26

Phylogeny and the Tree of Life

▲ **Figure 26.1 What is this organism?**

OVERVIEW

Investigating the Tree of Life

Look closely at the organism in **Figure 26.1**. Although it resembles a snake, this animal is actually an Australian legless lizard known as the common scaly-foot (*Pygopus lepidopodus*). Why isn't the scaly-foot considered a snake? More generally, how do biologists distinguish and categorize the millions of species on Earth?

An understanding of evolutionary relationships suggests one way to address these questions: We can decide in which "container" to place a species by comparing its traits with those of potential close relatives. For example, the scaly-foot does not have a fused eyelid, a highly mobile jaw, or a short tail posterior to the anus, three traits shared by all snakes. These and other characteristics suggest that despite a superficial resemblance, the scaly-foot is not a snake. Furthermore, a survey of the lizards reveals that the scaly-foot is not alone; the legless condition has evolved independently in several different groups of lizards. Most legless lizards are burrowers or live in grasslands, and like snakes, these species lost their legs over generations as they adapted to their environments.

Snakes and lizards are part of the continuum of life extending from the earliest organisms to the great variety of species alive today. In this unit, we will survey this diversity and describe hypotheses regarding how it evolved. As we do so, our emphasis will shift from the *process* of evolution (the evolutionary mechanisms described in Unit Four) to its *pattern* (observations of evolution's products over time).

To set the stage for surveying life's diversity, in this chapter we consider how biologists trace **phylogeny**, the evolutionary history of a species or group of species. A phylogeny of lizards and snakes, for example, indicates that both the scaly-foot and snakes evolved from lizards with legs—but that they evolved from different lineages of legged lizards. Thus, it appears that their legless conditions evolved independently.

To construct phylogenies, biologists utilize **systematics**, a discipline focused on classifying organisms and determining their evolutionary relationships. Systematists use data ranging from fossils to molecules and genes to infer evolutionary relationships **(Figure 26.2)**. This information is enabling biologists to construct a tree of all life, which will continue to be refined as additional data are collected.
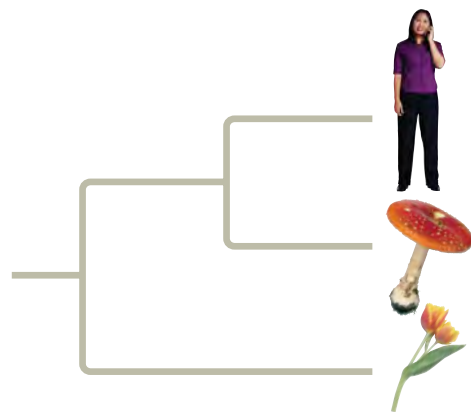
▲ **Figure 26.2 An unexpected family tree.** What are the evolutionary relationships between a human, a mushroom, and a tulip? A phylogeny based on DNA data reveals that—despite appearances—animals (including humans) and fungi (including mushrooms) are more closely related to each other than either is to plants.

Phylogenies show evolutionary relationships

As we discussed in Chapter 22, organisms share many characteristics because of common ancestry. As a result, we can learn a great deal about a species if we know its evolutionary history. For example, an organism is likely to share many of its genes, metabolic pathways, and structural proteins with its close relatives. We'll consider practical applications of such information at the close of this section, but first we'll examine how organisms are named and classified, the scientific discipline of **taxonomy**. We'll also look at how we can interpret and use diagrams that represent evolutionary history.

Binomial Nomenclature

Common names for organisms—such as monkey, finch, and lilac—convey meaning in casual usage, but they can also cause confusion. Each of these names, for example, refers to more than one species. Moreover, some common names do not accurately reflect the kind of organism they signify. Consider these three "fishes": jellyfish (a cnidarian), crayfish (a small lobsterlike crustacean), and silverfish (an insect). And of course, a given organism has different names in different languages.

To avoid ambiguity when communicating about their research, biologists refer to organisms by Latin scientific names. The two-part format of the scientific name, commonly called a **binomial**, was instituted in the 18th century by Carolus Linnaeus (see Chapter 22). The first part of a binomial is the name of the **genus** (plural, *genera*) to which the species belongs. The second part, called the specific epithet, is unique for each species within the genus. An example of a binomial is *Panthera pardus*, the scientific name for the large cat commonly called the leopard. Notice that the first letter of the genus is capitalized and the entire binomial is italicized. (Newly created scientific names are also "latinized": You can name an insect you discover after a friend, but you must add a Latin ending.) Many of the more than 11,000 binomials assigned by Linnaeus are still used today, including the optimistic name he gave our own species—*Homo sapiens*, meaning "wise man."
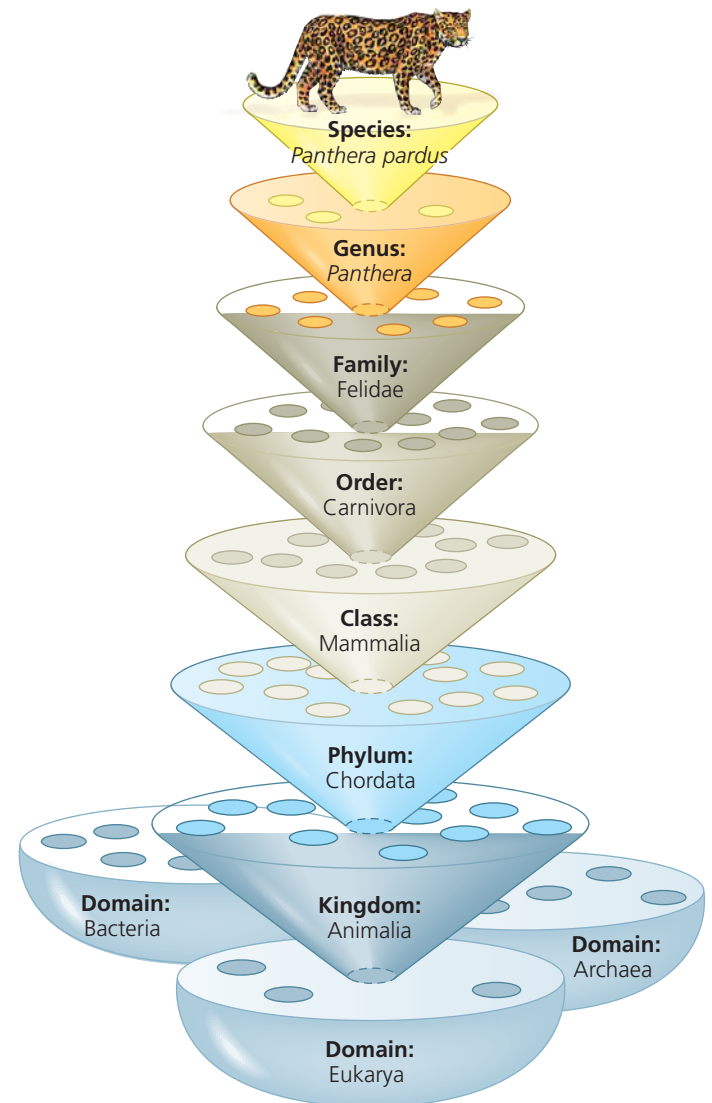
Hierarchical Classification

In addition to naming species, Linnaeus also grouped them into a hierarchy of increasingly inclusive categories. The first grouping is built into the binomial: Species that appear to be closely related are grouped into the same genus. For example, the leopard (*Panthera pardus*) belongs to a genus that also includes the African lion (*Panthera leo*), the tiger (*Panthera tigris*), and the jaguar (*Panthera onca*). Beyond genera, taxonomists employ progressively more comprehensive categories of classification. The taxonomic system named after Linnaeus, the

Linnaean system, places related genera in the same **family**, families into **orders**, orders into **classes**, classes into **phyla** (singular, *phylum*), phyla into **kingdoms**, and, more recently, kingdoms into **domains** (Figure 26.3). The resulting biological classification of a particular organism is somewhat like a postal address identifying a person in a particular apartment, in a building with many apartments, on a street with many apartment buildings, in a city with many streets, and so on.

The named taxonomic unit at any level of the hierarchy is called a **taxon** (plural, *taxa*). In the leopard example, *Panthera* is a taxon at the genus level, and Mammalia is a taxon at the class level that includes all the many orders of mammals. Note that in the Linnaean system, taxa broader than the genus are not italicized, though they are capitalized.

Classifying species is a way to structure our human view of the world. We lump together various species of trees to which we give the common name of pines and distinguish them from other trees that we call firs. Taxonomists have decided

Species:
Panthera pardus

Genus:
Panthera

Family:
Felidae

Order:
Carnivora

Class:
Mammalia

Phylum:
Chordata

Domain:
Bacteria

Kingdom:
Animalia

Domain:
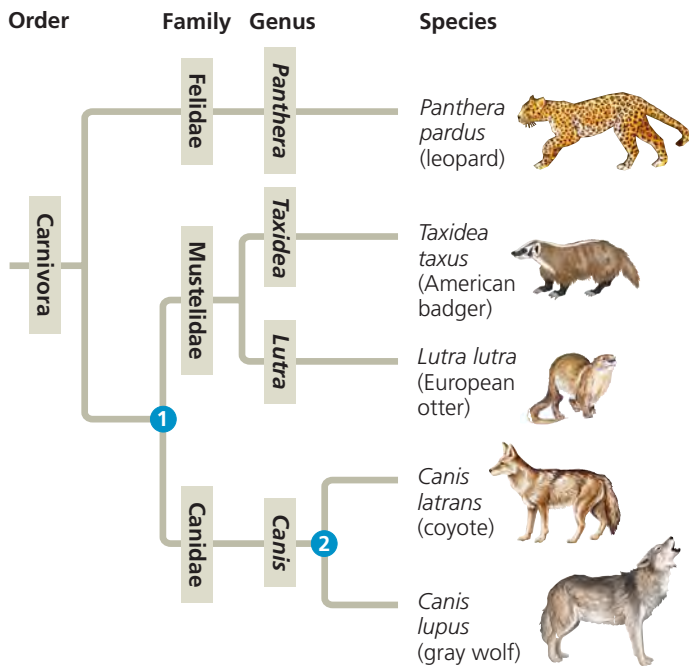Archaea

Domain:
Eukarya

▲ **Figure 26.3 Linnaean classification.** At each level, or "rank," species are placed in groups within more inclusive groups.

that pines and firs are different enough to be placed in separate genera, yet similar enough to be grouped into the same family, Pinaceae. As with pines and firs, higher levels of classification are usually defined by particular characters chosen by taxonomists. However, characters that are useful for classifying one group of organisms may not be appropriate for other organisms. For this reason, the larger categories often are not comparable between lineages; that is, an order of snails does not exhibit the same degree of morphological or genetic diversity as an order of mammals. Furthermore, as we'll see, the placement of species into orders, classes, and so on, does not necessarily reflect evolutionary history.

Linking Classification and Phylogeny

The evolutionary history of a group of organisms can be represented in a branching diagram called a **phylogenetic tree**. As in **Figure 26.4**, the branching pattern often matches how taxonomists have classified groups of organisms nested within more inclusive groups. Sometimes, however, taxonomists have placed a species within a genus (or other group) to which it is *not* most closely related. One reason for misclassification might be that over the course of evolution, a species has lost a key feature shared by its close relatives. If DNA or other new evidence indicates that such a mistake has occurred, the organism may be reclassified to accurately reflect its evolutionary
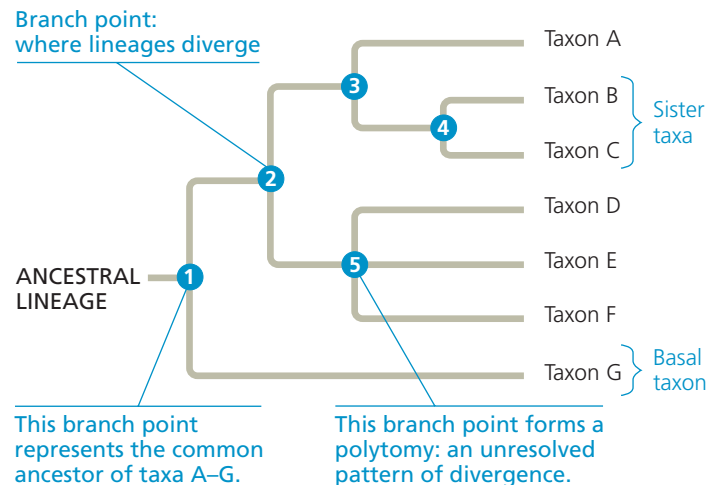
history. Another issue is that while the Linnaean system may distinguish groups, such as mammals, reptiles, birds, and other classes of vertebrates, it tells us nothing about these groups' evolutionary relationships to one another.

In fact, such difficulties in aligning Linnaean classification with phylogeny have led some systematists to propose that classification be based entirely on evolutionary relationships. A system called **PhyloCode**, for example, only names groups that include a common ancestor and all of its descendants. While PhyloCode would change the way taxa are defined and recognized, the taxonomic names of most species would remain the same. But groups would no longer have "ranks" attached to them, such as family or class. Also, some commonly recognized groups would become part of other groups previously of the same rank. For example, because birds evolved from a group of reptiles, Aves (the Linnaean class to which birds are assigned) would be considered a subgroup of Reptilia (also a class in the Linnaean system). Although PhyloCode is controversial, many systematists are adopting the phylogenetic approach on which it is based.

Whether groups are named according to PhyloCode or according to Linnaean classification, a phylogenetic tree represents a hypothesis about evolutionary relationships. These relationships often are depicted as a series of dichotomies, or two-way **branch points**. Each branch point represents the divergence of two evolutionary lineages from a common ancestor. In **Figure 26.5**, for example, branch point ❸ represents the common ancestor of taxa A, B, and C. The position of branch point ❹ to the right of ❸ indicates that taxa B and C diverged after their shared lineage split from that of taxon A. (Tree branches can be rotated around a branch point without changing their evolutionary relationships.)

In Figure 26.5, taxa B and C are **sister taxa**, groups of organisms that share an immediate common ancestor (branch

▲ **Figure 26.4 The connection between classification and phylogeny.** Hierarchical classification can reflect the branching patterns of phylogenetic trees. This tree traces possible evolutionary relationships between some of the taxa within order Carnivora, itself a branch of class Mammalia. The branch point ❶ represents the most recent common ancestor of all members of the weasel (Mustelidae) and dog (Canidae) families. The branch point ❷ represents the most recent common ancestor of coyotes and gray wolves.

▲ **Figure 26.5 How to read a phylogenetic tree.**
DRAW IT *Redraw this tree, rotating the branches around branch points* ❷ *and* ❹. *Does your new version tell a different story about the evolutionary relationships between the taxa? Explain.*

point ④) and hence are each other's closest relatives. Note also that this tree, like most of the phylogenetic trees in this book, is **rooted**, which means that a branch point within the tree (often drawn farthest to the left) represents the most recent common ancestor of all taxa in the tree. The term **basal taxon** refers to a lineage that diverges early in the history of a group and hence, like taxon G in Figure 26.5, lies on a branch that originates near the common ancestor of the group. Finally, the lineage leading to taxa D–F includes a **polytomy**, a branch point from which more than two descendant groups emerge. A polytomy signifies that evolutionary relationships among the taxa are not yet clear.

What We Can and Cannot Learn from Phylogenetic Trees

Let's summarize three key points about phylogenetic trees. First, they are intended to show patterns of descent, not phenotypic similarity. Although closely related organisms often resemble one another due to their common ancestry, they may not if their lineages have evolved at different rates or faced very different environmental conditions. For example, even though crocodiles are more closely related to birds than to lizards (see Figure 22.17), they look more like lizards because morphology has changed dramatically in the bird lineage.

Second, the sequence of branching in a tree does not necessarily indicate the actual (absolute) ages of the particular species. For example, the tree in Figure 26.4 does not indicate that the wolf evolved more recently than the European otter; rather, the tree shows only that the most recent common ancestor of the wolf and otter (branch point ①) lived before the most recent common ancestor of the wolf and coyote (②). To indicate when wolves and otters evolved, the tree would need to include additional divergences in each evolutionary lineage, as well as the dates when those splits occurred. Generally, unless given specific information about what the branch lengths in a phylogenetic tree mean—for example, that they are proportional to time—we should interpret the diagram solely in terms of patterns of descent. No assumptions should be made about when particular species evolved or how much change occurred in each lineage.

Third, we should not assume that a taxon on a phylogenetic tree evolved from the taxon next to it. Figure 26.4 does not indicate that wolves evolved from coyotes or vice versa. We can infer only that the lineage leading to wolves and the lineage leading to coyotes both evolved from the common ancestor ②. That ancestor, which is now extinct, was neither a wolf nor a coyote. However, its descendants include the two *extant* (living) species shown here, wolves and coyotes.

Applying Phylogenies

Understanding phylogeny can have practical applications. Consider maize (corn), which originated in the Americas and is now an important food crop worldwide. From a phylogeny of maize based on DNA data, researchers have been able to identify two species of wild grasses that may be maize's closest living relatives. These two close relatives may be useful as "reservoirs" of beneficial alleles that can be transferred to cultivated maize by cross-breeding or genetic engineering (see Chapter 20).

A different use of phylogenetic trees is described in **Figure 26.6**: investigating whether whale meat samples had
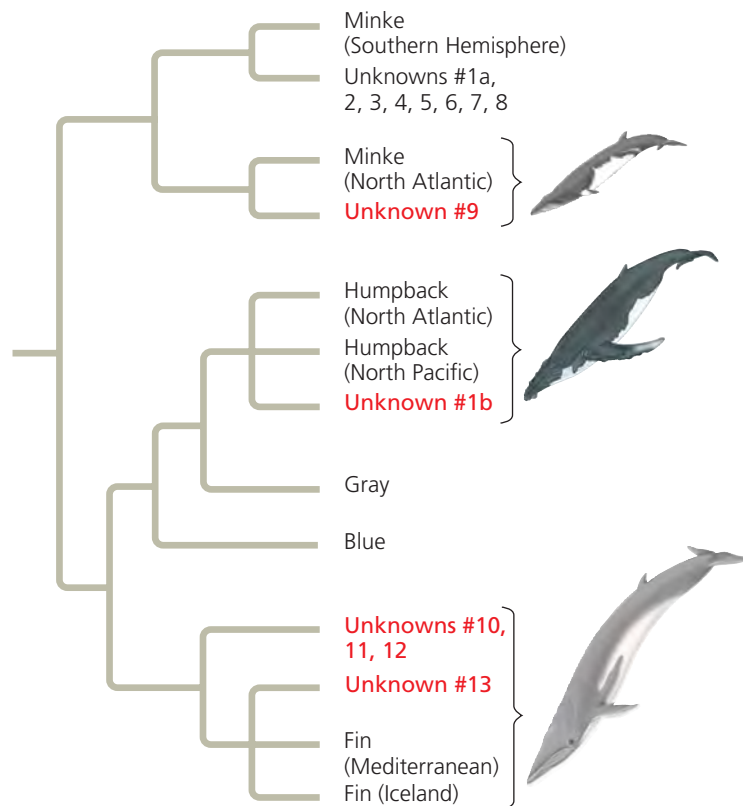
▼ **Figure 26.6**

INQUIRY

What is the species identity of food being sold as whale meat?

EXPERIMENT C. S. Baker, then at the University of Auckland, New Zealand, and S. R. Palumbi, then at the University of Hawaii, purchased 13 samples of "whale meat" from Japanese fish markets. They sequenced a specific part of the mitochondrial DNA from each sample and compared their results with the comparable DNA sequence from known whale species. To infer the species identity of each sample, Baker and Palumbi constructed a *gene tree*, a phylogenetic tree that shows patterns of relatedness among DNA sequences rather than among taxa.

RESULTS The analysis yielded the following gene tree:

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

Minke
(North Atlantic)
Unknown #9

Humpback
(North Atlantic)
Humpback
(North Pacific)
Unknown #1b

Gray

Blue

Unknowns #10,
11, 12
Unknown #13
Fin
(Mediterranean)
Fin (Iceland)

CONCLUSION This analysis indicated that DNA sequences of six of the unknown samples (in red) were most closely related to DNA sequences of whales that are not legal to harvest.

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

WHAT IF? What different results would have indicated that the whale meat had *not* been illegally harvested?

been illegally harvested from whale species protected under international law—rather than from species that can be harvested legally, such as Minke whales caught in the Southern Hemisphere. This phylogeny indicated that meat from humpback, fin, and Minke whales caught in the Northern Hemisphere was being sold illegally in some Japanese fish markets.

How do researchers construct trees like those we've considered here? In the next section, we'll begin to answer that question by examining the data used to determine phylogenies.

1. Which levels of the classification in Figure 26.3 do humans share with leopards?
2. What does the phylogenetic tree in Figure 26.4 indicate about the evolutionary relationships of the leopard, badger, and wolf?
3. Which of the trees shown here depicts an evolutionary history different from the other two? Explain.

(a)   (b)   (c)
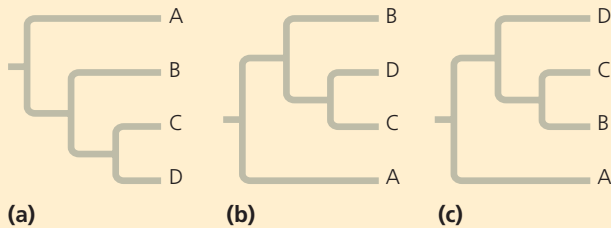
4. **WHAT IF?** Suppose new evidence indicates that taxon E in Figure 26.5 is the sister taxon of a group consisting of taxa D and F. Redraw the tree to accommodate this new finding.

For suggested answers, see Appendix A.

Phylogenies are inferred from morphological and molecular data

To infer phylogeny, systematists must gather as much information as possible about the morphology, genes, and biochemistry of the relevant organisms. It is important to focus on features that result from common ancestry, because only such features reflect evolutionary relationships.

Morphological and Molecular Homologies

Recall that phenotypic and genetic similarities due to shared ancestry are called *homologies*. For example, the similarity in the number and arrangement of bones in the forelimbs of mammals is due to their descent from a common ancestor with the same bone structure; this is an example of a morphological homology (see Figure 22.15). In the same way, genes or other DNA sequences are homologous if they are descended from sequences carried by a common ancestor.

In general, organisms that share very similar morphologies or similar DNA sequences are likely to be more closely related than organisms with vastly different structures or sequences. In some cases, however, the morphological divergence between related species can be great and their genetic divergence small (or vice versa). Consider the Hawaiian silversword plants discussed in Chapter 25. These species vary dramatically in appearance throughout the islands. Some are tall, twiggy trees, and others are dense, ground-hugging shrubs (see Figure 25.20). But despite these striking phenotypic differences, the silverswords' genes are very similar. Based on these small molecular divergences, scientists estimate that the silversword group began to diverge 5 million years ago, which is also about the time when the oldest of the current islands formed. We'll discuss how scientists use molecular data to estimate such divergence times later in this chapter.

Sorting Homology from Analogy

A potential red herring in constructing a phylogeny is similarity due to convergent evolution—called **analogy**—rather than to shared ancestry (homology). As you read in Chapter 22, convergent evolution occurs when similar environmental pressures and natural selection produce similar (analogous) adaptations in organisms from different evolutionary lineages. For example, the two mole-like animals illustrated in **Figure 26.7** are very similar in their external appearance. However, their internal anatomy, physiology, and reproductive systems are very dissimilar. Australian "moles" are marsupials; their young complete their embryonic development in a pouch on the outside of the mother's body. North American moles, in contrast, are eutherians; their young complete

▲ **Figure 26.7 Convergent evolution of analogous burrowing characteristics.** An elongated body, enlarged front paws, small eyes, and a pad of thickened skin that protects a tapered nose all evolved independently in the marsupial Australian "mole" (top) and a eutherian North American mole (bottom).

their embryonic development in the uterus within the mother's body. Indeed, genetic comparisons and the fossil record provide evidence that the common ancestor of these moles lived 140 million years ago, about the time the marsupial and eutherian mammals diverged. This common ancestor and most of its descendants were not mole-like, but analogous characteristics evolved independently in these two mole lineages as they became adapted to similar lifestyles.

Distinguishing between homology and analogy is critical in reconstructing phylogenies. To see why, consider bats and birds, both of which have adaptations that enable flight. This superficial resemblance might imply that bats are more closely related to birds than they are to cats, which cannot fly. But a closer examination reveals that a bat's wing is far more similar to the forelimbs of cats and other mammals than to a bird's wing. Bats and birds descended from a common tetrapod ancestor that lived about 320 million years ago. This common ancestor could not fly. Thus, although the underlying skeletal systems of bats and birds are homologous, their *wings* are not. Flight is enabled in different ways—stretched membranes in the bat wing versus feathers in the bird wing. Fossil evidence also documents that bat wings and bird wings arose independently from the forelimbs of different tetrapod ancestors. Thus, with respect to flight, a bat's wing is *analogous*, not homologous, to a bird's wing. Analogous structures that arose independently are also called **homoplasies** (from the Greek, meaning "to mold in the same way").

Besides corroborative similarities and fossil evidence, another clue to distinguishing between homology and analogy is the complexity of the characters being compared. The more elements that are similar in two complex structures, the more likely it is that they evolved from a common ancestor. For instance, the skulls of an adult human and an adult chimpanzee both consist of many bones fused together. The compositions of the skulls match almost perfectly, bone for bone. It is highly improbable that such complex structures, matching in so many details, have separate origins. More likely, the genes involved in the development of both skulls were inherited from a common ancestor. The same argument applies to comparisons at the gene level. Genes are sequences of thousands of nucleotides, each of which represents an inherited character in the form of one of the four DNA bases: A (adenine), G (guanine), C (cytosine), or T (thymine). If genes in two organisms share many portions of their nucleotide sequences, it is likely that the genes are homologous.

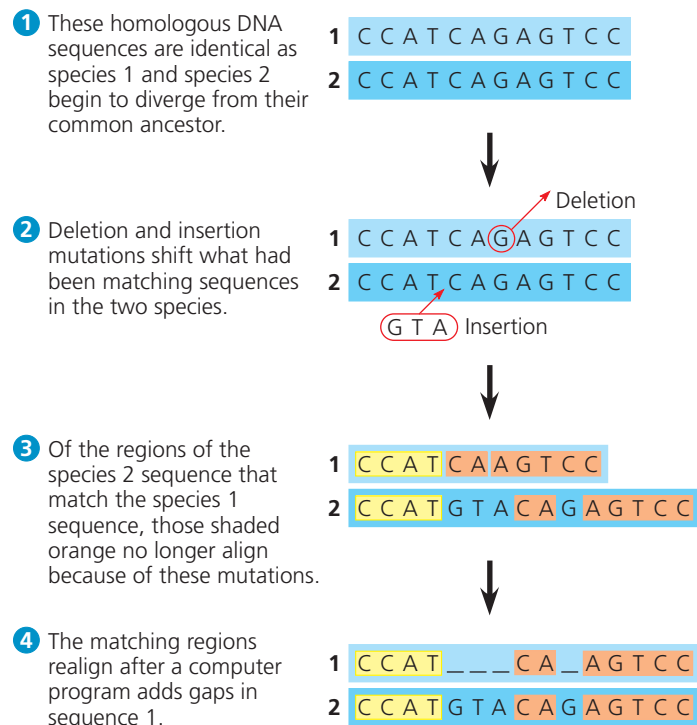## Evaluating Molecular Homologies

Comparisons of DNA molecules often pose technical challenges for researchers. The first step after sequencing the molecules is to align comparable sequences from the species being studied. If the species are very closely related, the sequences probably differ at only one or a few sites. In contrast, comparable nucleic acid sequences in distantly related species usually have different bases at many sites and may have different lengths. This is because insertions and deletions accumulate over long periods of time.

Suppose, for example, that certain noncoding DNA sequences near a particular gene are very similar in two species, except that the first base of the sequence has been deleted in one of the species. The effect is that the remaining sequence shifts back one notch. A comparison of the two sequences that does not take this deletion into account would overlook what in fact is a very good match. To address such problems, researchers have developed computer programs that estimate the best way to align comparable DNA segments of differing lengths **(Figure 26.8)**.

Such molecular comparisons reveal that many base substitutions and other differences have accumulated in the comparable genes of an Australian mole and a North American mole. The many differences indicate that their lineages have diverged greatly since their common ancestor; thus, we say that the living species are not closely related. In contrast, the high degree of gene sequence similarity among the silverswords indicates that they are all very closely related, in spite of their considerable morphological differences.

Just as with morphological characters, it is necessary to distinguish homology from analogy in evaluating molecular similarities for evolutionary studies. Two sequences that resemble each other at many points along their length most likely are

1 These homologous DNA sequences are identical as species 1 and species 2 begin to diverge from their common ancestor.

1 C C A T C A G A G T C C
2 C C A T C A G A G T C C

2 Deletion and insertion mutations shift what had been matching sequences in the two species.

Deletion
1 C C A T C A G A G T C C
2 C C A T C A G A G T C C
G T A Insertion

3 Of the regions of the species 2 sequence that match the species 1 sequence, those shaded orange no longer align because of these mutations.

1 C C A T C A A G T C C
2 C C A T G T A C A G A G T C C

4 The matching regions realign after a computer program adds gaps in sequence 1.

1 C C A T _ _ _ C A _ A G T C C
2 C C A T G T A C A G A G T C C

▲ **Figure 26.8 Aligning segments of DNA.** Systematists search for similar sequences along DNA segments from two species (only one DNA strand is shown for each species). In this example, 11 of the original 12 bases have not changed since the species diverged. Hence, those portions of the sequences still align once the length is adjusted.

```
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
```

▲ **Figure 26.9 A molecular homoplasy.** These two DNA sequences from organisms that are not closely related coincidentally share 25% of their bases. Statistical tools have been developed to determine whether DNA sequences that share more than 25% of their bases do so because they are homologous.

? *Why might you expect organisms that are not closely related to nevertheless share roughly 25% of their bases?*

homologous (see Figure 26.8). But in organisms that do not appear to be closely related, the bases that their otherwise very different sequences happen to share may simply be coincidental matches, called molecular homoplasies **(Figure 26.9)**. Scientists have developed statistical tools that can help distinguish "distant" homologies from such coincidental matches in extremely divergent sequences.

To date, researchers have sequenced more than 110 billion bases of DNA from thousands of species. This enormous collection of data has fueled a boom in the study of phylogeny. The new data have supported earlier hypotheses regarding many evolutionary relationships, such as that between Australian and North American moles, and have clarified other relationships, such as those between the various silverswords. In the rest of this unit, you will see how our understanding of phylogeny has been transformed by **molecular systematics**, the discipline that uses data from DNA and other molecules to determine evolutionary relationships.
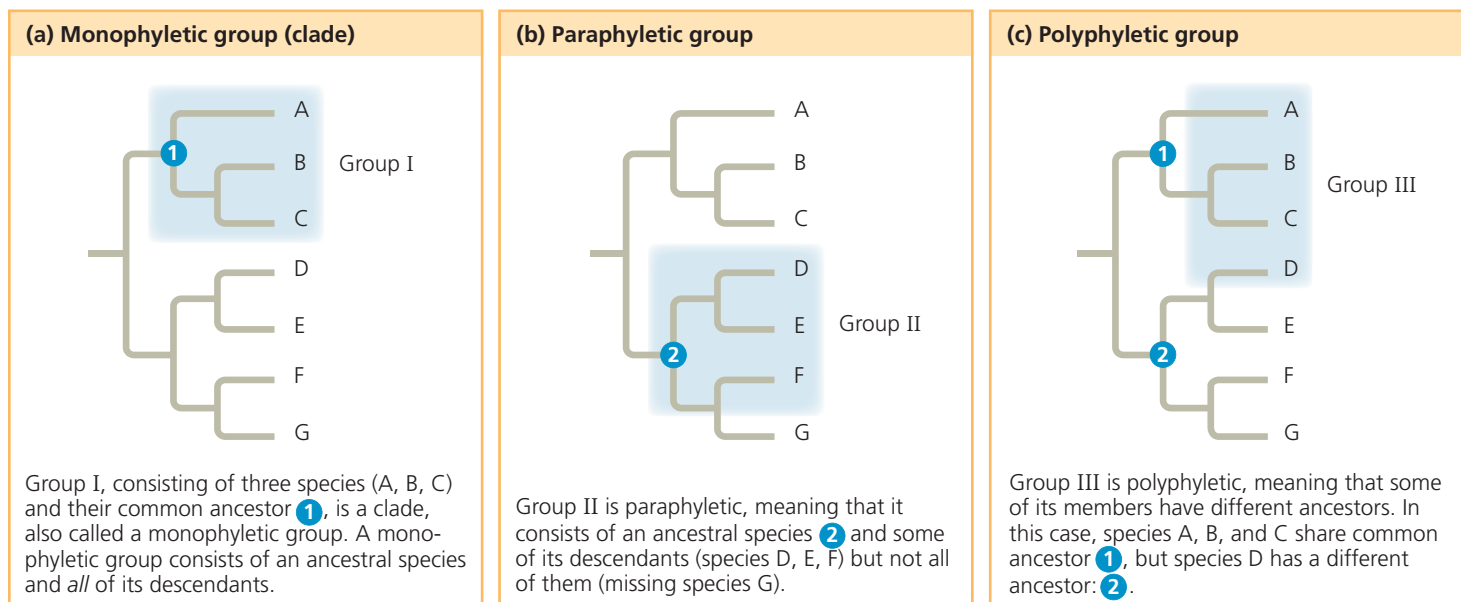
CONCEPT 26.3

Shared characters are used to construct phylogenetic trees

In reconstructing phylogenies, the first step is to distinguish homologous features from analogous ones (since only homology reflects evolutionary history). Next we must choose a method of inferring phylogeny from these homologous characters. A widely used set of methods is known as cladistics.

Cladistics

In the approach to systematics called **cladistics**, common ancestry is the primary criterion used to classify organisms. Using this methodology, biologists attempt to place species into groups called **clades**, each of which includes an ancestral species and all of its descendants **(Figure 26.10a)**. Clades, like

▼ **Figure 26.10 Monophyletic, paraphyletic, and polyphyletic groups.**

(a) Monophyletic group (clade)

Group I, consisting of three species (A, B, C) and their common ancestor **1**, is a clade, also called a monophyletic group. A monophyletic group consists of an ancestral species and *all* of its descendants.

(b) Paraphyletic group

Group II is paraphyletic, meaning that it consists of an ancestral species **2** and some of its descendants (species D, E, F) but not all of them (missing species G).

(c) Polyphyletic group

Group III is polyphyletic, meaning that some of its members have different ancestors. In this case, species A, B, and C share common ancestor **1**, but species D has a different ancestor: **2**.

taxonomic ranks, are nested within larger clades. In Figure 26.4, for example, the cat group (Felidae) represents a clade within a larger clade (Carnivora) that also includes the dog group (Canidae). However, a taxon is equivalent to a clade only if it is **monophyletic** (from the Greek, meaning "single tribe"), signifying that it consists of an ancestral species and all of its descendants (see Figure 26.10a). Contrast this with a **paraphyletic** ("beside the tribe") group, which consists of an ancestral species and some, but not all, of its descendants **(Figure 26.10b)**, or a **polyphyletic** ("many tribes") group, which includes taxa with different ancestors **(Figure 26.10c)**. Next we'll discuss how clades are identified using shared derived characters.

Shared Ancestral and Shared Derived Characters

As a result of descent with modification, organisms both share characteristics with their ancestors and differ from them. For example, all mammals have backbones, but a backbone does not distinguish mammals from other vertebrates because *all* vertebrates have backbones. The backbone predates the branching of mammals from other vertebrates. Thus for mammals, the backbone is a **shared ancestral character**, a character that originated in an ancestor of the taxon. In contrast, hair is a character shared by all mammals but *not* found in their ancestors. Thus, in mammals, hair is considered a **shared derived character**, an evolutionary novelty unique to a clade.
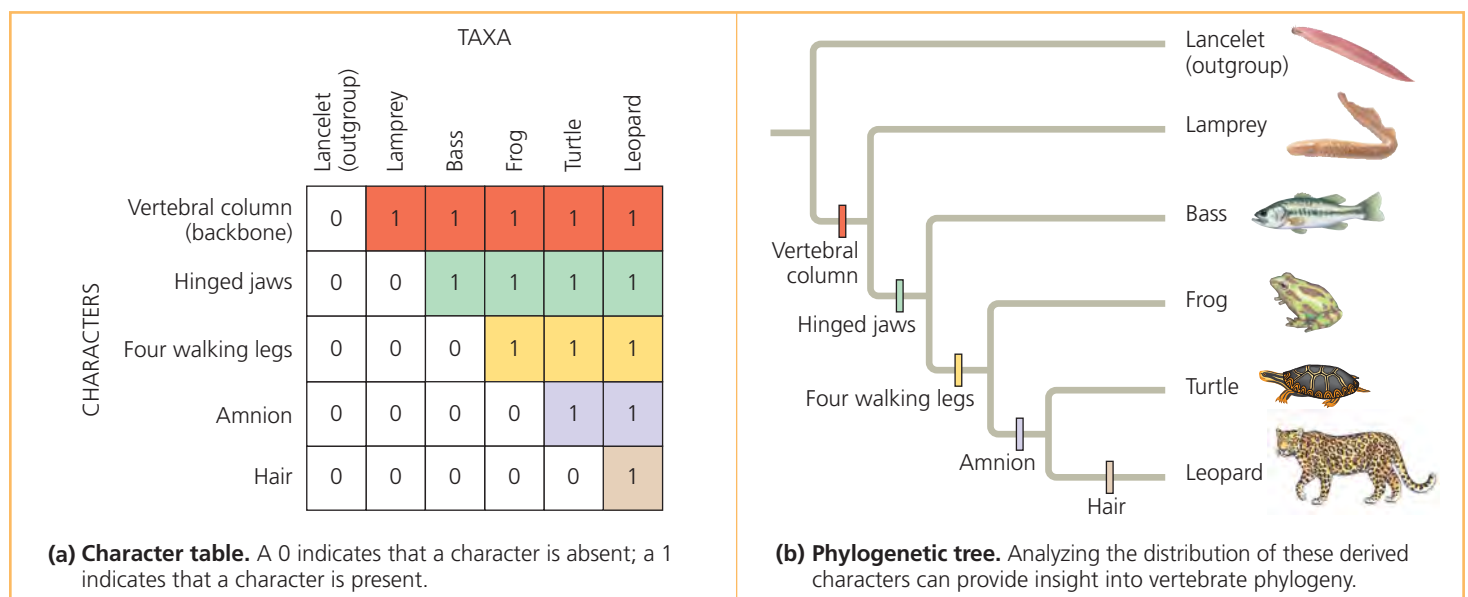
Note that it is a relative matter whether a particular character is considered ancestral or derived. A backbone can also qualify as a shared derived character, but only at a deeper branch point that distinguishes all vertebrates from other animals. Among mammals, a backbone is considered a shared ancestral character because it was present in the ancestor common to all mammals.

Inferring Phylogenies Using Derived Characters

Shared derived characters are unique to particular clades. Because all features of organisms arose at some point in the history of life, it should be possible to determine the clade in which each shared derived character first appeared and to use that information to infer evolutionary relationships.

To see how this analysis is done, consider the set of characters shown in **Figure 26.11a** for each of five vertebrates—a leopard, turtle, frog, bass, and lamprey (a jawless aquatic vertebrate). As a basis of comparison, we need to select an outgroup. An **outgroup** is a species or group of species from an evolutionary lineage that is known to have diverged before the lineage that includes the species we are studying (the **ingroup**). A suitable outgroup can be determined based on evidence from morphology, paleontology, embryonic development, and gene sequences. An appropriate outgroup for our example is the lancelet, a small animal that lives in mudflats and (like vertebrates) is a member of Chordata. Unlike the vertebrates, however, the lancelet does not have a backbone.

By comparing members of the ingroup with each other and with the outgroup, we can determine which characters were derived at the various branch points of vertebrate evolution. For example, *all* of the vertebrates in the ingroup have backbones: This character was present in the ancestral vertebrate, but not in the outgroup. Now note that hinged jaws are a character absent in lampreys but present in other members of the ingroup; this character helps us to identify an early branch point in the vertebrate clade. Proceeding in this way, we can translate the data in our table of characters into a phylogenetic tree that groups all the ingroup taxa into a hierarchy based on their shared derived characters **(Figure 26.11b)**.

	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

TAXA — CHARACTERS

(a) Character table. A 0 indicates that a character is absent; a 1 indicates that a character is present.

(b) Phylogenetic tree. Analyzing the distribution of these derived characters can provide insight into vertebrate phylogeny.

▲ **Figure 26.11 Constructing a phylogenetic tree.** The characters used here include the amnion, a membrane that encloses the embryo inside a fluid-filled sac (see Figure 34.25).

DRAW IT *In (b), circle the most inclusive clade for which a hinged jaw is a shared ancestral character.*
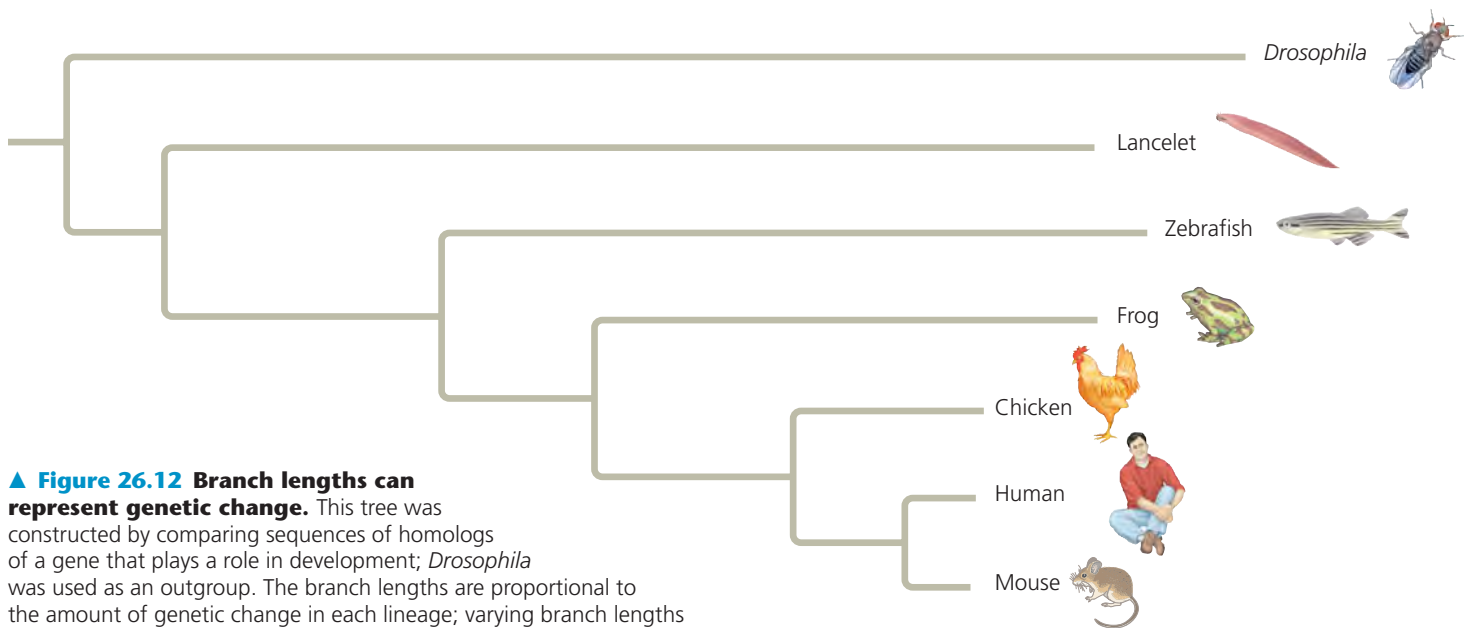
▲ **Figure 26.12 Branch lengths can represent genetic change.** This tree was constructed by comparing sequences of homologs of a gene that plays a role in development; *Drosophila* was used as an outgroup. The branch lengths are proportional to the amount of genetic change in each lineage; varying branch lengths indicate that the gene has evolved at different rates in different lineages.

? *In which vertebrate lineage has the studied gene evolved most rapidly? Explain.*

Phylogenetic Trees with Proportional Branch Lengths

In the phylogenetic trees we have presented so far, the lengths of the tree's branches do not indicate the degree of evolutionary change in each lineage. Furthermore, the chronology represented by the branching pattern of the tree is relative (earlier versus later) rather than absolute (how many millions of years ago). But in some tree diagrams, branch lengths are proportional to amount of evolutionary change or to the times at which particular events occurred.

In **Figure 26.12**, for example, the branch length of the phylogenetic tree reflects the number of changes that have taken place in a particular DNA sequence in that lineage. Note that the total length of the horizontal lines from the base of the tree to the mouse is less than that of the line leading to the outgroup species, the fruit fly *Drosophila*. This implies that in the time since the mouse and fly diverged from a common ancestor, more genetic changes have occurred in the *Drosophila* lineage than in the mouse lineage.

Even though the branches of a phylogenetic tree may have different lengths, among organisms alive today, all the different lineages that descend from a common ancestor have survived for the same number of years. To take an extreme example, humans and bacteria had a common ancestor that lived over 3 billion years ago. Fossils and genetic evidence indicate that this ancestor was a single-celled prokaryote. Even though bacteria have apparently changed little in their morphology since that common ancestor, there have nonetheless been 3 billion years of evolution in the bacterial lineage, just as there have been 3 billion years of evolution in the eukaryotic lineage that includes humans.

These equal spans of chronological time can be represented in a phylogenetic tree whose branch lengths are proportional to time **(Figure 26.13)**. Such a tree draws on fossil data to place branch points in the context of geologic time. Additionally, it is possible to combine these two types of trees by labeling branch points with information about rates of genetic change or dates of divergence.

Maximum Parsimony and Maximum Likelihood

As the growing database of DNA sequences enables us to study more species, the difficulty of building the phylogenetic tree that best describes their evolutionary history also grows. What if you are analyzing data for 50 species? There are 3×10^{76} different ways to arrange 50 species into a tree! And which tree in this huge forest reflects the true phylogeny? Systematists can never be sure of finding the most accurate tree in such a large data set, but they can narrow the possibilities by applying the principles of maximum parsimony and maximum likelihood.

According to the principle of **maximum parsimony**, we should first investigate the simplest explanation that is consistent with the facts. (The parsimony principle is also called "Occam's razor" after William of Occam, a 14th-century English philosopher who advocated this minimalist problem-solving approach of "shaving away" unnecessary complications.) In the case of trees based on morphology, the most parsimonious tree requires the fewest evolutionary events, as measured by the origin of shared derived morphological characters. For phylogenies based on DNA, the most parsimonious tree requires the fewest base changes.
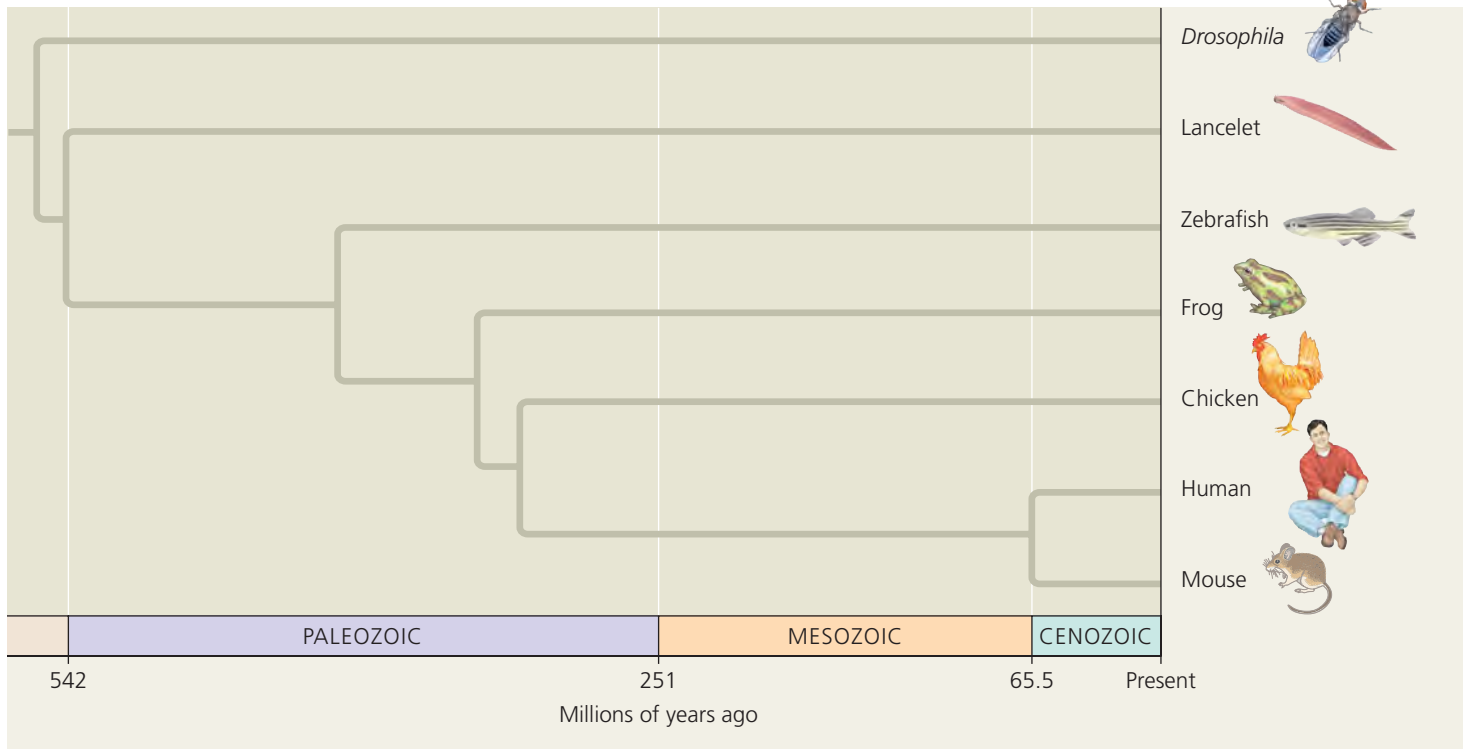
▲ **Figure 26.13 Branch lengths can indicate time.** This tree is based on the same molecular data as the tree in Figure 26.12, but here the branch points are mapped to dates based on fossil evidence. Thus, the branch lengths are proportional to time. Each lineage has the same total length from the base of the tree to the branch tip, indicating that all the lineages have diverged from the common ancestor for equal amounts of time.

	Human	Mushroom	Tulip
Human	0	30%	40%
Mushroom		0	40%
Tulip			0

(a) Percentage differences between sequences

Tree 1: More likely   Tree 2: Less likely

(b) Comparison of possible trees

The principle of **maximum likelihood** states that given certain probability rules about how DNA sequences change over time, a tree can be found that reflects the most likely sequence of evolutionary events. Maximum-likelihood methods are complex, but as a simple example, let us return to the phylogenetic relationships between a human, a mushroom, and a tulip. **Figure 26.14** shows two possible, equally parsimonious trees for this trio. Tree 1 is more likely if we assume that DNA changes have occurred at equal rates along all the branches of the tree from the common ancestor. Tree 2 requires assuming that the rate of evolution slowed greatly in the mushroom lineage and sped up greatly in the tulip lineage. Thus, assuming that equal rates are more common than unequal rates, tree 1 is more likely. We will soon see that many genes do evolve at approximately equal rates in different lineages. But note that if we find new evidence of unequal rates, tree 2 might be more likely! The likelihood of a tree depends on the assumptions on which it is based.

Scientists have developed many computer programs to search for trees that are parsimonious and likely. When a large amount of accurate data is available, the methods used in these programs usually yield similar trees. As an example of one
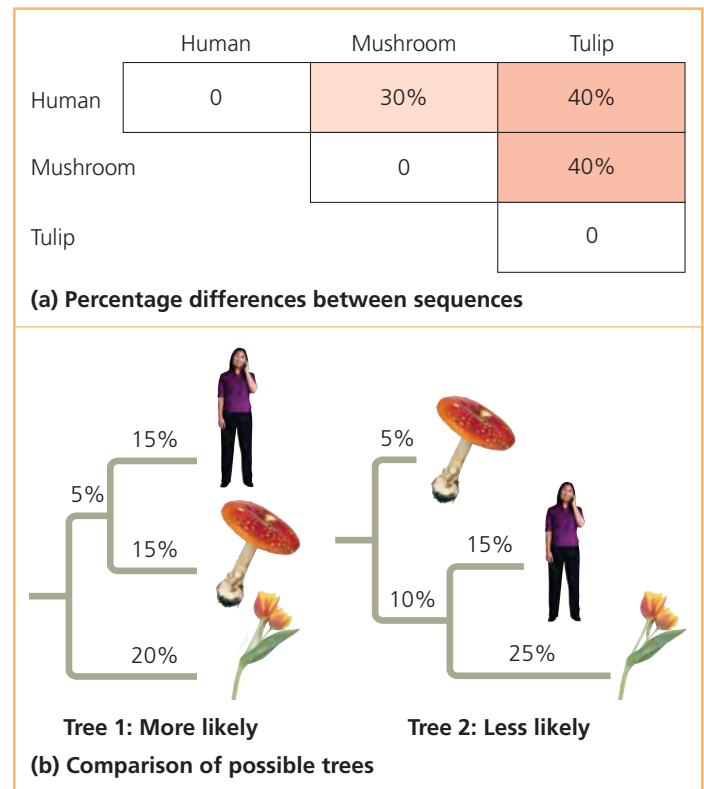
▲ **Figure 26.14 Trees with different likelihoods.** Based on percentage differences between genes in a human, a mushroom, and a tulip **(a)**, there are two phylogenetic trees with the same total branch length **(b)**. The sum of the percentages from a point of divergence in a tree equals the percentage differences in (a). For example, in tree 1, the human–tulip divergence is 15% + 5% + 20% = 40%. In tree 2, this divergence also equals 40% (15% + 25%). If the genes have evolved at the same rate in the different branches, tree 1 is more likely.

RESEARCH METHOD

Applying Parsimony to a Problem in Molecular Systematics

APPLICATION In considering possible phylogenies for a group of species, systematists compare molecular data for the species. An efficient way to begin is by identifying the most parsimonious hypothesis—the one that requires the fewest evolutionary events (molecular changes) to have occurred.

TECHNIQUE Follow the numbered steps as we apply the principle of parsimony to a hypothetical phylogenetic problem involving three closely related bird species.

Species I   Species II   Species III

1 First, draw the three possible phylogenies for the species. (Although only 3 trees are possible when ordering 3 species, the number of possible trees increases rapidly with the number of species: There are 15 trees for 4 species and 34,459,425 trees for 10 species.)

Three phylogenetic hypotheses:
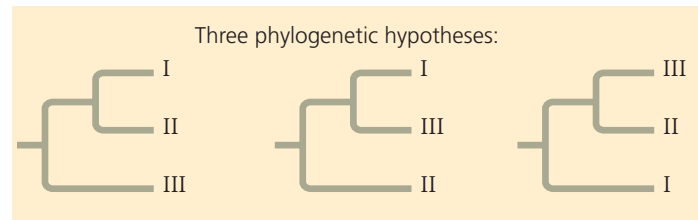
2 Tabulate the molecular data for the species. In this simplified example, the data represent a DNA sequence consisting of just four nucleotide bases. Data from several outgroup species (not shown) were used to infer the ancestral DNA sequence.

|	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 Now focus on site 1 in the DNA sequence. In the tree on the left, a single base-change event, represented by the purple hatchmark on the branch leading to species I and II (and labeled 1/C, indicating a change at site 1 to nucleotide C), is sufficient to account for the site 1 data. In the other two trees, two base-change events are necessary.
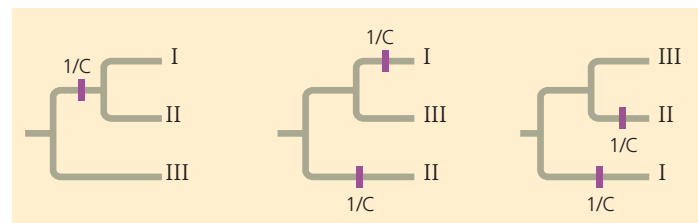
4 Continuing the comparison of bases at sites 2, 3, and 4 reveals that each of the three trees requires a total of five additional base-change events (purple hatchmarks).
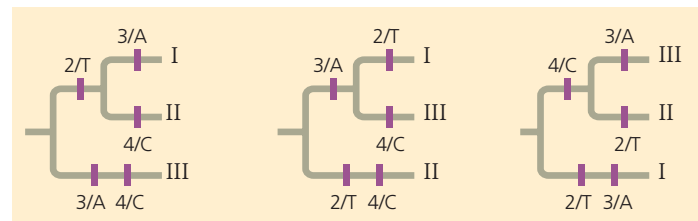
RESULTS To identify the most parsimonious tree, we total all of the base-change events noted in steps 3 and 4. We conclude that the first tree is the most parsimonious of the three possible phylogenies. (In a real example, many more sites would be analyzed. Hence, the trees would often differ by more than one base-change event.)

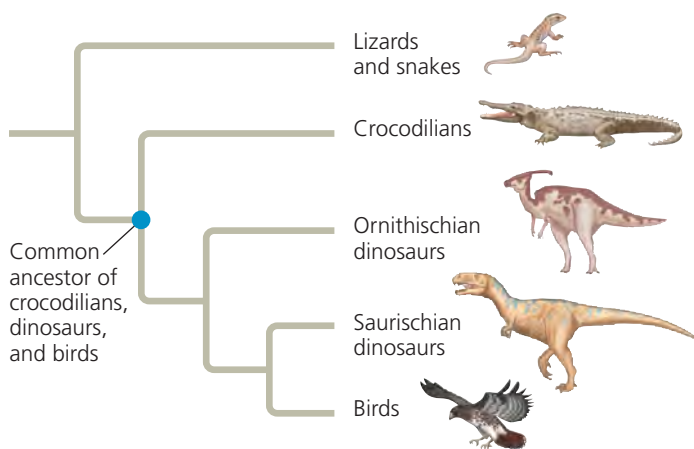6 events   7 events   7 events
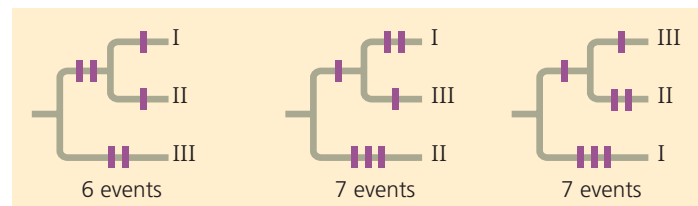
method, **Figure 26.15**, on the facing page, walks you through the process of identifying the most parsimonious molecular tree for a three-species problem. Computer programs use the principle of parsimony to estimate phylogenies in a similar way: They examine large numbers of possible trees and select the tree or trees that require fewest evolutionary changes.

Phylogenetic Trees as Hypotheses

This is a good place to reiterate that any phylogenetic tree represents a hypothesis about how the various organisms in the tree are related to one another. The best hypothesis is the one that best fits all the available data. A phylogenetic hypothesis may be modified when new evidence compels systematists to revise their trees. Indeed, while many older phylogenetic hypotheses have been supported by new morphological and molecular data, others have been changed or rejected.

Thinking of phylogenies as hypotheses also allows us to use them in a powerful way: We can make and test predictions based on the assumption that a phylogeny—our hypothesis—is correct. For example, in an approach known as *phylogenetic bracketing*, we can predict (by parsimony) that features shared by two groups of closely related organisms are present in their common ancestor and all of its descendants unless independent data indicate otherwise. (Note that "prediction" can refer to unknown past events as well as to evolutionary changes yet to occur.)

This approach has been used to make novel predictions about dinosaurs. For example, there is evidence that birds descended from the theropods, a group of bipedal saurischian dinosaurs. As seen in **Figure 26.16**, the closest living relatives of birds are crocodiles. Birds and crocodiles share numerous features: They have four-chambered hearts, they "sing" to defend territories and attract mates (although a crocodile's "song" is more l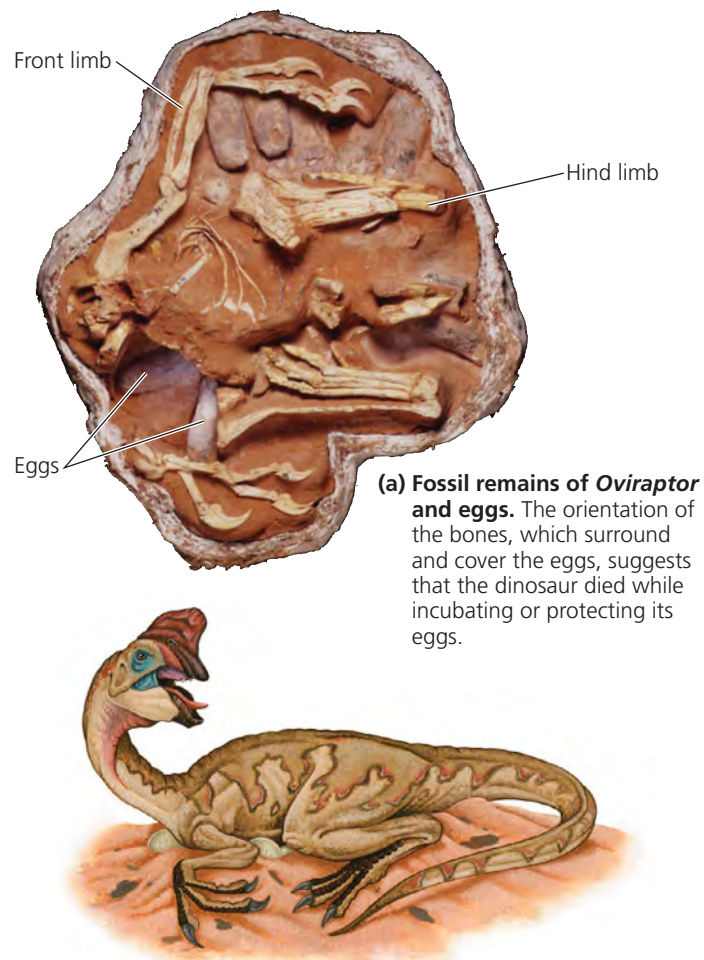ike a bellow), and they build nests. Both birds and crocodiles also care for their eggs by *brooding*, a behavior in which a parent warms the eggs with its body. Birds brood by sitting on their eggs, whereas crocodiles cover their eggs with their neck. Reasoning that any feature shared by birds and crocodiles is likely to have been present in their common ancestor (denoted by the blue dot in Figure 26.16) and *all* of its descendants, biologists predicted that dinosaurs had four-chambered hearts, sang, built nests, and exhibited brooding.

Internal organs, such as the heart, rarely fossilize, and it is, of course, difficult to test whether dinosaurs sang to defend territories and attract mates. However, fossilized dinosaur eggs and nests have provided evidence supporting the prediction of brooding in dinosaurs. First, a fossil embryo of an *Oviraptor* dinosaur was found, still inside its egg. This egg was identical to those found in another fossil, one that showed an *Oviraptor* adult crouching over a group of eggs in a posture similar to that in brooding birds today **(Figure 26.17)**. Researchers suggested that the *Oviraptor* dinosaur preserved in this second fossil died while incubating or protecting its eggs. The broader conclusion that emerged from this work—that

(a) **Fossil remains of *Oviraptor* and eggs.** The orientation of the bones, which surround and cover the eggs, suggests that the dinosaur died while incubating or protecting its eggs.

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

▲ **Figure 26.17** Fossil support for a phylogenetic prediction: Dinosaurs built nests and brooded their eggs.

▲ **Figure 26.16** A phylogenetic tree of birds and their close relatives.

? *What is the most basal taxon represented in this tree?*

dinosaurs built nests and exhibited brooding—has since been strengthened by additional fossil discoveries that show that other species of dinosaurs built nests and sat on their eggs. Finally, by supporting predictions based on the phylogenetic hypothesis shown in Figure 26.16, fossil discoveries of nests and brooding in dinosaurs provide independent data that suggest that the hypothesis is correct.

CONCEPT CHECK 26.3

1. To distinguish a particular clade of mammals within the larger clade that corresponds to class Mammalia, would hair be a useful character? Why or why not?
2. The most parsimonious tree of evolutionary relationships can be inaccurate. How can this occur?
3. **WHAT IF?** Draw a phylogenetic tree that includes the relationships from both Figure 25.6 and Figure 26.16. Traditionally, all the taxa shown besides birds and mammals were classified as reptiles. Would a cladistic approach support that classification? Explain.

For suggested answers, see Appendix A.

CONCEPT 26.4

An organism's evolutionary history is documented in its genome

As you have seen in this chapter, molecular systematics—using comparisons of nucleic acids or other molecules to deduce relatedness—is a valuable approach for tracing evolutionary history. The molecular approach helps us understand phylogenetic relationships that cannot be determined by nonmolecular methods such as comparative anatomy. For example, molecular systematics helps us uncover evolutionary relationships between groups that have little common ground for morphological comparison, such as animals and fungi. And molecular methods allow us to reconstruct phylogenies among groups of present-day organisms for which the fossil record is poor or lacking entirely. Overall, molecular biology has helped to extend systematics to evolutionary relationships far above and below the species level, ranging from the major branches of the tree of life to its finest twigs.

Different genes evolve at different rates, even in the same evolutionary lineage. As a result, molecular trees can represent short or long periods of time, depending on which genes are used. For example, the DNA that codes for ribosomal RNA (rRNA) changes relatively slowly. Therefore, comparisons of DNA sequences in these genes are useful for investigating relationships between taxa that diverged hundreds of millions of years ago. Studies of rRNA sequences indicate, for instance, that fungi are more closely related to animals than to green plants (see Figure 26.2). In contrast, mitochondrial DNA (mtDNA) evolves relatively rapidly and can be used to ex-

plore recent evolutionary events. One research team has traced the relationships among Native American groups through their mtDNA sequences. The molecular findings corroborate other evidence that the Pima of Arizona, the Maya of Mexico, and the Yanomami of Venezuela are closely related, probably descending from the first of three waves of immigrants that crossed the Bering land bridge from Asia to the Americas about 15,000 years ago.

Gene Duplications and Gene Families

What does molecular systematics reveal about the evolutionary history of genome change? Consider gene duplication, which plays a particularly important role in evolution because it increases the number of genes in the genome, providing more opportunities for further evolutionary changes. Molecular techniques now allow us to trace the phylogenies of gene duplications and the influence of these duplications on genome evolution. These molecular phylogenies must account for repeated duplications that have resulted in *gene families*, groups of related genes within an organism's genome (see Figure 21.11). Accounting for such duplications leads us to distinguish two types of homologous genes: orthologous genes and paralogous genes. **Orthologous genes** (from the Greek *orthos*, exact) are those found in different species, and their divergence traces back to the speciation events that produced the species **(Figure 26.18a)**. The cytochrome *c* genes (which code for an electron transport chain protein) in humans and dogs are orthologous. In **paralogous genes** (from the Greek *para*, in parallel), the homology results from gene duplication; hence, multiple copies of these genes have diverged from one another within a species **(Figure 26.18b)**. In Chapter 23, you encountered the example of olfactory receptor genes, which have undergone many gene duplications in vertebrates. Humans and mice both have huge families of more than 1,000 of these paralogous genes.
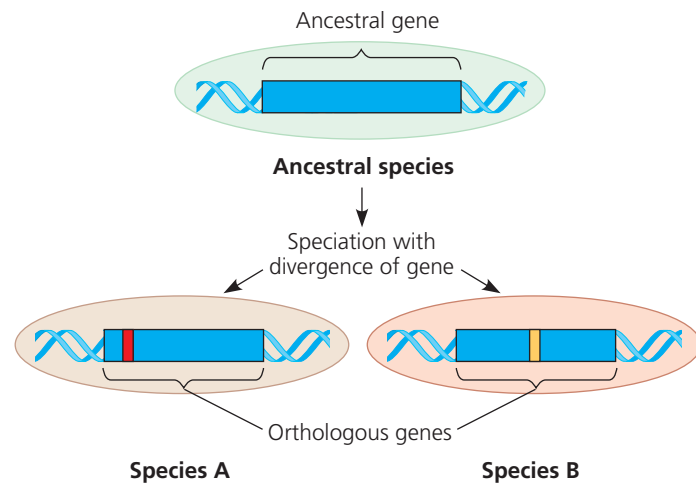
Note that orthologous genes can only diverge after speciation has taken place, that is, after the genes are found in separate gene pools. For example, although the cytochrome *c* genes in humans and dogs serve the same function, the gene's sequence in humans has diverged from that in dogs in the time since these species last shared a common ancestor. Paralogous genes, on the other hand, can diverge within a species because they are present in more than one copy in the genome. The paralogous genes that make up the olfactory receptor gene family in humans have diverged from each other during our long evolutionary history. They now specify proteins that confer sensitivity to a wide variety of molecules, ranging from food odors to sex pheromones.
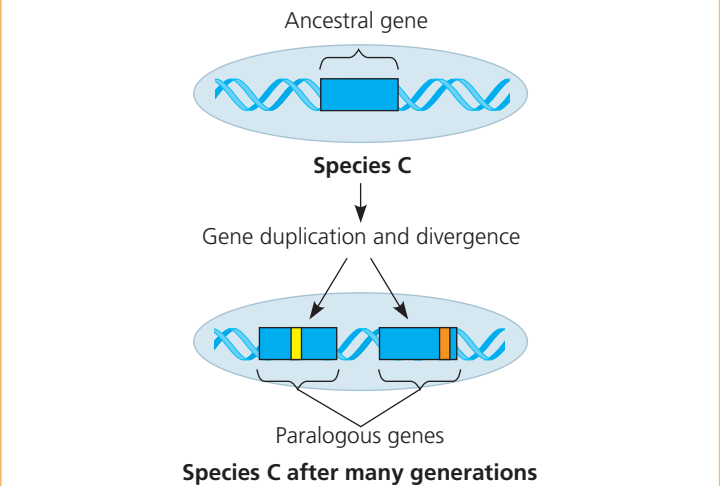
Genome Evolution

Now that we can compare the entire genomes of different organisms, including our own, two patterns have emerged. First, lineages that diverged long ago can share orthologous genes. For

▼ **Figure 26.18** **Two types of homologous genes.** Colored bands mark regions of the genes where differences in base sequences have accumulated.

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

Ancestral gene

Ancestral species

Speciation with divergence of gene

Species A   Species B

Orthologous genes

(b) Formation of paralogous genes: within a species

Ancestral gene

Species C

Gene duplication and divergence

Species C after many generations

Paralogous genes

---

example, though the human and mouse lineages diverged about 65 million years ago, 99% of the genes of humans and mice are orthologous. And 50% of human genes are orthologous with those of yeast, despite 1 billion years of divergent evolution. Such commonalities explain why disparate organisms nevertheless share many biochemical and developmental pathways.

Second, the number of genes a species has doesn't seem to increase through duplication at the same rate as perceived phenotypic complexity. Humans have only about four times as many genes as yeast, a single-celled eukaryote, even though—unlike yeast—we have a large, complex brain and a body with more than 200 different types of tissues. Evidence is emerging that many human genes are more versatile than those of yeast: A single human gene can encode multiple proteins that perform different tasks in various body tissues. Unraveling the mechanisms that cause this genomic versatility and phenotypic variation is an exciting challenge.

CONCEPT CHECK 26.4

1. Explain how comparing proteins of two species can yield data about the species' evolutionary relationship.
2. **WHAT IF?** Suppose gene A is orthologous in species 1 and species 2, and gene B is paralogous to gene A in species 1. Suggest a sequence of two evolutionary events that could result in the following: Gene A differs considerably between species, yet gene A and gene B show little divergence from each other.
3. **MAKE CONNECTIONS** Review Figure 18.13 (p. 363); then suggest how a particular gene could have different functions in different tissues within an organism.

For suggested answers, see Appendix A.

Molecular clocks help track evolutionary time

One goal of evolutionary biology is to understand the relationships among all organisms, including those for which there is no fossil record. However, if we attempt to determine the timing of molecular phylogenies that extend beyond the fossil record, we must rely on an assumption about how change occurs at the molecular level.

Molecular Clocks

We stated earlier that researchers have estimated that the common ancestor of Hawaiian silversword plants lived about 5 million years ago. How did they make this estimate? They relied on the concept of a **molecular clock**, a yardstick for measuring the absolute time of evolutionary change based on the observation that some genes and other regions of genomes appear to evolve at constant rates. The assumption underlying the molecular clock is that the number of nucleotide substitutions in orthologous genes is proportional to the time that has elapsed since the species branched from their common ancestor (divergence time). In the case of paralogous genes, the number of substitutions is proportional to the time since the ancestral gene was duplicated.

We can calibrate the molecular clock of a gene that has a reliable average rate of evolution by graphing the number of genetic differences—for example, nucleotide, codon, or amino acid differences—against the dates of evolutionary branch points that are known from the fossil record
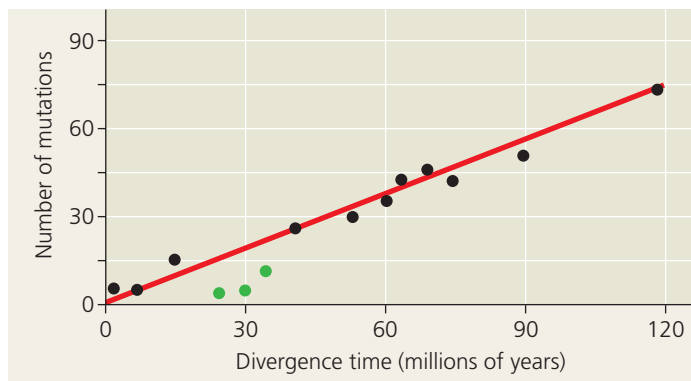
▲ **Figure 26.19** **A molecular clock for mammals.** The number of accumulated mutations in seven proteins has increased over time in a consistent manner for most mammal species. The three green data points represent primate species, whose proteins appear to have evolved more slowly than those of other mammals. The divergence time for each data point was based on fossil evidence.

? *Use the graph to estimate the divergence time for a mammal with a total of 30 mutations in the seven proteins.*

(Figure 26.19). The average rates of genetic change inferred from such graphs can then be used to estimate the dates of events that cannot be discerned from the fossil record, such as the origin of the silverswords discussed earlier.

Of course, no gene marks time with complete precision. In fact, some portions of the genome appear to have evolved in irregular bursts that are not at all clocklike. And even those genes that seem to act as reliable molecular clocks are accurate only in the statistical sense of showing a fairly smooth *average* rate of change. Over time, there may still be deviations from that average rate. Furthermore, the same gene may evolve at different rates in different groups of organisms. And even among genes that are clocklike, the rate of the clock may vary greatly from one gene to another; some genes evolve a million times faster than others.

Neutral Theory

The observed regularity of change that enables us to use some genes as molecular clocks raises the possibility that many of the changes in these sequences result from mutations that have become fixed in a population by genetic drift (see Chapter 23) and that the changes are selectively neutral—neither beneficial nor detrimental. In the 1960s, Motoo Kimura, at the Japanese National Institute of Genetics, and Jack King and Thomas Jukes, at the University of California, Berkeley, independently published papers describing this **neutral theory**— that much evolutionary change in genes and proteins has no effect on fitness and therefore is not influenced by natural selection. Kimura pointed out that many new mutations are harmful and are removed quickly. But if most of the rest are neutral and have little or no effect on fitness, then the rate of molecular change should indeed be regular, like a clock. Differences in the clock rate for different genes are a function of how important a gene is. If the exact sequence of amino

acids that a gene specifies is essential to survival, most of the mutational changes will be harmful and only a few will be neutral. As a result, such genes change only slowly. But if the exact sequence of amino acids is less critical, fewer of the new mutations will be harmful and more will be neutral. Such genes change more quickly.

Problems with Molecular Clocks

In fact, molecular clocks do not run as smoothly as neutral theory predicts. Many irregularities are likely to be the result of natural selection in which certain DNA changes are favored over others. Consequently, some scientists question the utility of molecular clocks for timing evolution. Their skepticism is part of a broader debate about the extent to which neutral genetic variation can account for some kinds of DNA diversity. Indeed, evidence suggests that almost half the amino acid differences in proteins of two *Drosophila* species, *D. simulans* and *D. yakuba*, are not neutral but have resulted from directional natural selection. But because the direction of natural selection may change repeatedly over long periods of time (and hence may average out), some genes experiencing selection can nevertheless serve as approximate markers of elapsed time.

Another question arises when researchers attempt to extend molecular clocks beyond the time span documented by the fossil record. Although some fossils are more than 3 billion years old, these are very rare. An abundant fossil record extends back only about 550 million years, but molecular clocks have been used to date evolutionary divergences that occurred a billion or more years ago. These estimates assume that the clocks have been constant for all that time. Such estimates are highly uncertain.

In some cases, problems may be avoided by calibrating molecular clocks with many genes rather than just one or a few genes (as is often done). By using many genes, fluctuations in evolutionary rate due to natural selection or other factors that vary over time may average out. For example, one group of researchers constructed molecular clocks of vertebrate evolution from published sequence data for 658 nuclear genes. Despite the broad period of time covered (nearly 600 million years) and the fact that natural selection probably affected some of these genes, their estimates of divergence times agreed closely with fossil-based estimates.

Applying a Molecular Clock: The Origin of HIV

Researchers have used a molecular clock to date the origin of HIV infection in humans. Phylogenetic analysis shows that HIV, the virus that causes AIDS, is descended from viruses that infect chimpanzees and other primates. (Most of these viruses do not cause AIDS-like diseases in their native hosts.) When did HIV jump to humans? There is no simple answer, because the virus has spread to humans more than once. The multiple origins of HIV are reflected in the variety of strains

▲ **Figure 26.20 Dating the origin of HIV-1 M with a molecular clock.** The black data points are based on DNA sequences of an HIV gene in blood samples collected from patients. (The dates when these individual HIV gene sequences arose are not known with certainty because a person can harbor the virus for years before symptoms occur.) Projecting the gene's rate of change in the 1980s and 1990s backward in time suggests that the virus originated in the 1930s.

(genetic types) of the virus. HIV's genetic material is made of RNA, and like other RNA viruses, it evolves quickly.

The most widespread strain in humans is HIV-1 M. To pinpoint the earliest HIV-1 M infection, researchers compared samples of the virus from various times during the epidemic, including a sample from 1959. A comparison of gene sequences showed that the virus has evolved in a clocklike fashion **(Figure 26.20)**. Extrapolating backward in time using the molecular clock indicates that the HIV-1 M strain first spread to humans during the 1930s.

CONCEPT CHECK 26.5

1. What is a molecular clock? What assumption underlies the use of a molecular clock?
2. **MAKE CONNECTIONS** Review Concept 17.5 (pp. 344–346). Then explain how numerous base changes could occur in an organism's DNA yet have no effect on its fitness.
3. **WHAT IF?** Suppose a molecular clock dates the divergence of two taxa at 80 million years ago, but new fossil evidence shows that the taxa diverged at least 120 million years ago. Explain how this could happen.

For suggested answers, see Appendix A.

CONCEPT 26.6

New information continues to revise our understanding of the tree of life

The discovery that the scaly-foot in Figure 26.1 evolved from a different lineage of legless lizards than did snakes is one example of how systematics is used to reconstruct the evolutionary relationships of life's diverse forms. In recent decades, we have gained insight into even the very deepest branches of the tree of life through molecular systematics.
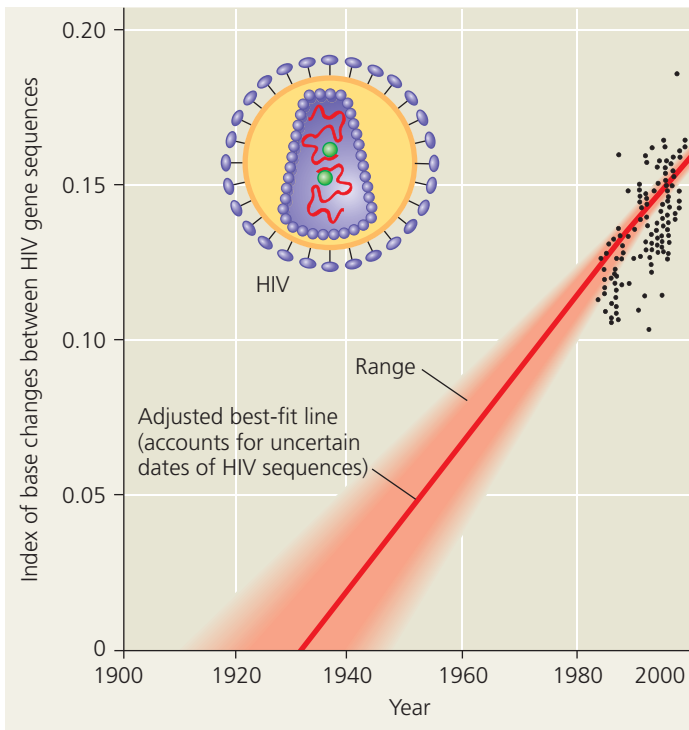
From Two Kingdoms to Three Domains

Early taxonomists classified all known species into two kingdoms: plants and animals. Even with the discovery of the diverse microbial world, the two-kingdom system persisted: Noting that bacteria had a rigid cell wall, taxonomists placed them in the plant kingdom. Eukaryotic unicellular organisms with chloroplasts were also considered plants. Fungi, too, were classified as plants, partly because most fungi, like most plants, are unable to move about (never mind the fact that fungi are not photosynthetic and have little in common structurally with plants!). In the two-kingdom system, unicellular eukaryotes that move and ingest food—protozoans—were classified as animals. Those such as *Euglena* that move and are photosynthetic were claimed by both botanists and zoologists and showed up in both kingdoms.

Taxonomic schemes with more than two kingdoms gained broad acceptance in the late 1960s, when many biologists recognized five kingdoms: Monera (prokaryotes), Protista (a diverse kingdom consisting mostly of unicellular organisms), Plantae, Fungi, and Animalia. This system highlighted the two fundamentally different types of cells, prokaryotic and eukaryotic, and set the prokaryotes apart from all eukaryotes by placing them in their own kingdom, Monera.

However, phylogenies based on genetic data soon began to reveal a problem with this system: Some prokaryotes differ as much from each other as they do from eukaryotes. Such difficulties have led biologists to adopt a three-domain system. The three domains—Bacteria, Archaea, and Eukarya—are a taxonomic level higher than the kingdom level. The validity of these domains is supported by many studies, including a recent study that analyzed nearly 100 completely sequenced genomes.

The domain Bacteria contains most of the currently known prokaryotes, including the bacteria closely related to chloroplasts and mitochondria. The second domain, Archaea, consists of a diverse group of prokaryotic organisms that inhabit a wide variety of environments. Some archaea can use hydrogen as an energy source, and others were the chief source of the natural gas deposits that are found throughout Earth's crust. As you will read in Chapter 27, bacteria differ from

archaea in many structural, biochemical, and physiological characteristics. The third domain, Eukarya, consists of all the organisms that have cells containing true nuclei. This domain includes many groups of single-celled organisms (see Chapter 28) as well as multicellular plants (Chapters 29 and 30), fungi (Chapter 31), and animals (Chapters 32–34). **Figure 26.21** represents one possible phylogenetic tree for the three domains and the many lineages they encompass.

The three-domain system highlights the fact that much of the history of life has been about single-celled organisms. The two prokaryotic domains consist entirely of single-celled organisms, and even in Eukarya, only the branches shown in red (plants, fungi, and animals) are dominated by multicellular organisms. Of the five kingdoms previously recognized by taxonomists, most biologists continue to recognize Plantae, Fungi, and Animalia, but not Monera and Protista. The kingdom Monera is obsolete because it would have members in two different domains. As you'll read in Chapter 28, the kingdom Protista has also crumbled because it is polyphyletic—it includes members that are more closely related to plants, fungi, or animals than to other protists.

A Simple Tree of All Life

The evolutionary relationships shown in Figure 26.21 can be summarized in a simpler tree (see the figure legend question). In this tree, the first major split in the history of life occurred when bacteria diverged from other organisms. If this tree is correct, eukaryotes and archaea are more closely related to each other than either is to bacteria.

This reconstruction of the tree of life is based largely on sequence comparisons of rRNA genes, which code for the RNA components of ribosomes. Because ribosomes are fundamental to the workings of the cell, rRNA genes have evolved so slowly that homologies between distantly related organisms
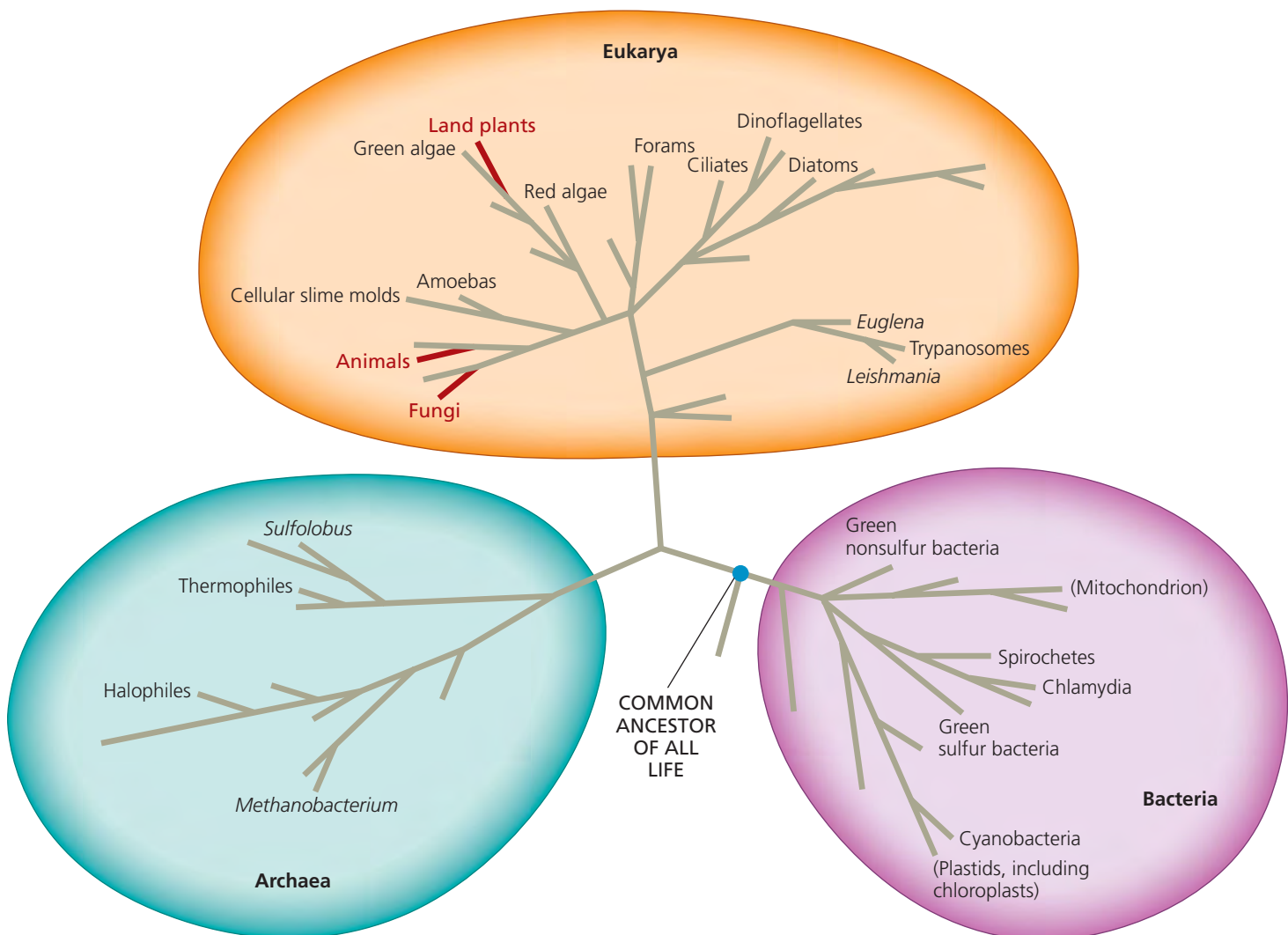
▲ **Figure 26.21 The three domains of life.** The phylogenetic tree shown here is based on rRNA gene sequences. Branch lengths are proportional to the amount of genetic change in each lineage. (To simplify the figure, only some branches are labeled.) In this diagram, the lineages within Eukarya that are dominated by multicellular organisms (plants, fungi, and animals) are shown in red. All other lineages consist solely or primarily of single-celled organisms.
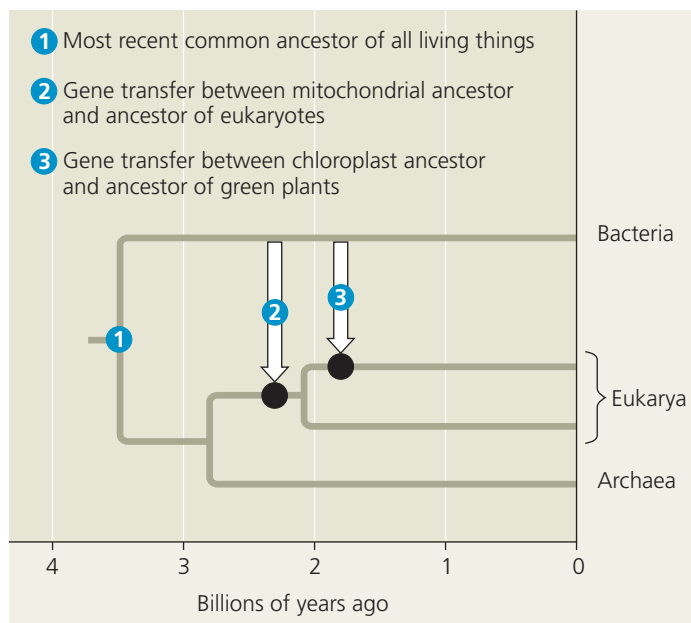
DRAW IT *Redraw this tree as a horizontal tree that has just three branches, one for each domain. Which domain was the first to diverge? Which is the sister domain to Eukarya?*

can still be detected—making these genes very useful for determining evolutionary relationships between deep branches in the history of life. However, other genes reveal a different set of relationships. For example, researchers have found that many of the genes that influence metabolism in yeast (a unicellular eukaryote) are more similar to genes in the domain Bacteria than they are to genes in the domain Archaea—a finding that suggests that the eukaryotes may share a more recent common ancestor with bacteria than with archaea.

Comparisons of complete genomes from the three domains show that there have been substantial movements of genes between organisms in the different domains **(Figure 26.22)**. These took place through **horizontal gene transfer**, a process in which genes are transferred from one genome to another through mechanisms such as exchange of transposable elements and plasmids, viral infection (see Chapter 19), and perhaps fusions of organisms. Recent research reinforces the view that horizontal gene transfer is important. For example, a 2008 analysis indicated that, on average, 80% of the genes in 181 prokaryotic genomes had moved between species at some point during the course of evolution. Because phylogenetic trees are based on the assumption that genes are passed vertically from one generation to the next, the occurrence of such horizontal transfer events helps to explain why trees built using different genes can give inconsistent results.

Is the Tree of Life Really a Ring?

Some biologists, including W. Ford Doolittle, interviewed on pages 534–535, have argued that horizontal gene transfer was

1 Most recent common ancestor of all living things

2 Gene transfer between mitochondrial ancestor and ancestor of eukaryotes

3 Gene transfer between chloroplast ancestor and ancestor of green plants

▲ **Figure 26.22 The role of horizontal gene transfer in the history of life.** This tree shows two major episodes of horizontal gene transfer, the dates of which are uncertain. It is known that many more such events occurred. (Because the tree is horizontal, the arrows representing "horizontal" transfer are vertical here.)
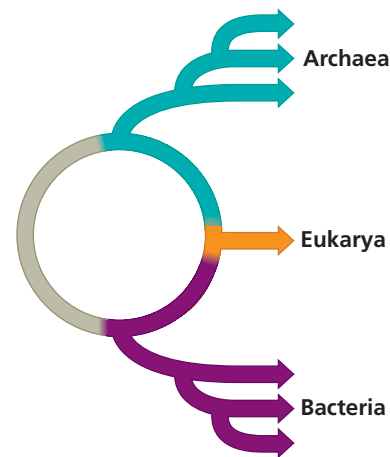
▲ **Figure 26.23 A ring of life.** In this hypothesis, the eukaryote lineage (orange) arose when an early archaean (teal) fused with an early bacterium (purple). Such an event is consistent with a "ring of life" but not with a tree of life. Three great domains (Archaea, Eukarya, and Bacteria) emerged from the ring and gave rise to the tremendous diversity of life we observe today.

so common that the early history of life should be represented as a tangled network of connected branches—not a simple, dichotomously branching tree like that in Figure 26.22. Others have suggested that relationships among early organisms are best represented by a ring, not a tree **(Figure 26.23)**. In an analysis based on hundreds of genes, these researchers hypothesized that eukaryotes arose as a fusion between an early bacterium and an early archaean. If correct, eukaryotes are simultaneously most closely related to bacteria *and* archaea—an evolutionary relationship that cannot be depicted in a tree of life, but can be depicted in a *ring* of life.

Although scientists continue to debate whether early steps in the history of life are best represented as a tree, a ring, or a tangled web, in recent decades there have been many exciting discoveries about evolutionary events that occurred later in time. We'll explore such discoveries in the rest of this unit's chapters, beginning with Earth's earliest inhabitants, the prokaryotes.

CONCEPT CHECK 26.6

1. Why is the kingdom Monera no longer considered a valid taxon?
2. Explain why phylogenies based on different genes can yield different branching patterns for the tree of all life.
3. **WHAT IF?** Draw the three possible dichotomously branching trees showing evolutionary relationships for the domains Bacteria, Archaea, and Eukarya. Two of these trees have been supported by genetic data. Is it likely that the third tree might also receive such support? Explain your answer.

For suggested answers, see Appendix A.

SUMMARY OF KEY CONCEPTS

CONCEPT 26.1

Phylogenies show evolutionary relationships (pp. 537–540)

- Linnaeus's **binomial** classification system gives organisms two-part names: a **genus** plus a specific epithet.
- In the Linnaean system, species are grouped in increasingly broad taxa: Related genera are placed in the same family, families in orders, orders in classes, classes in phyla, phyla in kingdoms, and (more recently) kingdoms in domains.
- Systematists depict evolutionary relationships as branching **phylogenetic trees**. Many systematists propose that classification be based entirely on evolutionary relationships.
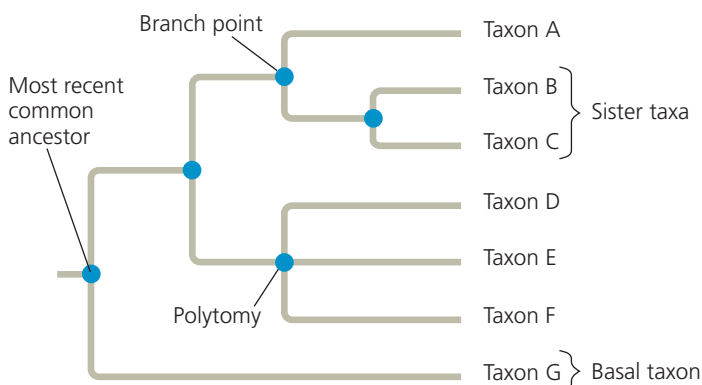
- Unless branch lengths are proportional to time or amount of genetic change, a phylogenetic tree indicates only patterns of descent.
- Much information can be learned about a species from its evolutionary history; hence, phylogenies are useful in a wide range of applications.

? *Humans and chimpanzees are sister species. Explain what that means.*

CONCEPT 26.2

Phylogenies are inferred from morphological and molecular data (pp. 540–542)

- Organisms with similar morphologies or DNA sequences are likely to be more closely related than organisms with very different structures and genetic sequences.
- To infer phylogeny, **homology** (similarity due to shared ancestry) must be distinguished from **analogy** (similarity due to convergent evolution).
- Computer programs are used to align comparable DNA sequences and to distinguish molecular homologies from coincidental matches between taxa that diverged long ago.
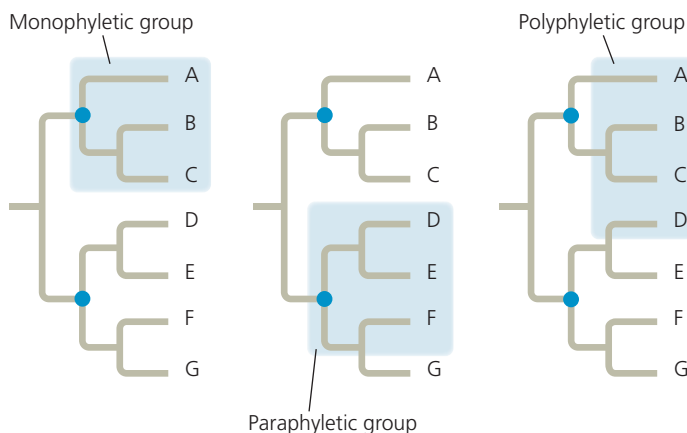
? *Why is it necessary to distinguish homology from analogy to infer phylogeny?*

CONCEPT 26.3

Shared characters are used to construct phylogenetic trees (pp. 542–548)

- A **clade** is a monophyletic grouping that includes an ancestral species and all of its descendants.

- Clades can be distinguished by their **shared derived characters**.

- Branch lengths can be drawn proportional to the amount of evolutionary change or time.
- Among phylogenies, the most parsimonious tree is the one that requires the fewest evolutionary changes. The most likely tree is the one based on the most likely pattern of changes.
- Well-supported phylogenetic hypotheses are consistent with a wide range of data.

? *Explain the logic of using shared derived characters to infer phylogeny.*

CONCEPT 26.4

An organism's evolutionary history is documented in its genome (pp. 548–549)

- **Orthologous genes** are homologous genes found in different species as a result of speciation. **Paralogous genes** are homologous genes within a species that result from gene duplication; such genes can diverge and potentially take on new functions.
- Distantly related species can have orthologous genes. The small variation in gene number in organisms of varying complexity suggests that genes are versatile and may have multiple functions.

? *When reconstructing phylogenies, is it better to compare orthologous or paralogous genes? Explain.*

CONCEPT 26.5

Molecular clocks help track evolutionary time (pp. 549–551)

- Some regions of DNA change at a rate consistent enough to serve as a **molecular clock**, in which the amount of genetic change is used to estimate the date of past evolutionary events. Other DNA regions change in a less predictable way.
- A molecular clock analysis suggests that the most common strain of HIV jumped from primates to humans in the 1930s.

? *Describe some assumptions and limitations of molecular clocks.*

CONCEPT 26.6

New information continues to revise our understanding of the tree of life (pp. 551–553)

- Past classification systems have given way to the current view of the tree of life, which consists of three great **domains**: Bacteria, Archaea, and Eukarya.

- Phylogenies based on rRNA genes suggest that eukaryotes are most closely related to archaea, while data from some other genes suggest a closer relationship to bacteria.
- Other genetic analyses suggest that eukaryotes arose as a fusion between a bacterium and an archaean, leading to a "ring of life" in which eukaryotes are equally closely related to bacteria and archaea.

**?** *Why was the five-kingdom system abandoned for a three-domain system?*

TESTING YOUR UNDERSTANDING

LEVEL 1: KNOWLEDGE/COMPREHENSION

1. In Figure 26.4, which similarly inclusive taxon descended from the same common ancestor as Canidae?
   - a. Felidae
   - b. Mustelidae
   - c. Carnivora
   - d. *Canis*
   - e. *Lutra*

2. Three living species X, Y, and Z share a common ancestor T, as do extinct species U and V. A grouping that consists of species T, X, Y, and Z (but not U or V) makes up
   - a. a valid taxon.
   - b. a monophyletic clade.
   - c. an ingroup, with species U as the outgroup.
   - d. a paraphyletic group.
   - e. a polyphyletic group.

3. In a comparison of birds and mammals, having four limbs is
   - a. a shared ancestral character.
   - b. a shared derived character.
   - c. a character useful for distinguishing birds from mammals.
   - d. an example of analogy rather than homology.
   - e. a character useful for sorting bird species.

4. To apply parsimony to constructing a phylogenetic tree,
   - a. choose the tree that assumes all evolutionary changes are equally probable.
   - b. choose the tree in which the branch points are based on as many shared derived characters as possible.
   - c. base phylogenetic trees only on the fossil record, as this provides the simplest explanation for evolution.
   - d. choose the tree that represents the fewest evolutionary changes, in either DNA sequences or morphology.
   - e. choose the tree with the fewest branch points.

LEVEL 2: APPLICATION/ANALYSIS

5. Based on this tree, which statement is *not* correct?
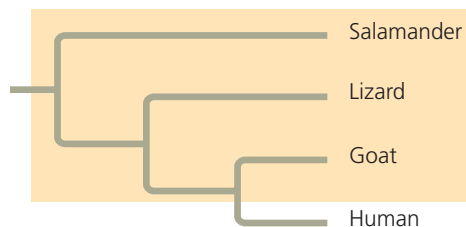
   - a. The salamander lineage is a basal taxon.
   - b. Salamanders are a sister group to the group containing lizards, goats, and humans.
   - c. Salamanders are as closely related to goats as to humans.
   - d. Lizards are more closely related to salamanders than to humans.
   - e. The group highlighted by shading is paraphyletic.

6. If you were using cladistics to build a phylogenetic tree of cats, which of the following would be the best outgroup?
   - a. lion   b. domestic cat   c. wolf   d. leopard   e. tiger

7. The relative lengths of the frog and mouse branches in the phylogeny in Figure 26.12 indicate that
   - a. frogs evolved before mice.
   - b. mice evolved before frogs.
   - c. the genes of frogs and mice have only coincidental homoplasies.
   - d. the homolog has evolved more slowly in mice.
   - e. the homolog has evolved more rapidly in mice.

LEVEL 3: SYNTHESIS/EVALUATION

8. **EVOLUTION CONNECTION**
   Darwin suggested looking at a species' close relatives to learn what its ancestors may have been like. How does his suggestion anticipate recent methods, such as phylogenetic bracketing and the use of outgroups in cladistic analysis?

9. **SCIENTIFIC INQUIRY**
   DRAW IT (a) Draw a phylogenetic tree based on the first five characters in the table below. Place hatch marks on the tree to indicate the origin(s) of each of the six characters. (b) Assume that tuna and dolphins are sister species and re-draw the phylogenetic tree accordingly. Place hatch marks on the tree to indicate the origin(s) of each of the six characters. (c) How many evolutionary changes are required in each tree? Which tree is most parsimonious?

Character	Lancelet (outgroup)	Lamprey	Tuna	Salamander	Turtle	Leopard	Dolphin
Backbone	0	1	1	1	1	1	1
Hinged jaw	0	0	1	1	1	1	1
Four limbs	0	0	0	1	1	1	1*
Amnion	0	0	0	0	1	1	1
Milk	0	0	0	0	0	1	1
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.

10. **WRITE ABOUT A THEME**
   The Cellular Basis of Life; The Genetic Basis of Life In a short essay (100–150 words), explain how these two themes—along with the process of descent with modification (see Chapter 22)—enable scientists to construct phylogenies that extend hundreds of millions of years back in time.

For selected answers, see Appendix A.

Mastering BIOLOGY® www.masteringbiology.com

1. MasteringBiology® Assignments
Video Tutor Sessions Phylogenetic Trees • Survey of Biodiversity
Tutorial Constructing Phylogenetic Trees
Activity Classification Schemes
Questions Student Misconceptions • Reading Quiz • Multiple Choice • End-of-Chapter

2. eText
Read your book online, search, take notes, highlight text, and more.

3. The Study Area
Practice Tests • Cumulative Test • *BioFlix®* 3-D Animations • MP3 Tutor Sessions • Videos • Activities • Investigations • Lab Media • Audio Glossary • Word Study Tools • Art