Phylogenetics Laboratory: Reconstructing Evolutionary History

Kefyn M. Catley
Western Carolina University

Laura R. Novick
Vanderbilt University

Acknowledgments

The development of this laboratory 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.

This laboratory was designed for use in an introductory biology class for science majors. It can be completed during the standard 3-hour laboratory section with students working in small groups. The laboratory materials include a student laboratory manual (appended here), an instructors' guide with answer key (appended here), and various specimens. These are the typical specimens found in a general undergraduate biology laboratory and include preserved wet specimens, dried material, and cased skeletons of the nine taxa used in the lab (Annelida, Arthropoda, Chordata, Cnidaria, Echinodermata, Mollusca, Nematoda, Platyhelminthes, and Porifera). One or more undergraduate biology textbooks with which students are familiar should be made available. Students may also be encouraged to search the internet for information concerning the characters. Alternatively (or in addition), instructors may choose to provide pictures of some of the characters from such sources. The specimens, organismal chapters of the textbooks, and internet sources are resources students may use to complete the character matrix in Part I of the lab.

Phylogenetics Laboratory: Reconstructing Evolutionary History

Kefyn M. Catley
Western Carolina University

Laura R. Novick
Vanderbilt University

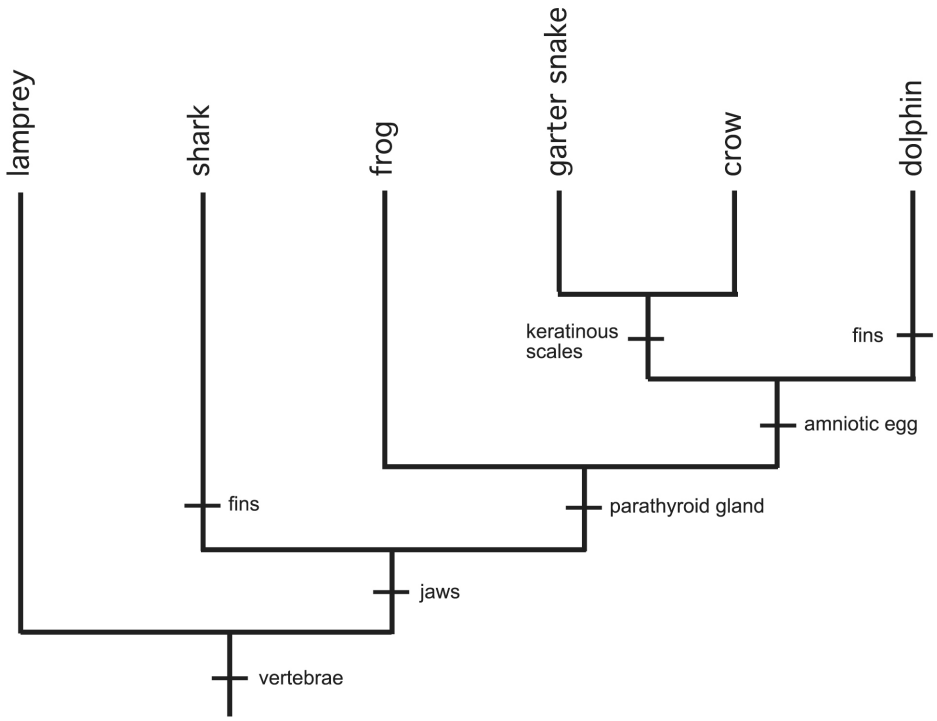
Version 1 © February, 2010
Version 2 © August, 2010
Version 3.1 © June, 2011
Version 3.2 © August, 2012
Version 3.3 © June, 2016

Introduction

Phylogenetics is the study of the history of life as reflected in the evolutionary relationships among taxa. 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. Phylogenetics is a very powerful tool that allows biologists in all areas to organize and retrieve information about biodiversity and to make predictions based on most recent common ancestry. This methodology is based on cladistics and attempts to reconstruct the evolutionary history of taxa based on observable and testable evidence.

Rather than grouping taxa based on overall similarity, cladistics proposes to group taxa based on a special type of similarity called a *synapomorphy*—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* (MRCA) and (b) the ancestor of that ancestor did not possess that character (i.e., the character is new in the MRCA of the taxa in question). For example, having jaws is a synapomorphy that defines fish, amphibians, reptiles, and mammals as belonging to a single group, labeled *Gnathostomata.* This character was new (i.e., derived) in the MRCA of these taxa; in contrast, the MRCA of all these taxa (i.e., of *Gnathostomata*) plus lampreys did not have jaws. Note that the character jaws has two states (i.e., possible values): present and absent. Other characters may have more than two states. 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 organizational and predictive tool in modern biology.

Phylogenies are most often depicted in a type of diagram called a cladogram, as shown in the figure at the top of the next page. Before a cladogram can be drawn, characters (i.e., possible synapomorphies) have to be proposed and organized in a matrix of taxa by characters. A cladogram is then constructed that best fits the character data. That is, a cladogram, in its most basic form, is a diagram showing the distribution of characters among taxa.



Part I: Identifying character states

In the first part of this lab, you will examine representatives of 9 major groups of animal taxa (see the rows of the matrix below) to determine how the different possible states of each of 11 characters (see the columns of the matrix) are distributed among the groups. You should also use your textbook to help you fill in the matrix. The table at the top of the next page describes the possible states for each character. For most characters, you just need to determine whether that character is present (yes) or absent (no). For other characters, you need to determine what state occurs (e.g., bilateral vs. radial symmetry). Notice that in some cases it is possible to respond that a particular character does not apply (e.g., a particular taxon does not exhibit any kind of symmetry). To help you get started, some character states have already been entered in the character matrix.

	rDNA sequence	Type of skeleton	Segmentation	Notochord	Trochophore larvae	Elongate cylindrical body	Molt	Radial cleavage	Multicellular	Symmetry	Germ layers
Annelida	y				y						3
Arthropoda	n							n			
Chordata	n							y			
Cnideria	n									r	
Echinodermata	n										
Mollusca	y										
Nematoda	n				n						
Platyhelminthes	y										
Porifera	n									–	–

Character	State 1	State 2	State 3	
Unique 18S rDNA sequence	yes (y)	no (n)		
Type of skeleton	hydrostatic (h)	ecto (ec)	endo (en)	does not apply (–)
Segmentation	yes (y)	no (n)		does not apply (–)
Notochord	yes (y)	no (n)		
Trochophore larvae	yes (y)	no (n)		
Elongate cylindrical body	yes (y)	no (n)		
Molt	yes (y)	no (n)		
Radial cleavage	yes (y)	no (n)		
Multicellular	yes (y)	no (n)		
Symmetry	bilateral (b)	radial (r)		does not apply (–)
Germ layers	2	3		does not apply (–)

Part II: Mapping character states onto alternative cladogram topologies

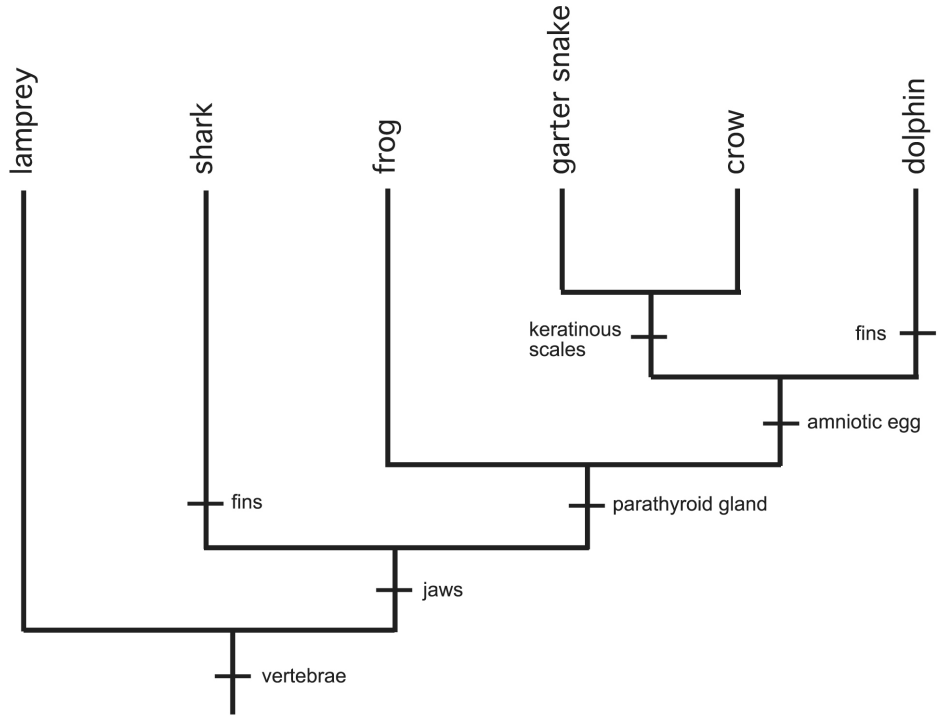
Mapping the character states used to derive a particular topology back onto the cladogram is a powerful way to answer such questions as (a) which characters are due to descent from a MRCA and which represent convergent or parallel evolution (i.e., independent evolution from different ancestors) and (b) what is the best evidence concerning the historical evolutionary relationships among the set of taxa. To answer these questions, evolutionary biologists try to find the simplest way to map the character states (e.g., the *yes* and *no* responses in the cells of the character matrix) onto a cladogram; that is, the way that requires the *fewest* number of steps or changes.

Consider the following character matrix, which shows the distribution of six characters (amniotic egg, jaws, vertebrae, fins, keratinous scales, and a parathyroid gland) among six vertebrate taxa (crow, shark, dolphin, lamprey, frog, and garter snake). For example, you can see that both crows and garter snakes have keratinous scales, but none of the other taxa do.

	amniotic egg	jaws	vertebrae	fins	keratinous scales	parathyroid gland
crow	y	y	y	n	y	y
shark	n	y	y	y	n	n
dolphin	y	y	y	y	n	y
lamprey	n	n	y	n	n	n
frog	n	y	y	n	n	y
garter snake	y	y	y	n	y	y

y = yes, n = no

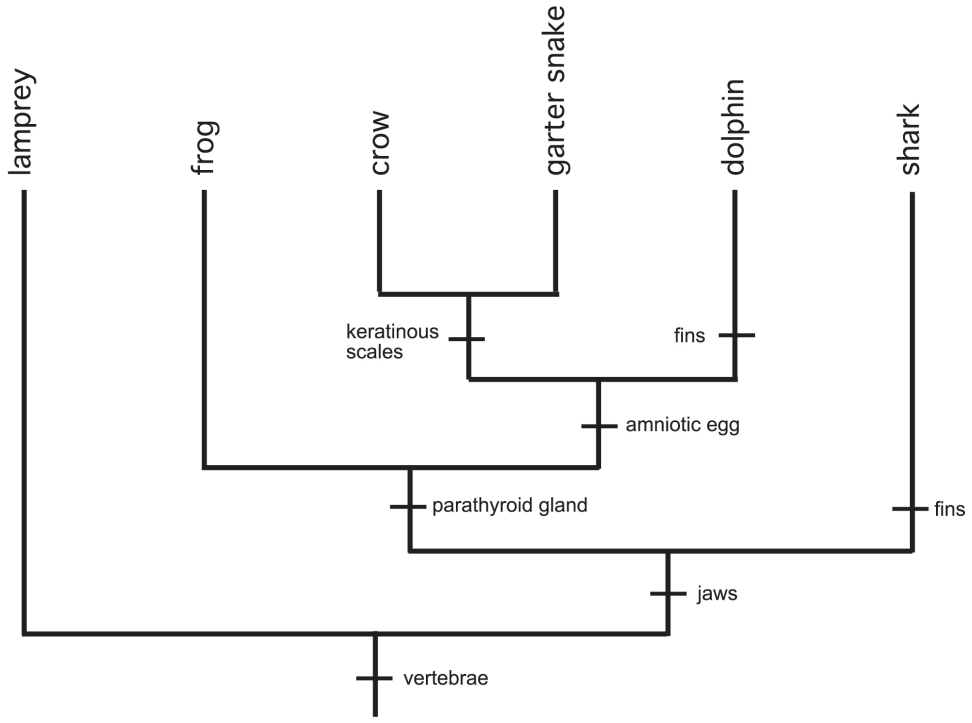
The cladogram at the top of the next page shows the portion of the complete tree of life that only includes these six taxa. The topology (branching structure) of this cladogram is supported by the characters in the above matrix, as well as many other characters. This is just a simple example to illustrate how to map character states back onto a cladogram.



Notice how the character states have been placed onto the branching points and terminal branches of the cladogram in the simplest way possible. For example, the position of *parathyroid gland* on the cladogram conveys the information that frogs, garter snakes, crows, and dolphins all possess this character but sharks do not. We hypothesize that frogs, garter snakes, crows, and dolphins share this character because they share a MRCA in which that character was newly derived. This is a simpler explanation than assuming it was derived independently four times—i.e., once in each taxon. Thus, it is unnecessary to place *parathyroid gland* on each of the separate branches for these four taxa. In contrast, notice that *fins* is located at two places on the cladogram—on the branch for shark and on the branch for dolphin. We hypothesize that these two taxa share this character due to convergent evolution and not as result of most recent common ancestry.

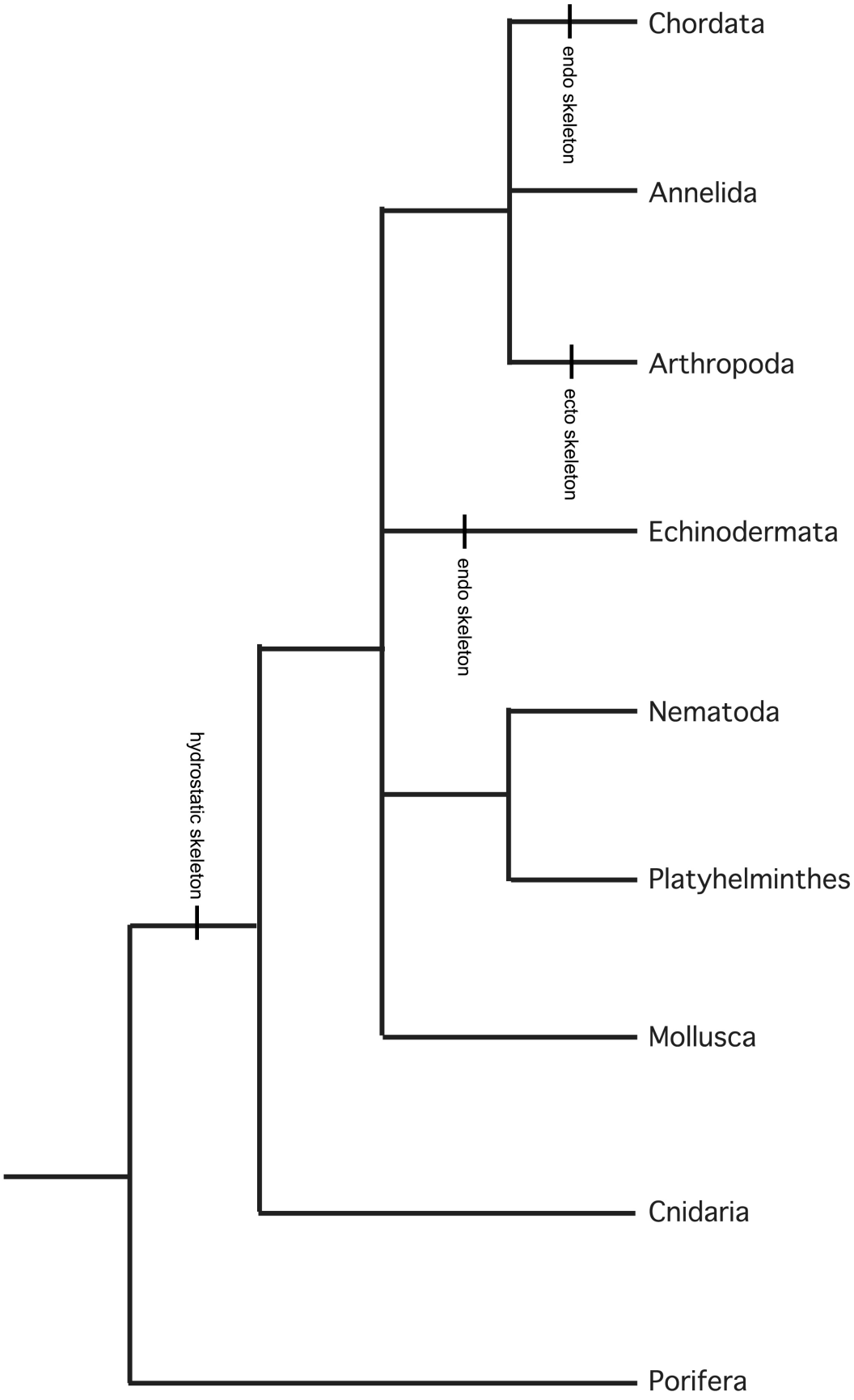
In this cladogram, frogs, garter snakes, crows, and dolphins comprise a *clade*. A clade is a group that contains a most recent common ancestor (in this case the hypothesized ancestor in which *parathyroid gland* was derived) and *all* of its descendants. Clades are also known as monophyletic groups. A cladogram, then, is a diagram that shows nested clades, representing nested levels of most recent common ancestry.

An important point to keep in mind about cladograms is that they are like mobiles. If you turn the cladogram on this page upside down and hang it by the line on which *vertebrae* is located, it will turn in the wind. Any rotation of the branches due to this turning will preserve the branching structure depicted and thus is the same cladogram. For example, the cladogram at the top of the next page is one rotation of the cladogram shown on this page. Notice how the branching structure, supported by the characters, remains unchanged.

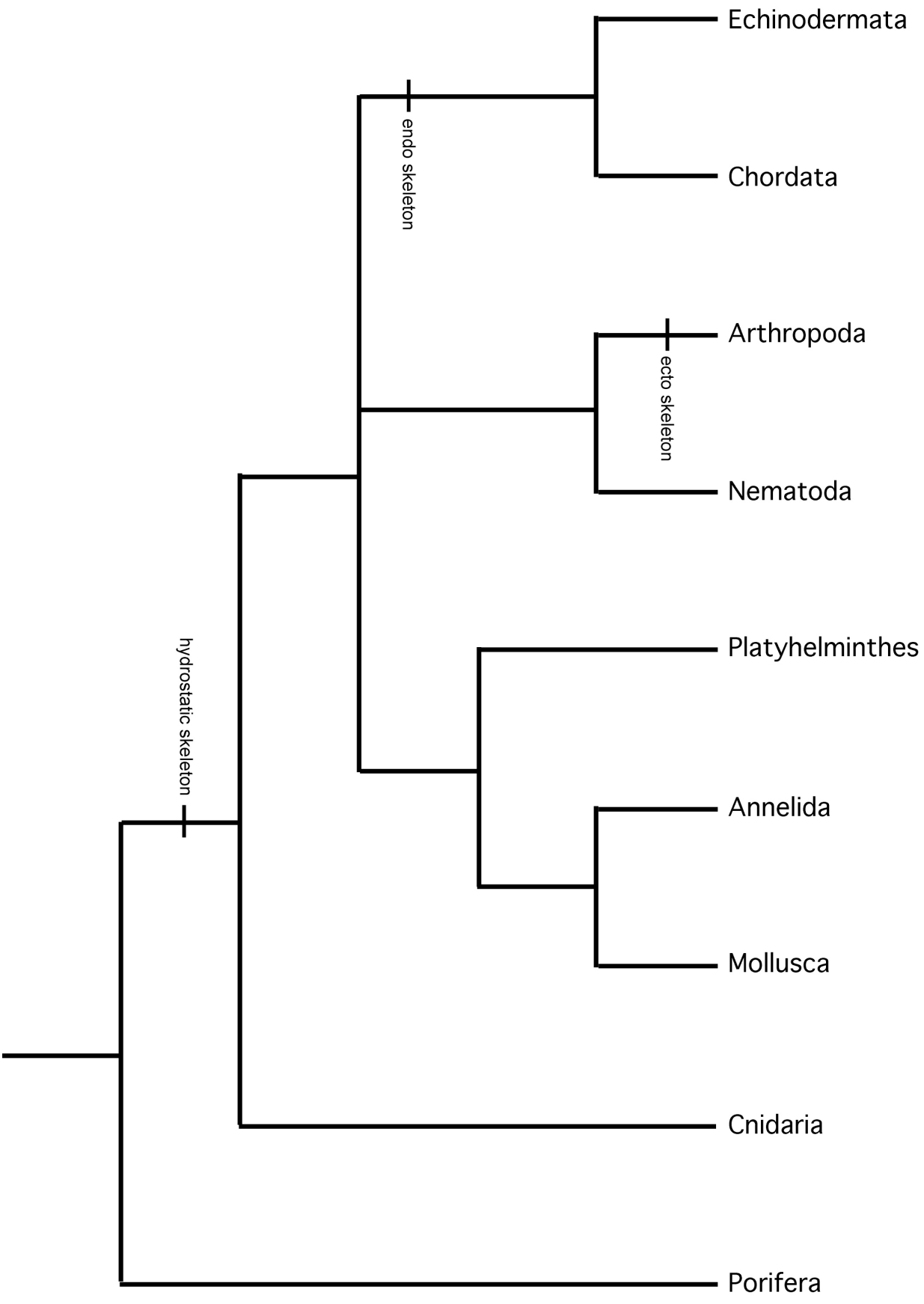


In this part of the lab, your task is to map the character states you identified for the nine major groups of animal taxa onto each of three separate cladograms. Each of these cladograms shows a different topology, representing a different hypothesis concerning the evolutionary history of these taxa; but all are based on the same character matrix. The character states for one of the 11 characters (type of skeleton) have been mapped onto each cladogram to help you get started. As in the above example, it is important to make sure that (a) every character state for every taxon that is specified in the character matrix is represented on the cladogram and (b) the character states are placed on the cladogram in as efficient (i.e., parsimonious) a manner as possible (i.e., do NOT repeat character states on multiple branches unless you have to). After you have finished mapping the character states onto the alternative cladograms in this part of the lab, you will be asked to compare the three cladograms to determine which is the best hypothesis in the next part of the lab.

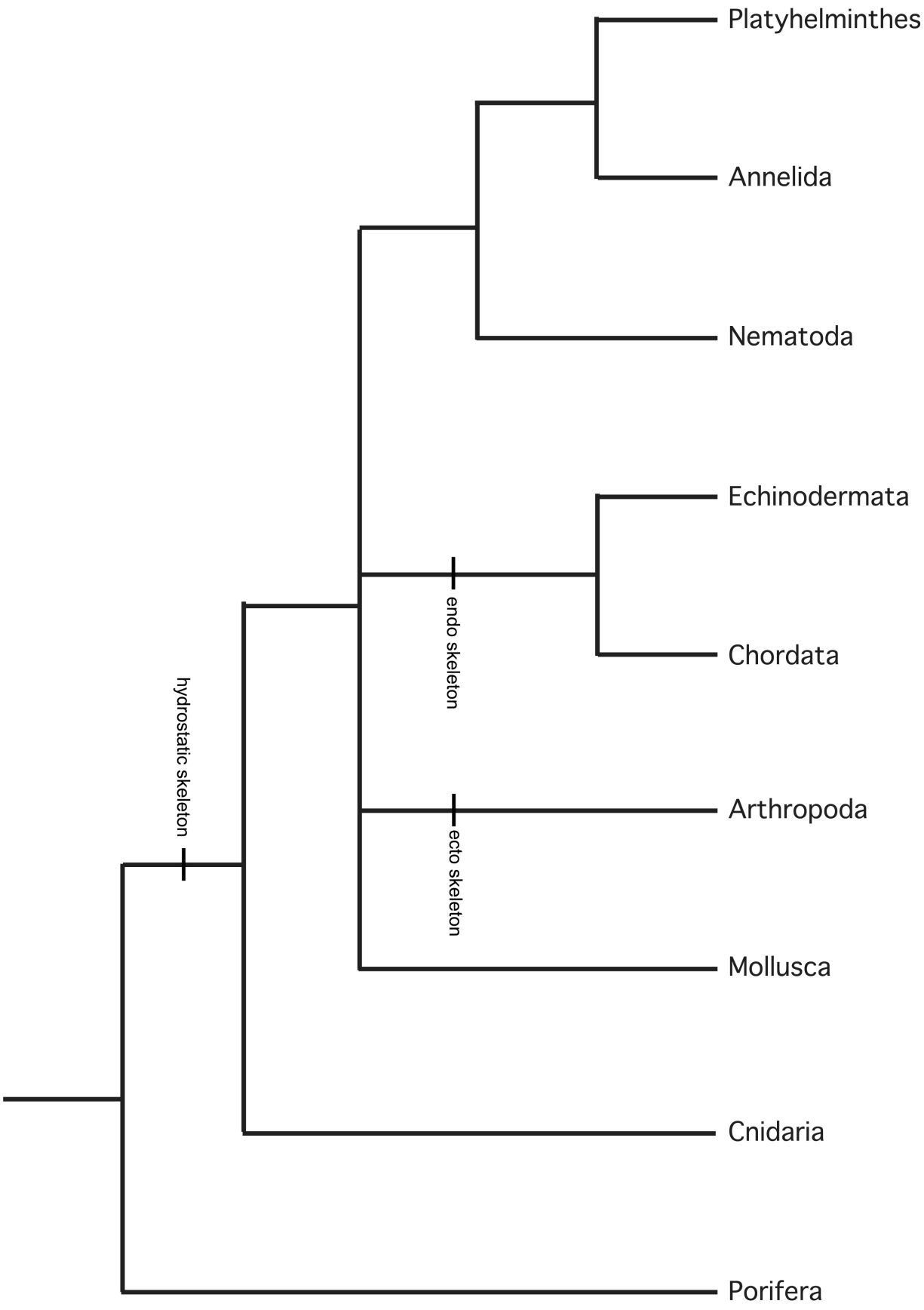
Hypothesis 1



Hypothesis 2



Hypothesis 3

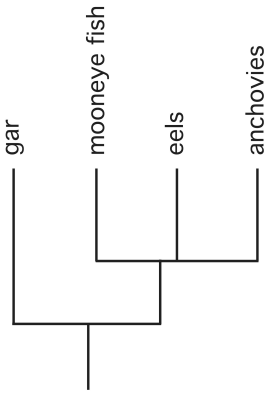


Part III: Which cladogram best represents the historical evolutionary relationships?

As noted earlier, (a) the three cladograms represent different hypotheses concerning the evolutionary history (phylogeny) of these taxa, and (b) these hypotheses are based on the same data (character states). A primary goal of evolutionary biology is to determine which hypothesis is best supported by existing data. Consider the following two criteria for evaluating the three hypotheses:

1. *The number of character states you had to place on the cladogram in order to accurately represent the character matrix.* Scientists usually prefer to propose and test simpler hypotheses over more complex ones, other factors being equal. This method is known as parsimony and is used in science generally, not just in phylogenetics. That is, the solution that requires the fewest number of ad hoc explanations, that minimizes the number of character states or "steps", is the one to choose according to the criterion of parsimony. Therefore, placing fewer character states rather than more on the cladogram is better because it represents a hypothesis that is easier to test (i.e., falsify). The cladogram at the top of page 4 shows seven character states (each character is represented once, except for "fins," which is depicted twice). (Note: An ad hoc explanation is one that applies to a particular situation only and is not applicable more generally.)

2. *How much of the topology is resolved.* When the topology is completely resolved, there are no more than two taxa stemming from any given MRCA. Sometimes, however, the data are not strong enough to support such a structure. In that case, the cladogram may contain one or more polytomies. A *polytomy* is a group of three or more taxa that stem from a particular MRCA. For example, in the cladogram shown to the right, mooneye fish, eels, and anchovies form a polytomy. These three taxa share a more recent common ancestor with each other than they do with gar, the other fish on the cladogram. However, there is insufficient evidence to know which two of these three taxa share a MRCA in respect to the third taxon. Resolved structures are preferred over polytomies because they describe a set of relationships that are fully testable. A cladogram containing one or more polytomies requires more data to resolve the polytomies into testable sets of relationships. In situations for which the data (character states) do not support complete resolution, it is best to minimize both the number of polytomies and their size (i.e., the number of taxa involved in the polytomy).



Evaluate the three cladograms (hypotheses of relationships) using these criteria. In the space below, explain which hypothesis concerning the evolutionary history of the nine animal taxa is best supported by the data, and why.

Part IV: Homology vs. Homoplasy

Homology refers to any similarity between taxa that is due to shared ancestry. In the example cladogram in Part II, the parathyroid gland is a homologous character. Homoplasy, on the other hand, refers to any similarity between taxa that is due to convergent evolution from separate (nonshared) ancestors.

The cladogram from Part III that represents the best hypothesis concerning the historical evolutionary relationships among nine major groups of animal taxa suggests that one or more character states from the data matrix are homoplasious (i.e., represent examples of convergent evolution). Which character state(s) are shared by taxa due to convergent evolution? What evidence from the cladogram supports this conclusion?

Part V: Using cladograms to make inferences

Cladograms are useful for a lot of reasons. One reason is that they are a powerful tool to support inferences. For example consider a new character on which the nine animal taxa differ—whether the digestive system has one opening or two. The character states for all taxa except Annelida are shown in the table to the right:

	Digestive system openings
Annelida	???
Arthropoda	2
Chordata	2
Cnideria	1
Echinodermata	2
Mollusca	2
Nematoda	2
Platyhelminthes	1
Porifera	–

Using just the information in the cladogram that best represents the evolutionary relationships among these nine taxa (i.e., without looking at your annelid specimen and without looking up the answer in your textbook), predict how many digestive system openings Annelida has. What evidence supports your inference?

Part VI: Questions for discussion

1. Consider the different characters on which you evaluated the nine groups of animal taxa, represented by your nine specimens.
    a. Most of the characters used in the matrix are morphological. One—the rDNA sequence—however, is molecular. Discuss the advantages and disadvantages of morphological and molecular characters.
    b. Do you consider the character "elongate cylindrical body" to be a useful informative character for phylogenetic analysis? Why or why not?
2. What are the implications of most recent common ancestry versus simply common ancestry when considering phylogenies?
3. How would we know if a character is a synapomorphy? What simple comparative method is used to determine this? Hint: It does not involve fossils.
4. How could epidemiologists use a phylogeny to combat emergent diseases or pathogens?

Acknowledgments

The development of this laboratory 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.

Phylogenetics Laboratory: Reconstructing Evolutionary History
Instructors' Guide and Answer Key

Kefyn M. Catley
Western Carolina University

Laura R. Novick
Vanderbilt University

Version 1 © February, 2010
Version 2.1 © August, 2010
Version 2.2 © August, 2012
Version 2.3 © June, 2016

Acknowledgments

The development of this laboratory 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.

Part I: Identifying character states.

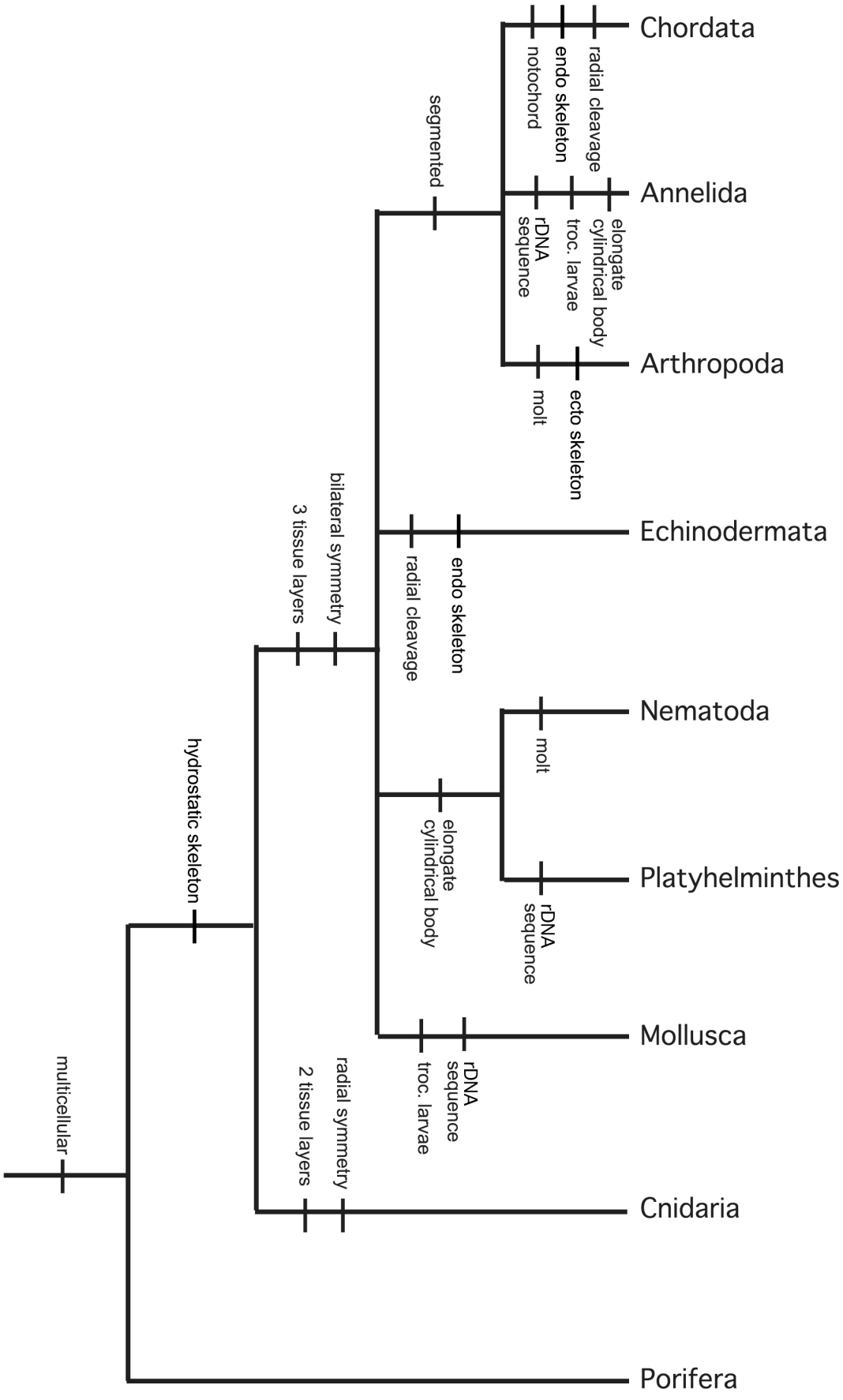
Here is the correct answer for the character matrix:

	rDNA sequence	Type of skeleton	Segmentation	Notochord	Trocohphore larvae	Elongate cylindrical body	Molt	Radial cleavage	Multicellular	Symmetry	Germ layers
Annelida	y	h	y	n	y	y	n	n	y	b	3
Arthropoda	n	ec	y	n	n	n	y	n	y	b	3
Chordata	n	en	y	y	n	n	n	y	y	b	3
Cnidaria	n	h	n	n	n	n	n	n	y	r	2
Echinodermata	n	en	n	n	n	n	n	y	y	b	3
Mollusca	y	h	n	n	y	n	n	n	y	b	3
Nematoda	n	h	n	n	n	y	y	n	y	b	3
Platyhelminthes	y	h	n	n	n	y	n	n	y	b	3
Porifera	n	-	-	n	n	n	n	n	y	-	-

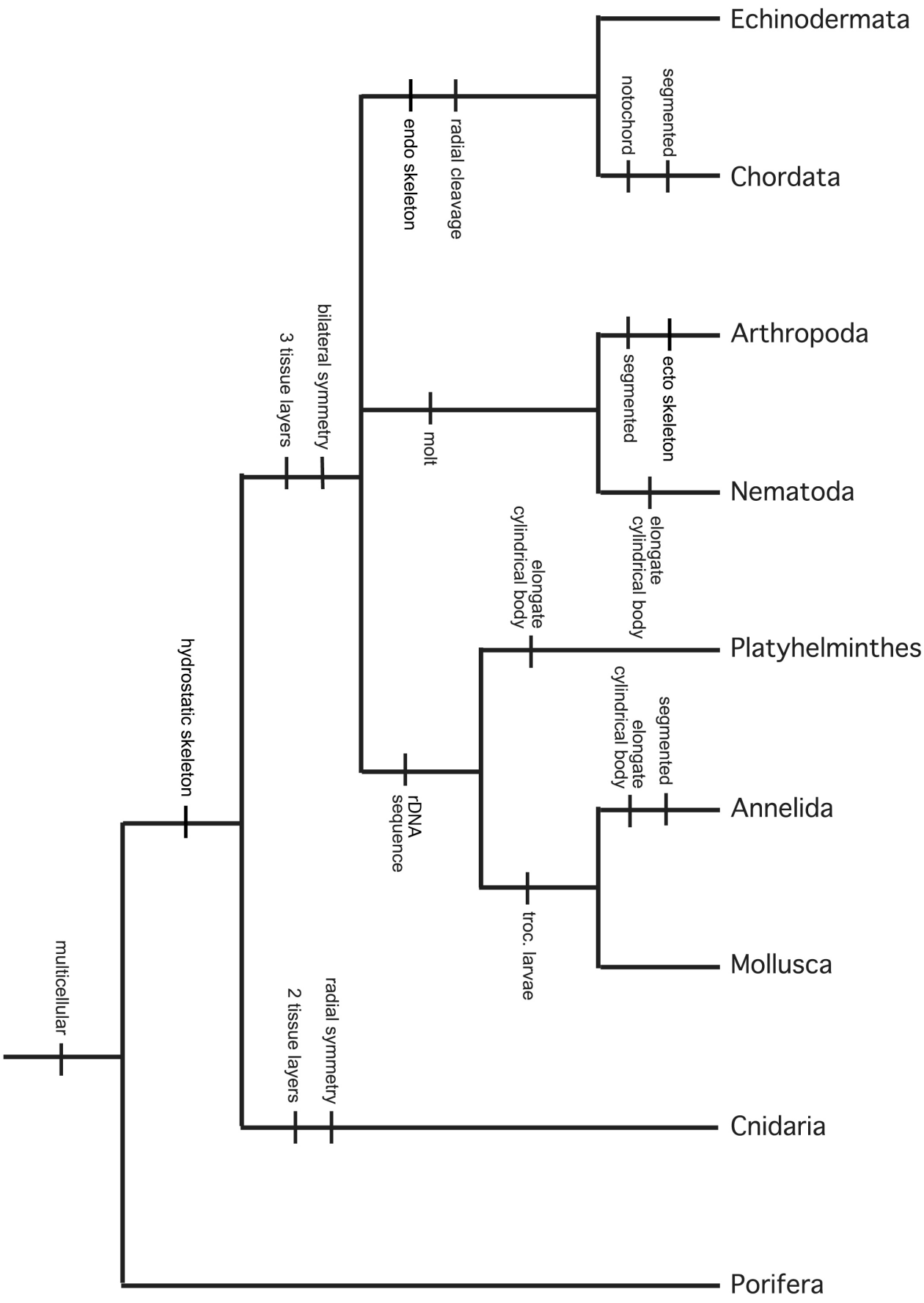
Part II: Mapping character states onto alternative cladogram topologies.

The next three pages show the three cladograms (hypotheses) with the characters correctly mapped onto the cladograms. Note that when multiple characters appear on a particular branch, they are in no particular order because no evidence was provided that pertains to ascertaining the order in which the characters arose. In students' responses, any ordering of the correct characters on a branch is acceptable.

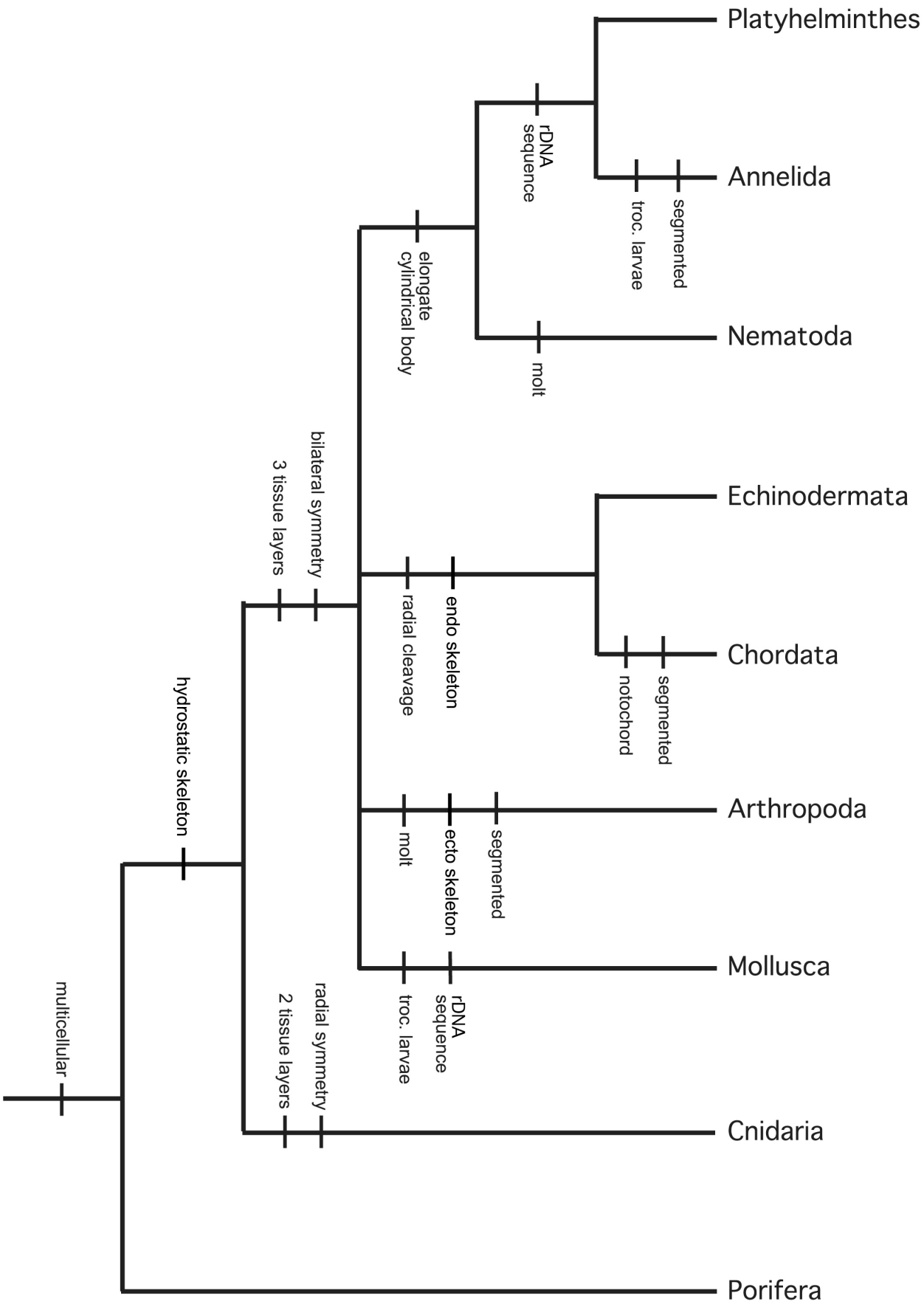
Hypothesis 1



Hypothesis 2



Hypothesis 3



Part III: Which cladogram best represents the historical evolutionary relationships?

Hypothesis 2 is best supported by the data because:

(1) It is supported by the fewest number of characters (states represented on the cladogram). Thus, it is the most parsimonious solution. Specifically, it has 19 character states, compared with 22 for hypothesis 1 and 20 for hypothesis 3.

(2) It is also the most resolved and therefore the easiest to test. This is an application of Occam's razor: among competing hypotheses, the one with the fewest assumptions (or steps on the tree in this case) should be selected. By minimizing the number of ad hoc explanations that are required to explain the character data, the hypothesis can be more easily tested (or falsified) than one that is supported by a larger number of potential explanations. Hypothesis 2 includes only one polytomy with three branches. In contrast, hypothesis 1 has two polytomies, one with three branches and one with four; and hypothesis 3 has one polytomy with four branches.

Part IV: Homology vs. Homoplasy

Hypothesis 2 suggests that there are two homoplasious character states for the nine taxa:

(1) Segmentation appears to have evolved independently in Chordata, Arthropoda, and Annelida. This character is most parsimoniously mapped three times on the cladogram, once for each taxon.

(2) Elongate cylindrical body appears to have evolved independently in Nematoda, Platyhelminthes, and Annelida. This character also is most parsimoniously mapped three times on the cladogram, once for each taxon.

Part V: Using cladograms to make inferences

The most strongly-supported inference is that Annelida has two digestive openings. The most parsimonious mapping of the character states for this character onto the cladogram representing Hypothesis 2 is one opening for Cnidaria and for Platyhelminthes (convergent evolution) and two openings as a synapomorphy for the clade comprising all taxa except Cnidaria and Porifera.

Part VI: Questions for discussion

1. Consider the different characters on which you evaluated the nine groups of animal taxa, represented by your nine specimens.

    a. Most of the characters used in the matrix are morphological. One—the rDNA sequence—however, is molecular. Discuss the advantages and disadvantages of morphological and molecular characters.

    Encourage discussion that considers the following: the effect of variation in molecular and morphological data, the costs involved collecting each type of data, the possibility that a morphological character may reflect a large number and variety of different genes, the fact that fossils typically provide no molecular data, issues of determining homology in molecular and morphological data, the role of convergent evolution, rates of evolution. And, come up with your own!

b. Do you consider the character "elongate cylindrical body" to be a useful informative character for phylogenetic analysis? Why or why not?

   This type of character is almost impossible to classify into discrete character states. The condition "elongate cylindrical body" is highly variable (how would you decide where to draw the boundaries between character states?) and not likely to have a discrete genetic basis. It is likely to involve a large number of genes making it a potentially messy and unreliable character.

2. What are the implications of most recent common ancestry versus simply common ancestry when considering phylogenies?

   Detecting the most recent common ancestor based on a synapomorphy allows us to reconstruct the tree of life in small, nested subsets of three taxon statements—i.e., that A and B are more closely related to each other than either is to C. Therefore, most recent common ancestry is very informative, whereas common ancestry is not because it does not allow us to create these hierarchical subsets

3. How would we know if a character is a synapomorphy? What simple comparative method is used to determine this? Hint: It does not involve fossils.

   Often students think firstly of using fossils to determine the polarity of characters—i.e., which is plesiomorphic (ancestral) versus synapomorphic (derived). The most widely used method, however, is that of outgroup comparison. If a closely-related taxon (i.e., not a member of the ingroup but closely related enough that meaningful comparisons can be made) shares the same character as the taxa in the in-group, then that character is more likely to be plesiomorphic (ancestral and more widely distributed). However, should the outgroup not share the character, then it is more likely to be a synapomorphy (i.e., derived and more restricted in distribution). For example, when considering the polarity of a character in placental mammals, marsupial mammals would make a good choice of an outgroup whereas reptiles would not. This comparison tells us that although milk is not a synapomorphy for the ingroup, having a placenta is.

4. How could epidemiologists use a phylogeny to combat emergent diseases or pathogens?

   There are many examples for which knowing the sister group of a pathogen, or at least which clade it is in, has led to successful predictions about the spread of a disease or novel treatment options based on most recent common ancestry (e.g., West Nile virus, Severe Acute Respiratory Syndrome (SARS), and various influenza strains).