Green fluorescent protein (GFP) was discovered in 1962 when Osamu Shimomura, an organic chemist and marine biologist, led a team that was able to purify the protein from the tissues of the bioluminescent jellyfish Aequorea victoria. Some 30 years after its initial discovery, Martin Chalfie had the idea (and the technology) to link the gene for GFP to other protein-coding genes, so that the expression of specific genes of interest could be visualized in glowing green within cells and tissues of living organisms (see Figure 13.6). This work was extended by Roger Tsien, who changed some of the amino acids within GFP to create proteins of several distinct colors. Different colored proteins meant that the expression of a number of different proteins could be visualized and studied in the same organism at the same time. These three scientists were awarded the 2008 Nobel Prize in Chemistry for the isolation and development of GFP for visualizing gene expression.

Although Tsien was able to produce different colored proteins, he could not produce a red protein. This was frustrating; a red fluorescent protein would be particularly useful to biologists because red light penetrates tissues more easily than do other colors. Tsien’s work stimulated Mikhail Matz to look for new fluorescent proteins in corals (which are relatives of the jellyfishes). Among the different species he studied, Matz found coral proteins that fluoresced in various shades of green, cyan (blue-green)—and red.

How had fluorescent red pigments evolved among the corals, given that the necessary molecular changes had eluded Tsien? To answer this question, Matz sequenced the genes of the fluorescent proteins and used these sequences to reconstruct the evolutionary history of the amino acid changes that produced different colors in different species of corals.

Matz’s work showed that the ancestral fluorescent protein in corals was green, and that red fluorescent proteins evolved in a series of gradual steps. His analysis allowed him to retrace these steps in detail. Such an evolutionary history, as depicted in a tree of relationships among lineages, is called a phylogeny.

The evolution of many aspects of an organism’s biology can be reconstructed using phylogenetic methods. This information is used in all fields of biology to understand the structure, function, and behavior of organisms.

How are phylogenetic methods used to resurrect protein sequences from extinct organisms?
The sequencing of complete genomes from many diverse species has confirmed what biologists have long suspected: all of life is related through a common ancestor. The common ancestry of life explains why the general principles of biology apply to all organisms. Thus we can learn much about how the human genome works by studying the biology of model organisms because we share a common evolutionary history with those organisms. The evolutionary history of these relationships is known as **phylogeny**, and a **phylogenetic tree** is a diagrammatic reconstruction of that history.

Phylogenetic trees are commonly used to depict the evolutionary history of species, populations, and genes. For many years such trees have been constructed based on physical structures, behaviors, and biochemical attributes. Now, as genomes are sequenced for more and more organisms, biologists are able to reconstruct the history of life in ever greater detail.

In Chapter 15, we discussed why we expect populations of organisms to evolve over time. Such a series of ancestor and descendant populations forms a **lineage**, which we can depict as a line drawn on a time axis:

![Lineage diagram](image)

What happens when a single lineage divides into two? For example, a geographic barrier may divide an ancestral population into two descendant populations that no longer interact with one another. We depict such an event as a split, or **node**, in a phylogenetic tree. Each of the descendant populations give rise to a new lineage, and as these independent lineages evolve, new traits arise in each:

![Node diagram](image)

As the lineages continue to split over time, this history can be represented in the form of a branching tree that can be used to trace the evolutionary relationships from the ancient common ancestor of a group of species, through the various lineage splits, up to the present populations of the organisms:

![Phylogenetic tree diagram](image)

A phylogenetic tree may portray the evolutionary history of all life forms; of a major evolutionary group (such as the insects); of a small group of closely related species; or in some cases, even the history of individuals, populations, or genes within a species. The common ancestor of all the organisms in the tree forms the **root** of the tree. The depictions of phylogenetic trees in this book are rooted at the left, with time flowing from left (earliest) to right (most recent):

![Rooted tree diagram](image)

The timing of splitting events in lineages is shown by the position of nodes on a time axis. These splits represent events where one lineage diverged into two, such as a **speciation event** (for a tree of species), a **gene duplication event** (for a tree of genes), or a **transmission event** (for a tree of viral lineages transmitted through a host population). The time axis may have an explicit scale, or it may simply show the relative timing of divergence events.

In this book’s illustrations, the order in which nodes are placed along the horizontal (time) axis has meaning, but the vertical
distance between the branches does not. Vertical distances have been adjusted for legibility and clarity of presentation; they do not correlate with the degree of similarity or difference between groups. Note too that lineages can be rotated around nodes in the tree, so the vertical order of lineages is also largely arbitrary:

Any group of species that we designate with a name is a taxon (plural taxa). Examples of familiar taxa include humans, primates, mammals, and vertebrates; in this series, each taxon is also a member of the next, more inclusive taxon. Any taxon that consists of all the evolutionary descendants of a common ancestor is called a clade. Clades can be identified by picking any point on a phylogenetic tree and from that point tracing all the descendant lineages to the tips of the terminal branches (FIGURE 16.1). Two species that are each other’s closest relatives are called sister species; similarly, any two clades that are each other’s closest relatives are sister clades.

Before the 1980s, phylogenetic trees tended to be seen only in the literature on evolutionary biology, especially in the area of systematics—the study and classification of biodiversity. But almost every journal in the life sciences published during the last few years contains phylogenetic trees. Trees are widely used in molecular biology, biomedicine, physiology, behavior, ecology, and virtually all other fields of biology. Why have phylogenetic studies become so widespread?

Phylogenetic trees are the basis of comparative biology

In biology, we study life at all levels of organization—from genes, cells, organisms, populations, and species to the major divisions of life. In most cases, however, no individual gene or organism (or other unit of study) is exactly like any other gene or organism that we investigate.

Consider the individuals in your biology class. We recognize each person as an individual human, but we know that no two are exactly alike. If we knew everyone’s family tree in detail, the genetic similarity of any pair of students would be more predictable. We would find that more closely related students have many more traits in common (from the color of their hair to their susceptibility or resistance to diseases). Likewise, biologists use phylogenies to make comparisons and predictions about shared traits across genes, populations, and species.

The evolutionary relationships among species, as represented in the tree of life, form the basis for biological classification. Biologists estimate that there are tens of millions of species on Earth. So far, however, only about 1.8 million species have been classified—that is, formally described and named. New species are being discovered all the time and phylogenetic analyses are constantly reviewed and revised, so our knowledge of the tree of life is far from complete. Yet knowledge of evolutionary relationships is essential for making comparisons in biology, so biologists build phylogenies for groups of interest as the need arises. The tree of life’s evolutionary framework allows us to make many predictions about the behavior, ecology, physiology, genetics, and morphology of species that have not yet been studied in detail.

When biologists compare species, they observe traits that differ within the group of interest and try to ascertain when these traits evolved. In many cases, investigators are interested in how the evolution of a trait relates to environmental conditions or selective pressures. For instance, scientists have used phylogenetic analyses to discover changes in the genome of human immunodeficiency virus that confer resistance to particular drug treatments. The association of a particular genetic change in HIV with a particular treatment provides a hypothesis about the evolution of resistance that can be tested experimentally.

Any features shared by two or more species that have been inherited from a common ancestor are said to be homologous. Homologous features may be any heritable traits, including DNA sequences, protein structures, anatomical structures, and even some behavior patterns. For example, all living vertebrates have a vertebral column, as did the ancestral vertebrate. Therefore, the vertebral column is judged to be homologous in all vertebrates.

FIGURE 16.1 Clades Represent All the Descendants of a Common Ancestor All clades are subsets of larger clades, with all of life as the most inclusive taxon. In this example, the groups called mammals, amniotes, tetrapods, and vertebrates represent successively larger clades. Only a few species within each clade are represented on this tree.
Derived traits provide evidence of evolutionary relationships

In tracing the evolution of a character, biologists distinguish between ancestral and derived traits. Each character of an organism evolves from one condition (the ancestral trait) to another condition (the derived trait). Derived traits that are shared among a group of organisms and are also viewed as evidence of the common ancestry of the group are called synapomorphies (syn, “shared”; apo, “derived”; morph, “form,” referring to the “form” of a trait). Thus the vertebral column is considered a synapomorphy—a shared, derived trait of the vertebrates. (The ancestral trait was an undivided supporting rod.)

Not all similar traits are evidence of relatedness. Similar traits in unrelated groups of organisms can develop for either of the following reasons:

• Independently evolved traits subjected to similar selection pressures may become superficially similar, a phenomenon called convergent evolution. For example, although the wing bones of bats and birds are homologous, having been inherited from a common tetrapod ancestor, the wings of bats and birds are not homologous because they evolved independently from the forelimbs of different nonflying ancestors (FIGURE 16.2).

• A character may revert from a derived state back to an ancestral state in an event called an evolutionary reversal. For example, the derived limbs of terrestrial tetrapods evolved from the ancestral fins of their aquatic ancestors. Then, within the mammals, the ancestors of modern cetaceans (whales and dolphins) returned to the ocean, and cetacean limbs evolved to once again resemble their ancestral state—fins. The superficial similarity of cetacean and fish fins does not suggest a close relationship between these groups; the similarity arises from evolutionary reversal.

Similar traits generated by convergent evolution and evolutionary reversals are called homoplastic traits or homoplasies.

A particular trait may be ancestral or derived, depending on our point of reference. For example, all birds have feathers. We infer from this that feathers (which are highly modified scales) were present in the common ancestor of modern birds. Therefore, we consider the presence of feathers to be an ancestral trait for any particular group of modern birds, such as the songbirds. However, feathers are not present in any other living animals. In reconstructing a phylogeny of all living vertebrates, the presence of feathers is a derived trait found only among birds, and thus is a synapomorphy of the birds.

Do You Understand Concept 16.1?

• What biological processes are represented in a phylogenetic tree?

• Why is it important to consider only homologous characters in reconstructing phylogenetic trees?

• What are some reasons that similar traits might arise independently in species that are only distantly related? Can you think of examples among familiar organisms?

Phylogenetic analyses of evolutionary history have become increasingly important to many types of biological research in recent years, and they are the basis for the comparative nature of biology. For the most part, however, evolutionary history cannot be observed directly. How, then, do biologists reconstruct the past?

Phylogeny Can Be Reconstructed from Traits of Organisms

Consider the eight vertebrate animals listed in Table 16.1: lamprey, perch, salamander, lizard, crocodile, pigeon, mouse, and chimpanzee. To illustrate how the phylogenetic tree in FIGURE 16.3 is constructed, we assume initially that a given derived trait arose only once during the evolution of these animals (there has been no convergent evolution), and that no derived traits were lost from any of the descendant groups (there has been no evolutionary reversal). For simplicity, we have selected traits that are either present (+) or absent (–).

In a phylogenetic study, the group of organisms of primary interest is called the ingroup. As a point of reference, an ingroup is compared with an outgroup: a species or group known to be closely related to but phylogenetically outside the group of interest. If the outgroup is known to have diverged before the ingroup, the outgroup can be used to determine which traits of the
Phylogeny Can Be Reconstructed from Traits of Organisms

5.2

Ingroup are derived (i.e., evolved within the ingroup) and which are ancestral (i.e., evolved before the origin of the ingroup). The lamprey belongs to a group of jawless fishes thought to have separated from the lineage leading to the other vertebrates before the jaw arose. Therefore, we have specified the lamprey as the outgroup for our analysis. Because derived traits were acquired by other members of the vertebrate lineage after they diverged from the outgroup, any trait that is present in both the lamprey and the other vertebrates is judged to be ancestral.

We begin by noting that the chimpanzee and mouse share two derived traits—mammary glands and fur—that are absent in both the outgroup and in the other species of the ingroup. Therefore, we infer that mammary glands and fur are derived traits that evolved in a common ancestor of chimpanzees and mice after that lineage separated from the lineages leading to the other vertebrates. These characters are synapomorphies that unite chimpanzees and mice (as well as all other mammals, although we have not included other mammalian species in this example). By the same reasoning, we can infer that the other shared derived traits are synapomorphies for the various groups in which they are expressed. For instance, keratinous scales are a synapomorphy of the lizard, crocodile, and pigeon.

### TABLE 16.1 Eight Vertebrates and the Presence or Absence of Some Shared Derived Traits

<table>
<thead>
<tr>
<th>TAXON</th>
<th>JAWS</th>
<th>LUNGS</th>
<th>CLAWS OR NAILS</th>
<th>GIZZARD</th>
<th>FEATHERS</th>
<th>FUR</th>
<th>MAMMARY GLANDS</th>
<th>KERATINOUS SCALES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lamprey (outgroup)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Perch</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Salamander</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Lizard</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Crocodile</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Pigeon</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Mouse</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

FIGURE 16.3 Inferring a Phylogenetic Tree  This phylogenetic tree was constructed from the information given in Table 16.1 using the parsimony principle. Each clade in the tree is supported by at least one shared derived trait, or synapomorphy.
Table 16.1 also tells us that, among the animals in our in-group, the pigeon has a unique trait: the presence of feathers. Feathers are a synapomorphy of birds and their extinct relatives. However, because we only have one bird in this example, the presence of feathers provides no clues concerning relationships among these eight species of vertebrates. However, gizzards are found in both birds and crocodiles, so this trait is evidence of a close relationship between birds and crocodilians.

By combining information about the various synapomorphies, we can construct the phylogenetic tree in Figure 16.3. We infer from our information that mice and chimpanzees—the only two animals that share fur and mammary glands—share a more recent common ancestor with each other than they do with pigeons and crocodiles. Otherwise, we would need to assume that the ancestors of pigeons and crocodiles also had fur and mammary glands but subsequently lost them; these additional assumptions are unnecessary in this case.

This particular tree was easy to construct because it is based on a very small sample of traits, and the derived traits we examined evolved only once and were never lost after they appeared. Had we included a snake in the group, our analysis would not have been as straightforward. We would need to examine additional characters to determine that snakes evolved from a group of lizards that had limbs. In an evolutionary reversal, limbs were lost in the ancestors of snakes as an adaptation to a subterranean existence.

Typically, biologists construct phylogenetic trees using hundreds or thousands of traits. With larger data sets, we would expect to observe traits that have changed more than once, and thus would expect to see convergence and evolutionary reversal. How do we determine which traits are synapomorphies and which are homoplasies? One way is to invoke the principle of parsimony.

Parsimony provides the simplest explanation for phylogenetic data

In its most general form, the parsimony principle states that the preferred explanation of observed data is the simplest explanation. Applying the principle of parsimony to the reconstruction of phylogenies entails minimizing the number of evolutionary changes that need to be assumed over all characters in all groups in the tree. In other words, the best hypothesis under the parsimony principle is one that requires the fewest homoplasies. This application of parsimony is a specific case of a general principle of logic called Occam’s razor: The best explanation is the one that best fits the data while making the fewest assumptions. More complicated explanations are accepted only when the evidence requires them. Phylogenetic trees represent our best estimates about evolutionary relationships, given the evidence available.

### APPLY THE CONCEPT

**Phylogeny can be reconstructed from traits of organisms**

The matrix below supplies data for seven land plants and an outgroup (an aquatic plant known as a stonewort). Each trait is scored as either present (+) or absent (−) in each of the plants. Use this data matrix to reconstruct the phylogeny of land plants and answer the questions.

- 1. Which two of these taxa are most closely related?
- 2. Plants that produce seeds are known as seed plants. What is the sister group to the seed plants among these taxa?
- 3. Which two traits evolved along the same branch of your reconstructed phylogeny?
- 4. Are there any homoplasies in your reconstructed phylogeny?

<table>
<thead>
<tr>
<th>TAXON</th>
<th>PROTECTED EMBRYOS</th>
<th>TRUE ROOTS</th>
<th>PERSISTENTLY GREEN SPOROPHYTE</th>
<th>VASCULAR CELLS</th>
<th>STOMATA</th>
<th>MEGAPHYLLS (TRUE LEAVES)</th>
<th>SEEDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stonewort (outgroup)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Liverwort</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Pine tree</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Bracken fern</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>Club moss</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Sphagnum moss</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Hornwort</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Sunflower</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
As with all analyses in science, phylogenetic trees are continually modified as additional information is obtained.

**Phylogenies are reconstructed from many sources of data**

Naturalists have constructed various forms of phylogenetic trees for more than 150 years. In fact, the only figure in the first edition of Darwin’s *Origin of Species* was a conceptual diagram of a phylogenetic tree. Phylogenetic tree construction has been revolutionized by the advent of computer software that allows us to consider far more data and analyze far more traits than could ever before be processed. Combining these advances in methodology with the massive comparative data sets being generated through genome sequencing and other molecular studies, biologists are learning details about the tree of life at a remarkable pace.

Any trait that is genetically determined—and therefore heritable—can be used in a phylogenetic analysis. Evolutionary relationships can be revealed through studies of morphology, development, the fossil record, behavioral traits, and molecular traits such as DNA and protein sequences.

**MORPHOLOGY** An important source of phylogenetic information is morphology: the presence, size, shape, and other attributes of body parts. Since living organisms have been observed, depicted, and studied for millennia, we have a wealth of recorded morphological data as well as extensive museum and herbarium collections of organisms whose traits can be measured. New technological tools, such as the electron microscope and computed tomography (CT) scans, enable systematists to examine and analyze the structures of organisms at much finer scales than was formerly possible.

Most species are described and known primarily by their morphology, and morphology still provides the most comprehensive data set available for many taxa. The morphological features that are important for phylogenetic analysis are often specific to a particular group. For example, the presence, development, shape, and size of various features of the skeletal system are important in vertebrate phylogeny, whereas the structures of the floral organs (petals, carpels, stamens and sepals; see p. 796) are important for studying the relationships among flowering plants.

Reconstructing the evolutionary relationships of most extinct species depends almost exclusively on morphological comparisons. Fossils show us where and when organisms lived in the past and give us an idea of what they looked like. Fossils provide important evidence that helps distinguish ancestral from derived traits. The fossil record can also reveal when lineages diverged and began their independent evolutionary histories. Furthermore, in groups with few species that have survived to the present, information on extinct species is often critical to an understanding of the large divergences among the surviving species. The fossil record is limited, however; for some groups, few or no fossils have been found, and for others the fossil record is fragmentary.

Although useful, morphological approaches to phylogenetic analysis do have limitations. Some taxa exhibit little morphological diversity, despite great species diversity. For example, the phylogeny of the leopard frogs of North and Central America would be difficult to infer from morphological differences alone, because the many species look very similar, despite important differences in their behavior and physiology. At the other extreme, few morphological traits can be compared across distantly related species (an earthworm and a mammal, for instance). Furthermore, some morphological variation has an environmental rather than a genetic basis and so must be excluded from phylogenetic analyses. For these reasons, an accurate phylogenetic analysis often requires information beyond that supplied by morphology.

**DEVELOPMENT** Observations of similarities in developmental patterns may reveal evolutionary relationships. Some organisms exhibit similarities in early developmental stages only. The larvae of marine creatures called sea squirts, for example, have a flexible gelatinous rod in the back—the notochord—that disappears as the larvae develop into adults. All vertebrate animals also have a notochord at some time during their development (FIGURE 16.4). This shared structure is one of the reasons for inferring that sea squirts are more closely related to vertebrates than would be suspected if only adult sea squirts were examined.

**BEHAVIOR** Some behavioral traits are culturally transmitted (learned from other individuals); others have a genetic basis (see Chapter 41). If a particular behavior is culturally transmitted, it may not accurately reflect evolutionary relationships (but may nonetheless reflect cultural connections). Bird songs, for instance, are often learned and may be inappropriate traits for phylogenetic analysis. Frog calls, however, are genetically determined and appear to be acceptable sources of information for reconstructing phylogenies.

**MOLECULAR DATA** All heritable variation is encoded in DNA, and so the complete genome of an organism contains an enormous set of traits (the individual nucleotide bases of DNA) that can be used in phylogenetic analyses. In recent years, DNA sequences have become among the most widely used sources of data for constructing phylogenetic trees. Comparisons of nucleotide sequences are not limited to the DNA in the cell nucleus. Eukaryotes have genes in their mitochondria as well as in their nuclei; plant cells have genes in their chloroplasts as well.

The chloroplast genome (cpDNA) is used extensively in phylogenetic studies of plants because it has changed slowly over evolutionary time and can thus be used to study relatively ancient phylogenetic relationships. Most animal mitochondrial DNA (mtDNA) has changed more rapidly, so mitochondrial genes are used to study evolutionary relationships among closely related animal species (the mitochondrial genes of plants evolve more slowly). Many nuclear gene sequences are also commonly analyzed, and now that entire genomes have been sequenced from many species, they too are used to construct phylogenetic trees. Information on gene products (such as the amino acid sequences of proteins) is also widely used for phylogenetic analyses.
Mathematical models expand the power of phylogenetic reconstruction

As biologists began to use DNA sequences to infer phylogenies, they developed explicit mathematical models describing how DNA sequences change over time. These models account for multiple changes at a given position in a DNA sequence. They also take into account different rates of change at different positions in a gene, at different positions in a codon, and among different nucleotides. For example, transitions (changes between two purines or between two pyrimidines) are usually more likely than are transversions (changes between a purine and pyrimidine).

Mathematical models can be used to compute how a tree might evolve given the observed data. A maximum likelihood method will identify the tree that most likely produced the observed data, given the assumed model of evolutionary change. Maximum likelihood methods can be used for any kind of characters, but they are most often used with molecular data, for which explicit mathematical models of evolutionary change are easier to develop. The principal advantages to maximum likelihood analyses are that they incorporate more information about evolutionary change than do parsimony methods, and they are easier to treat in a statistical framework. The principal disadvantages are that they are computationally intensive and require explicit models of evolutionary change (which may not be available for some kinds of character change).

The accuracy of phylogenetic methods can be tested

If phylogenetic trees represent reconstructions of past events, and if many of these events occurred before any humans were around to witness them, how can we test the accuracy of phylogenetic methods? Biologists have conducted experiments both in living organisms and with computer simulations that have demonstrated the effectiveness and accuracy of phylogenetic methods.

In one experiment designed to test the accuracy of phylogenetic analysis, a single viral culture of bacteriophage T7 was used as a starting point, and lineages were allowed to evolve from this ancestral virus in the laboratory (FIGURE 16.5). The initial culture was split into two separate lineages, one of which became the ingroup for analysis and the other of which became the outgroup for rooting the tree. The lineages in the ingroup were split in two after every 400 generations, and samples of the virus were saved for analysis at each branching point. The lineages were allowed to evolve until there were eight lineages in the ingroup. Mutagens were added to the viral cultures to increase the mutation rate so that the amount of change and the degree of homoplasy would be typical of the organisms analyzed in average phylogenetic analyses. The investigators then sequenced samples from the end points of the eight lineages, as well as from the ancestors at the branching points. They then gave the sequences from the end points of the lineages to other
INVESTIGATION

16.5 The Accuracy of Phylogenetic Analysis

To test whether analysis of gene sequences can accurately reconstruct evolutionary phylogeny, we must have an unambiguously known phylogeny to compare against the reconstruction. Will the observed phylogeny match the reconstruction?

HYPOTHESIS

A phylogeny reconstructed from analysis of the DNA sequences of living organisms can accurately match the known evolutionary history of the organisms.

METHOD

In the laboratory, one group of investigators produced an unambiguous phylogeny of 9 viral lineages, enhancing the mutation rate to increase variation among the lineages.

1. Select single plaque (source of common ancestor).
2. Split each lineage every 400 generations, sequencing each ancestor at time of split.
3. Present final genes (blue dots) to a second group of investigators who are unaware of the history of the lineages or the gene sequences of the ancestral viruses. These "blind" investigators then determine the sequences of the descendant genes and use these sequences to reconstruct the evolution of these lineages in the form of a phylogenetic tree.

RESULTS

The true phylogeny and ancestral DNA sequences were accurately reconstructed solely from the DNA sequences of the viruses at the tips of the tree.

CONCLUSION

Phylogenetic analysis of DNA sequences can accurately reconstruct evolutionary history.

ANALYZE THE DATA

The full DNA sequences for the T7 strains in this experiment are thousands of nucleotides long. The nucleotides ("characters") at 23 DNA positions are given in the table.

<table>
<thead>
<tr>
<th>Character at position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outgroup</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>E</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>F</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>H</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
</tbody>
</table>

A. Construct a phylogenetic tree from these nucleotide DNA positions using the parsimony method. Use the outgroup to root your tree. Assume that all changes among nucleotides are equally likely.

B. Using your tree, reconstruct the DNA sequences of the ancestral lineages.

For more, go to Working with Data 16.1 at yourBioPortal.com.

Go to yourBioPortal.com for original citations, discussions, and relevant links for all INVESTIGATION figures.
investigators to analyze, without revealing the known history of the lineages or the sequences of the ancestral viruses.

After the phylogenetic analysis was completed, the investigators asked two questions. Did phylogenetic methods reconstruct the known history correctly? And, were the sequences of the ancestral viruses reconstructed accurately? The answer in both cases was “yes.” The branching order of the lineages was reconstructed exactly as it had occurred; more than 98% of the nucleotide positions of the ancestral viruses were reconstructed correctly; and 100% of the amino acid changes in the viral proteins were reconstructed correctly.

The experiment shown in Figure 16.5 demonstrated that phylogenetic analysis was accurate under the conditions tested, but it did not examine all possible conditions. Other experimental studies have taken other factors into account, such as the sensitivity of phylogenetic analysis to convergent environments and highly variable rates of evolutionary change. In addition, computer simulations based on evolutionary models have been used extensively to study the effectiveness of phylogenetic analysis. These studies have also confirmed the accuracy of phylogenetic methods and have been used to refine those methods and extend them to new applications.

The reconstructed phylogeny suggests that self-incompatibility is the ancestral state and that self-compatibility evolved three times within this group of Leptosiphon. The change to self-compatibility eliminated the plants’ dependence on an outside pollinator and has been accompanied by the evolution of reduced petal size. Indeed, the striking morphological similarity of the flowers in the self-compatible groups once led to their being classified as members of a single species. Phylogenetic analysis, however, shows them to be members of three distinct lineages.

Why do biologists expend the time and effort necessary to reconstruct phylogenies? In fact, information about the evolutionary relationships among organisms is a useful source of data for scientists investigating a wide variety of biological questions. Next we will describe how phylogenetic trees are used to answer questions about the past, and to predict and compare traits of organisms in the present.

Once a phylogeny is reconstructed, what do we do with it? What beyond an understanding of evolutionary history does phylogeny offer us?

Reconstructing the past is important for understanding many biological processes

Phylogeny often clarifies the origin and evolution of traits that are of great interest in understanding fundamental biological processes. This information is then widely applied in fields such as agriculture and medicine.

SELF-COMPATIBILITY Like most animals, most flowering plants (angiosperms) reproduce by mating with another individual of the same species. But in many angiosperm species, the same individual produces both male and female gametes (contained within pollen and ovules, respectively). Outcrossing species have mechanisms to prevent fertilization of the ovule by the individual’s own pollen, and so are referred to as self-incompatible. Individuals of some species, however, regularly fertilize their ovules using their own pollen; they are self-fertilizing or selfing species, and their gametes are self-compatible.

The evolution of angiosperm fertilization mechanisms was examined in Leptosiphon (formerly called Linanthus; see page 323), a genus in the phlox family that exhibits a diversity of breeding systems and pollination mechanisms. The outcrossing (self-incompatible) species of Leptosiphon have long petals and are pollinated by long-tongued flies. In contrast, self-pollinating species have short petals and do not require insect pollinators to reproduce successfully. Using nuclear ribosomal DNA sequences, investigators reconstructed a phylogeny for twelve species in the genus (FIGURE 16.6). They then determined whether each species was self-compatible by artificially pollinating flowers with the plant’s own pollen or with pollen from other individuals and observing whether viable seeds formed.

The reconstructed phylogeny suggests that self-incompatibility is the ancestral state and that self-compatibility evolved three times within this group of Leptosiphon. The change to self-compatibility eliminated the plants’ dependence on an outside pollinator and has been accompanied by the evolution of reduced petal size. Indeed, the striking morphological similarity of the flowers in the self-compatible groups once led to their being classified as members of a single species. Phylogenetic analysis, however, shows them to be members of three distinct lineages.

ZOONOTIC DISEASES Many infectious pathogens, particularly viruses, affect individuals of only one species. However, zoonotic diseases, or zoonoses, are caused by infectious organisms transmitted from an infected animal of a different species. Approximately 150 zoonotic diseases are known to affect humans; rabies and
bubonic plague are historically well-known examples. Lyme disease and acquired immunodeficiency syndrome (AIDS) are more recent examples of zoonotic transfers.

In dealing with zoonotic diseases, it is important to understand when, where, and how the disease first entered a human population. Phylogenetic analyses have become important for studying the transmission of zoonotic pathogens, including the human immunodeficiency virus (HIV) that causes AIDS. Phylogenies are also important for understanding the present global diversity of HIV and for determining the virus’s origins in human populations. A broader phylogenetic analysis of immunodeficiency viruses shows that humans acquired these viruses from two different hosts: HIV-1 from chimpanzees, and HIV-2 from sooty mangabeys (FIGURE 16.7).

HIV-1 is the common form of the virus in human populations in central Africa, where chimpanzees are hunted for food. HIV-2 is the common form in human populations in western Africa, where sooty mangabeys are hunted for food. These viruses apparently entered human populations through hunters who cut themselves while skinning their primate prey. The relatively recent global pandemic of AIDS occurred when these infections in local African populations spread through human populations around the world. By understanding how such viruses first entered human populations, health care workers can know what steps need to be taken to prevent other zoonotic diseases from spreading into human populations.

FIGURE 16.7  Phylogenetic Tree of Immunodeficiency Viruses
Immunodeficiency viruses have been transmitted to humans from two different simian hosts: HIV-1 from chimpanzees and HIV-2 from sooty mangabeys. SIV stands for “simian immunodeficiency virus.”
Understanding and developing new treatments for many viral diseases depends in large part on information from phylogenetic analyses. Different strains of dengue virus require different forms of treatment, but the virus changes quickly and new variants are found on a regular basis. Phylogenetic analyses show the relationships of a new dengue variant to known strains, which is an effective way to predict the most effective treatment.

**Phylogenies allow us to understand the evolution of complex traits**

Biologists are constantly confronted with evolutionary adaptations that at first seem puzzling, and many of these have to do with mating behavior and sexual selection. The tails of male wid-owbirds (see Figure 15.8) are one of many examples; another is found among the swordtail fishes.

Male swordtail fishes have a long, colorful tail extension (FIGURE 16.8A), and their reproductive success is closely associated with this appendage. Males with a long sword are more likely to mate successfully than are males with a short sword. Several explanations have been advanced for the evolution of this structure, including the hypothesis that the sword exploits a preexisting bias in the sensory system of the female fish. This sensory exploitation hypothesis suggests that female swordtails had a bias to prefer males with long tails even before this trait evolved (perhaps because females assess the size of males by their total body length—including the tail—and prefer larger males).

To test the sensory exploitation hypothesis, phylogenetic analysis identified the swordtail relatives that had split most recently from their lineage before the evolution of sword extensions. These closest relatives turned out to be the platyfishes (FIGURE 16.8B). Even though male platyfishes do not normally have swords, when researchers attached artificial swordlike structures to the tails of some male platyfishes, female platyfishes preferred the males with an artificial sword, thus providing support for the hypothesis that female swordtails had a preexisting sensory bias favoring tail extensions even before the trait evolved (FIGURE 16.8C). Thus, a long tail became a sexually selected trait because of the preexisting preference of the females.

**Ancestral states can be reconstructed**

In addition to using phylogenetic methods to infer evolutionary relationships, biologists can use these techniques to reconstruct the morphology, behavior, or nucleotide and amino acid sequences of ancestral species (as was demonstrated for the ancestral sequence of bacteriophage T7 in Figure 16.5). For instance, a phylogenetic analysis was used to reconstruct an opsin protein in the ancestral archosaur (the most recent common ancestor of birds, dinosaurs, and crocodiles).

**Opsins** are pigment proteins involved in vision; different opsins (with different amino acid sequences) are excited by different wavelengths of light (see Figure 35.18). Investigators used phylogenetic analysis of opsin from living vertebrates to predict the amino acid sequence of the visual pigment that existed in the ancestral archosaur. A protein with the predicted sequence was then constructed in the laboratory. Investigators tested the reconstructed opsin and found a significant shift toward the red end of the spectrum in the light sensitivity of this protein compared with that of most modern opsins. Modern species that exhibit similar red (and infrared) sensitivity are adapted for nocturnal vision; thus the investigators inferred that the ancestral archosaur was likely to have been active at night. These findings may remind you of the movie Jurassic Park, although here the extinct species are being “brought back to life” one protein at a time.

**Molecular clocks help date evolutionary events**

For many applications, biologists want to know not only the order in which evolutionary lineages split but also the timing of those splits. In 1965, Emile Zuckerkandl and Linus Pauling hypothesized that rates of molecular change were constant enough that they could be used to predict evolutionary divergence times—an idea that has become known as the molecular clock hypothesis.

Different genes evolve at different rates. In addition, there are differences in evolutionary rates among species related to differing generation times, environments, efficiencies of DNA repair systems, and other biological factors. Nonetheless, among closely related species, a given gene usually evolves at a reasonably
Phylogeny makes Biology Comparative and Predictive

Therefore, the protein encoded by the gene accumulates amino acid replacements at a relatively constant rate (FIGURE 16.9). A molecular clock uses the average rate at which a given gene or protein accumulates changes to gauge the time of divergence for a particular split in the phylogeny. Molecular clocks must be calibrated using independent data such as the fossil record (see Concept 18.1), known times of divergence, or biogeographic dates (e.g., the time of separations of continents; see Concept 18.3). Using such calibrations, times of divergence have been estimated for many groups of species that have diverged over millions of years.

Besides being useful for dating ancient events, molecular clocks can also be used to study the timing of comparatively recent events. Although HIV-1 samples have generally been collected from humans only since the early 1980s, a few isolates from medical biopsies are available from as early as the 1950s. But biologists can use the observed changes in HIV-1 over the past three decades to project back to the common ancestor of all HIV-1 isolates, and thus can estimate when HIV-1 first entered human populations from chimpanzees (FIGURE 16.10). The clock can be calibrated using the samples from the 1980s and 1990s, then tested using the samples from the 1950s. As shown in Figure 16.10C, a sample from a 1959 biopsy is dated by molecular clock analysis at 1957 ± 10 years. The molecular clock was also used to project back to the common ancestor of this group of HIV-1 samples. Extrapolation suggests a date of origin for this group of viruses of about 1930. Although AIDS was unknown to Western medicine until the 1980s, this analysis shows that HIV-1 was present (probably at very low frequency) in human populations in Africa for at least 50 years before its emergence as a global pandemic. Biologists have used similar analyses to conclude that immunodeficiency viruses have been transmitted repeatedly into human populations from multiple primates for more than a century.
Do You Understand Concept 16.3?

- How can phylogenetic trees can help determine the number of times a particular trait evolved?
- How does the reconstruction of ancestral traits help biologists explain the biology of extinct species?
- What is the importance of adding a time dimension to phylogenetic trees, and how do biologists accomplish this?

All of life is connected through evolutionary history, and the relationships among organisms provide a natural basis for making biological comparisons. For these reasons, biologists use phylogenetic relationships as the basis for organizing life into a coherent classification system.

Concept 16.4 Phenylogy Is the Basis of Biological Classification

The biological classification system in widespread use today is derived from that developed by the Swedish biologist Carolus Linnaeus in the mid-1700s. Linnaeus developed a system of binomial nomenclature. Linnaeus gave each species two names, one identifying the species itself and the other the group of closely related species, or genus (plural, genera) to which it belongs. Optionally, the name of the taxonomist who first proposed the species name may be added at the end. Thus Homo sapiens Linnaeus is the name of the modern human species. Homo is the genus, sapiens identifies the particular species in the genus Homo, and Linnaeus is the person who proposed the name Homo sapiens.

You can think of Homo as equivalent to your surname and sapiens as equivalent to your first name. The first letter of the genus name is capitalized, and the specific name is lowercase. Both of these formal designations are italicized. Rather than repeating the name of a genus when it is used several times in the same discussion, biologists often spell it out only once and abbreviate it to the initial letter thereafter (e.g., D. melanogaster rather than Drosophila melanogaster).

As we noted earlier, any group of organisms that is treated as a unit in a biological classification system, such as all species in the genus Drosophila, or all insects, or all arthropods, is called a taxon. In the Linnaean system, species and genera are further grouped into a hierarchical system of higher taxonomic categories. The taxon above the genus in the Linnaean system is the family. The names of animal families end in the suffix “-idae.” Thus Formicidae is the family that contains all ant species, and the family Hominidae contains humans and our recent fossil relatives, as well as our closest living relatives, the chimpanzees and gorillas. Family names are based on the name of a member genus; Formicidae is based on the genus Formica, and Hominidae is based on Homo. The same rules are used in classifying plants, except that the suffix “-aceae” is used for plant family names instead of “-idae.” Thus Rosaceae is the family that includes the genus Rosa (roses) and its relatives.

In the Linnaean system, families are grouped into orders, orders into classes, and classes into phyla (singular phylum), and phyla into kingdoms. However, the ranking of taxa within Linnaean classification is subjective. Whether a particular taxon is considered, say, an order or a class is informative only with respect to the relative ranking of other related taxa. Although families are always grouped within orders, orders within classes, and so forth, there is nothing that makes a “family” in one group equivalent (in number of genera or in evolutionary age, for instance) to a “family” in another group.

Today, the Linnaean terms above the genus level are used largely for convenience. Linnaeus recognized the overarching hierarchy of life, but he developed his system before evolutionary thought had become widespread. Biologists today recognize the tree of life as the basis for biological classification and often name taxa without placing them into the various Linnaean ranks. But, regardless of whether they rank organisms into Linnaean categories or use unranked taxon names, modern biologists use...
evolutionary relationships as the basis for distinguishing, naming, and classifying biological groups.

Evolutionary history is the basis for modern biological classification

Today’s biological classifications express the evolutionary relationships of organisms. Taxa are expected to be monophyletic, meaning that the taxon contains an ancestor and all descendants of that ancestor, and no other organisms. In other words, a monophyletic taxon is a historical group of related species, or a complete branch on the tree of life. As noted earlier, this is also the definition of a clade. A true monophyletic group can be removed from a phylogenetic tree by a single “cut” in the tree, as shown in Figure 16.11.

Note that there are many monophyletic groups on any phylogenetic tree, and that these groups are successively smaller subsets of larger monophyletic groups. This hierarchy of biological taxa, with all of life as the most inclusive taxon and many smaller taxa within larger taxa, down to the individual species, is the modern basis for biological classification.

Although biologists seek to describe and name only monophyletic taxa, the detailed phylogenetic information needed to do so is not always available. A group that does not include its common ancestor is polyphyletic; a group that does not include all the descendants of a common ancestor is referred to as paraphyletic (see Figure 16.11). Virtually all taxonomists now agree that polyphyletic and paraphyletic groups are inappropriate as taxonomic units because they do not correctly reflect evolutionary history. Some classifications still contain such groups because some organisms have not been evaluated phylogenetically. As mistakes in prior classifications are detected, taxonomic names are revised and polyphyletic and paraphyletic groups are eliminated from the classifications.

Several codes of biological nomenclature govern the use of scientific names

Several sets of explicit rules govern the use of scientific names. Biologists around the world follow these rules voluntarily to facilitate communication and dialogue. There may be dozens of common names for an organism in many different languages, and the same common name may refer to more than one species (Figure 16.12). The rules of biological nomenclature are designed so that there is only one correct scientific name for any single recognized taxon and (ideally) a given scientific name applies only to a single taxon (that is, each scientific name is unique). Sometimes the same species is named more than once (when more than one taxonomist has taken up the task); the rules specify that the valid name is the first name that was proposed (as in the case of the Leptosiphon/Linanthus genus in Figure 16.6). If the same name is inadvertently given to two different species, then a replacement name must be given to the species that was named second.

Because of the historical separation of the fields of zoology, botany (which originally included mycology, the study of fungi), and microbiology, different sets of taxonomic rules were developed for each of these groups. Yet another set of rules emerged later for classifying viruses. This separation of fields resulted in duplicated taxon names in groups governed by the different sets of rules; Drosophila, for instance, is both a genus of fruit flies and a genus of fungi, and some species in both groups have identical names. Until recently these duplicated names caused little
confusion, since traditionally biologists who studied fruit flies were unlikely to read the literature on fungi (and vice versa). Today, given the prevalence of large, universal biological databases (such as GenBank, which includes DNA sequences from across all life), it is increasingly important that each taxon have a unique and unambiguous name.

Having described some of the mechanisms by which evolution occurs and how phylogenies can be used to study evolutionary relationships, we are now ready to consider the subject of speciation. Speciation is the process that leads to the splitting events (nodes) on the tree of life and eventually results in the millions of distinctive species that constitute Earth’s biodiversity.

Most genes and proteins of organisms that lived millions of years ago have decomposed in the fossil remains of these species. Nonetheless, sequences of many ancient genes and proteins can be reconstructed using the methods described in this chapter. As we discussed in Concept 16.3, just as we can reconstruct the morphological features of a clade’s ancestors, we can also reconstruct their DNA and protein sequences—if we have enough information about the genomes of their descendants.

Biologists have now reconstructed gene sequences from many species that have been extinct for millions of years. Using this information, a laboratory can reconstruct real proteins that correspond to the proteins that were present in long-extinct species. This is how Mikhail Matz and his colleagues were able to resurrect fluorescent proteins from the extinct ancestors of modern corals and then visualize the colors produced by these proteins in the laboratory and recreate the probable evolutionary path of the different pigments (Figure 16.13).

Biologists have even used phylogenetic analysis to reconstruct some protein sequences that were present in the common ancestor of life described in Concept 16.1. These hypothetical

---

**APPLY THE CONCEPT**

Phylogeny is the basis of biological classification

Consider the above phylogeny and three possible classifications of the taxa.

1. Which of these classifications contains a paraphyletic group?
2. Which of these classifications contains a polyphyletic group?
3. Which of these classifications is consistent with the goal of including only monophyletic groups in a biological classification?
4. Starting with the classification you named in Question 3, how many additional group names would you need to include all the clades shown in this phylogenetic tree?

**Do You Understand Concept 16.4?**

- What is the difference between monophyletic, paraphyletic, and polyphyletic groups?
- Why do biologists prefer monophyletic groups in formal classifications?
- What advantages or disadvantages do you see to having separate sets of taxonomic rules for animals, plants, bacteria, and viruses?
5.2 Phylogeny Is the Basis of Biological Classification

Mikhail Matz and his colleagues used phylogenetic analysis to reconstruct the sequences of extinct fluorescent proteins that were present in the ancestors of modern corals. They then expressed these proteins in bacteria and plated the bacteria in the form of a phylogenetic tree to show how the colors evolved over time.

Protein sequences can then be resurrected into actual proteins in the laboratory. When biologists measured the temperature optima for these resurrected proteins, they found the proteins functioned best in the range of 55ºC–65ºC. This analysis is consistent with hypotheses that life evolved in a high-temperature environment.

To reconstruct protein sequences from species that have been extinct for millions or even billions of years, biologists use detailed mathematical models that take into account much of what we have learned of molecular evolution, as described in Concept 16.2. These models incorporate information on rates of replacement among different amino acid residues in proteins, information on different substitution rates among nucleotides, and changes in the rate molecular evolution among the major lineages of life.