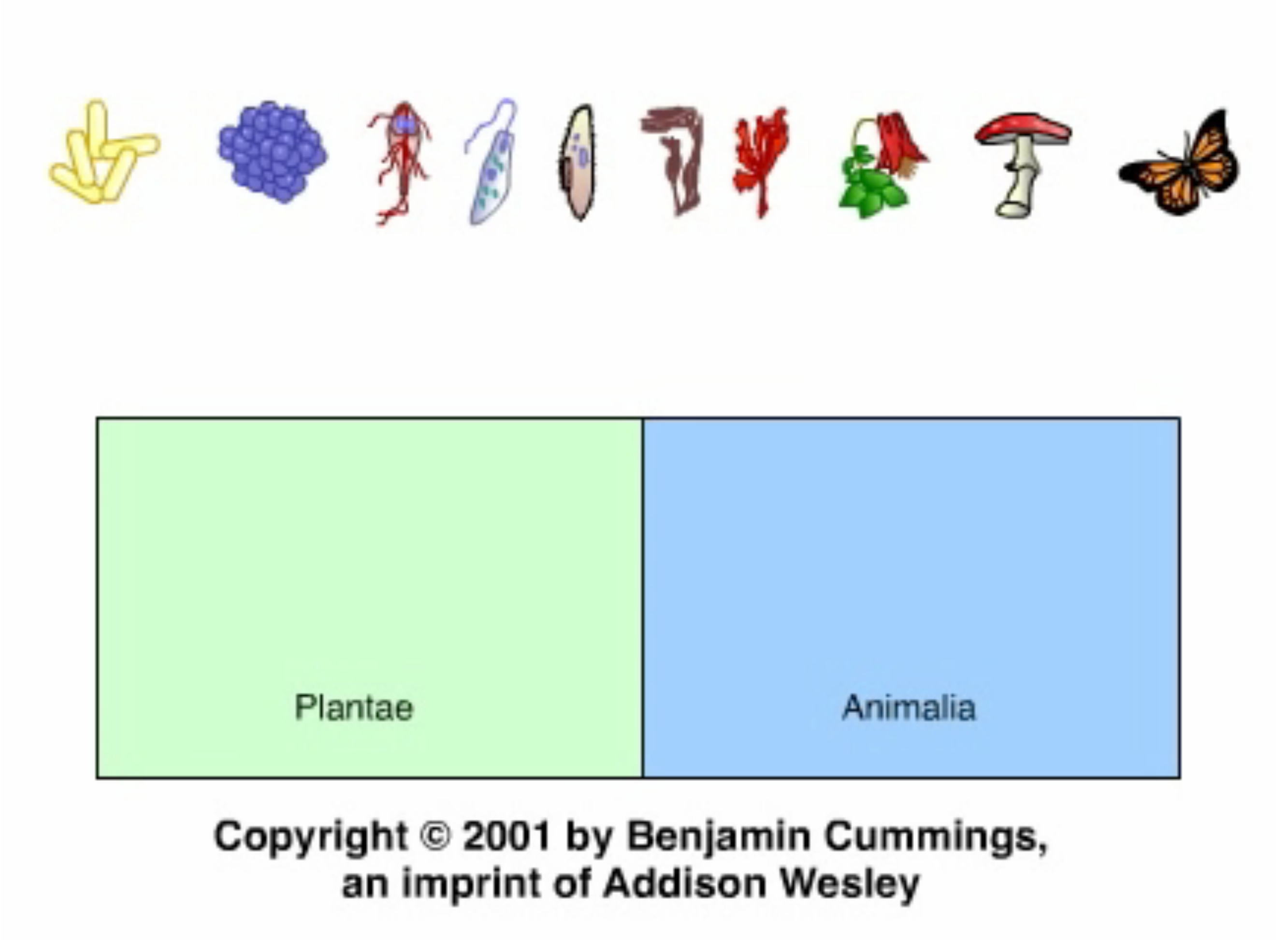
Hierarchical Classification

- Linnaeus introduced a system for grouping species in increasingly inclusive categories
- The taxonomic groups from broad to narrow are **domain**, **kingdom**, **phylum** (plural, *phyla*), **class**, **order**, **family**, genus, and species
- A group at any level of hierarchy is called a **taxon**

Figure 26.3


Species:
Panthera pardus
Genus:
Panthera
Family:
Felidae
Order:
Carnivora
Class:
Mammalia
Phylum:
Chordata
Domain:
Bacteria
Kingdom:
Animalia
Domain:
Archaea
Domain:
Eukarya

Animation: Classification Schemes



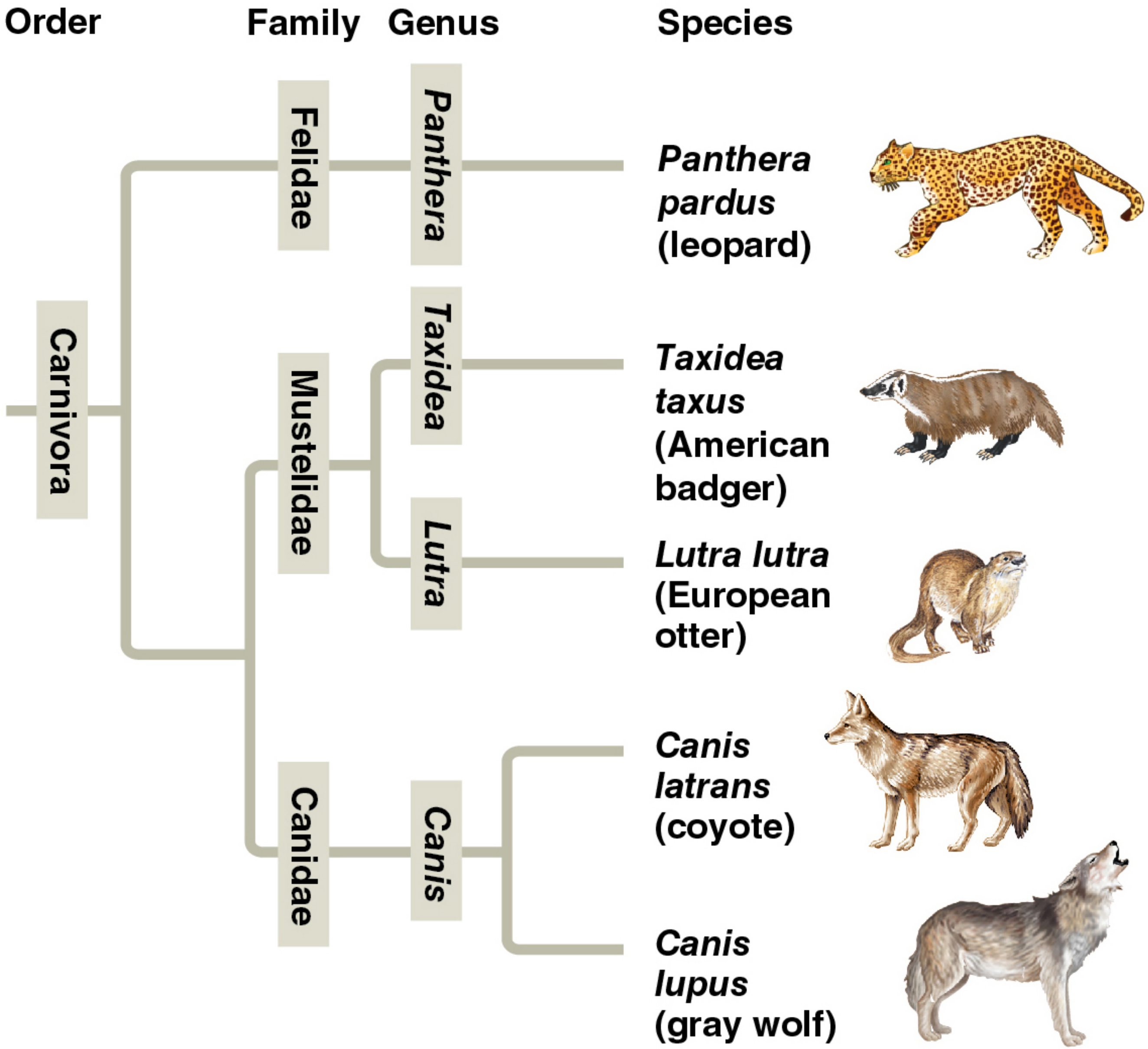
Copyright © 2001 by Benjamin Cummings,
an imprint of Addison Wesley

- The broader taxa are not comparable between lineages
  - For example, an order of snails does not exhibit the same degree of morphological or genetic diversity as an order of mammals

Linking Classification and Phylogeny

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

Figure 26.4



Order

Family

Genus

Species

Carnivora

Felidae

Panthera

Panthera pardus
(leopard)

Mustelidae

Taxidea

Taxidea taxus
(American badger)

Lutra

Lutra lutra
(European otter)

Canidae

Canis

Canis latrans
(coyote)

Canis lupus
(gray wolf)

- Placement of groups within the Linnaean system of classification does not always reflect evolutionary relationships
  - For example, if a species has lost a key feature shared by its close relatives, it may be misclassified
  - DNA or other new evidence may support reclassification to reflect evolutionary history

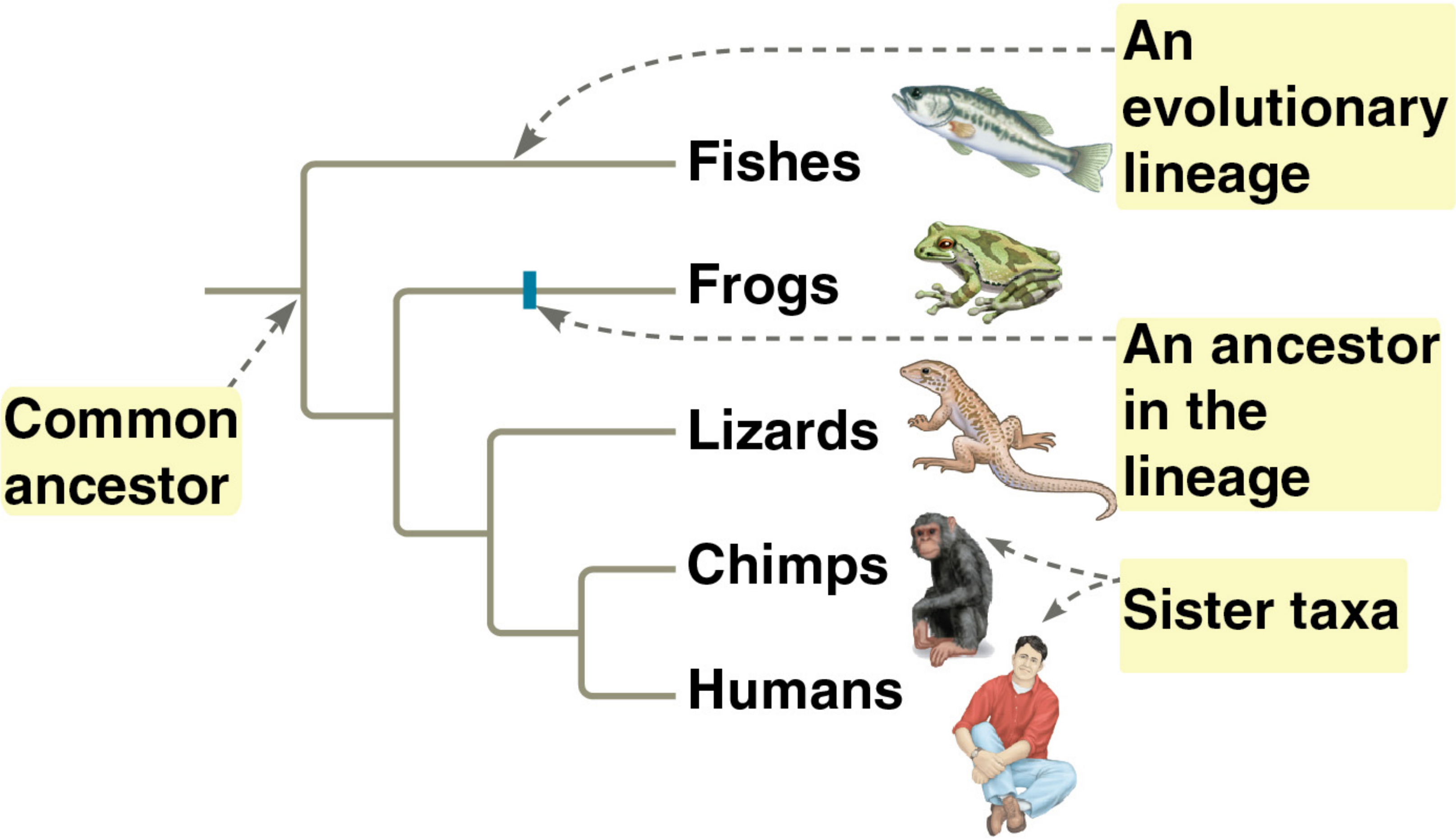
- Systematists have proposed that classification be based only on evolutionary relationships
- In this system, only groups that include a common ancestor and all its descendants would be named

What We Can and Cannot Learn from Phylogenetic Trees

- A phylogenetic tree represents a hypothesis about evolutionary relationships
- Each **branch point** represents the divergence of two evolutionary lineages from a common ancestor
- An **evolutionary lineage** is a sequence of ancestral organisms leading to a particular descendant taxon
- **Sister taxa** are groups that share a common ancestor that is not shared by any other group

Figure 26.5

Parts of a Phylogenetic Tree



An evolutionary lineage

Fishes

Frogs

An ancestor in the lineage

Lizards

Common ancestor

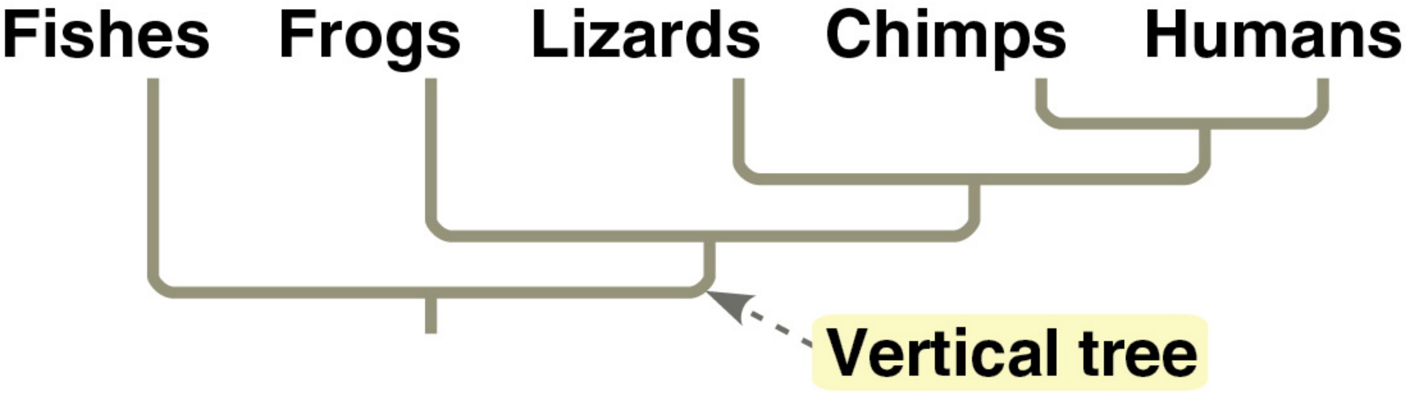
Chimps

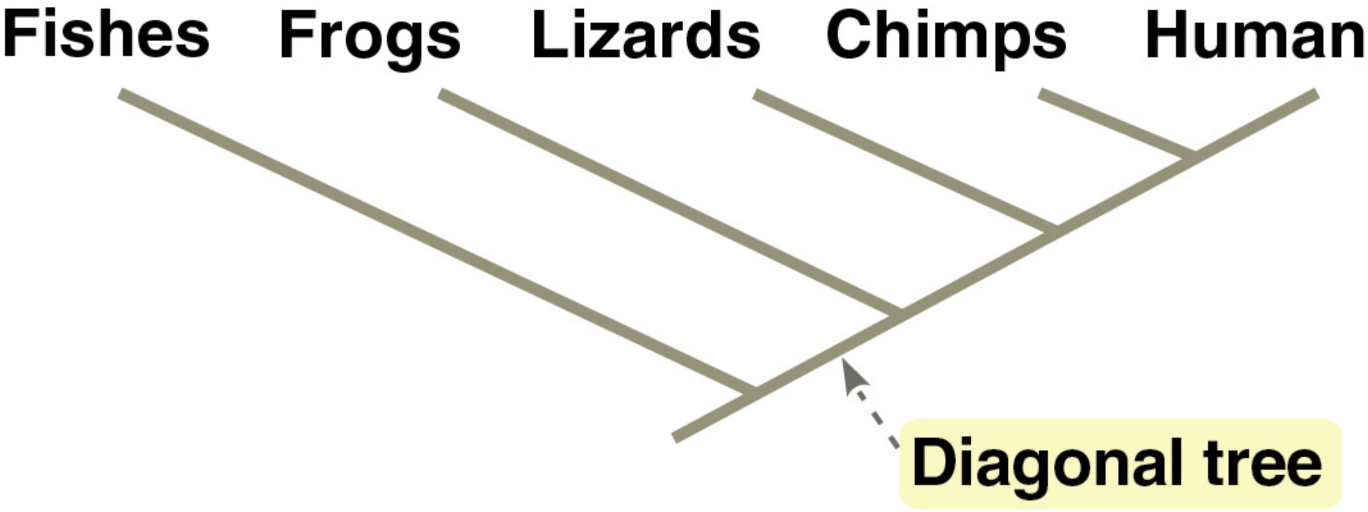
Humans

Sister taxa

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

Alternative Forms of Tree Diagrams

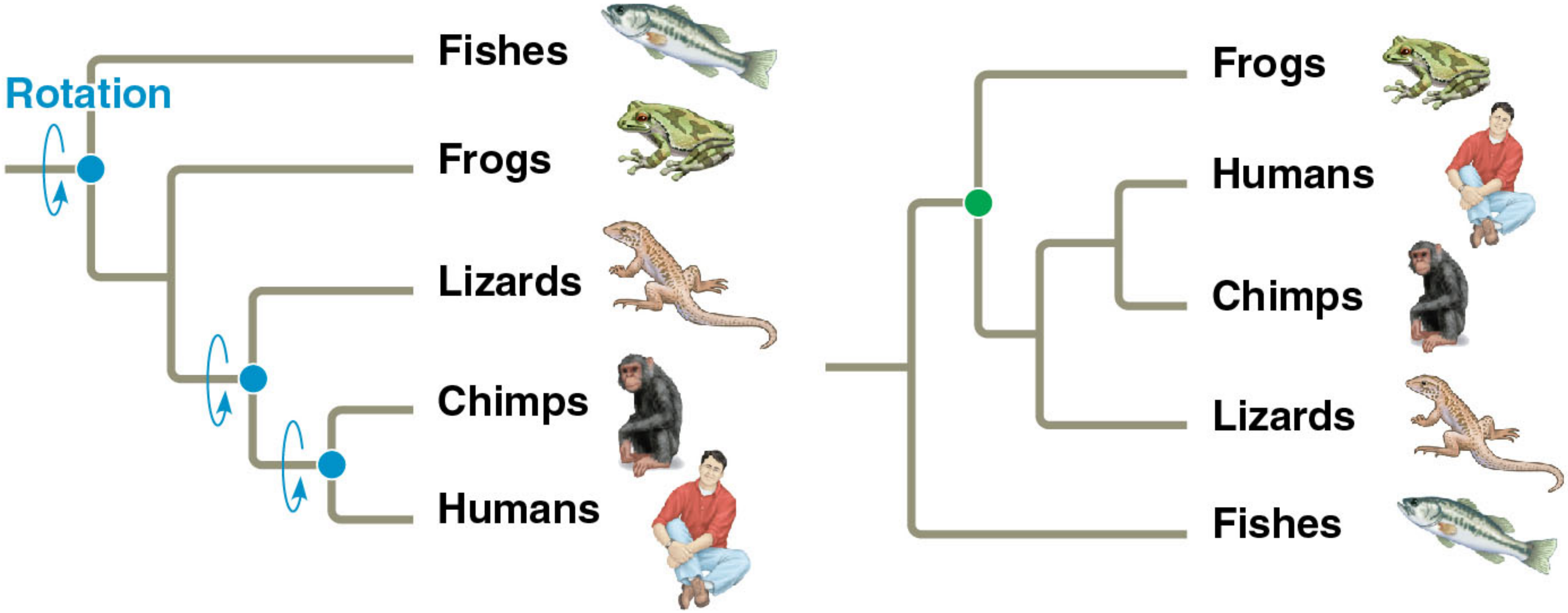




- Tree branches can be rotated around a branch point without changing what they convey about evolutionary relationships
- The order in which taxa appear at the branch tips is not significant
- The branching pattern signifies the order in which the lineages diverged from common ancestors

Figure 26.5

Rotating Around Branch Points



- A **rooted** tree includes a branch to represent the most recent common ancestor of all taxa in the tree
- A lineage that diverges from all other members of the group early in the history of the group is called a **basal taxon**

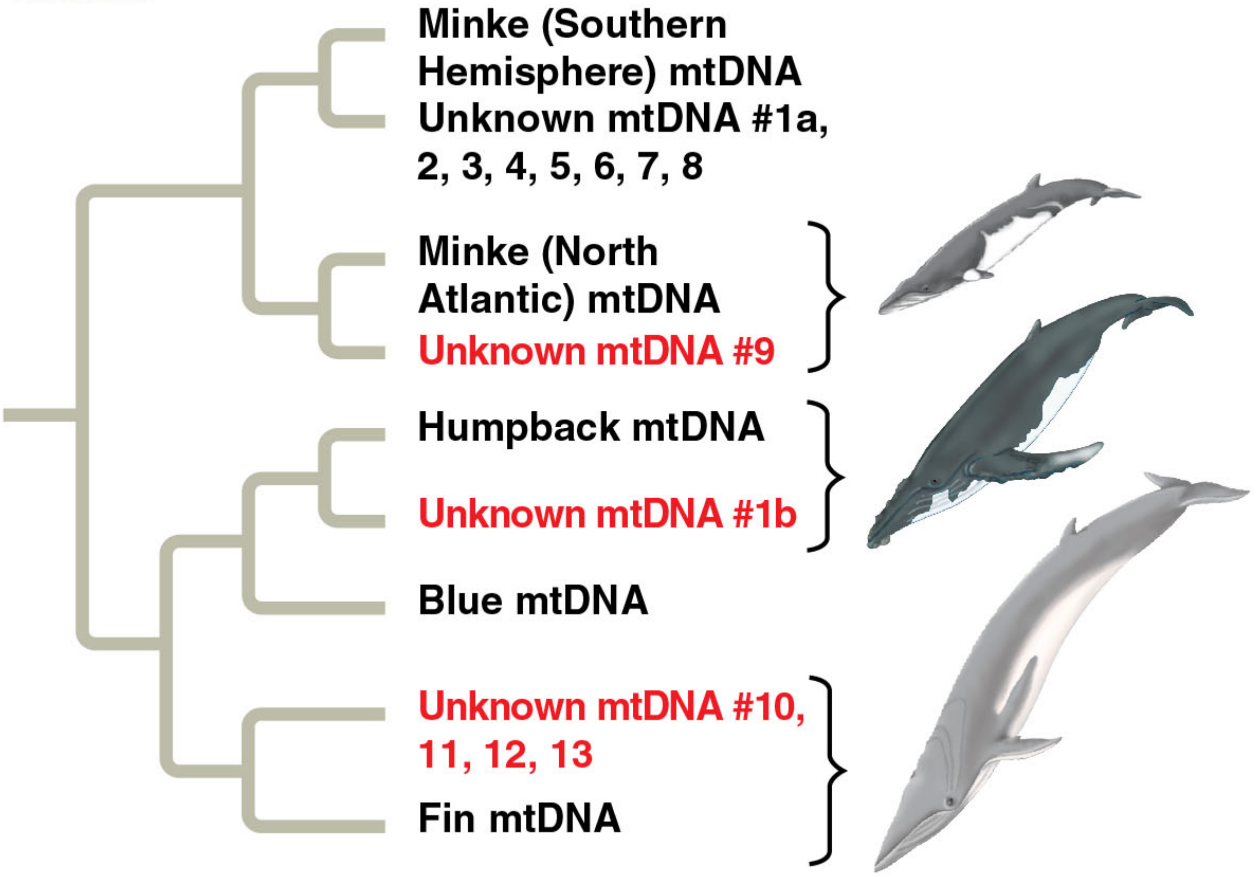
- Phylogenetic trees show patterns of descent, not phenotypic similarity
- Phylogenetic trees do not indicate when species evolved or how much change occurred in a lineage
- It should not be assumed that a taxon evolved from the taxon next to it

Applying Phylogenies

- Phylogenies can have practical applications
  - For example, researchers used a phylogenetic tree to determine whether several samples of “whale meat” came from legally harvested species

Figure 26.6

Results



Minke (Southern Hemisphere) mtDNA
Unknown mtDNA #1a, 2, 3, 4, 5, 6, 7, 8

Minke (North Atlantic) mtDNA
Unknown mtDNA #9

Humpback mtDNA

Unknown mtDNA #1b

Blue mtDNA

Unknown mtDNA #10, 11, 12, 13
Fin mtDNA

Data from C. S. Baker and S. R. Palumbi, Which whales are hunted? A molecular genetic approach to monitoring whaling, *Science* 265:1538–1539 (1994). Reprinted with permission from AAAS.

CONCEPT 26.2: Phylogenies are inferred from morphological and molecular data

- Systematists use data on the morphology, genes, and biochemistry of organisms to infer phylogeny
- Only features resulting from common ancestry are useful for determining evolutionary relationships

Morphological and Molecular Homologies

- Phenotypic and genetic similarities due to shared ancestry are called **homologies**
- Organisms with similar morphology or DNA sequence are likely to be more closely related than those that vastly differ in structure and sequence

- In some cases, morphological divergence can be great between species that are genetically similar (or vice versa)
  - For example, Hawaiian silverswords, though all genetically similar, range from tall, twiggy trees to dense, ground-hugging shrubs

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

- Unrelated species evolve superficial similarities through convergent evolution in response to natural selection to similar environmental conditions
  - For example, similarities between the Australian “mole” and the African golden mole result from adaptation to similar lifestyles, not shared ancestry

Figure 26.7

Australian “mole”

African golden mole

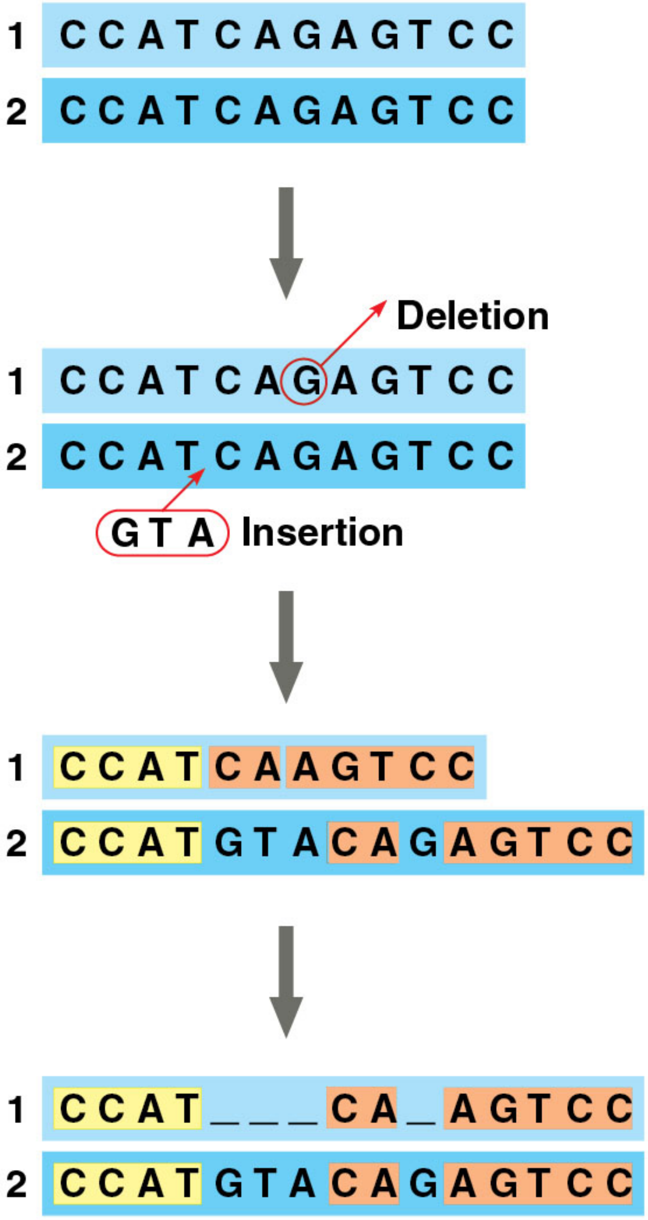
- The complexity of characters being compared can help to distinguish between homology and analogy
- The more similarity between complex structures, the more likely they evolved from a common ancestor
  - For example, if the genes in two organisms share many portions of nucleotide sequence, it is likely the genes are homologous

Evaluating Molecular Homologies

- The first step after sequencing DNA is to align comparable sequences from the study species
  - Closely related species are likely to differ at only one or a few sites along the DNA sequence
  - Distantly related species usually have different bases at many sites and may have different lengths

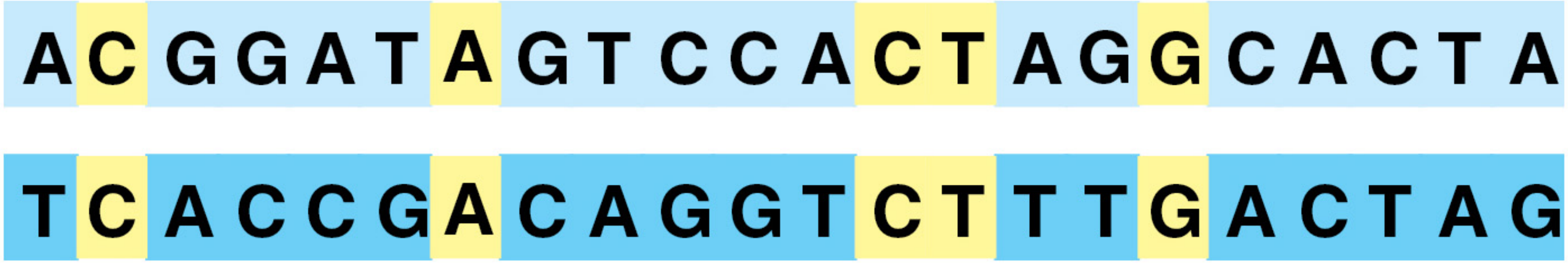
- Insertions and deletions are point mutations that shift the entire DNA sequence following the mutation
- Failure to take these mutations into account would overlook otherwise good sequence matches
- Computer programs are used to identify such matches by testing possible sequence alignments

Figure 26.8


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

- In most cases, many points of sequence similarity reflect homology, but coincidental matches between unrelated organisms can occur
- Statistical tools are used to distinguish between distant homologies and coincidental matches

Figure 26.9



A C G G A T A G T C C A C T A G G C A C T A

T C A C C G A C A G G T C T T T G A C T A G

CONCEPT 26.3: Shared characters are used to construct phylogenetic trees

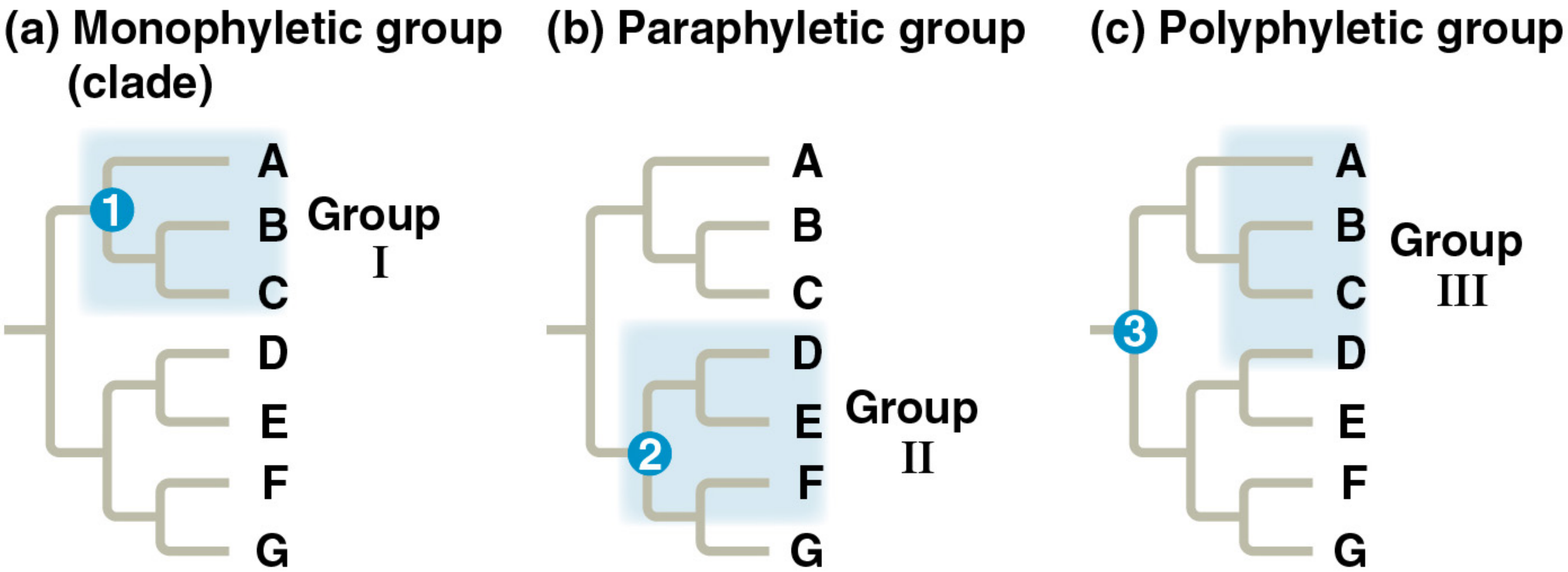
- Once homologous characters have been identified, they can be used to infer a phylogeny

Cladistics

- In **cladistics** organisms are grouped primarily by common ancestry
- A **clade** is a group of species that includes an ancestral species and all its descendants
- Clades can be nested within larger clades

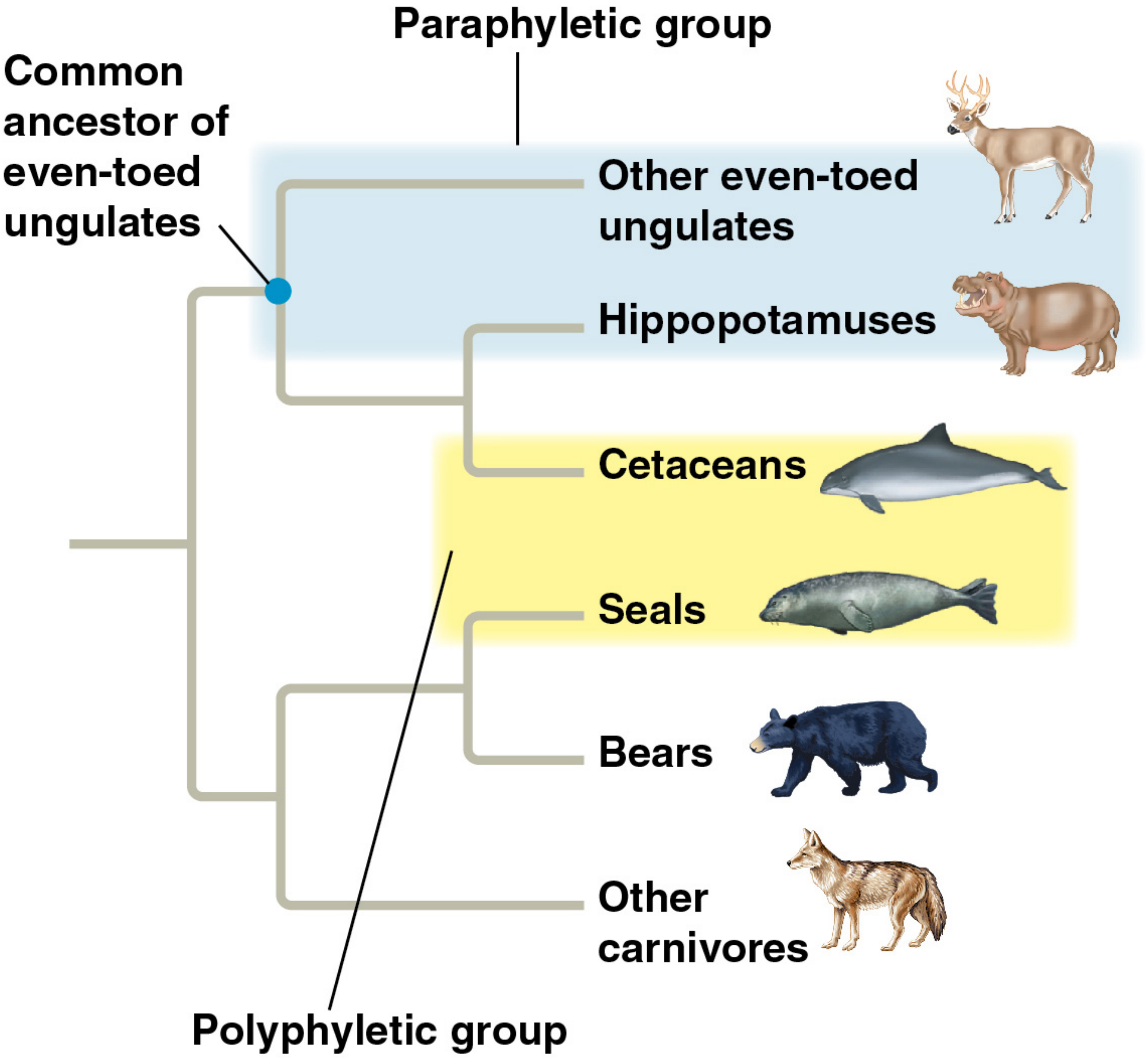
- Only **monophyletic** groups, consisting of the ancestor and all of its descendants, are clades
- A **paraphyletic** group consists of an ancestral species and some, but not all, of the descendants
- A **polyphyletic** group includes distantly related species but not their most recent common ancestor

Figure 26.10


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

- In paraphyletic groups, unlike polyphyletic groups, the common ancestor to all members of the group is part of the group
- When new evidence indicates that a group is polyphyletic, its members are typically reclassified

Figure 26.11


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

Shared Ancestral and Shared Derived Characters

- Organisms share some characters with their ancestors, but also have many unique characters
- A **shared ancestral character** is a character that originated in an ancestor of the taxon
- A **shared derived character** is an evolutionary novelty unique to a particular clade

- Lost features, as well as new features, can be shared derived characters
  - For example, the loss of limbs in snakes or whales

- Whether a character is considered ancestral or derived is relative
  - For example, having a backbone is ancestral within the vertebrate clade, but it is also a derived character distinguishing vertebrates from other animals

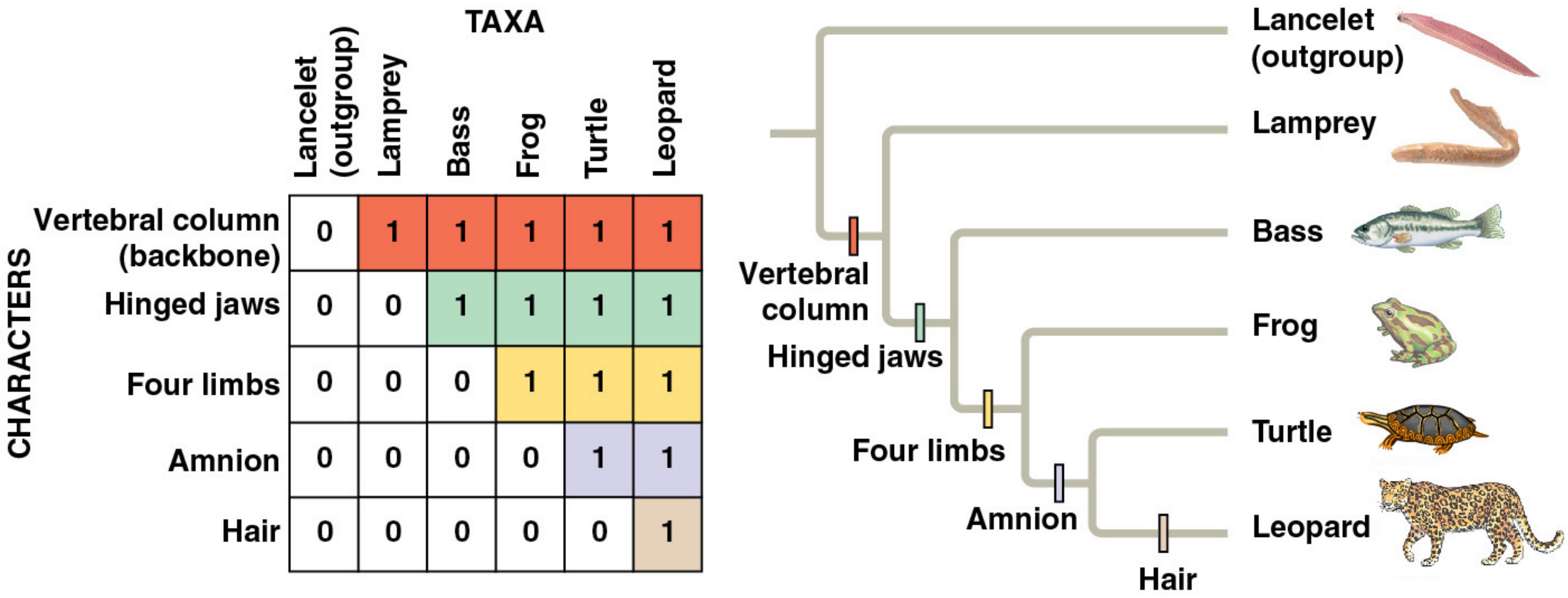
Inferring Phylogenies Using Derived Characters

- Identification of the clades in which shared derived characters first appeared can be used to infer evolutionary relationships

- An **outgroup** is a species or group of species closely related to, but not part of the group of species being studied (the **ingroup**)
- Each ingroup species is compared with the outgroup to differentiate between shared derived and shared ancestral characters

- Characters shared by the outgroup and ingroup are assumed to be ancestral
- Each derived character is assumed to have arisen only once in the ingroup
- The point at which characters were derived can be determined by comparing members of the ingroup with each other and the outgroup

Figure 26.12



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 limbs	0	0	0	1	1	1
Amnion	0	0	0	0	1	1
Hair	0	0	0	0	0	1

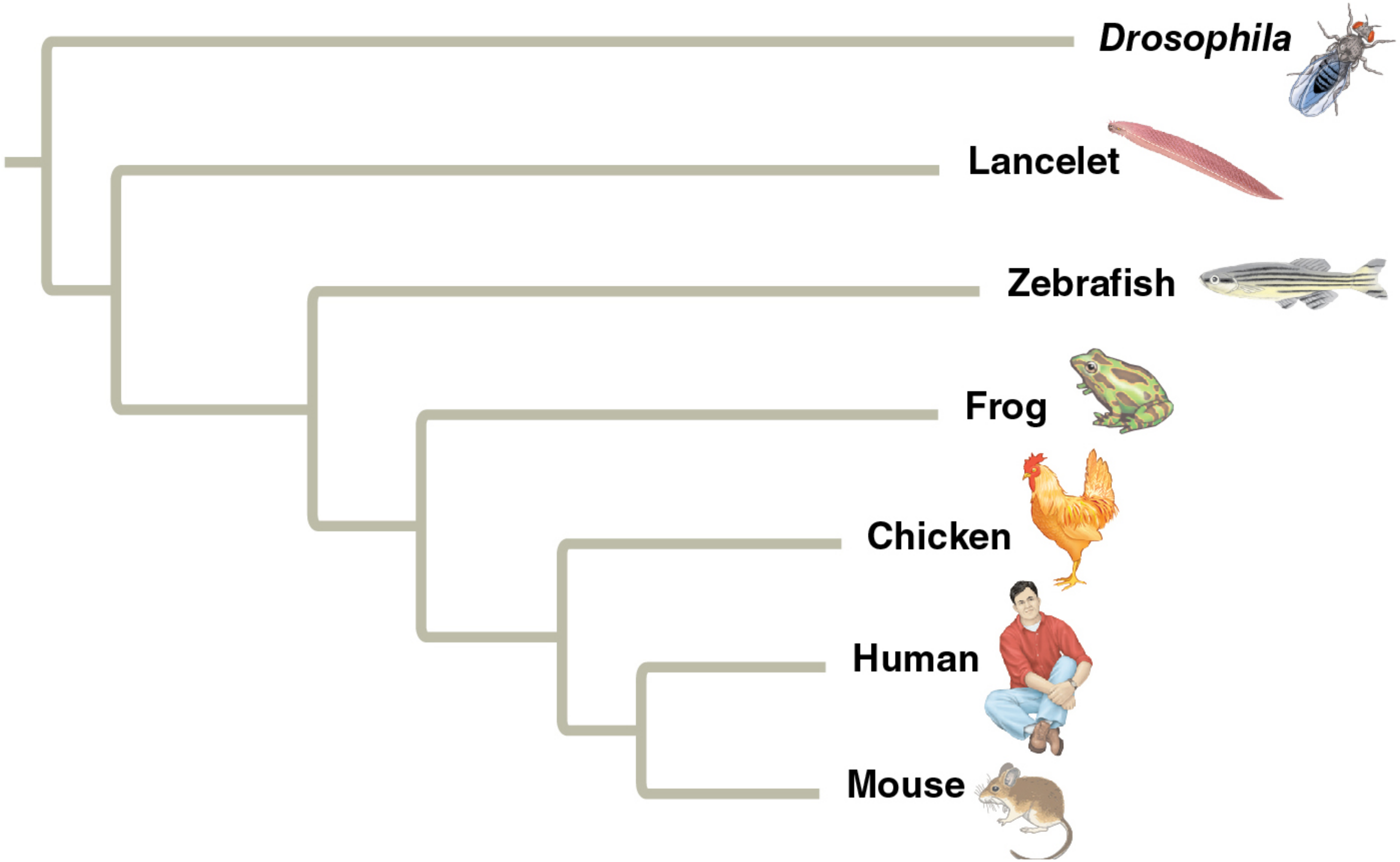
(a) Character table

(b) Phylogenetic tree

Phylogenetic Trees with Proportional Branch Lengths

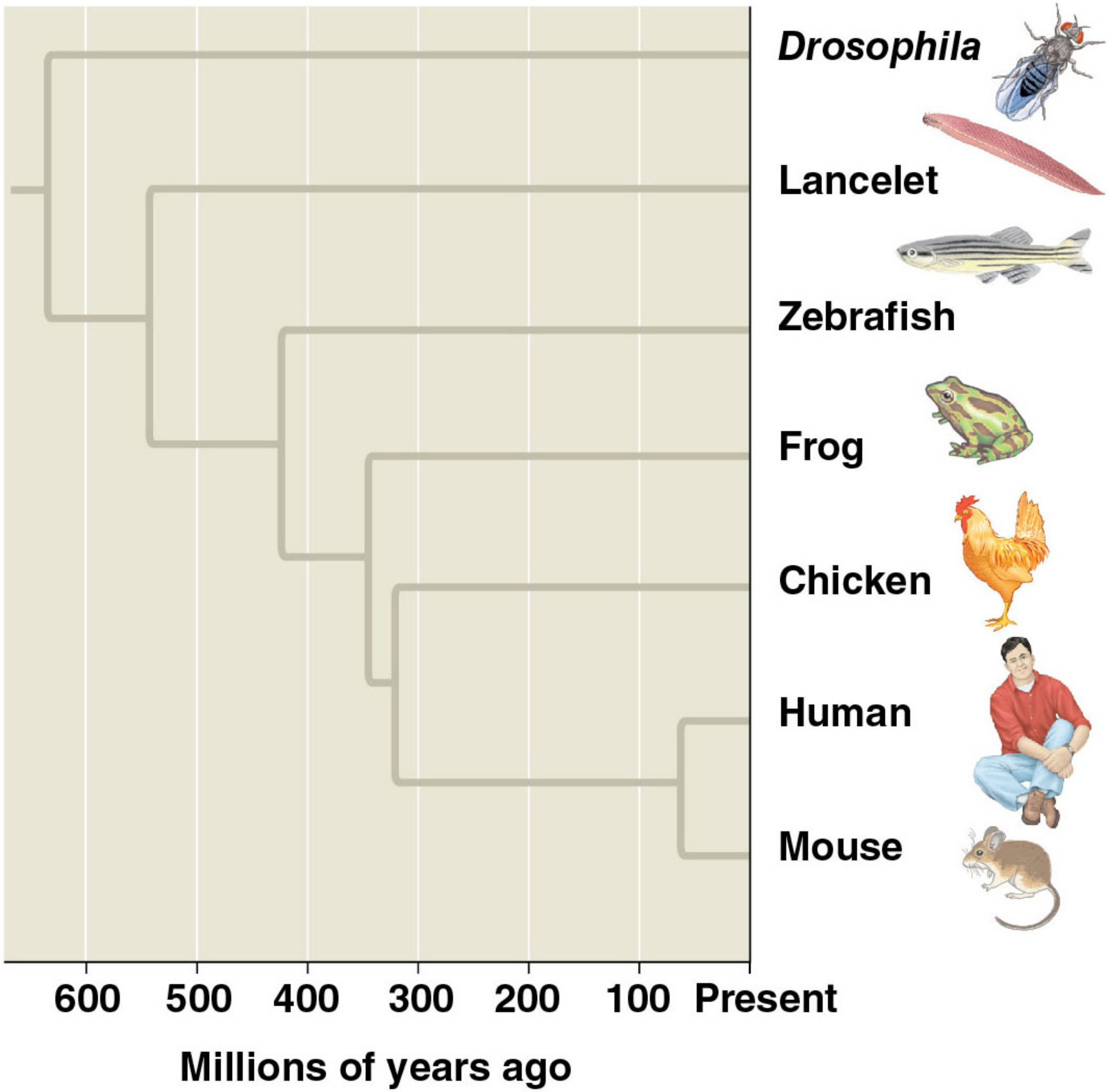
- In some trees, branch length reflects the number of genetic changes that have occurred in each lineage
- Lineages with shorter branches reflect fewer genetic changes than those with longer branches

Figure 26.13



- In some trees, branch length is proportional to time
- Fossil data is used to place branch points in the context of geological time

Figure 26.14


Drosophila
Lancelet
Zebrafish
Frog
Chicken
Human
Mouse
600
500
400
300
200
100
Present
Millions of years ago

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** assumes that the most likely tree is the one that requires the fewest evolutionary events (appearances of shared derived characters)
- In phylogenies based on DNA, the most parsimonious tree has the fewest base changes

- **Maximum likelihood** identifies the tree most likely to have produced a given set of DNA data based on probability rules about how DNA changes over time
  - For example, the probability rules could be based on the assumption that all nucleotide substitutions are equally likely

- Many computer programs have been developed to search for trees that are parsimonious and likely
- Programs based on parsimony estimate phylogenies by examining many possible trees and identifying those that require the fewest evolutionary changes

Figure 26.15

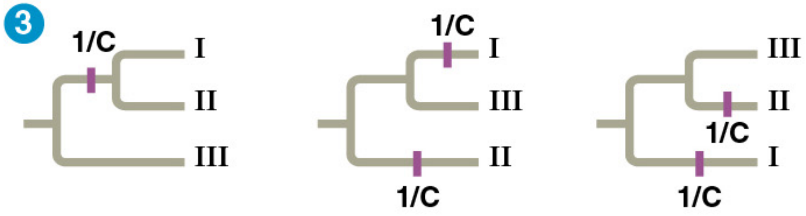
Technique

① **Three phylogenetic hypotheses:**

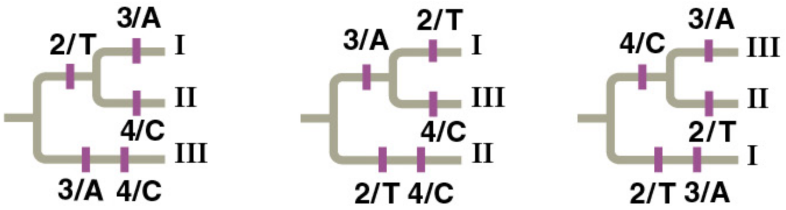
②

	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

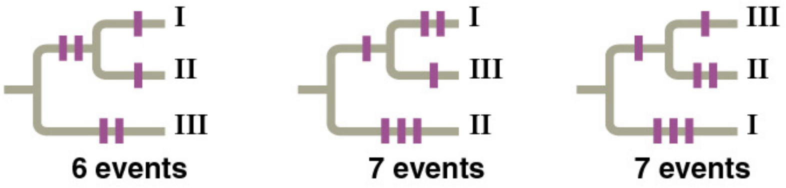
③



④



Results



Phylogenetic Trees as Hypotheses

- Predictions can be made and tested based on the assumption that a particular phylogeny is correct
  - For example, *phylogenetic bracketing* predicts that features shared by two closely related groups will be present in their ancestor and all its descendants

- Phylogenetic bracketing has been used to make predictions about dinosaurs
  - Birds and crocodiles share several features: four-chambered hearts, song, nest building, and brooding
  - If they are shared by the common ancestor and all its descendants, then dinosaurs also had these features
  - The fossil record supports nest building and brooding in dinosaurs

Figure 26.16

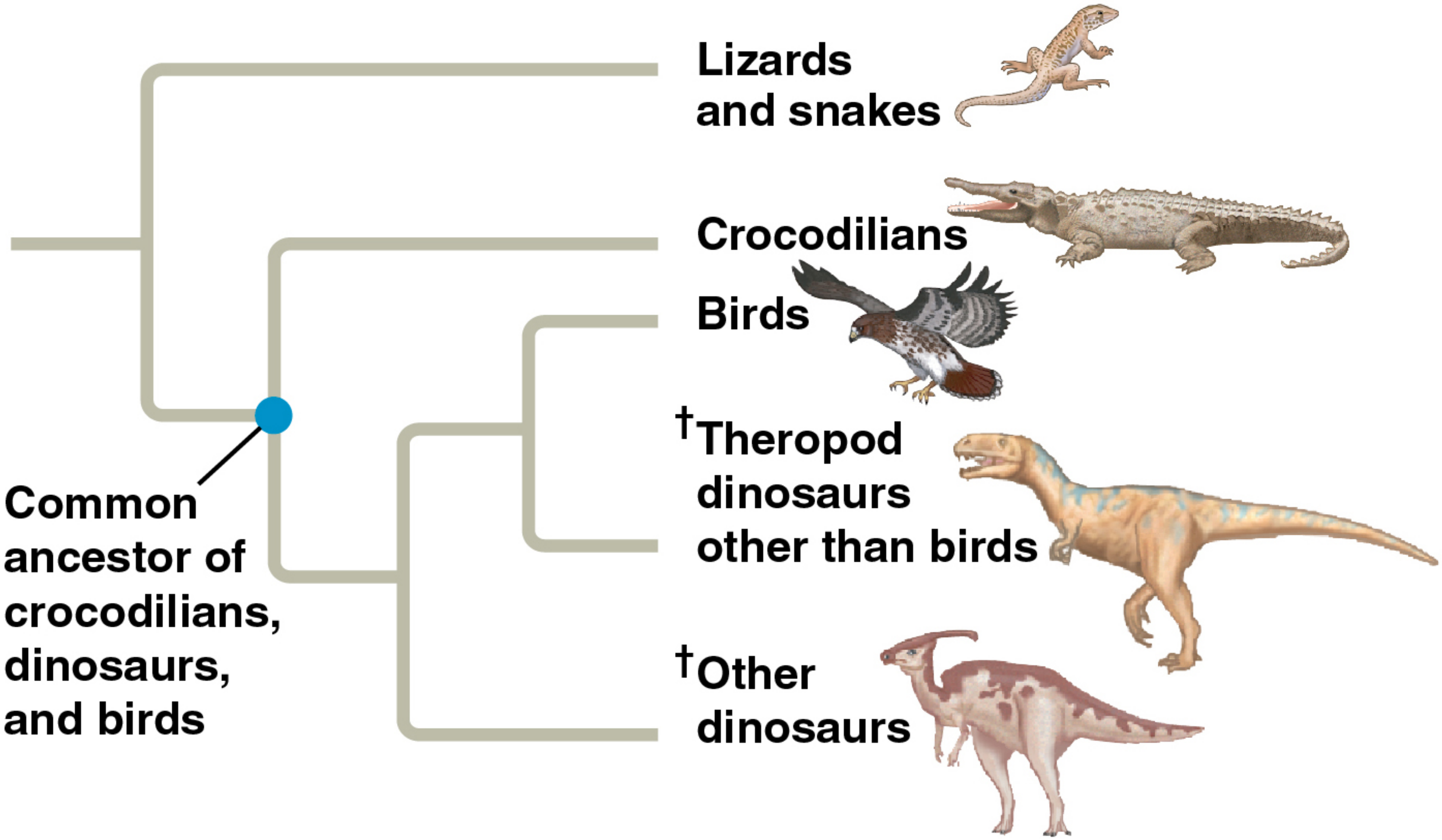
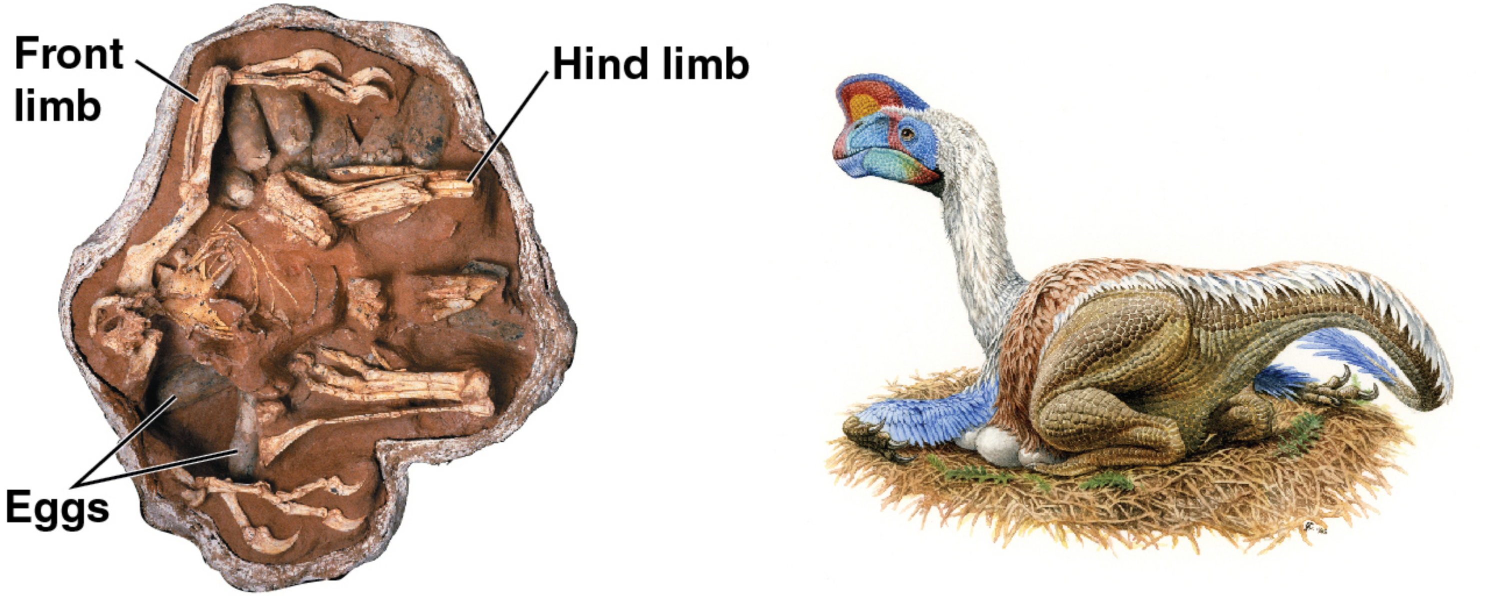

Lizards and snakes
Crocodilians
Birds
†Theropod dinosaurs other than birds
†Other dinosaurs
Common ancestor of crocodilians, dinosaurs, and birds

Figure 26.17



(a) Fossil remains of *Oviraptor* and eggs

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

CONCEPT 26.4: An organism's evolutionary history is documented in its genome

- Molecular comparisons can be used to deduce relatedness between groups that are morphologically dissimilar or poorly represented in the fossil record

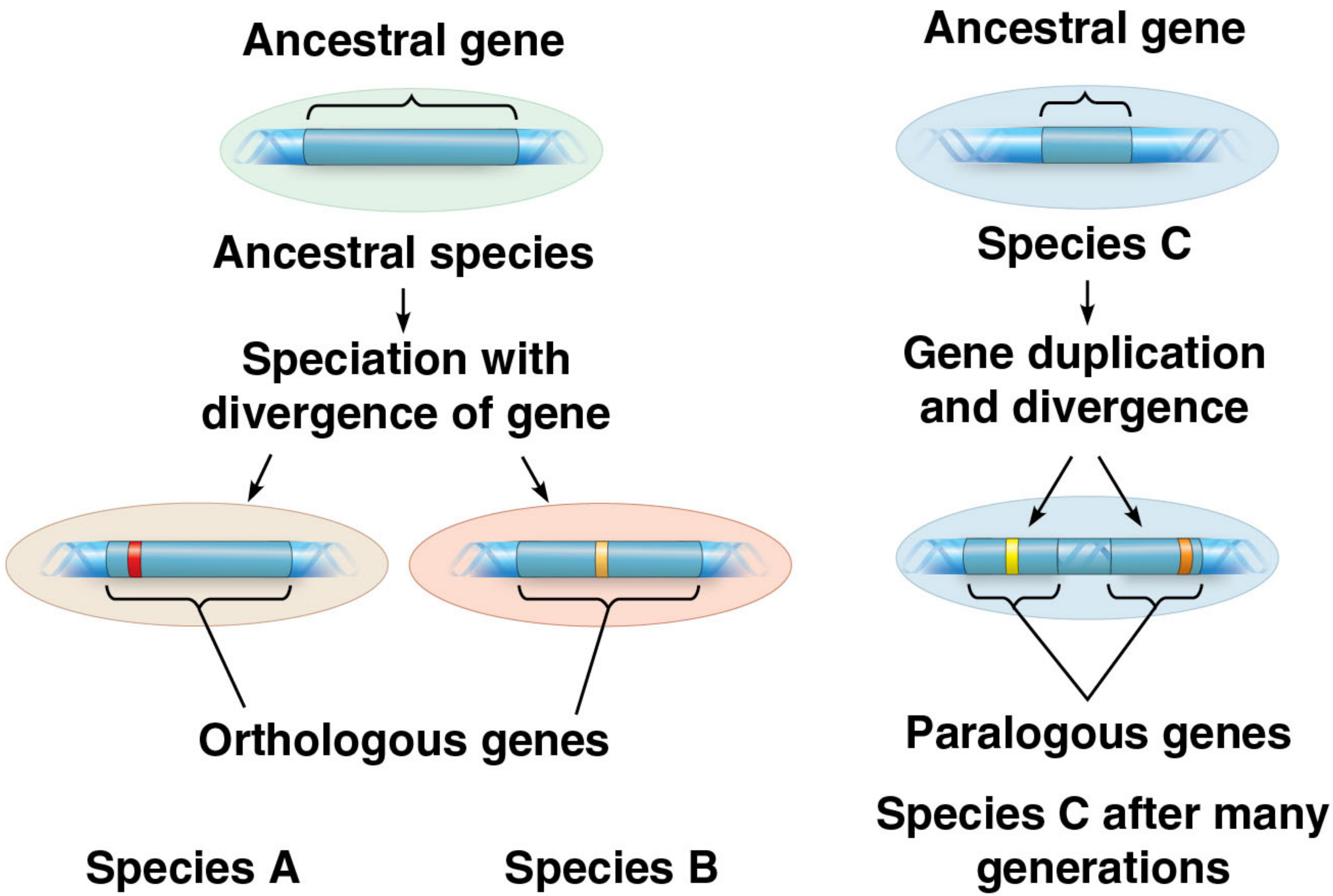
- Different genes can evolve at different rates in the same lineage
  - For example, DNA coding for ribosomal RNA changes slowly; it can be used to detect ancient relationships
  - Mitochondrial DNA evolves relatively rapidly; it can be used to explore more recent evolutionary events

Gene Duplications and Gene Families

- Gene duplication increases the number of genes in the genome, providing more opportunities for evolutionary change
- Repeated gene duplications result in gene families, groups of related genes within a genome

- There are two types of homologous genes: orthologous and paralogous genes
  - In **orthologous genes**, the homology is the result of a speciation event and occurs between genes found in different species
  - In **paralogous genes**, the homology results from gene duplication and occurs between gene copies within a species

Figure 26.18



(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

- Orthologous genes are used for inferring phylogeny because they reflect the history of speciation events
  - For example, cytochrome *c* genes serve the same function in humans and dogs, but the sequence has diverged since they last shared an ancestor

- Paralogous genes can diverge within a species because they are present in more than one copy
  - For example, genes in the human olfactory receptor gene family have diverged to produce proteins that respond to a wide variety of molecules

Genome Evolution

- Lineages that diverged long ago often share many orthologous genes
  - For example, humans and mice diverged about 65 million years ago, and 99% of our genes are orthologous

- Gene number does not increase through duplication at the same rate as phenotypic complexity
  - For example, humans have only four times as many genes as yeast, a single-celled eukaryote
  - Many human genes are more versatile than those in yeast; single human genes can encode multiple proteins with different functions in various tissues

CONCEPT 26.5: 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

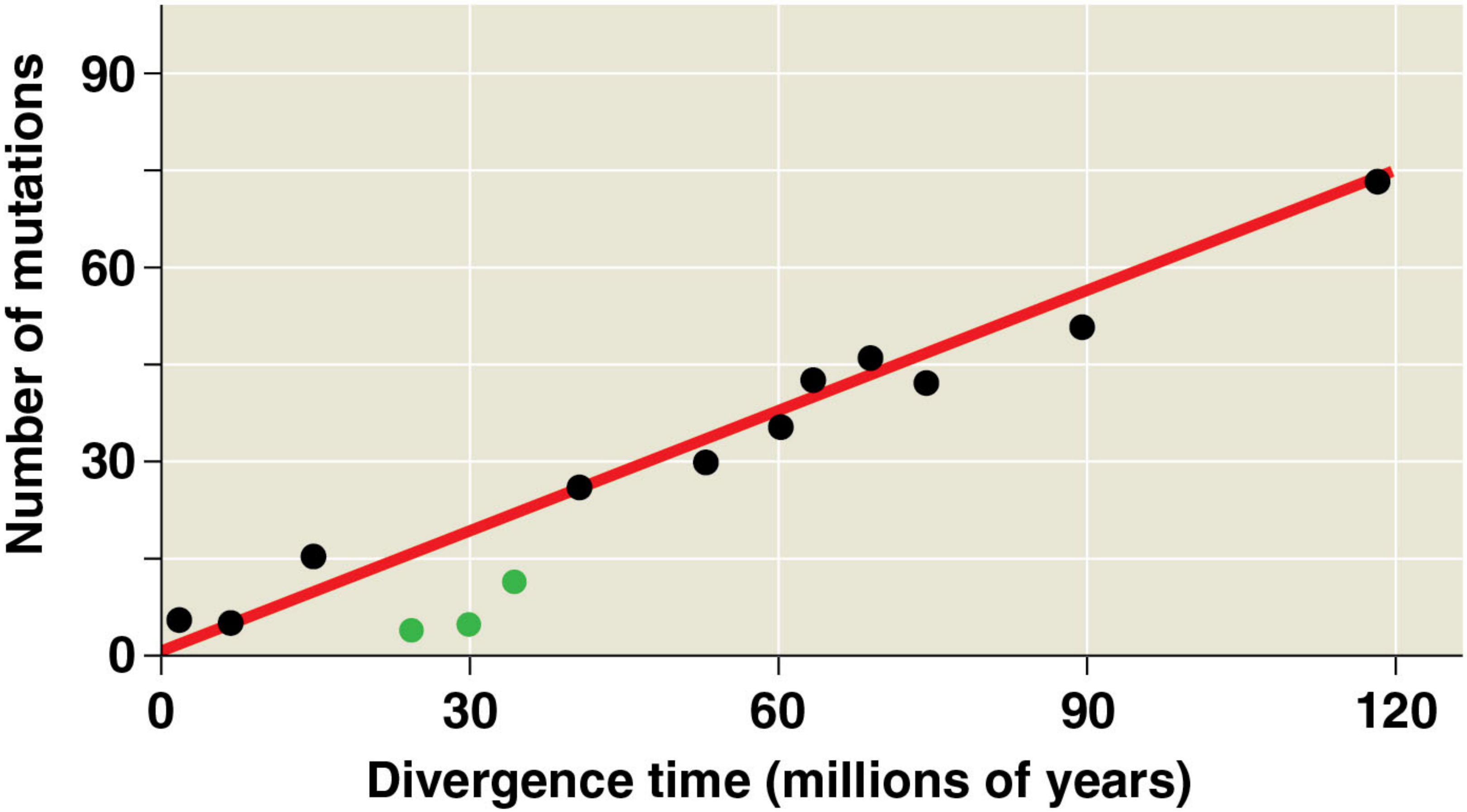
Molecular Clocks

- A **molecular clock** is an approach used to estimate the absolute time of evolutionary change
- The estimate is based on the observation that some genes appear to evolve at a constant rate

- An assumption of the molecular clock concept is that the number of nucleotide substitutions is proportional to the time since
  - the split from a common ancestor for orthologous genes
  - gene duplication for paralogous genes

- A molecular clock is calibrated by graphing the number of genetic differences in a gene against dates of branch points known from the fossil record
- The inferred rate of change is used to estimate the dates of events not discernable from the fossil record

Figure 26.19



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

- There are limits to the molecular clock approach
  - Some genes evolve in irregular bursts, rather than clocklike precision
  - The rate of evolution deviates from the average periodically, even in reliable clocklike genes
  - The same gene may evolve at different rates in different taxa
  - Some clocklike genes evolve at dramatically different rates from each other

Differences in Clock Speed

- Some mutations are selectively neutral—neither beneficial or detrimental
- If most mutations are neutral, then the rate of molecular change should be regular, like a clock

- Differences in clock rate among genes are related to how critical the exact amino acid sequence is
  - If the sequence is critical to survival, then most mutations will be harmful and change will be slow
  - If it is less critical, more mutations will be neutral and change will be faster

Potential Problems with Molecular Clocks

- Molecular clocks do not run as smoothly as would be expected if mutations were neutral
- Irregularities result from natural selection favoring some mutations over others
- If the direction of natural selection changes repeatedly, it may average out over time, making genes more clocklike

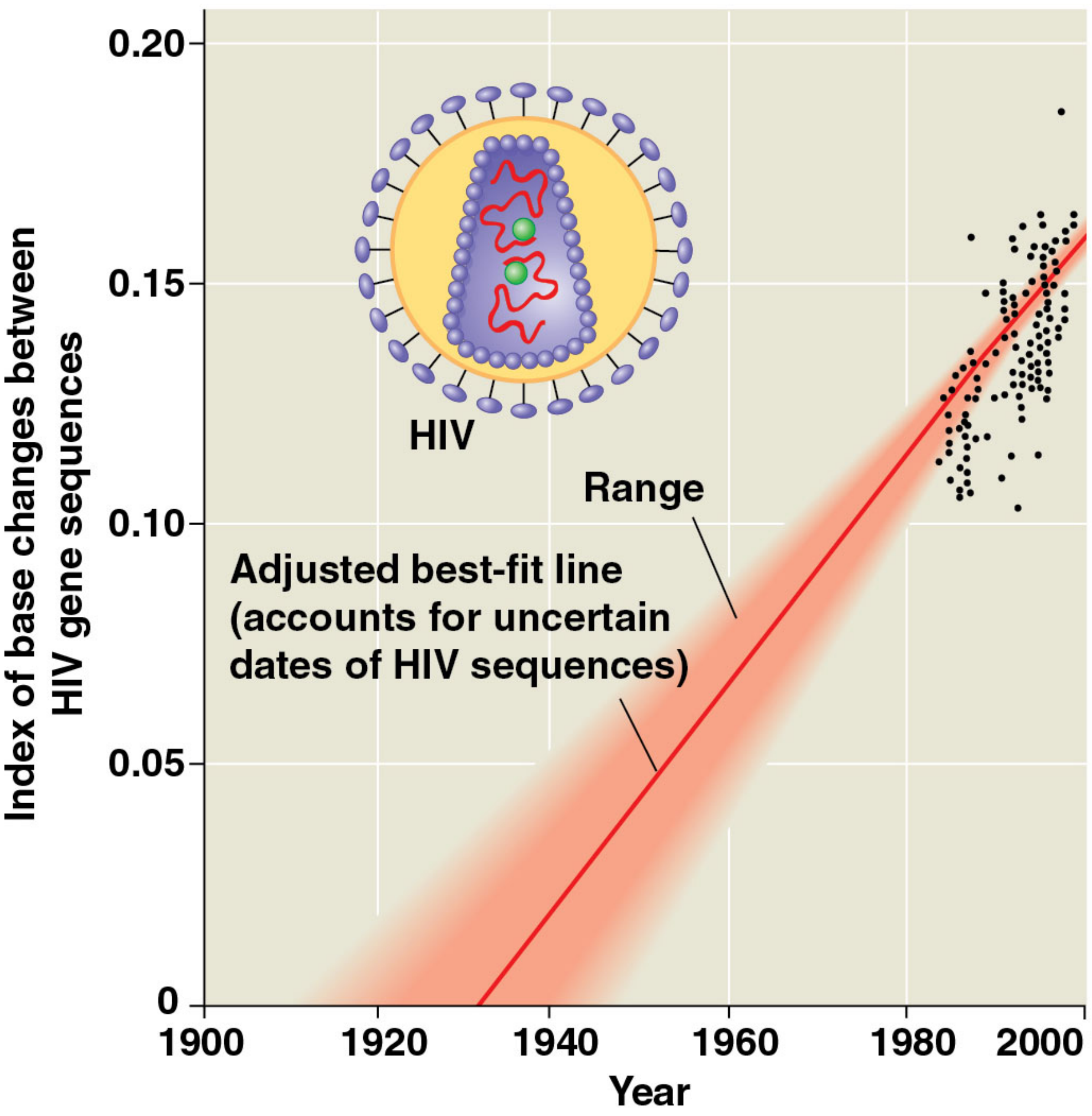
- Estimates of evolutionary divergences older than the fossil record have a high degree of uncertainty
- The use of multiple genes or genes that evolved in different taxa may improve estimates

Applying a Molecular Clock: Dating the Origin of HIV

- HIV, the virus that causes AIDS, 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

- HIV-1 M is the most widespread strain in humans
- Application of a molecular clock indicates that it spread to humans around 1930
- A more advanced molecular clock approach estimated the first spread to humans around 1910

Figure 26.20



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

CONCEPT 26.6: Our understanding of the tree of life continues to change based on new data

- Systematics has informed our understanding of life’s diversity
- Analysis of DNA sequence data has provided insight into the very deepest branches of the tree of life

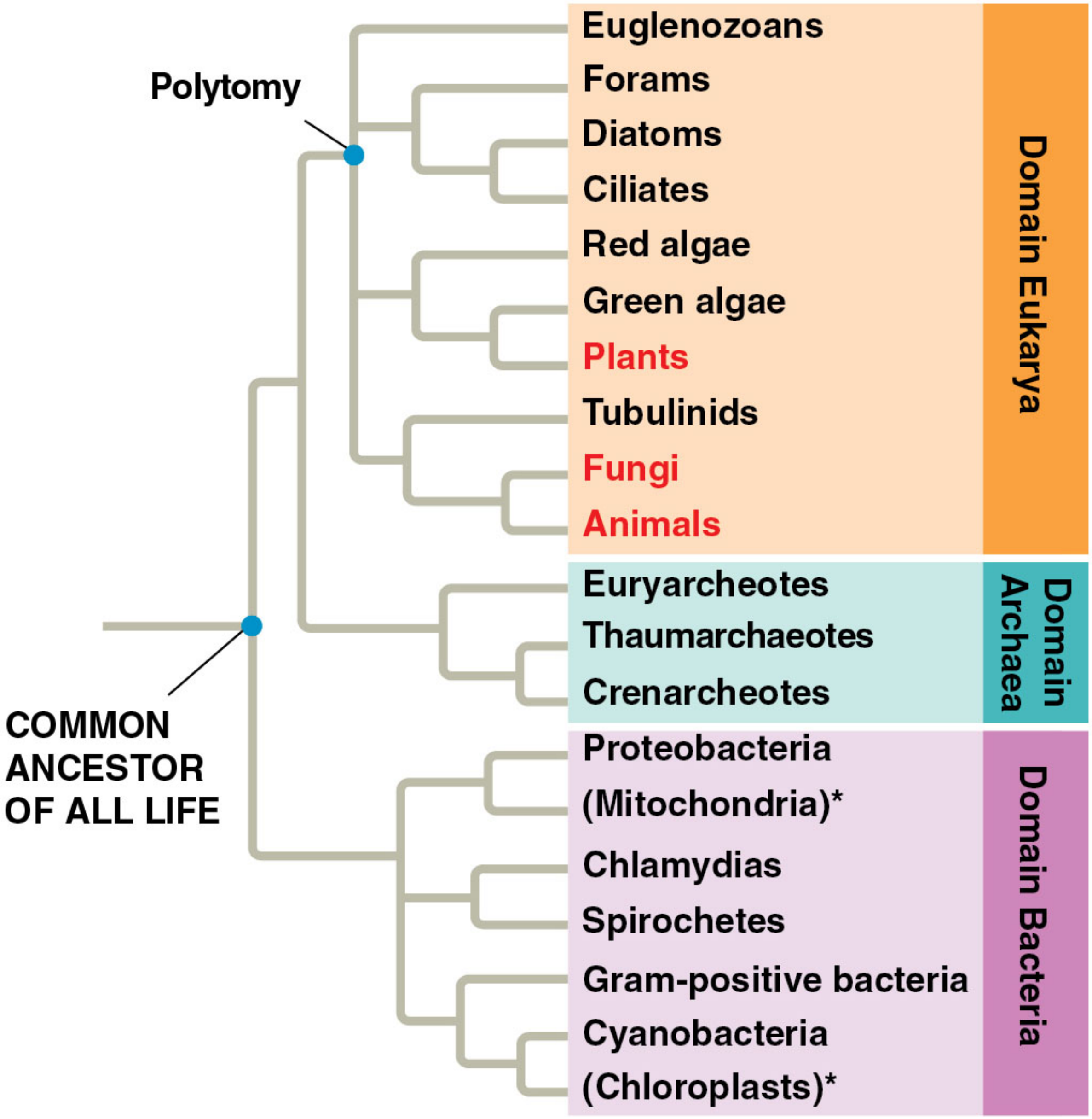
From Two Kingdoms to Three Domains

- Biologists once classified all species into two kingdoms: plants and animals
- By the 1960s there were five kingdoms: Monera (prokaryotes), Protista, Plantae, Fungi, and Animalia

- The three-domain system was recently adopted: Bacteria, Archaea, and Eukarya
  - Bacteria includes most of the known prokaryotes
  - Archaea consists of diverse prokaryotes that inhabit a wide variety of environments
  - Eukarya consists of single-celled and multicellular organisms that have cells containing true nuclei

- Prokaryotes form two domains because some differ as much from each other as they do from eukaryotes
- Eukarya includes three kingdoms: Plantae, Fungi, and Animalia
- Protista is no longer a kingdom because some of its members are more closely related to plants, fungi, or animals than other protists

Figure 26.21


Polytomy
COMMON ANCESTOR OF ALL LIFE
Euglenozoans
Forams
Diatoms
Ciliates
Red algae
Green algae
Plants
Tubulinids
Fungi
Animals
Euryarcheotes
Thaumarchaeotes
Crenarcheotes
Proteobacteria
(Mitochondria)*
Chlamydias
Spirochetes
Gram-positive bacteria
Cyanobacteria
(Chloroplasts)*
Domain Eukarya
Domain Archaea
Domain Bacteria

- New research discoveries continue to change our understanding of the tree of life
  - For example, a recently discovered group of archaea, the lokiarcheaota, may be the sister group of the eukaryotes

The Important Role of Horizontal Gene Transfer

- In the tree of life based on rRNA genes, eukaryotes and archaea are more closely related to each other than either is to bacteria
- Trees based on other genes reveal different relationships

- Disparities between different trees may result from movement of genes between the domains
- This is facilitated through **horizontal gene transfer**, the movement of genes from one genome to another
- Horizontal gene transfer can occur by exchange of transposable elements and plasmids, viral infection, and possibly fusion of organisms

- Horizontal gene transfer has played a key role in the evolution of both prokaryotes and eukaryotes
- Eukaryotes can acquire nuclear genes from bacteria and archaea
  - For example, the alga *Galdieria sulphuraria* acquired about 5% of its genes from bacterial and archaeal species

Figure 26.22

- Some biologists argue that horizontal gene transfer was commonplace in the early history of life
- They suggest representing the base of the tree of life as a tangled network of connected branches rather than a dichotomously branching tree

Figure 26.23

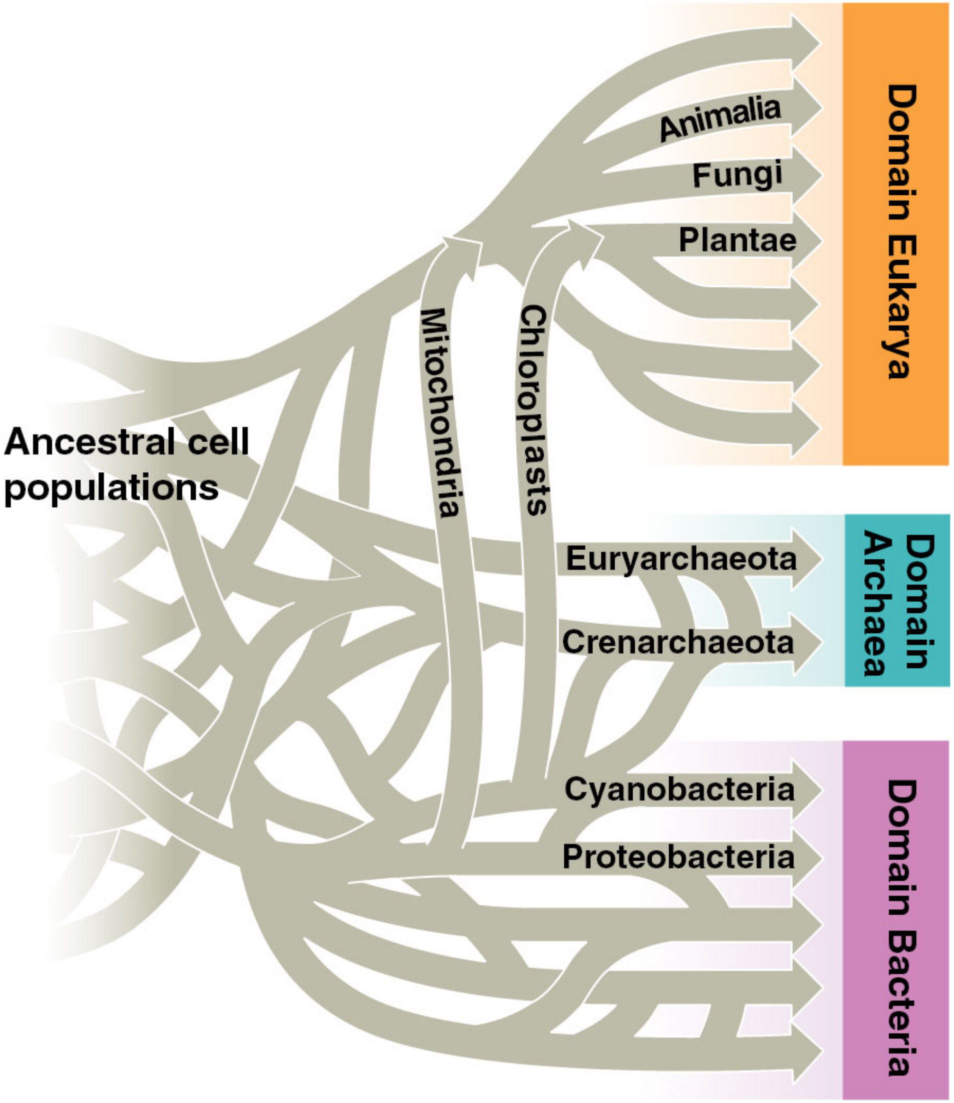
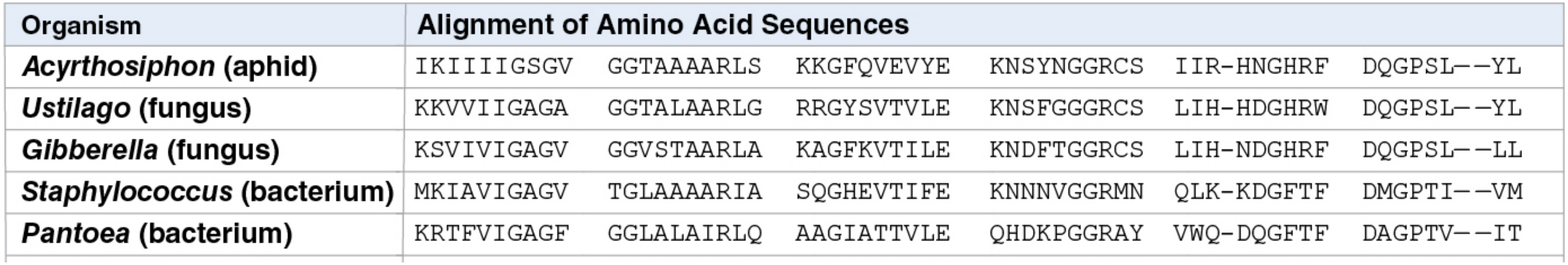

Domain Eukarya
Animalia
Fungi
Plantae
Ancestral cell populations
Mitochondria
Chloroplasts
Euryarchaeota
Crenarchaeota
Domain Archaea
Cyanobacteria
Proteobacteria
Domain Bacteria

Figure 26.UN01



Organism	**Alignment of Amino Acid Sequences**
***Acyrthosiphon* (aphid)**	`IKIIIIGSGV GGTAAAARLS KKGFQVEVYE KNSYNGGRCS IIR-HNGHRF DQGPSL--YL`
***Ustilago* (fungus)**	`KKVVIIGAGA GGTALAARLG RRGYSVTVLE KNSFGGGRCS LIH-HDGHRW DQGPSL--YL`
***Gibberella* (fungus)**	`KSVIVIGAGV GGVSTAARLA KAGFKVTILE KNDFTGGRCS LIH-NDGHRF DQGPSL--LL`
***Staphylococcus* (bacterium)**	`MKIAVIGAGV TGLAAAARIA SQGHEVTIFE KNNNVGGRMN QLK-KDGFTF DMGPTI--VM`
***Pantoea* (bacterium)**	`KRTFVIGAGF GGLALAIRLQ AAGIATTVLE QHDKPGGRAY VWQ-DQGFTF DAGPTV--IT`

Data from Nancy A. Moran, Yale University. See N. A. Moran and T. Jarvik, Lateral transfer of genes from fungi underlies carotenoid production in aphids, *Science* 328:624–627 (2010).

Figure 26.UN01a

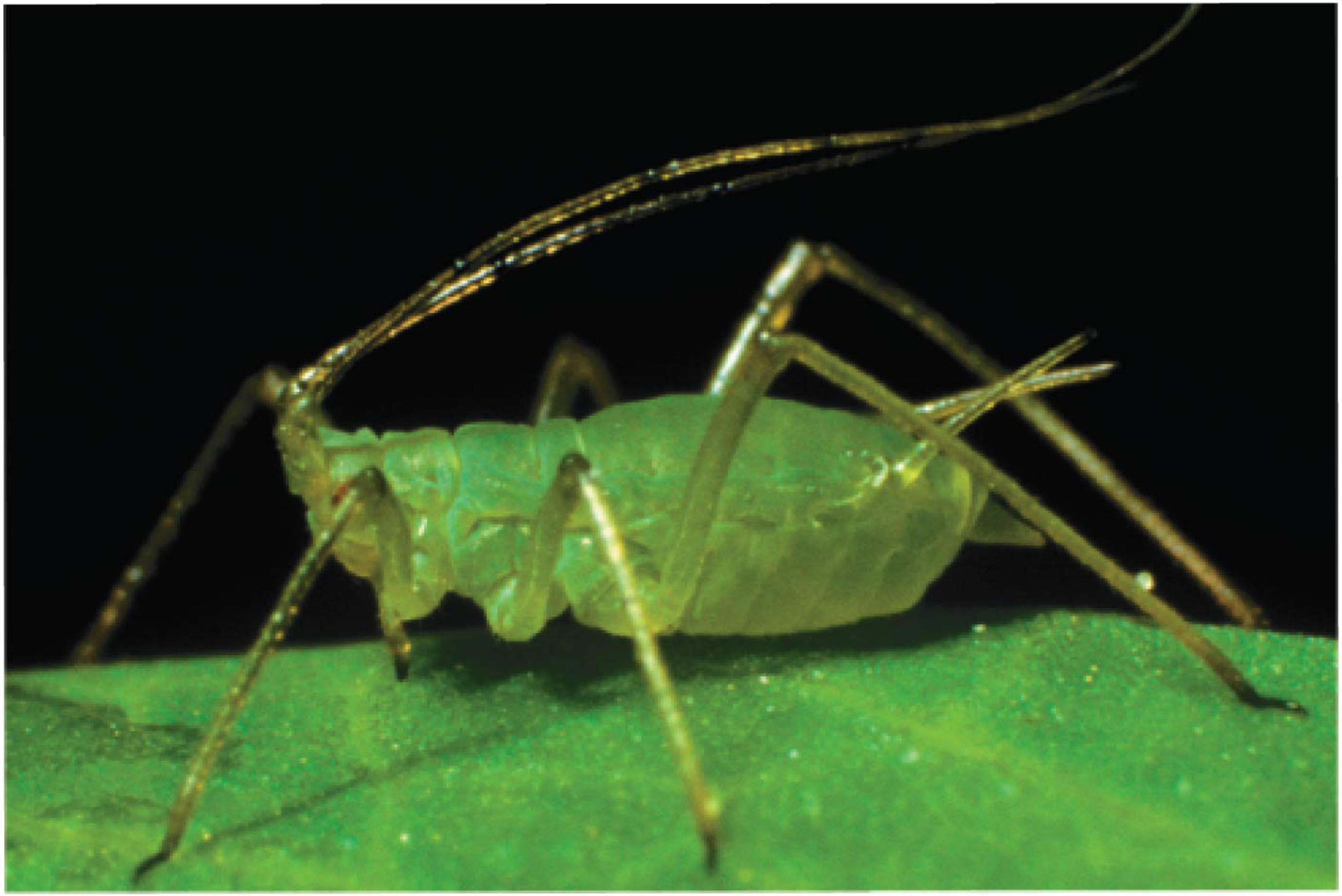


Figure 26.UN02

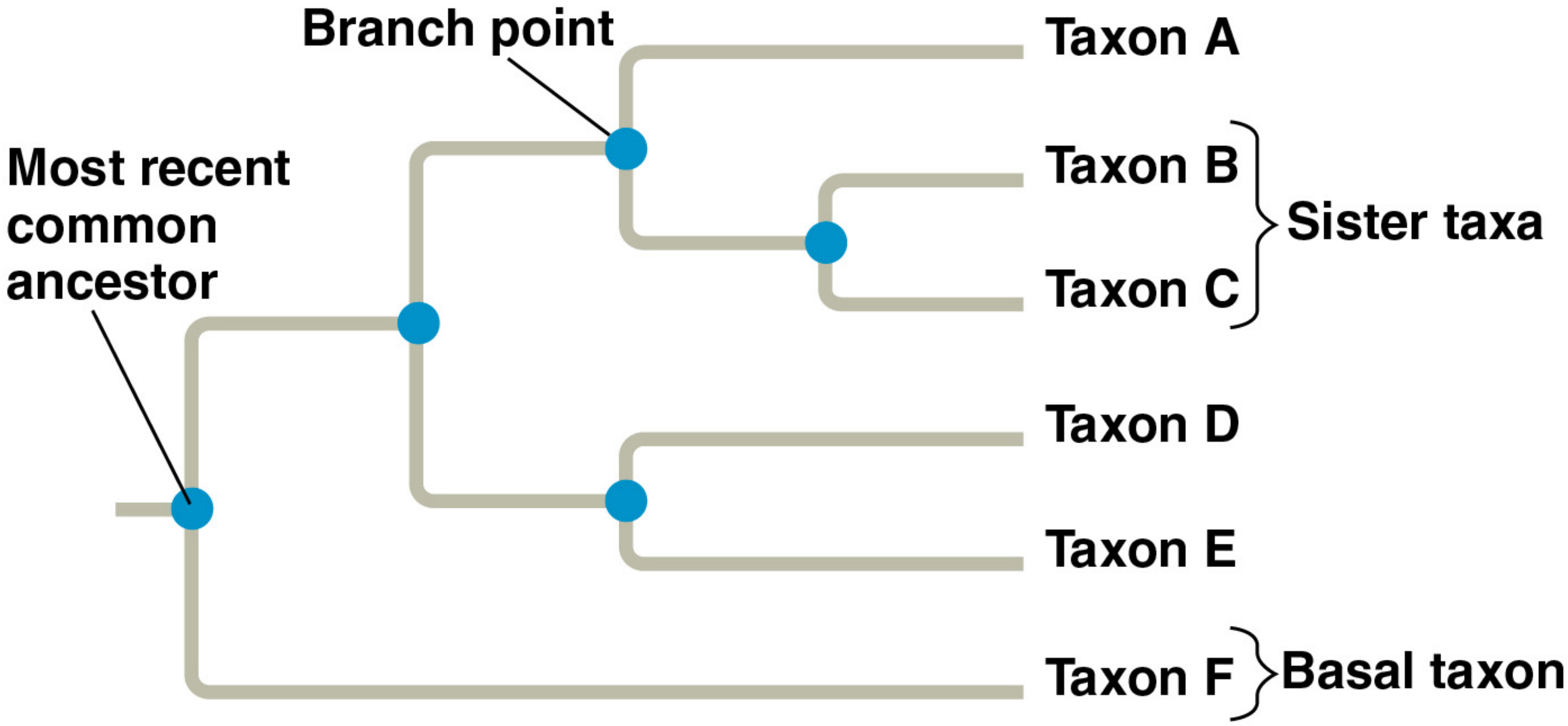


Figure 26.UN03

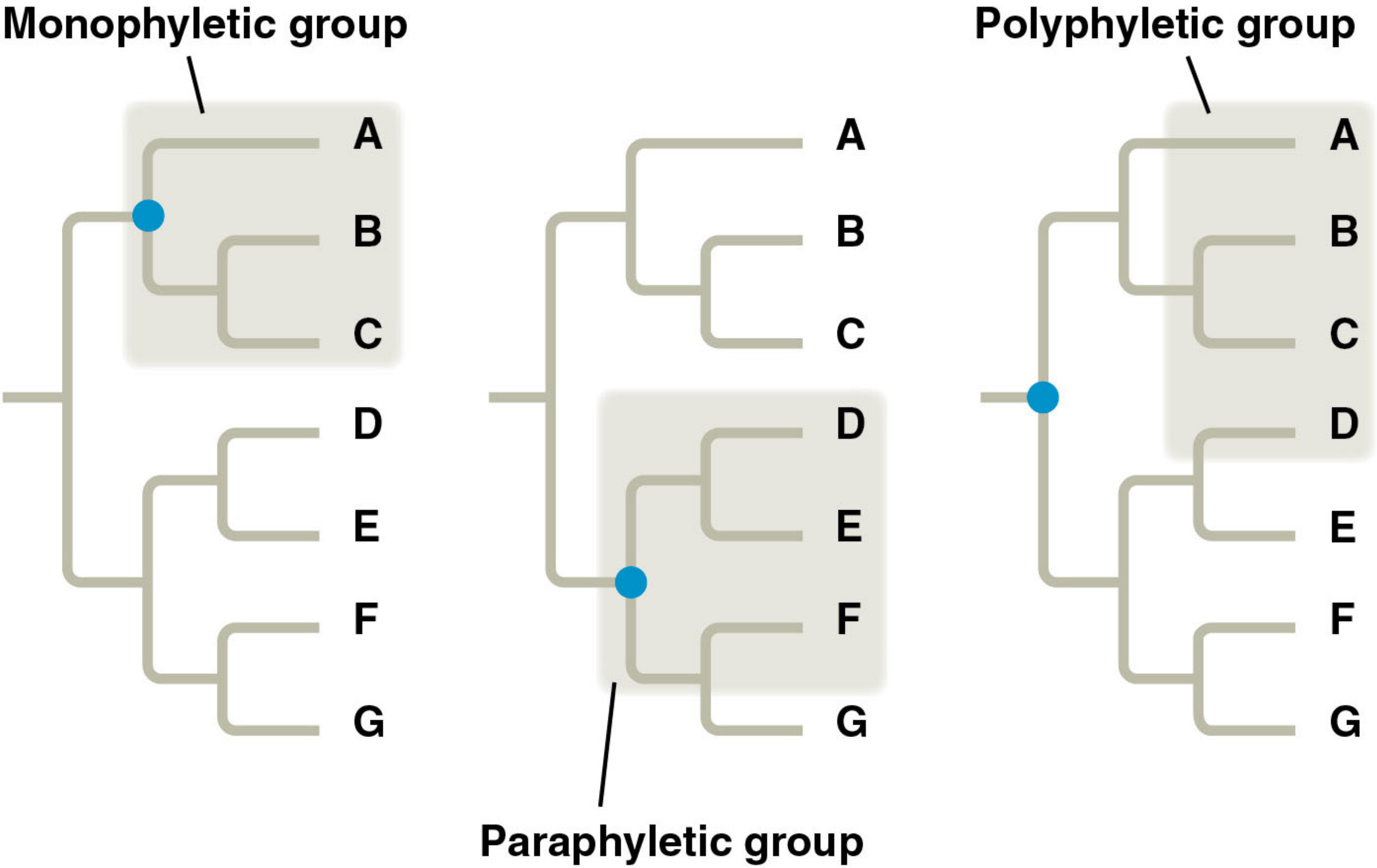


Figure 26.UN04

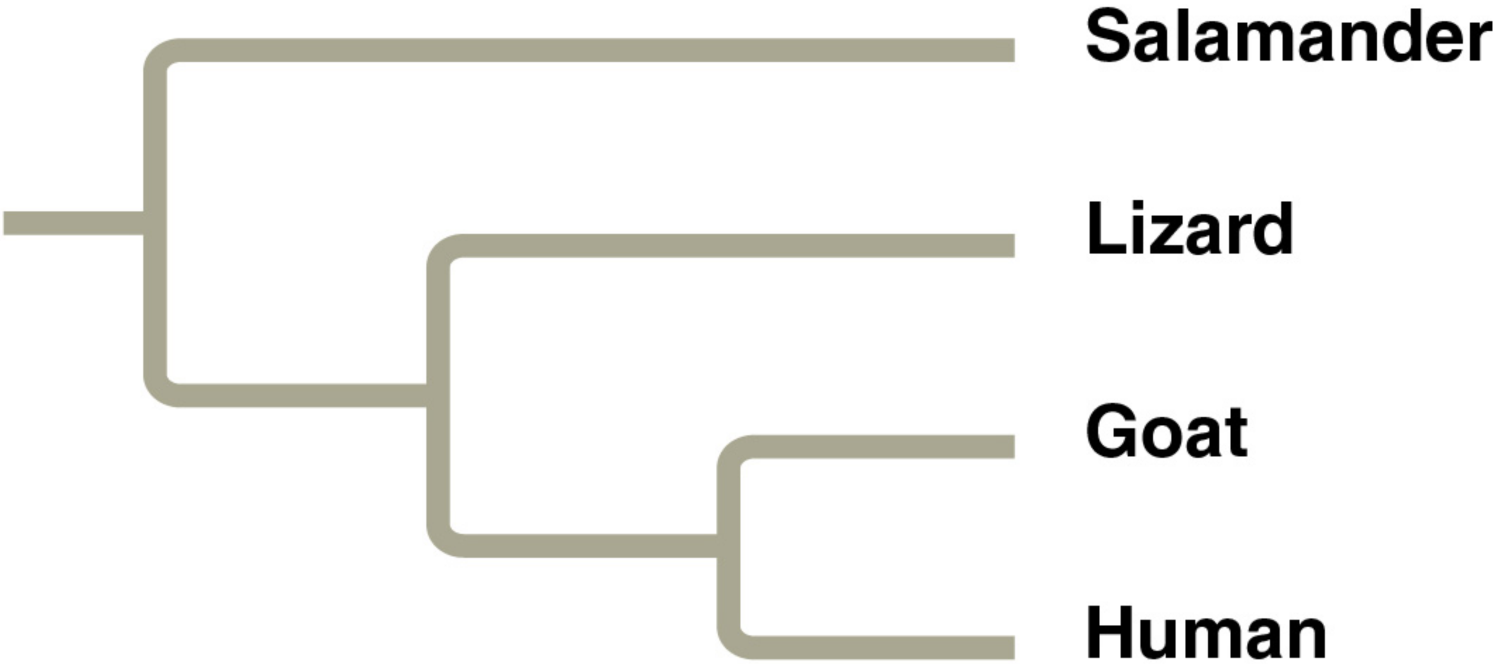


Figure 26.UN05

Character	Lancelet (outgroup)	Lamprey	Tuna	Salamander	Turtle	Leopard	Dolphin
(1) Backbone	0	1	1	1	1	1	1
(2) Hinged jaw	0	0	1	1	1	1	1
(3) Four limbs	0	0	0	1	1	1	1*
(4) Amnion	0	0	0	0	1	1	1
(5) Milk	0	0	0	0	0	1	1
(6) Dorsal fin	0	0	1	0	0	0	1

*Although adult dolphins have only two obvious limbs (their flippers), as embryos they have two hind-limb buds, for a total of four limbs.

Figure 26.UN06