

# Understanding Evolutionary History: An Introduction to Tree Thinking

**Laura R. Novick**  
Vanderbilt University

**Kefyn M. Catley**  
Western Carolina University

**Emily G. Schreiber**  
Vanderbilt University

**Version 1 © February, 2009**

**Version 2 © February, 2010**

**Version 3.1 © August, 2010**

**Version 3.2 © August, 2012**

## Basic Concepts

### Introduction

Imagine the evolutionary relationships among everything that has ever lived being shown in a single diagram. If life evolved only once (and there is no evidence to the contrary) and if we accept the tenet of “descent with modification” (Darwin, 1859), then such a diagram is theoretically possible. This diagram—an *evolutionary tree*—would be incredibly large, and it would contain an unimaginable amount of information that would be of unprecedented value to science. Such a tree would show the evolutionary history of every type of living thing. Because it is impossible to know everything that lived in the past, let alone their precise relationships, science has developed a methodology called *cladistics* that can be used to reconstruct the evolutionary history of *taxa* based on observable and testable evidence (Hennig, 1966; Thanukos, 2009). A *taxon* is any taxonomic category ranging from a species (e.g., blue jay) to a higher-order group (e.g., birds, amniotes, vertebrates). The plural of *taxon* is *taxa*.

It is common in everyday life to group things together that are similar in some way. Of course, there are lots of ways to determine whether two things are similar. For example, one person might say that a crow is similar to a bat because both are black and can fly or that a dolphin and a shark are similar because they both live in the ocean. Another person might say that a crow is similar to an ostrich and a hummingbird because all have feathers and bird DNA even though they do not look much alike; and a dolphin is similar to a bat and a lion because all have an amniotic egg and nurse their young. Cladistics proposes to group taxa together based on a special type of similarity called a *synapomorphy*, which is a shared, derived character. A character is considered to be derived if (a) two or more taxa inherited it from their most recent common ancestor and (b) the ancestor of that ancestor did not possess that character (i.e., the character appeared for the first time in the most recent common ancestor of the taxa in question). For example, having a skull is a synapomorphy that defines fish, amphibians, reptiles, and mammals as belonging to a single group, labeled *Craniata*. This character was new to the most recent common ancestor of these taxa; the most recent common ancestor of the *Craniata* group plus the echinoderms (starfish, sea urchins, etc.) did not have a skull. Again, the critical difference between a synapomorphy and the more general notion of similarity is the concept of a character not only being shared, but also being *derived*. This simple technique for defining biologically-meaningful groups has proved to be a powerful organizing and predictive tool in modern biology (e.g., Thanukos, 2009).

## Cladogram Structure

Using the principles of cladistics, biologists depict evolutionary relationships among taxa in a type of hierarchical branching diagram called a *cladogram*. In a cladogram, taxa are grouped into levels based on most recent common ancestry. Figure 1 shows a very simple cladogram involving only three taxa: lizards, bears, and felines. Cladograms can be read from the top down or from the bottom up. In Figure 1, the arrow shows historical time moving from the bottom of the cladogram to the top.

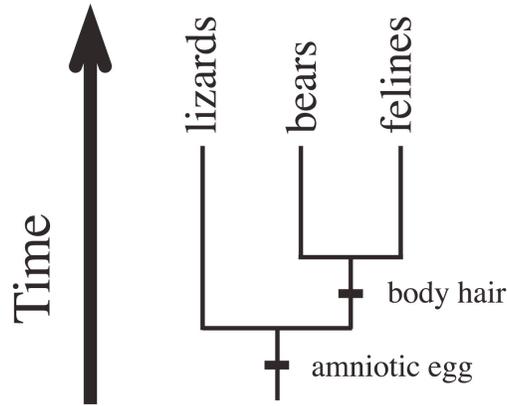


Figure 1

In the past, scientists relied on observed similarity (shared characters) as the basis for grouping taxa. “Similarity was similarity,” and there was no way to determine whether it was the result of independent evolution or a shared *most recent common ancestor (MRCA)*. With independent evolution, multiple taxa share a character because that character evolved separately multiple times. In contrast, when the shared character is a result of a shared MRCA, the character evolved only one time in the MRCA of the taxa in question, thus providing solid evidence to support grouping those taxa. As noted earlier, the cladistic approach is to group taxa only if their shared characters are synapomorphies—that is, are a result of a MRCA.

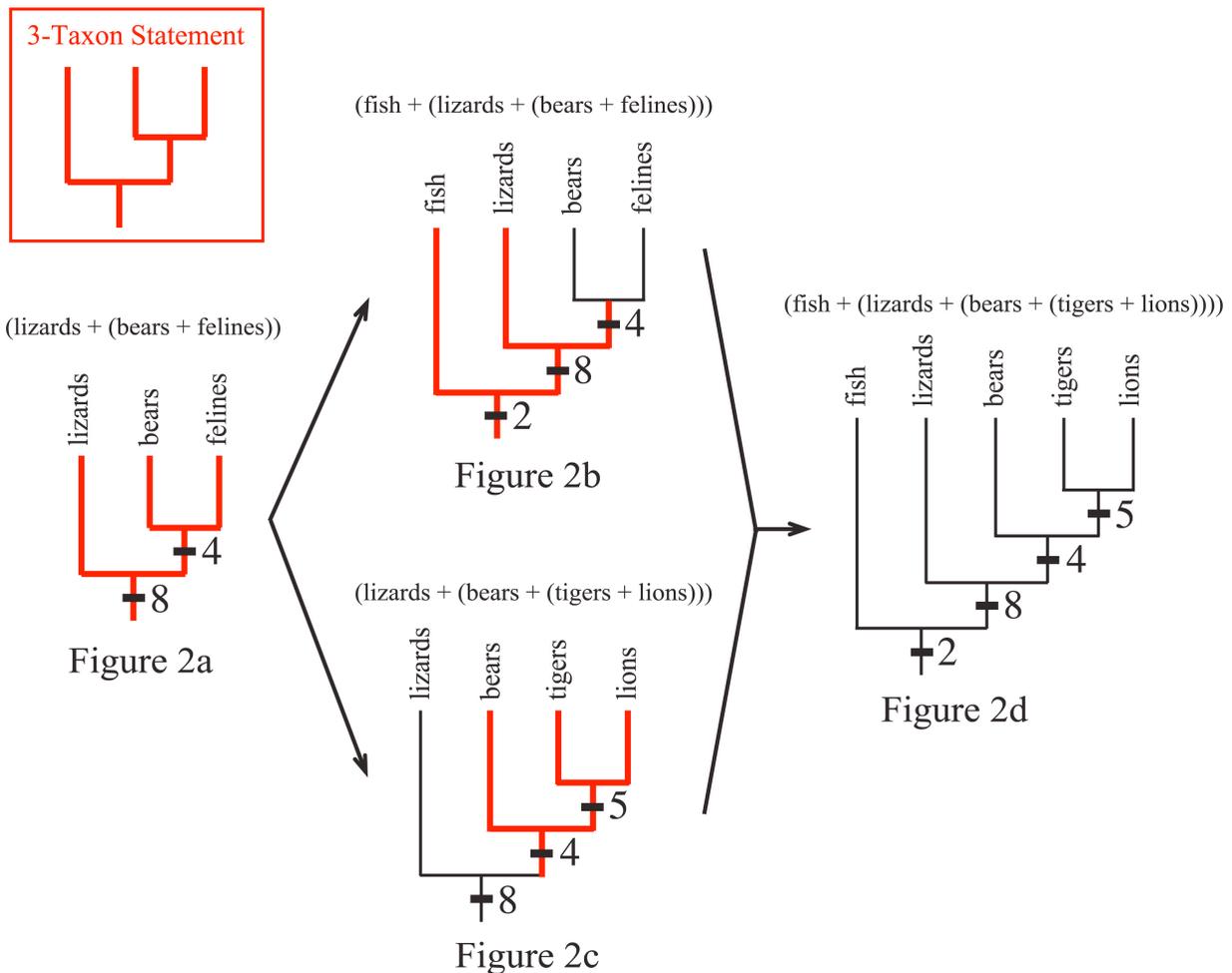
The cladogram in Figure 1 groups bears and felines together because they share a MRCA that possessed the novel (derived) character of having body hair. Thus, body hair is a synapomorphy that defines the group consisting of bears and felines (i.e., mammals). Lizards do not have body hair because that character was newly derived in the MRCA of bears and felines. Farther back in time, though, lizards, bears, and felines do share a MRCA, one that evolved the novel character of possessing an amniotic egg (i.e., an egg with inner membranes). Thus, these three taxa all have an amniotic egg because they share a most recent common ancestor at this point (see Figure 1). Amniotic egg is a synapomorphy that defines the group consisting of lizards, bears, and felines (biologists call this group *Amniota*). These nested levels of most recent common ancestry imply a time arrow running from earlier in historical time at the bottom of the cladogram to more recent historical times moving toward the top of the cladogram, as shown in Figure 1.

It is very important to appreciate the difference between a common ancestor of two or more taxa and the most recent common ancestor (MRCA) of those taxa. Saying that two taxa share a common ancestor provides no useful information about similarities between those taxa because all life ultimately shares a common ancestor. That is, every taxon has a common ancestor with every other taxon if one goes back far enough in evolutionary history. Thus, common ancestors *per se* are not informative when trying to reconstruct the evolutionary history of a group of taxa. MRCAs, on the other hand, are highly informative because they contain exactly the information biologists need to piece together the tree of life. MRCAs, defined by their synapomorphies, provide the evidence needed to create the hierarchical

structure of the tree of life. The pattern of most recent common ancestry among a set of taxa defines the structure of the cladogram depicting the evolutionary relationships among those taxa.

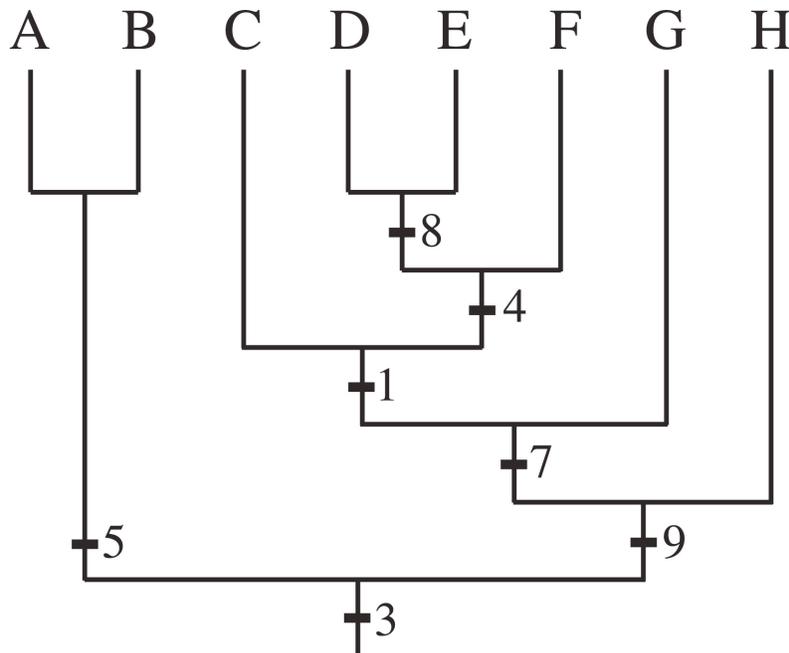
An important principle for understanding cladogram structure is the **3-taxa statement**. This principle states that given a set of three taxa, two taxa have to be more closely related to each other than either is to the third taxon. The cladogram in Figure 1 shows that bears and felines are more closely related to each other than either of these two taxa is to lizards. Therefore, using parenthetical notation, we group bears and felines together first, writing (bears + felines). The entire 3-taxa statement, then, can be written in parenthetical format as (lizards + (bears + felines)).

The cladogram in Figure 1 shows evolutionary relationships among only three taxa. As you might expect from the earlier discussion of the entire tree of life, cladograms can be constructed to depict such relationships among any number of taxa. These larger, more complicated cladograms are compiled from a series of 3-taxa statements and their supporting MRCAs. This is why the 3-taxa statement is critical for understanding cladogram structure. This is illustrated in Figure 2. Each new 3-taxa statement is colored in red in the cladograms. The cladogram in Figure 2b adds the 3-taxa statement (fish + (lizards + bears/felines)) to the original 3-taxa statement (lizards + (bears + felines)) shown in the cladogram in Figure 2a. In the cladogram in Figure 2c, felines are divided into tigers and lions, which yields the new 3-taxa statement (bears + (tigers + lions)). Figure 2d puts these two additions to the original 3-taxa statement together to create a cladogram depicting evolutionary relationships among five taxa. The numbers and corresponding black bars on the cladograms in Figure 2 represent synapomorphies. Notice that each synapomorphy is found in the same position in all four cladograms and supports a group containing the MRCA and all the taxa above it.



### Practice What You've Learned #1

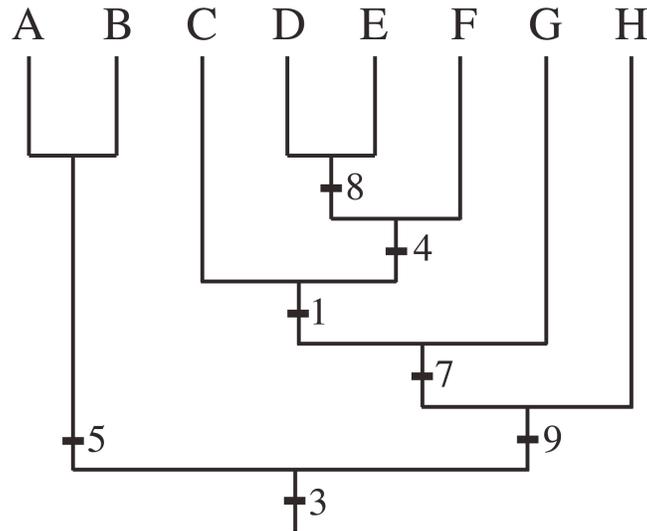
Answer the questions below without referring back to the previous pages. After you've completed the problems, you'll have a chance to check your answers against the correct answers. If anything is still unclear at that point, you may return to the earlier pages to revisit a topic.



1. Taxon A is more closely related to Taxon C than to Taxon B. True    False
2. Taxon D is more closely related to Taxon F than to Taxon C. True    False
3. Taxon G is more closely related to Taxon H than to Taxon E. True    False
4. Taxon C is more closely related to Taxon F than to Taxon H. True    False
5. Is  $(D + (E + F))$  a valid 3-taxon statement? If not, write the correct statement. Yes    No
6. Is  $(C + (D/E + F))$  a valid 3-taxon statement? If not, write the correct statement. Yes    No
7. What numbered character was possessed by the MRCA of Taxon D and Taxon E? \_\_\_\_\_
8. What numbered character was possessed by the MRCA of the taxa C, E, and G? \_\_\_\_\_
9. What numbered character was possessed by the MRCA of Taxon A and Taxon G? \_\_\_\_\_

**STOP!**  
**DO NOT TURN THE PAGE UNTIL YOU HAVE FINISHED**  
**“PRACTICE WHAT YOU’VE LEARNED #1”**

**Answers to Practice what you've learned:** Check your answers on the previous page with the correct answers below. Note that an explanation for each answer is printed in blue. If any of these explanations are unclear, please refer back to the earlier instructions.



1. Taxon A is more closely related to Taxon C than to Taxon B. True False  
*A shares a more recent common ancestor with B than with C.*
2. Taxon D is more closely related to Taxon F than to Taxon C. True False  
*D shares a more recent common ancestor with F than with C.*
3. Taxon G is more closely related to Taxon H than to Taxon E. True False  
*G shares a more recent common ancestor with E than with H.*
4. Taxon C is more closely related to Taxon F than to Taxon H. True False  
*C shares a more recent common ancestor with F than with H.*
5. Is (D + (E + F)) a valid 3-taxon statement? If not, write the correct statement. Yes No  
*This statement says that E and F are more closely related to each other than either is to D, but the cladogram shows that in fact D and E share a MRCA. So the correct 3-taxon statement is ((D + E) + F).*
6. Is (C + (D/E + F)) a valid 3-taxon statement? If not, write the correct statement. Yes No  
*The cladogram shows that (D + E) and F share a more recent common ancestor with each other than either shares with C.*
7. What numbered character was possessed by the MRCA of Taxon D and Taxon E? 8  
*D and E both evolved from an ancestor that had character 8, and this ancestor is the MRCA of these two taxa (i.e., it is at the highest branching point common to these two taxa).*
8. What numbered character was possessed by the MRCA of the taxa C, E, and G? 7  
*C, E, and G all evolved from an ancestor that had character 7, and this ancestor is the MRCA of these three taxa (i.e., it is at the highest branching point common to these three taxa).*
9. What numbered character was possessed by the MRCA of Taxon A and Taxon G? 3  
*A and G both evolved from an ancestor that had character 3, and this ancestor is the MRCA of these two taxa (i.e., it is at the highest branching point common to these two taxa).*

## Clades and Nested Hierarchical Structure

A **clade** (or *monophyletic group*) is a group of taxa that includes the MRCA of the group and all descendants of that ancestor. Groups that do not include the MRCA or that include only some of that ancestor's descendants are not clades. For example, the colloquial category *birds* is not a clade because it does not include the MRCA of all birds, which is understood to be a theropod dinosaur. For this reason, *birds* are not considered to be a *natural group*—i.e., a group that reflects the actual evolutionary history that created it.

Figure 3 illustrates that a cladogram is composed of nested sets of clades, each defined by its MRCA with its associated synapomorphy. First, tigers and lions at the far right comprise a clade, which is shown by the blue bracket surrounding these two taxa at the top of the cladogram. Bears, tigers, and lions also comprise a clade. Snakes and lizards comprise a clade too. Finally, snakes, lizards, bears, tigers, and lions constitute a clade. Note that the clades form nested levels in the cladogram, reflecting the nested levels of most recent common ancestry, as indicated by the nested brackets at the top of the figure. Cladistics tells us that clades are the meaningful (i.e., valid, supported by evolutionary evidence) groups of living things.

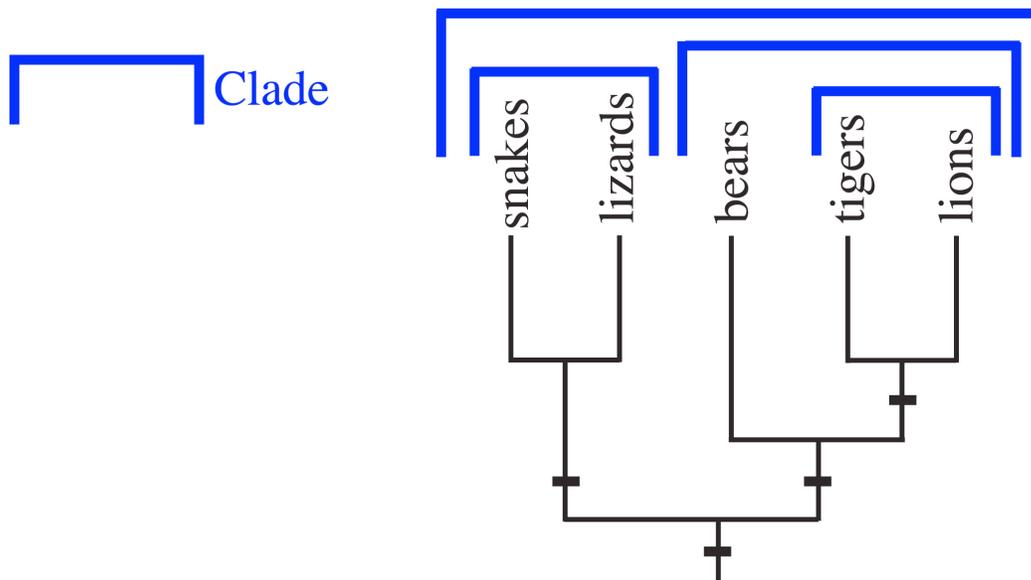


Figure 3

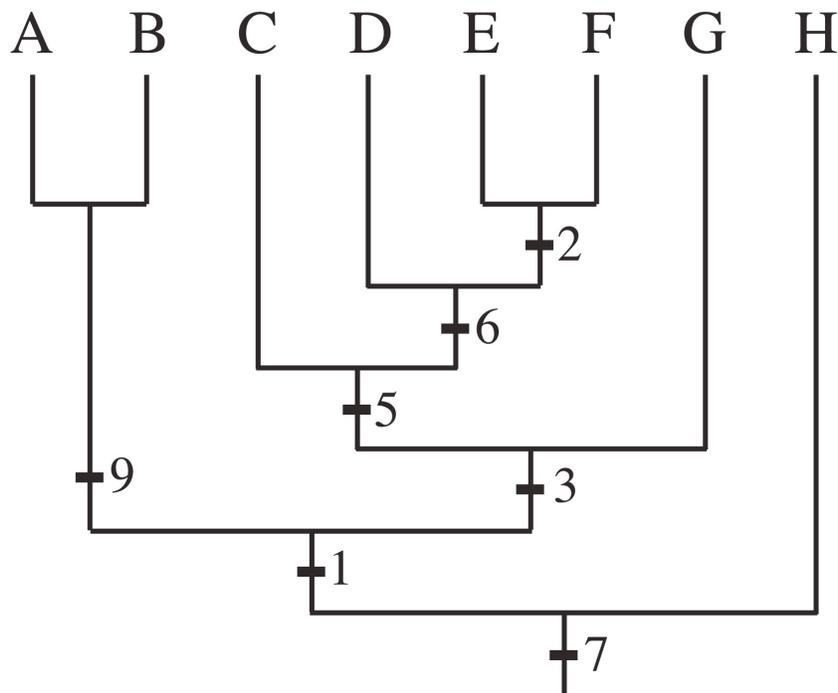
Note that bears and tigers in the cladogram in Figure 3 do not constitute a clade even though those taxa appear right next to each other in horizontal space. Although these two taxa share a MRCA, lions also descended from that ancestor. Remember, to be valid, a clade must include the MRCA and all of its descendants. Similarly, snakes, lizards, tigers, and lions in this cladogram do not constitute a clade because bears, which descended from the same MRCA, are not included in the group.

When a clade consists of only two taxa, it is given the special name **sister group**. In the cladogram in Figure 3, snakes and lizards are a sister group, and tigers and lions are a sister group. In the cladogram in Figure 1, bears and felines are a sister group.

Consider again the idea of the vast tree of life. By breaking apart this huge and impossibly complex set of relationships into meaningful groups (clades), biologists can then try to piece back together those groups that are well-supported by evidence into a synthesis that reflects the actual history of life on Earth.

## Practice What You've Learned #2

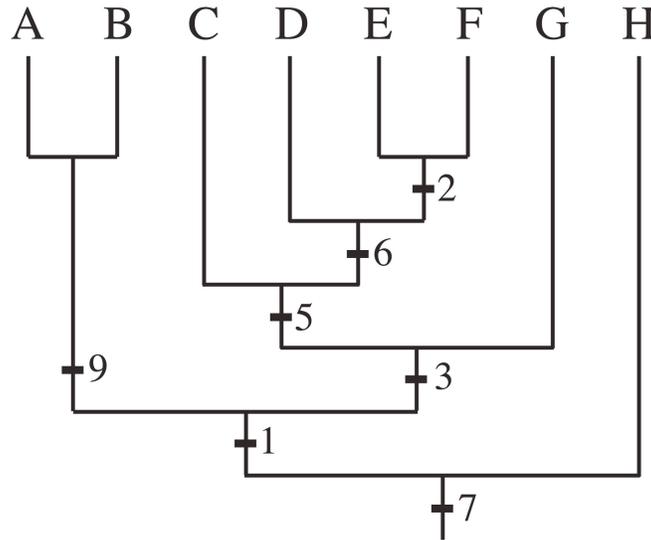
Answer the questions below without referring back to the previous pages. After you've completed the problems, you'll have a chance to check your answers against the correct answers. If anything is still unclear at that point, you may return to the earlier pages to revisit a topic.



1. Do the taxa C, D, E, and F form a clade? Yes      No
2. Do the taxa F, G, and H form a clade? Yes      No
3. Do Taxon A and Taxon B form a clade? Yes      No
4. Do the taxa A, B, C, D, E, and F form a clade? Yes      No
5. Do Taxon G and Taxon H form a sister group? Yes      No
6. Do Taxon E and Taxon F form a sister group? Yes      No
7. Which taxa belong to the clade defined by the MRCA that possessed character 5?  
\_\_\_\_\_
8. Which taxa belong to the clade defined by the MRCA that possessed character 3?  
\_\_\_\_\_

**STOP!**  
**DO NOT TURN THE PAGE UNTIL YOU HAVE FINISHED**  
**“PRACTICE WHAT YOU’VE LEARNED #2”**

**Answers to Practice what you've learned:** Check your answers on the previous page with the correct answers below. Note that an explanation for each answer is printed in blue. If any of these explanations are unclear, please refer back to the earlier instructions.



1. Do the taxa C, D, E, and F form a clade? Yes    No  
All the descendants of the MRCA of the named taxa are included in the group. The relevant MRCA for this question is the one that possessed character 5.
2. Do the taxa F, G, and H form a clade? Yes    No  
Many other taxa (A, B, C, D, and E) also are descendants of the MRCA of the three named taxa. The relevant MRCA for this question is the one that possessed character 7.
3. Do Taxon A and Taxon B form a clade? Yes    No  
All the descendants of the MRCA of the named taxa are included in the group. The relevant MRCA for this question is the one that possessed character 9. Because this clade contains only two taxa, it is also called a sister group.
4. Do the taxa A, B, C, D, E, and F form a clade? Yes    No  
One other taxon (G) also is a descendent of the MRCA of the six named taxa. The relevant MRCA for this question is the one that possessed character 1.
5. Do Taxon G and Taxon H form a sister group? Yes    No  
A sister group is a clade that contains only two taxa. G and H do not form a clade because many other taxa (A, B, C, D, E, and F) also are descendants of the MRCA of the two named taxa. The relevant MRCA for this question is the one that possessed character 7.
6. Do Taxon E and Taxon F form a sister group? Yes    No  
A sister group is a clade that contains only two taxa. E and F form a clade because all the descendants of the MRCA of the named taxa are included in the group. The relevant MRCA for this question is the one that possessed character 2.
7. Which taxa belong to the clade defined by the MRCA that possessed character 5?  
C, D, E, and F: All of these taxa evolved from the MRCA that possessed character 5, and no other taxa evolved from that ancestor.
8. Which taxa belong to the clade defined by the MRCA that possessed character 3?  
C, D, E, F, and G: All of these taxa evolved from the MRCA that possessed character 3, and no other taxa evolved from that ancestor.

## A Deeper Understanding of Evolutionary Trees

### The Basis for Determining Evolutionary Relatedness

**Most recent common ancestry.** Often scientists refer to a certain taxon as being more closely related to one taxon than to another. For example, lions are more closely related to bears than to monkeys, even though all three taxa belong to the mammal clade. In this section, we consider the basis for such a statement.

We first explain the correct basis for determining evolutionary relatedness, the one that biologists use. Then we discuss two incorrect methods that students often use instead. As part of that discussion, we will explain why these alternative methods are incorrect.

As you might expect, the correct basis for determining *evolutionary relatedness* is *most recent common ancestry* and not simply common ancestry. Taxa that share a more recent common ancestor are more closely related than taxa whose MRCA evolved less recently. Consider the cladogram shown in Figure 4a. According to the structure depicted, lizards are more closely related to bears than to sea urchins. This is because lizards share a more recent common ancestor with bears (marked by synapomorphy 3) than they do with sea urchins (marked by synapomorphy 8).

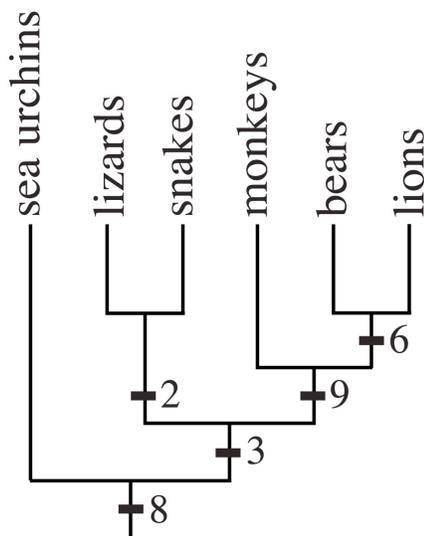


Figure 4a

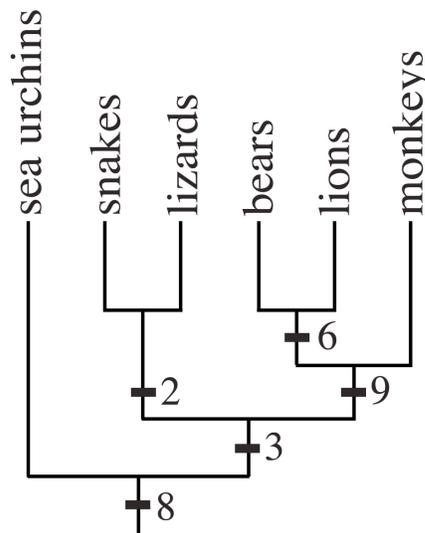


Figure 4b

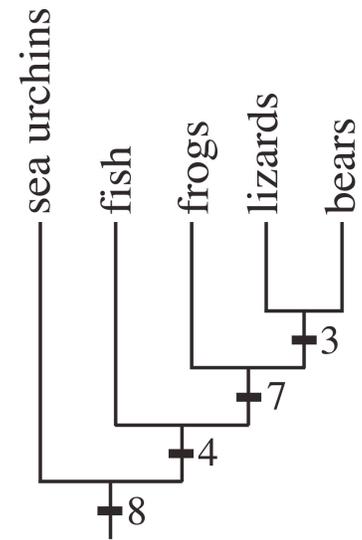


Figure 4c

**MISCONCEPTION: Horizontal distance.** A common misconception among students is that horizontal distance between the taxa at the top of the cladogram is important for determining evolutionary relatedness. In fact, horizontal distance between the taxa has no bearing on evolutionary relatedness. Even though lizards are closer to sea urchins than to bears in horizontal space in the cladogram shown in Figure 4a, they are actually more closely related to bears. You can see this because of the inherent relative time arrow running from the bottom to the top of the cladogram. The most recent ancestor shared by lizards and bears is more recent than the most recent ancestor shared by lizards and sea urchins.

One reason why horizontal distance in space is irrelevant is that cladograms are like mobiles. Imagine turning the cladogram in Figure 4a upside down and hanging it from the ceiling. If the horizontal lines are wooden dowels and the vertical lines are pieces of string, then the cladogram will turn in the breeze. As it rotates, the connections among the various clades will remain the same, but which taxa are next to each other will change. The cladogram in Figure 4b shows one possible rotation of the cladogram in Figure 4a. Notice that all the taxa are connected in exactly the same way (check the synapomorphies) but now they have different horizontal neighbors. In this depiction, lizards are now next to bears rather than to sea urchins, sea urchins are next to snakes, and monkeys are next to lions. The cladograms in Figures 4a and 4b are identical in terms of the evolutionary relationships they convey (i.e., they depict the same nested sets of clades); they are simply rotations of each other. Thus, the relationships among lizards, bears, and sea urchins, the three taxa we have been considering, is the same in the cladograms shown in Figures 4a and 4b: Lizards are more closely related to bears than to sea urchins.

**MISCONCEPTION: Number of steps.** Another common misconception is that the way to determine evolutionary relatedness is to count the number of steps between the taxa in question. In fact, the number of steps has no bearing on evolutionary relatedness. Even though there are fewer evolutionary steps (branching points) between lizards and sea urchins than between lizards and bears (see the cladograms in Figures 4a and 4b), lizards, as we have seen, are more closely related to bears. The reason why the number of steps is irrelevant is because any cladogram an evolutionary biologist might work with is only a tiny portion of the complete tree of life. A complete cladogram for animals would include all animal taxa known to evolutionary biologists—approximately 3-30 million species! An individual biologist, therefore, only studies a very small subset of these taxa. When a cladogram includes only a subset of taxa (as is always the case), the number of steps between any pair of taxa depends on which particular taxa happen to be included in the cladogram.

Consider now the cladogram shown in Figure 4c, which includes lizards, bears, and sea urchins in the context of fish and frogs rather than snakes, monkeys, and lions. In that cladogram, there are now fewer steps between lizards and bears (they comprise a sister group!) than between lizards and sea urchins. This is because the taxa “between” lizards and bears have been removed, and two new taxa have been added “between” lizards and sea urchins. With a little thought, it should be obvious that the relative evolutionary relatedness between pairs of taxa depends on their evolutionary history, not on which other taxa happen to be included in the cladogram. The one measure that is consistent across the cladograms in Figures 4a and 4b and the cladogram in Figure 4c is levels of *most recent common ancestry*. Regardless of the other taxa that are included in the cladogram, we can see that lizards share a more recent common ancestor with bears than with sea urchins. For this reason, lizards are more closely related to bears than they are to sea urchins.

## Understanding Polytomies

All the cladograms you have seen so far have been *resolved*. That is, there were never more than two branches stemming from each MRCA. However, biologists often construct cladograms that are not completely resolved; rather, they contain polytomies. Our research shows that students have more difficulty understanding the evolutionary relationships among taxa when the cladograms contain a polytomy (e.g., Novick & Catley, 2007).

A **polytomy** is an *unresolved* group of three or more taxa. For example, in the cladogram shown in Figure 5, fleas, flies, and scorpionflies form a polytomy. This relationship is highlighted in purple in the figure. These three taxa share a more recent common ancestor with each other than they do with any other taxon on the cladogram. However, because there is insufficient evidence to know which two of these three taxa share a MRCA in respect to the third taxon, these taxa cannot be drawn as a 3-taxon statement. Evolutionary biologists say that the relationships among these taxa are unresolved given current scientific evidence. The appearance of a polytomy in a cladogram points scientists to where more data are needed to resolve the relationships.

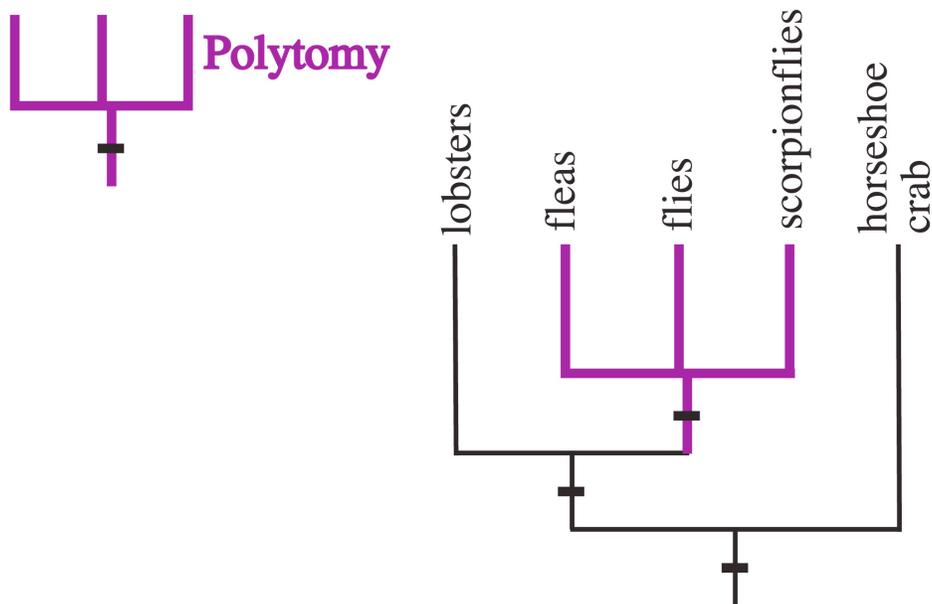


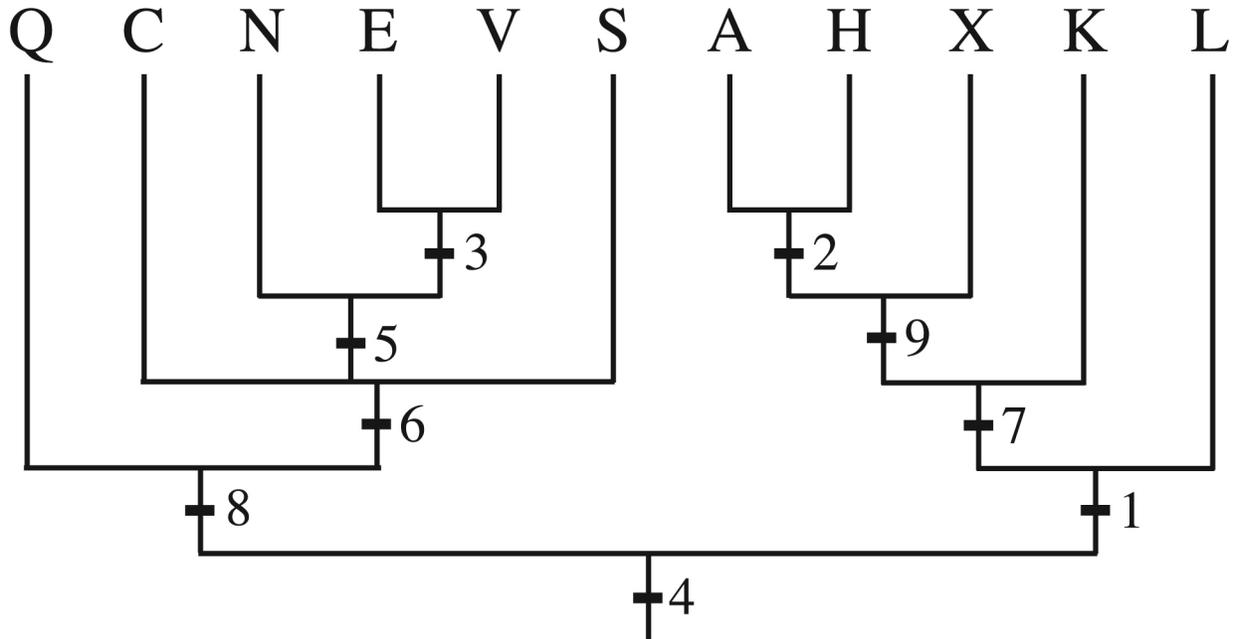
Figure 5

Because (a) all members of a polytomy have the same most recent common ancestor (given the current state of scientific evidence) and (b) the relationships among the polytomy members cannot be resolved into a 3-taxon statement, all members of a polytomy are considered to be equally closely related to each other. This is very important: **All taxa emanating from branches that form a polytomy are equally closely related to each other.** Remember, the only way to determine whether Taxon A is more closely related to Taxon B or to Taxon C, or whether those taxa are all equally closely related to each other, is to look at the levels of most recent common ancestry. If the level of most recent common ancestry differs, then evolutionary relatedness differs as well.

Correspondingly, if a set of taxa all have the same MRCA, and no more recent ancestor has been found to relate a subset of *the particular taxa under consideration*, then they are all equally closely related to each other. Consider, for example, that many species are included in the group *flies* in the cladogram shown in Figure 5 (e.g., hoverflies and houseflies). Given this cladogram, then, we would say that fleas, scorpionflies, and any one taxon in the group flies (e.g., houseflies) are equally closely related to each other: They share the same MRCA, and none of these three taxa shares a more recent common ancestor with one of the other two taxa under consideration.

### Practice What You've Learned #3

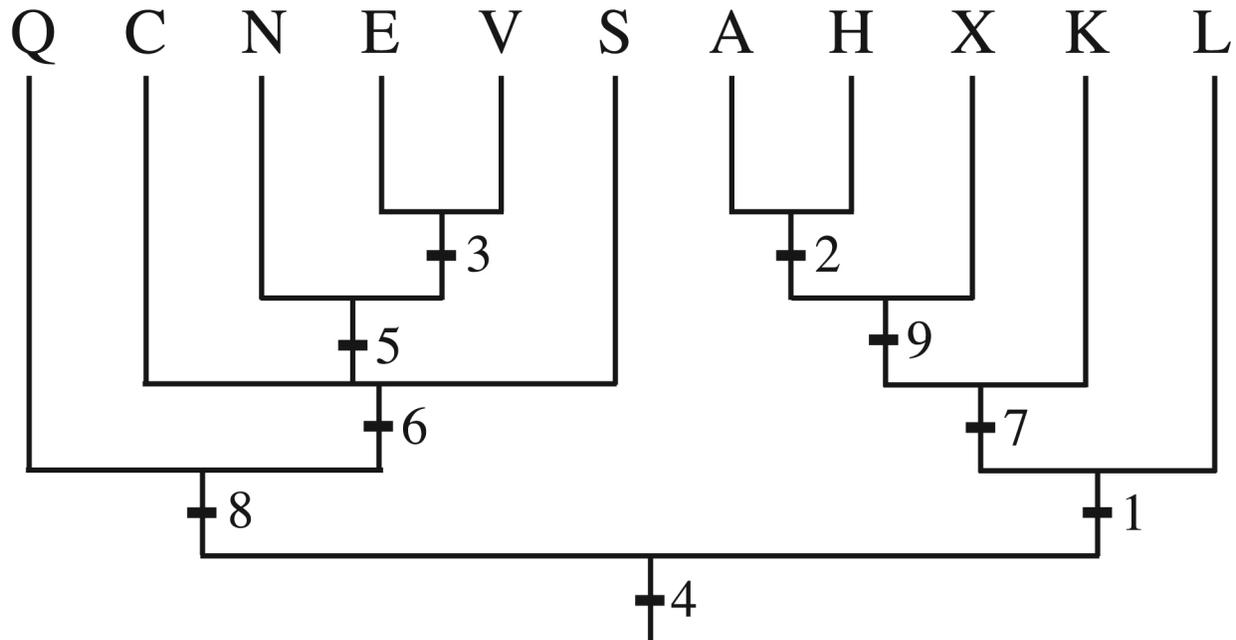
Answer the questions below without referring back to the previous pages. After you've completed the problems, you'll have a chance to check your answers against the correct answers. If anything is still unclear at that point, you may return to the earlier pages to revisit a topic.



1. Which taxa are more closely related: C and V or C and Q? \_\_\_\_\_
2. Which taxon/taxa is/are most closely related to A? \_\_\_\_\_
3. Which taxon, Q or H, is most closely related to V? \_\_\_\_\_
4. Which taxa are more closely related: C and E or C and S? \_\_\_\_\_
5. Which taxon, H or L, is most closely related to K? \_\_\_\_\_
6. Which taxon/taxa is/are most closely related to L? \_\_\_\_\_

**STOP!**  
**DO NOT TURN THE PAGE UNTIL YOU HAVE FINISHED**  
**“PRACTICE WHAT YOU’VE LEARNED #3”**

**Answers to Practice what you've learned:** Check your answers on the previous page with the correct answers below. Note that an explanation for each answer is printed in blue. If any of these explanations are unclear, please refer back to the earlier instructions.



1. Which taxa are more closely related: C and V or C and Q? C and V because they share a more recent common ancestor (the ancestor that possessed character 6) than do C and Q.
2. Which taxon/taxa is/are most closely related to A? H because it shares a MRCA with A (the one that possessed character 2). These two taxa form a sister group.
3. Which taxon, Q or H, is most closely related to V? Q because V shares a more recent common ancestor with Q (the ancestor that possessed character 8) than with H.
4. Which taxa are more closely related: C and E or C and S? C, E, and S are equally closely related because they all have the same MRCA, the one that possessed character 6. In this cladogram, no two of these three taxa possess a more recent common ancestor with each other than they do with the third taxon. These three taxa form a polytomy. More data are needed to resolve their relationships into a 3-taxon statement.
5. Which taxon, H or L, is most closely related to K? H because K shares a more recent common ancestor with H (the ancestor that possessed character 7) than with L.
6. Which taxon/taxa is/are most closely related to L? A, H, X, and K because they all share a MRCA with L (the one that possessed character 1). These five taxa form a clade.

## Thinking About Subsets of the Complete Tree of Life

It is important to understand that every cladogram is (a) necessarily only a restricted piece of the vast theoretical tree of life and (b) a testable hypothesis that reflects the best explanation of currently available data. Because they are hypotheses, tree topologies (i.e., patterns of arrangements of taxa) can and do change as newer techniques and the discovery of new (extant or fossil) taxa provide more data. Such testability is one of science's great strengths. Although the placement of major taxa (e.g., Mammalia) does not often change, the evolutionary relationships among taxa within some groups (e.g., Arthropoda) are still not completely known.

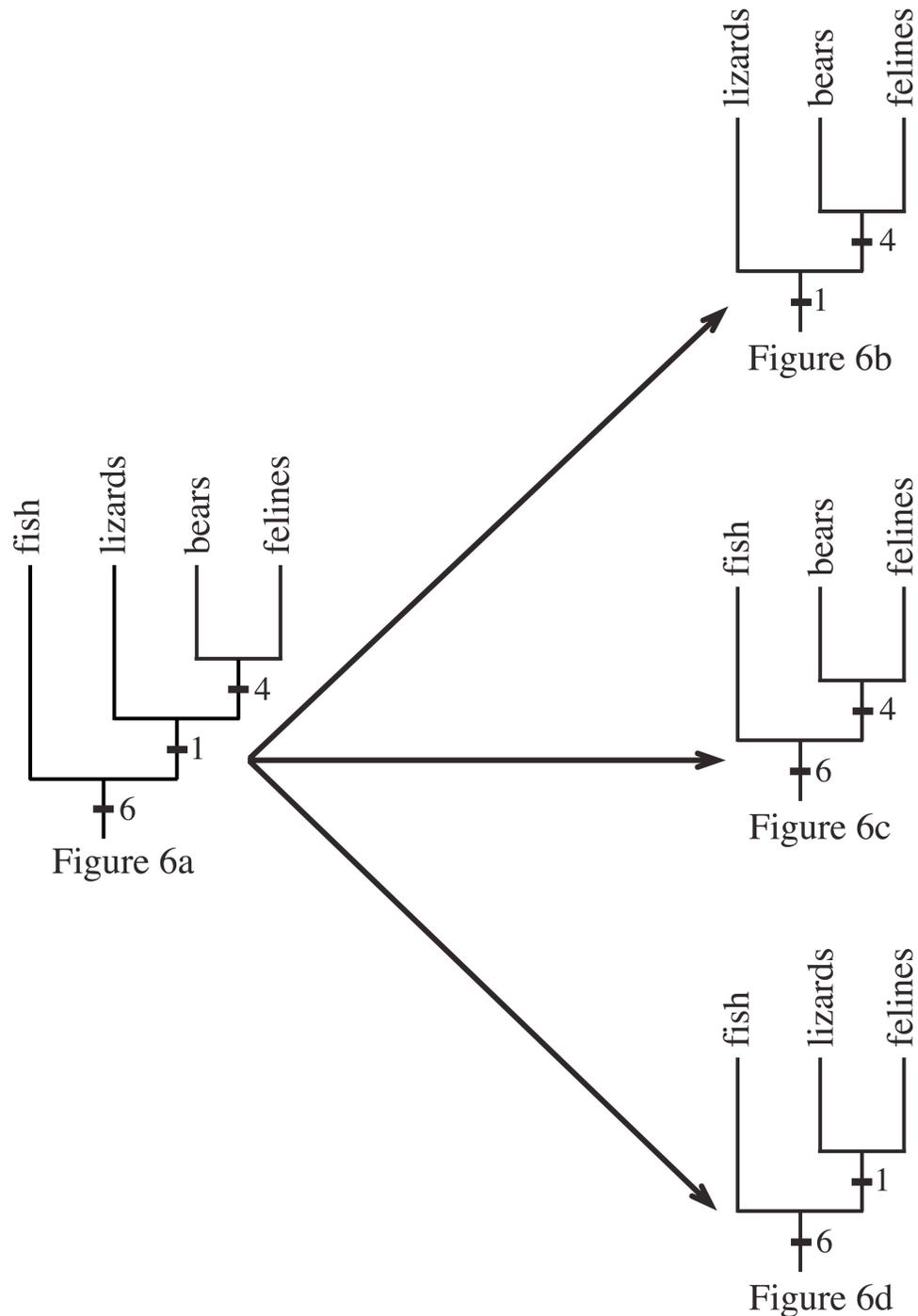
Because the complete tree of life would have millions of taxa on it, individual biologists only study a very small subset of those taxa. It is necessary, therefore, to be able to both (a) construct smaller cladograms that depict evolutionary relationships among the relevant subset of taxa from the (theoretical) complete tree of life and (b) combine evidence from smaller cladograms into a larger cladogram that depicts relationships among all taxa included in the separate cladograms.

To reiterate, a particular cladogram depicts evolutionary relationships among the set of taxa a researcher is studying or a set of taxa about which a biologist or teacher wishes to make some point. For example, we selected particular taxa for the cladograms in this booklet to help make certain points and to clarify certain concepts. Because in practice any cladogram depicts relationships among only a subset of taxa, it should be clear that two taxa shown as sister groups on a particular cladogram may not be the actual sister groups on the complete tree of life; there could be many other taxa (both living and extinct) that "fill in" the space between them. The cladograms in Figure 4 provided a simple illustration of this fact.

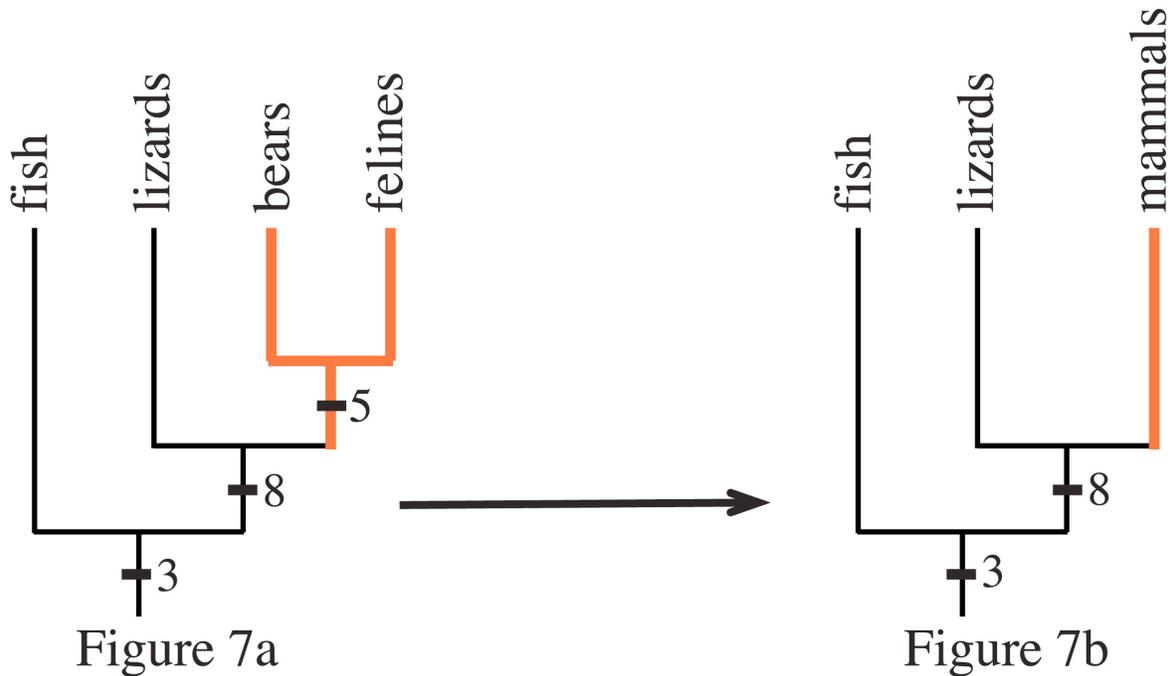
A cladogram that depicts evolutionary relationships among a subset of taxa can be constructed by pruning unneeded taxa off a larger tree and by collapsing lower-level taxa into higher-level taxa (e.g., robins, blue jays, and penguins can be collapsed into the category *birds*; hoverflies and houseflies can be collapsed into the category *flies*). Sometimes biologists need to remove taxa from trees (pruning) because newer evidence suggests that they are not part of the group in question after all and need to be placed, for example, in another family. Other times, biologists prune taxa that are not of current interest; they want to focus on the relationships among certain other taxa. Also, biologists sometimes collapse a set of taxa comprising a clade into a larger group supported by the MRCA of them all. In addition, you might imagine that a single taxon on a tree (say rats in a tree of other mammals) could represent all taxa of its kind (e.g., rodents in the previous example). Finally, it stands to reason that if the larger tree can be reduced to smaller versions by these methods, then smaller, overlapping trees proposed by different scientists studying overlapping subsets of taxa can be merged into a larger, combined tree. The remainder of this booklet illustrates these concepts of pruning and collapsing taxa and merging trees.

## Making Larger Cladograms Smaller

**Pruning** a taxon off a tree simply removes that taxon from a specific cladogram. This is illustrated in Figure 6. If we prune fish off the cladogram in Figure 6a, the cladogram becomes that shown in Figure 6b. If instead we prune lizards from the cladogram in Figure 6a, we get the cladogram shown in Figure 6c. Finally, if we prune bears off, we get the cladogram shown in Figure 6d. Notice that in all cases the pruned cladogram shows the same evolutionary relationships among the remaining taxa that were depicted in the original cladogram. This is very important and can be tested by checking the synapomorphies, which correspond across the original and pruned cladograms. For example, the relationship between lizards and felines is supported by synapomorphy 1 in all three cladograms (Figures 6a, 6b, and 6d) that include both of those taxa.



When pruning, the taxon in question is simply removed from the cladogram. When *collapsing*, two or more lower-level taxa that comprise a clade are combined into a single higher-level taxon. This is important: **You cannot collapse just any set of taxa; they have to comprise a clade supported by a MRCA.** Also, the relationship of the new higher-level taxon to the rest of the taxa in the cladogram must remain the same. This is illustrated with the taxa highlighted in orange in Figure 7. The clade (bears + felines) in the cladogram shown in Figure 7a can be collapsed into the taxon *mammals*, yielding the cladogram shown in Figure 7b. Notice that in both cladograms in Figure 7, the relationship between lizards and the mammals is supported by character 8.



## Merging Several Smaller Cladograms Into a Single Combined Cladogram

When two different researchers investigate overlapping sets of taxa, they may construct cladograms that partially overlap. These cladograms can be combined, or *merged*, into a single cladogram that (a) includes all the taxa investigated by both researchers and (b) preserves all of the evolutionary relationships depicted in the smaller cladograms. To merge two cladograms, find overlapping taxa and clades in the separate cladograms and put them together so as to preserve the relationships specified by the MRCAs. This is illustrated in Figure 8, with the common colors and numbered synapomorphies across the different cladograms indicating the preserved relationships. In the cladogram in Figure 8a, three taxa are supported by character 9: lizards, bears, and felines (i.e., character 9 is the synapomorphy that defines the MRCA of these three taxa). In the cladogram in Figure 8b, there are four taxa supported by this character: lizards, snakes, bears, and felines. This cladogram shows that snakes are a sister group to lizards (supported by character 6). The merged cladogram in Figure 8c must include all four taxa, plus it must preserve the relationships among those taxa that are shown in the cladograms in both Figures 8a and 8b. Finally, the cladogram in Figure 8a also shows that fish are a sister group to the clade consisting of (lizards + (bears + felines)). This relationship also must be preserved in the merged cladogram, as shown in Figure 8c.

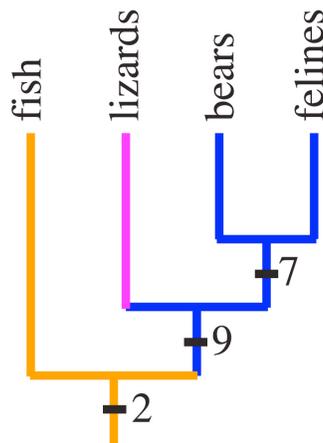


Figure 8a

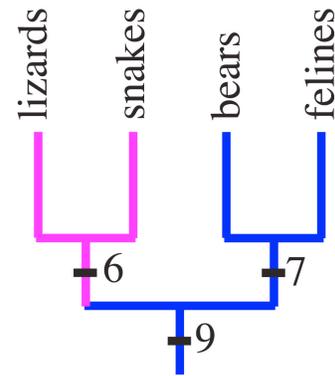


Figure 8b

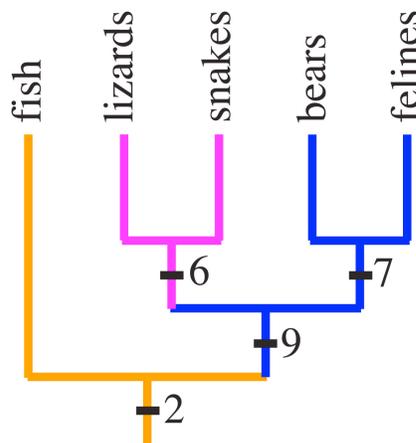


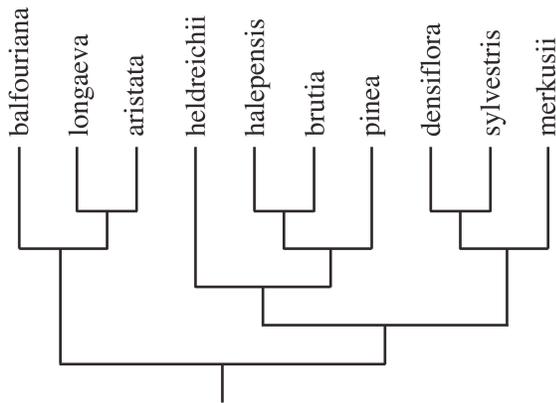
Figure 8c

### Practice What You've Learned #4

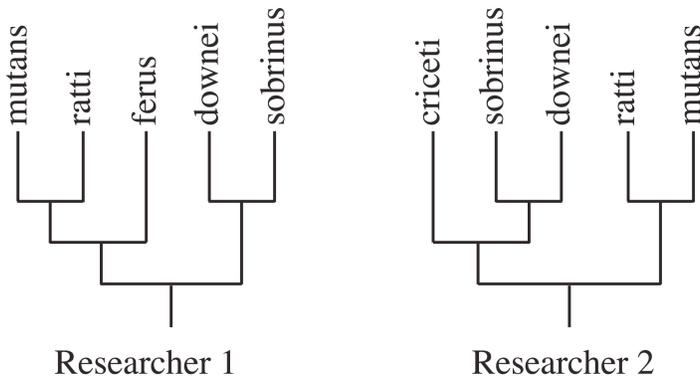
Answer the questions below without referring back to the previous pages. After you've completed the problems, you'll have a chance to check your answers against the correct answers. If anything is still unclear at that point, you may return to the earlier pages to revisit a topic.

1. The cladogram below shows evolutionary relationships among species of pine trees. Redraw this cladogram in the space to the right, making the following changes:

- Prune off *balfouriana* and *sylvestris*.
- Collapse the four taxa *heldreichii*, *halepensis*, *brutia*, *pinea* into a single group labeled *Pinaster*.



2. Researcher 3 is interested in the evolutionary relationships among particular species of *Streptococcus* bacteria that were studied by Researchers 1 and 2. Those two researchers studied some of the same taxa but some different taxa as well. Researcher 1 drew the cladogram below on the left to depict the evolutionary relationships among the 5 *Streptococcus* taxa she studied. Researcher 2 drew the cladogram below on the right to depict the evolutionary relationships among the 5 *Streptococcus* taxa he studied. Draw a single cladogram for Researcher 3 that depicts the evolutionary relationships among all 6 *Streptococcus* species studied by Researchers 1 and 2. Make sure the cladogram you draw preserves all of the relationships depicted in the two smaller cladograms.



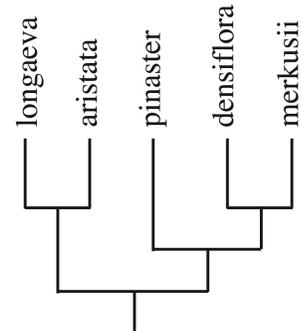
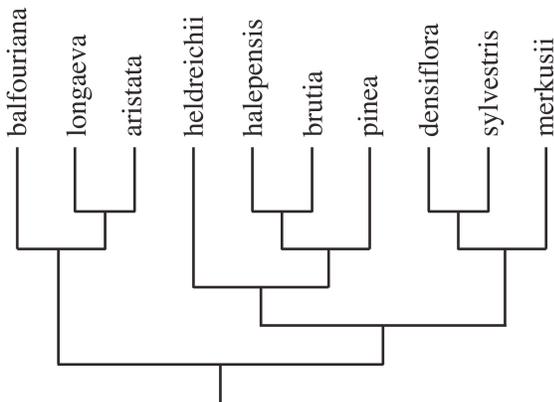
**STOP!**

**DO NOT TURN THE PAGE UNTIL YOU HAVE FINISHED  
"PRACTICE WHAT YOU'VE LEARNED #4"**

**Answers to Practice what you've learned:** Check your answers on the previous page with the correct answers below. Note that an explanation for each answer is printed in blue. If any of these explanations are unclear, please refer back to the earlier instructions.

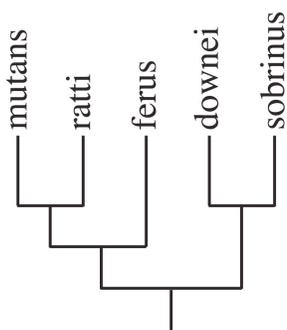
1. The cladogram below shows evolutionary relationships among species of pine trees. Redraw this cladogram in the space to the right, making the following changes:

- Prune off *balfouriana* and *sylvestris*.
- Collapse the four taxa *heldreichii*, *halepensis*, *brutia*, *pinea* into a single group labeled *Pinaster*.

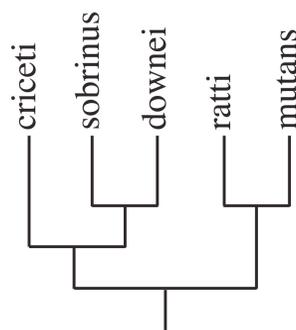


Correct Answer

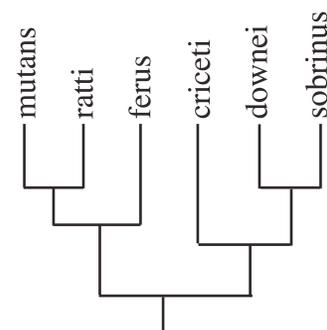
2. Researcher 3 is interested in the evolutionary relationships among particular species of *Streptococcus* bacteria that were studied by Researchers 1 and 2. Those two researchers studied some of the same taxa but some different taxa as well. Researcher 1 drew the cladogram below on the left to depict the evolutionary relationships among the 5 *Streptococcus* taxa she studied. Researcher 2 drew the cladogram below on the right to depict the evolutionary relationships among the 5 *Streptococcus* taxa he studied. Draw a single cladogram for Researcher 3 that depicts the evolutionary relationships among all 6 *Streptococcus* species studied by Researchers 1 and 2. Make sure the cladogram you draw preserves all of the relationships depicted in the two smaller cladograms.



Researcher 1



Researcher 2



Correct Answer

## Conclusion

You have now come to the end of the instructional booklet on understanding evolutionary trees known as cladograms. We have tried to teach you some of what biologists refer to as *tree thinking* (e.g., Baum, Smith, & Donovan, 2005; Catley, 2006, O'Hara, 1988). Cladograms are a vitally important tool within the biological sciences because they represent and organize existing knowledge about species and higher-order taxa. By using nested sets of clades supported by synapomorphies to systematize the 3.5 billion year history of life on Earth, cladograms enable evidence-based inference and provide a conceptual framework for basic and applied biology (e.g., Futuyma, 2004; Nickels & Nelson, 2005; Yates, Salazar-Bravo, & Dragoo, 2004).

Because cladograms reflect the underlying structure of the domain, namely evolutionary relationships among taxa, inferences based on them are likely to be informative and to have practical value. For example, inferring which antivenin to use to counteract the bite of a venomous king brown snake based on its close evolutionary relationship to the red-bellied black snake is more likely to lead to a successful outcome (survival!) than basing the choice of antivenin on the king brown snake's similar coloration to the western brown snake (Freeman, 2005).

A recent National Research Council (2009) report concluded that biology's tremendous potential for solving numerous problems of the modern world rests on the unifying concept of evolution. Many of these solutions will require phylogenetic analyses (e.g., Futuyma, 2004; Yates et al., 2004). Indeed, cladistics and macroevolution have come to play an increasingly important role in research in the life sciences. Such research has had, and will continue to have, important implications for society. For example, cladistic analyses were responsible for identifying the West Nile virus as the disease agent responsible for sickness/death in the initial outbreak in 1999; these analyses also provided health officials with information necessary to effectively treat the outbreak, as well as to make predictions about the spread of the disease (Yates et al., 2004). To take another example, because single-celled eukaryotes called microsporidians have recently been determined to be fungi, researchers are now testing whether fungicides can cure microsporidian infections in bee colonies (Freeman, 2011). These infections are a significant problem for world agriculture, and thus for combating world hunger, because many crops depend on bees for pollination. Indeed, every third bite we eat is a direct result of honeybee pollination. In sum, as O'Hara (1988) noted, answering almost all evolutionary "why" questions requires an understanding of cladograms—that is, the ability to engage in tree thinking.

## References

- Baum, D. A., Smith, S. D., & Donovan, S. S. (2005). The tree thinking challenge. *Science*, *310*, 979-980.
- Catley, K. M. (2006). Darwin's missing link: a new paradigm for evolution education. *Science Education*, *90*, 767-783.
- Darwin, C. (1859). *On the origin of species by means of natural selection* (First edition). London: John Murray.
- Freeman, S. (2005). *Biological science* (2<sup>nd</sup> ed.). Upper Saddle River (NJ): Prentice Hall.
- Freeman, S. (2011). *Biological Science: Evolution, Diversity, & Ecology* (Vol. 2; 4<sup>th</sup> ed.). Boston, MA: Benjamin Cummings.
- Futuyma, D. J. (2004). The fruit of the tree of life. In J. Cracraft & M. J. Donoghue (Eds.), *Assembling the tree of life* (pp. 25-39). New York, NY: Oxford University Press.
- Hennig, W. (1966). *Phylogenetic systematics*. Urbana, IL: University of Illinois Press.
- National Research Council. (2009). *A new biology for the 21st century: Ensuring the United States leads the coming biology revolution*. Washington, DC: National Academies Press.
- Nickels, M. K., & Nelson, C. E. (2005). Beware of nuts & bolts: Putting evolution into the teaching of biological classification. *The American Biology Teacher*, *67*, 283-289.
- Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. *Journal of Experimental Psychology: Applied*, *13*, 197-223.
- O'Hara, R. J. (1988). Homage to Clio, or, toward an historical philosophy for evolutionary biology. *Systematic Zoology*, *37*, 142-155.
- Thanukos, A. (2009). A name by any other tree. *Evolution: Education and Outreach*, *2*, 303-309.
- Yates, T. L., Salazar-Bravo, J., & Drago, J.W. (2004). The importance of the tree of life to society. In J. Cracraft & M. J. Donoghue (Eds.), *Assembling the tree of life* (pp. 7-17). New York: Oxford University Press.

## Acknowledgments

The development of this instructional booklet was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080621 to Vanderbilt University (L. R. Novick, Principal Investigator; K. M. Catley, Co-Investigator). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.