

Exploring Systematics and Phylogenetic Reconstruction Using Biological Models

Hans Lemke and Jeffrey Jensen

Department of Biology, University of Maryland, College Park MD 20742 USA

(hlemke@umd.edu; jensen@umd.edu)

Systematics and phylogenetic reconstructions are among the most difficult concepts for students to understand in introductory biology. They are also fundamental to understanding how biologists view the evolutionary history of the world around us. This exercise uses models of Caminalcules to teach basic tree construction and interpretation. Students first construct a tree of by hand, and then they then use the computer program Mesquite to construct and analyze a more complex tree.

Keywords: Phylogenetics, Systematics, Tree Thinking, Caminalcules, Mesquite, Teaching Models, Mold Making

Introduction

Tree thinking is one of the most fundamental concepts needed to understand evolutionary theory. This lab exercise uses models to provide students with a basic knowledge and understanding of systematics and phylogenetic reconstruction. It is designed for use in an introductory biology course for majors. For most first-year students, this will be their initial exposure to these techniques and ideas. In our course, the lecture and lab are tightly integrated and the basics of phylogenetic reconstruction are often introduced in lab. The goals of the exercise are to teach what phylogenies are, why it is important to construct and study them, the basics of out-group analysis and parsimony, and the steps involved in the

construction of a tree. With that in mind, the exercise takes the students step-by-step through the processes of creating and interpreting trees. The students use 3D models of five species of Caminalcules (imaginary organisms) to first build a tree by hand. They then use the computer program Mesquite to build a more complex tree with 10 species. For both of these exercises, there are questions to help students interpret their trees. This exercise can be easily adapted to other organisms. We have found that it works best when there are actual evolutionary relationships among the specimens, and have not had good luck using items such as paper clips, hardware or sports balls.

Student Outline

Objectives

- To understand the theoretical basis of phylogeny reconstruction, and to gain an understanding of how phylogenetic trees are constructed in practice
- To gain an appreciation of the value of phylogenetic reconstruction for understanding evolutionary patterns and processes
- To construct your own phylogenetic hypotheses

Perhaps the most important implication of the theory of evolution is that all species of organisms, including humans, are related to each other in a genealogical sense. Even organisms as distinct as fungi, humans, and bacteria have ancestors in common. Some species, such as chimps and humans, share a fairly recent common ancestor (less than 10 million years ago), while for others you have to go way back to find a common ancestor (e.g., the common ancestor of humans and jellyfish lived at least 600 million years ago). A very useful way of thinking about relationships among organisms is as a tree of life. Imagine that at the base of the tree is the original population of organisms. As they evolve through time, two things can happen: 1) This lineage of organisms can acquire new characteristics, and 2) the lineage of organisms can split into two separate lineages (speciation). These processes, repeated over and over, result in a “tree of life” that represents the relationship among organisms (Fig. 1). This tree of life is also referred to as a **phylogeny** or **phylogenetic tree**.¹

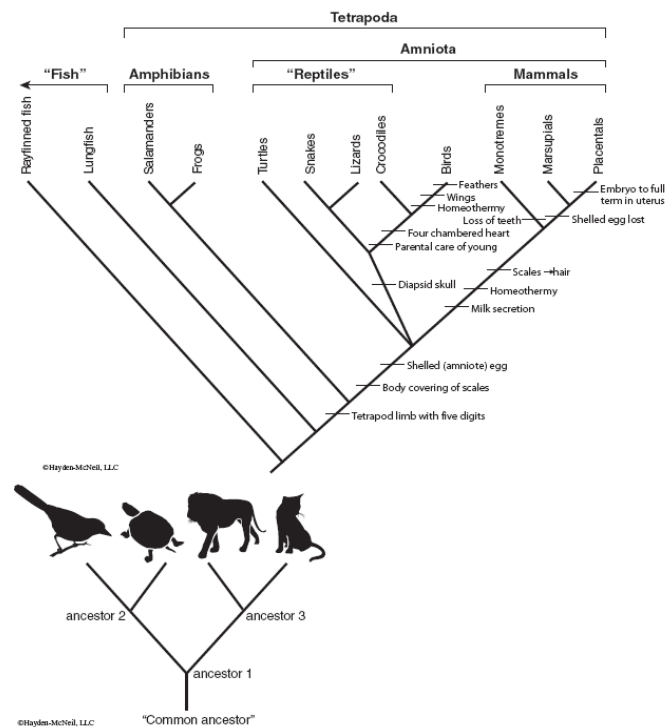


Figure 1. A simplified cladogram (phylogenetic tree) depicting the relationships of most living vertebrates

Introduction

The Tree of Life

Today we will investigate the theory and practice of reconstructing the phylogeny of organisms. As you can imagine, reconstructing events that occurred millions, or even billions, of years ago is not an easy task. Fortunately, the last 40 years have produced astounding theoretical and technical breakthroughs in phylogenetic reconstruction. It now plays a central role in all of comparative and historical biology, and advances in systematic knowledge and theory have led to tremendous progress in understanding the patterns and processes of evolution.

¹ The situation isn't always this simple. In some cases genes are shared across species (e.g., bacteria), or hybridization is common (e.g., oaks), and sometimes even very distantly related organisms merge to form a composite (e.g., eukaryotes). Each of these cases would cause the branches to rejoin. Fortunately, this doesn't appear to be a major problem for higher organisms.

Phylogenies—Why Are They Useful?

Before we discuss how phylogenies are reconstructed, it's worth asking the always relevant question “why bother?”

- (1) **Creating a non-arbitrary classification**—Systematics has traditionally been concerned with the identification and naming of species, and the grouping of these species into a classification. Carolinus Linnaeus (1707–1778), an early and very influential taxonomist, developed the hierarchical classification system used today in which species are arranged into ever more inclusive groups. The complete Linnean classification for humans is:

Table 1. Linnean classification for humans.

Taxonomic Rank	Taxon Name	Includes
Kingdom	Metazoa	All animals
Phylum	Chordata	Animals with dorsal, hollow nerve cords (among other characteristics)
Class	Mammalia	Chordates with lactation etc.
Order	Primates	
Family	Hominidae	Less human-like primates, including a number of if extinct species such as <i>Australopithecus afarensis</i>
Genus	<i>Homo</i>	Very human-like primates, including ourselves and several extinct species such as <i>Homo erectus</i>
Species	<i>sapiens</i>	“Wise humans” – our humble name for ourselves

While this seems very straightforward, confusion can arise about which groups of species to put together. For example, should we put insects, birds and bats in the same phylum because they all have wings (despite the fact that they are only very distantly related and their wings clearly have separate origins)? Should all warm-blooded vertebrates be placed in one class, even though birds and mammals clearly evolved warm-bloodedness independently? Does it make sense to lump the great diversity of remaining vertebrates (e.g., sharks, frogs, turtles, alligators) in another class because they are all cold-blooded? To avoid this confusion, and the tedious arguments that ensue, modern classifications tend to focus entirely on recency of shared ancestry in forming groups. Phrased another way, each taxon in a classification represents an entire branch of the phylogenetic tree. For example, birds are classified with crocodiles in the group Archosauria because they are each other's closest living relatives (despite the fact that birds are warm-blooded, feathered, and flying while crocodiles are cold-blooded, scaly, and walk on all four limbs). In order to make a classification that reflects recency of shared ancestry, we must have a way of reconstructing phylogeny.

- (2) **Understanding the process of evolution**—Knowledge of how organisms are related to each other is of tremendous value to evolutionary biologists for reasons too numerous to elaborate on here. Suffice it to say that armed with knowledge of the “pattern” of life's diversification, we can begin posing and attempting to answer questions about the “process” of evolution, i.e., the mechanisms driving these patterns of diversification. For example, “Why are birds so numerous and diverse, compared to their close relatives the crocodiles?” The use of phylogenetic trees to generate and test hypotheses about the evolutionary process has been called “tree thinking,” and has been very useful and influential in recent years. “Tree thinking is not a minor addition to the toolbox of evolutionary theory; it is a commitment, an abiding concern, to understand organisms through their history... It will change the way we do evolutionary biology” (Maddison and Maddison, 1992).
- (3) **Human Health: Identification of Pharmaceuticals and/or Disease Vectors**—As herbalists and native cultures have long known, and pharmaceutical companies are quick to recognize, the world's plants and animals provide a vast resource of valuable medicines. Conversely, many organisms are disease vectors and can be exceedingly destructive. Knowledge of the phylogeny of these organisms can be very useful in protecting human health. If a pharmaceutical company finds a medicinally valuable plant species, they would be wise to examine close relatives of that plant to determine if those plants can be obtained at lower cost, or can deliver the same beneficial compounds with fewer side effects. A prudent health worker, upon finding that a particular insect species carries a disease, would be well advised to determine whether or not that species' close relatives can also carry the disease before designing a control strategy.

Phylogenetic Reconstruction—Some Important Terms

Documenting Variation—What Is a Character?

When you compare any set of organisms, similarity relationships are obvious. Humans and great apes are clearly more similar to each other than either is to a goldfish; a lizard certainly appears to be more similar to a crocodile than either does to a hummingbird. This may sound very simple, but much of the progress in the understanding of evolution has depended on a successively deeper understanding of the concept of similarity. To study similarity between groups, systematists must first clearly define and tabulate the attributes of their study organisms. The term **character** is used to refer to any property of an object that can be quantified, such as its color, its size, the number of parts, etc. Examples of biological characters are particular DNA sequences, the number of petals in a flower, the type of skin covering (scales, feathers, hairs, or none), the length of the skull of a mammal, the type of tracheid in the wood of an angiosperm, etc. Characters are often said to show different **states**, thus “eye color” might be a character, with the states blue, brown, and green.

Organisms can be similar for a variety of different reasons. For our purposes, it is essential to distinguish among the following three types of similarity: **convergent** similarity, **ancestral** similarity, and **derived** similarity.

- **Convergent similarity** is similarity due to the independent appearance of a particular character in two different evolutionary lineages. Thus birds and bats are similar in both having the character “wings,” but this similarity (the joint possession of the character “wings”) is a convergent similarity because it was acquired independently in the branches of the evolutionary tree leading to bats, on the one hand, and to birds, on the other. Such characters are said to be **analogous**. Groups that are united by independently evolved (i.e. convergent) characters are called **polyphyletic** (Greek, meaning many races). Refer to Figure 2.
- **Derived similarity** (also called “synapomorphy”—syn = shared, apo = derived) is similarity due to the appearance of a new character or character state in some particular species, which is then transmitted to the evolutionary descendants of that species. The ancestor of all living birds developed feathers, and feathers persist in all the varied descendants of that ancestral species. Feathers are thus a derived similarity of birds. Groups that contain an ancestor and all of its descendant species are called **monophyletic** (one race).
- **Ancestral similarity** is the most conceptually difficult of the three types of similarity. Consider the vertebrate group known as the Amniota, all of the truly terrestrial vertebrates which are not dependent on water for reproduction (unlike fish and amphibians). The Amniota contains the animals commonly known as reptiles, birds, and mammals. Birds all share feathers as a body covering, and this is a shared *derived* character as described above. Mammals share hair as their body covering, and hair is also a shared *derived* character, one which unites all mammals. But what about the character “scales,” which all reptiles share? Since we are assuming knowledge of the true evolutionary tree in these examples, we know that both birds and mammals evolved from different lineages of reptiles, and so birds’ feathers are actually modified scales, and mammalian hair is also a scale derivative. Groups such as reptiles, which are based on ancestral similarity (in this case the absence of the derived features of birds and mammals), are called **paraphyletic**.

The shared similarity “scales,” found in all reptiles, is thus an ancestral similarity, which has been further modified in two branches of the amniote evolutionary tree, the bird branch and the mammal branch. The main point is that the shared possession of ancestral characters, such as scales, is not an indication of close phylogenetic relationship. We can see from the phylogeny above that crocodiles share a more recent common ancestor with birds than with lizards, even though both lizards and crocodiles have scaly bodies. The most important take-home lesson from this lab is that *only derived similarities identify branches of the evolutionary tree*.

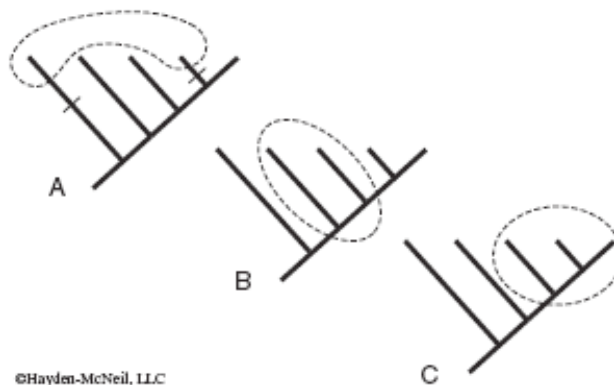


Figure 2. Polyphyletic (A), paraphyletic (B), and monophyletic (C) groups

Outgroup Analysis and Parsimony

We have defined a set of terms, and now understand the importance of shared derived characters in the construction of evolutionary trees, but clearly a practical problem remains: how does one tell whether a given similarity shared by two or more organisms is derived, and so characterizes a branch of a cladogram, or is convergent or ancestral, and so does not? Two principles are used to distinguish derived from ancestral and convergent similarities—**Outgroup Analysis** and **Parsimony**.

Shared derived characters are distinguished from shared ancestral characters by the principle of **outgroup** comparison. An outgroup is a group of organisms thought to be closely related to the group under study, but not part of it. The outgroup may be thought of as “the next branch out” on the evolutionary tree or cladogram, thus Crocodilia could serve as an outgroup for the study of phylogenetic relationships within the birds, or the amphibians could serve as an outgroup for investigating the phylogenetic relationships within the Amniota.

The principle of outgroup comparison may be stated as follows: *when two states of a character occur in a monophyletic study group the state which also occurs in the outgroup is the ancestral state*. As an example, within birds there are two different types of sperm morphology: one type has a simple, elongate head, while the other type has a very long and highly coiled head. All the 9,000 or so species of birds fall into two groups according to the type of sperm morphology they have. If we wish to tell which type is derived we may look to the reptiles as an outgroup. Reptilian sperm is of the simple type, so we conclude that the coiled type must be derived (with the simple type being ancestral), and that the birds with coiled sperm form a monophyletic group, which is the group known as the Passeriformes (“perching birds”).

Derived similarity is distinguished from convergent similarity by (among other things) the principle of parsimony. The principle of **parsimony** is just a version of the scientific and philosophical principle known as Occam’s Razor (after the Medieval philosopher William of Occam), and says that the simpler of two competing explanations of the same data should be preferred to the more complex. As an example, suppose we were trying to reconstruct the phylogeny of the Amniota. We find two groups of amniotes which have wings, the birds and the bats, and we have done outgroup comparison to amphibians so we know that wings is not an ancestral amniote character. The question then is: does birds + bats constitute a monophyletic group defined by the derived character “wings” or were wings independently (convergently) derived in these two lineages? We find, upon further morphological investigation, that birds and bats do indeed share wings, but that bats and all other mammals share hair, live birth, mammary glands, and a great many other derived characters. The **parsimony principle** says that we should accept the hypothesis that bats + other mammals is a monophyletic group, and reject the hypothesis that bats + birds is a monophyletic group, because it is simpler to assume that the single character “wings” evolved independently in birds and bats than it is to assume that hair, and live birth, and mammary glands, *and* all the other derived characters which bats and other mammals share all arose independently. The parsimony principle does *not* imply that evolution necessarily operates parsimoniously or “optimally” or “efficiently.” It is just a logical principle for choosing among different hypotheses which attempt to explain the same data. In a large study with many taxa and characters it is often very difficult to find the most parsimonious tree (the tree with the least number of character state changes going from the ancestral root to the tips). In such cases, computerized algorithms are often used to search among the possible trees to find the shortest. The development of such methods is a very active area of systematic research.

A Few Questions to Apply These Concepts

Most mammals give birth to live young, but a small number of mammals (the Prototheria) lay eggs. Answer the following set of questions, referring to Figure 1 if you need to, to make sure you are comfortable applying these concepts.

Is egg-laying a derived or primitive feature among mammals? How do you know?

Is the shared trait of egg-laying, by itself, an indication that egg-laying mammals are more closely related to each other than either is to other mammals?

For what group of amniotes is live birth a shared derived characteristic?

For what group of amniotes is shelled eggs a shared derived characteristic?

Reconstructing Phylogeny (In Three Steps)

1. Specifying the outgroup

As mentioned above, an outgroup allows you to determine which character states are primitive for your group of organisms, and which are derived (and therefore useful indicators of relationship). Suppose you are trying to reconstruct the phylogeny of a group of organisms, some of which have hair and some of which don't. If your outgroup has hair, you would conclude that baldness is a derived characteristic and an indicator of close relationship. Does that make sense?

2. Generating a data matrix

You will be using variation among organisms to generate a hypothesis of how they are related. As described above, you will need to choose characters and determine which species exhibit which character states. For example, you may have the character "Digit Number" and the character states "Three," "Four," and "Five." Tabulating this type of information for all your characters would give you a data matrix like the one shown below:

Table 2. Data Matrix

Taxon/Character	Digit Number	Skull Length	Teeth	Etc.
Outgroup	3	Short	Present	
Species 1	4	Long	Present	
Species 2	3	Long	Present	
Species 3	5	Long	Absent	
Species 4	5	Long	Absent	
etc.	etc.	etc.	etc.	

As an example, we will examine the example of egg-laying mammals. Egg-laying in mammals is a primitive feature (shared with many other vertebrates such as birds, "reptiles" and the ancestors of mammals that give live birth). It is not a feature that indicates modern egg-laying mammals have a unique ancestor in which it evolved.

Character	Outgroup (Shark)	Trout	Platypus	Porpoise	Human
Makes Milk	0	0	1	1	1

This distribution of characters suggests that platypuses, porpoises and humans are more closely related to each other than any are to trout. Why? Because they share a feature that must have evolved fairly recently (i.e. it is not in the outgroup).

No single character is going to give you all the information you need to reconstruct a phylogeny, and some characters may actually give contradictory information. In phylogenetic reconstruction we attempt to find the tree that best explains the data – *i.e.* requires the fewest independent evolutions of shared features. In the example above we concluded based on milk production that trout and humans were not each other's closest relatives. It's possible that they are closely related and that platypuses, porpoises and humans independently evolved milk production, but in the absence of any evidence to the contrary the simplest interpretation is that platypuses, porpoises and humans inherited milk production from a common ancestor not shared with trout (review the idea of parsimony). Let's expand the table above to illustrate this reasoning:

Table 3. Data matrix for vertebrates

Taxon/Character	Live Birth	Makes Milk	Has Lungs	Streamlined	Has Nipples
Trout (outgroup)	0	0	0	0	0
Lizard	0	0	1	1	0
Platypus	0	1	1	1	0
Porpoise	1	1	1	0	1
Human	1	1	1	1	1

3. Computing the shortest tree(s)

You will propose a hypothetical phylogenetic tree, called a **cladogram**, based on the set of characters you have chosen. For reasons discussed above, the shortest cladogram (i.e., the one requiring the fewest steps or evolutionary changes) is usually the preferred estimate of the true phylogeny. As you might imagine, generating a tree that most parsimoniously explains the evolution of a large number of characters among many species is a computationally daunting task. For a group of ten species there are millions of possible trees. For each of these possible trees, the number of changes in each character must be summed up and the total compared to all the other trees.

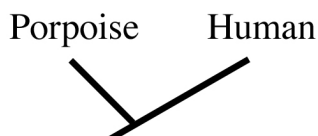
Now that you have created a data matrix for morphological variation you are ready to use this information to create a phylogenetic tree. Remember that primitive similarity is not an indication of close relationship, and we can distinguish primitive similarity from derived similarity using Outgroup Analysis. In outgroup analysis, the trait found in the outgroup is considered to be the state found in the common ancestor of the ingroup. Two species that share a primitive similarity probably inherited it from this common ancestor (shared with all the other members of the ingroup) rather than because they are most closely related to each other.

We can construct a phylogeny by hand if we cluster together the species that have the greatest number of derived features (indicated by a 1 in the above table). Let's make a matrix of number of shared derived features for our ingroup.

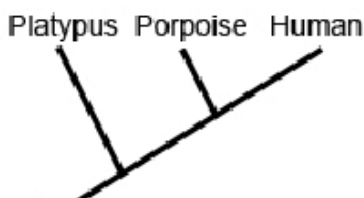
Table 4. Matrix of shared derived features of vertebrates

	Trout	Lizard	Platypus	Porpoise	Human
Trout		0	0	0	0
Lizard			2	1	2
Platypus				2	3
Porpoise					4
Human					

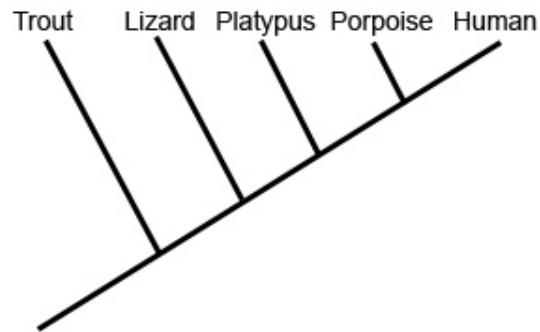
If we link the two species that have the greatest number of derived features (Humans and Porpoises share four derived traits), we get:



Now, find the species with the next greatest number of derived features (in this case Humans and Platypuses share three derived traits). If neither taxon belongs to a previously constructed group, place the new pair as nearest relatives. If either member or both are part of a previous group, join the two groups or the previously unplaced taxon within the group. In this case, Humans are already in a previous group, so join Platypus to that group:



Lizards share two derived traits with both humans and platypus so we add them next. The outgroup is the last one added:



Exercise 1: Phylogeny of Caminalcules

As the first biologist to arrive at the recently discovered island of Camin, you are excited to investigate its biological wonders. Arriving at an inland valley, you are delighted to find delightful animals cavorting along the valley floor (Fig. 3). You quickly make a collection of species A, B, C, and D. Fascinated by these small animals, called Caminalcules², you decide to investigate their evolution. You begin, quite reasonably, by recalling that in order to reconstruct their phylogeny, you need to undertake the following steps:

1. *Select an outgroup.*
2. *Generate a data matrix.*
3. *Find the shortest tree.*

With that in mind, we can work through the steps to generate a tree:

1. Selection of an outgroup

A Caminalcule found on an adjacent island is thought to be the most primitive species, and will serve as your **outgroup**. This Caminalcule is labeled **OG**.

2. Generation of a data matrix

Examine the Caminalcules carefully. What characters vary among these species? What are the character states? For this exercise, the characters and states are provided in Table I. You will find it helpful to refer to Figure 4 when referring to anatomical features of the Caminalcules. Use Table II below to compile your data. For convenience, code the character states with numbers and assign the number 0 to the state found in the outgroup and 1 for the different state found in the ingroup. If there are multiple derived states in the ingroup you can give them different numbers (e.g. 1 vs. 2 vs. 3 etc.)

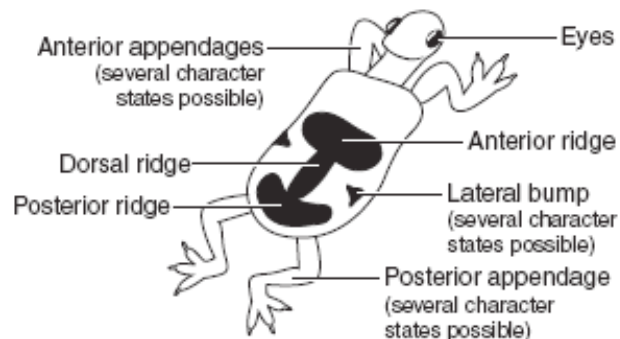


Figure 3. Anatomy of a representative Caminalcule

²H.H Camin developed a set of hypothetical organisms with a known phylogeny called Caminalcules. These organisms have been used to test various methods of phylogenetic reconstruction. The organisms in this lab are modified from that original set.

Table 5: Caminalcule characters and states

Character	State 0	State 1
Posterior Ridge	Absent	Present
Posterior Appendages	Absent	Present
Eyes	Present	Absent
Dorsal Ridge	Absent	Present
Lateral Bump	Circular	Triangular

Table 6. Data matrix for Caminalcules

Taxon/Char	<i>PosRid</i>	<i>PosApp</i>	<i>Eyes</i>	<i>DorRid</i>	<i>LatBmp</i>
Outgroup					
A					
B					
C					
D					

For the Caminalcules, we have chosen an outgroup for you, Species OG, to represent the ancestral set of traits for your in-group (species A, B, C, and D). For all the traits you investigate it will have “State 0”. The other species will have State 0 for some traits (indicating that they retained the primitive trait found in the outgroup) or state 1 (a derived trait not found in the outgroup).

3. Find the shortest tree

Construct a matrix of shared derived features for the Caminalcule data set (from Table II): Remember that this matrix is based on shared derived traits so you need to enter the total number of traits in which the two species share derived traits (i.e. they are coded with a 1).

Table 7. Matrix of shared derived features of Caminalcules

	OG	A	B	C	D
OG					
A					
B					
C					
D					

Creating a cladogram

Now cluster your Caminalcule species into a tree using the matrix above.

1. Start by finding the species that have the greatest number of shared derived features and join them. This amounts to making the hypothesis that they are each others’ closest relatives. Look back at the mammal example if you need to.
2. Then find the species with the next greatest number of shared derived features
 - a. If neither taxon belongs to a previously constructed group, place the new pair as nearest relatives
 - b. If either member or both are part of a previous group, join the two groups or the previously unplaced taxon within the group.
3. Continue in this way until all of the taxa have been placed.
4. The last group that you should add is the outgroup.

5. Draw out the whole cladogram. Create your tree in the space below.

Interpreting the cladogram.

Congratulations! You have constructed a cladogram of five species of Caminalcules. Now you can use the tree to answer some of the burning evolutionary questions that you have about these fascinating creatures.

Although cladograms seem like simple diagrams, a great number of hypotheses are embedded within them. For each species, the cladogram represents a hypothesis about how it is related to all of the other species. For each character, the cladogram represents a hypothesis for where (on the tree) it arose, and what happened later (e.g. did it undergo an evolutionary reversal (i.e. revert to the primitive state) or did the same trait evolve elsewhere on the tree?).

For each of the characters, diagram where it arose on your tree and indicate any instances of reversal or convergence. The easiest way to do this is to draw a short line across the tree where the trait arose and add a label (see Fig. 1 for examples).

How many total evolutionary changes (i.e. origins or reversals/losses) occur on your tree? _____

Can you see any ways to rearrange the branches to require fewer changes?

What do the “nodes” (i.e. the intersections of lines on the tree) represent? _____

The simplest way a trait could evolve would be to evolve once and be retained in all descendants – this requires only a single change (from the primitive to the derived state). Provide one example from your tree: _____

Exercise 2: Further Exploration of the Caminalcules

Upon further exploration of the island, you discover four additional species of Caminalcules. Since you are now an experienced taxonomist, you decide to add these new species to your phylogenetic analysis. As you might imagine, it is significantly harder to construct a tree with nine species than with five. Fortunately, you remembered to pack your laptop on the expedition so you can use software to help build your trees. But first you will need to reexamine the characters and states that you used previously. A quick glance at the new species shows that they have a wider variety of states than was accounted for.

Your first task is to define a new set of characters and character states. You may find it helpful to refer back to Figure 4 for more information about Caminalcule anatomy. Another thing to keep in mind is that characters can have more than 2 different states. When this occurs, pick one of the derived states to be State 1 and the other is State 2. Keep in mind that assigning a 1 or 2 does not imply judgment on your part about evolutionary history; they are simply codes to allow for easier data entry. Remember that 0 is whatever state is displayed in the outgroup.

Lateral bumps are a great example and this character is filled in on the table to get you started. Work with your group to create a new set of Characters and States and complete Table 8.

Table 8: Characters and states for Caminalcules (remember that you don't always need to use State 2)

Character	State 0	State 1	State 2
Lateral Bump	Circular	Rectangular	Triangular

Now that you have a new set of characters, use them to fill in Table 9 below:

Table 9: Data matrix for Caminalcules

Taxon/Char	<i>LatBmp</i>							
Outgroup								
A								
B								
C								
D								
E								
F								
G								
H								

Now that you have assigned character states to all of your Caminalcules, it is time to find the shortest tree. You will use a computer program to aid in tree construction. **Mesquite** is an interactive program that allows you to move branches around the tree, and automatically calculates the tree length (i.e. the total number of evolutionary changes required to explain the character distribution given that phylogeny). Mesquite's great strength is that it allows you to trace the evolution of individual characters on a tree. This is extremely valuable for evolutionary biologists interested in the evolution of particular characters. Although it is not designed to search among all the thousands or millions of possible alternatives for the shortest possible tree, it is sufficient for our purposes. Follow the directions below to enter, analyze and interpret your Caminalcule data using Mesquite.

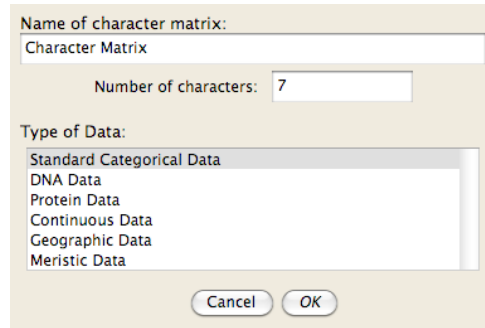
Data Entry with Mesquite

Click on Mesquite icon on the dock

In the top menu bar select File -> New

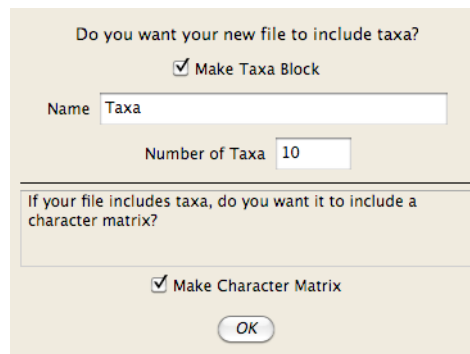
Enter a file name in the following format: _____?

A new window will open – Enter 10 as the number of taxa, and click the “Make Character Matrix” box as indicated in the following figure:



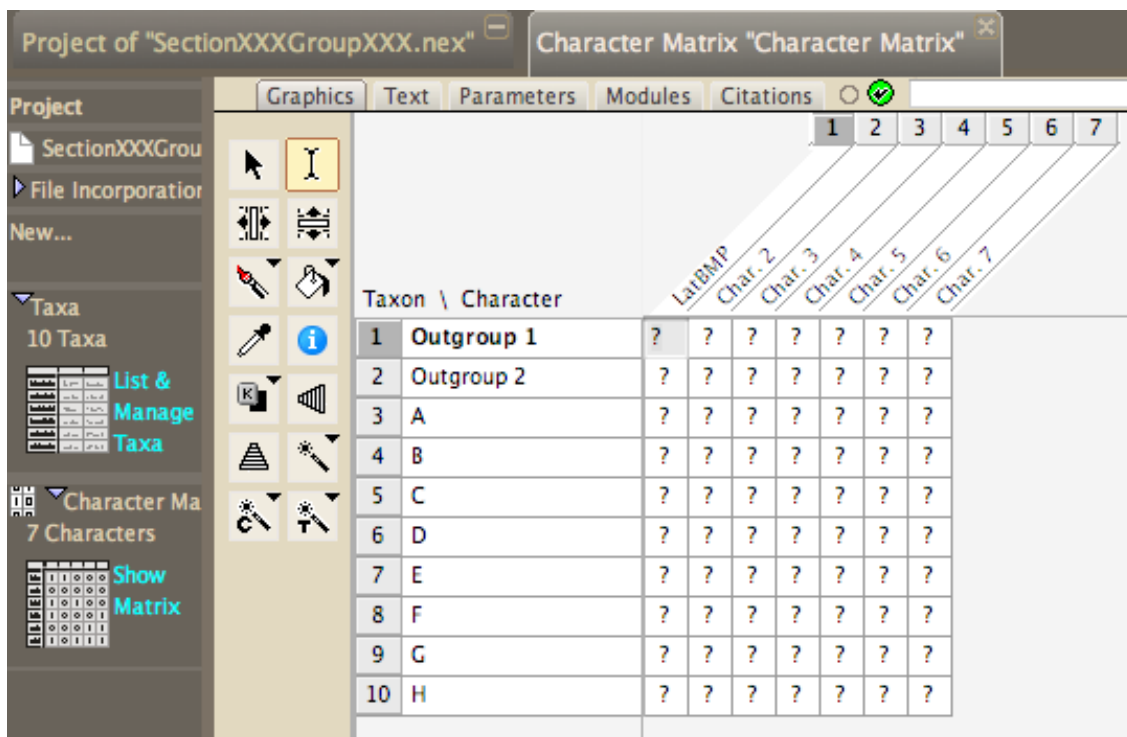
Click OK.

In the new window that appears, give a name to your character matrix, indicate the number of characters you have (e.g. 7), and select “Standard Categorical Data” as indicated below



Click OK.

A new window will appear showing your empty data matrix. Enter the names of your taxa, the names of the characters, and the characters states. A partly completed matrix (without the characters completed, and without the character states) is shown below.



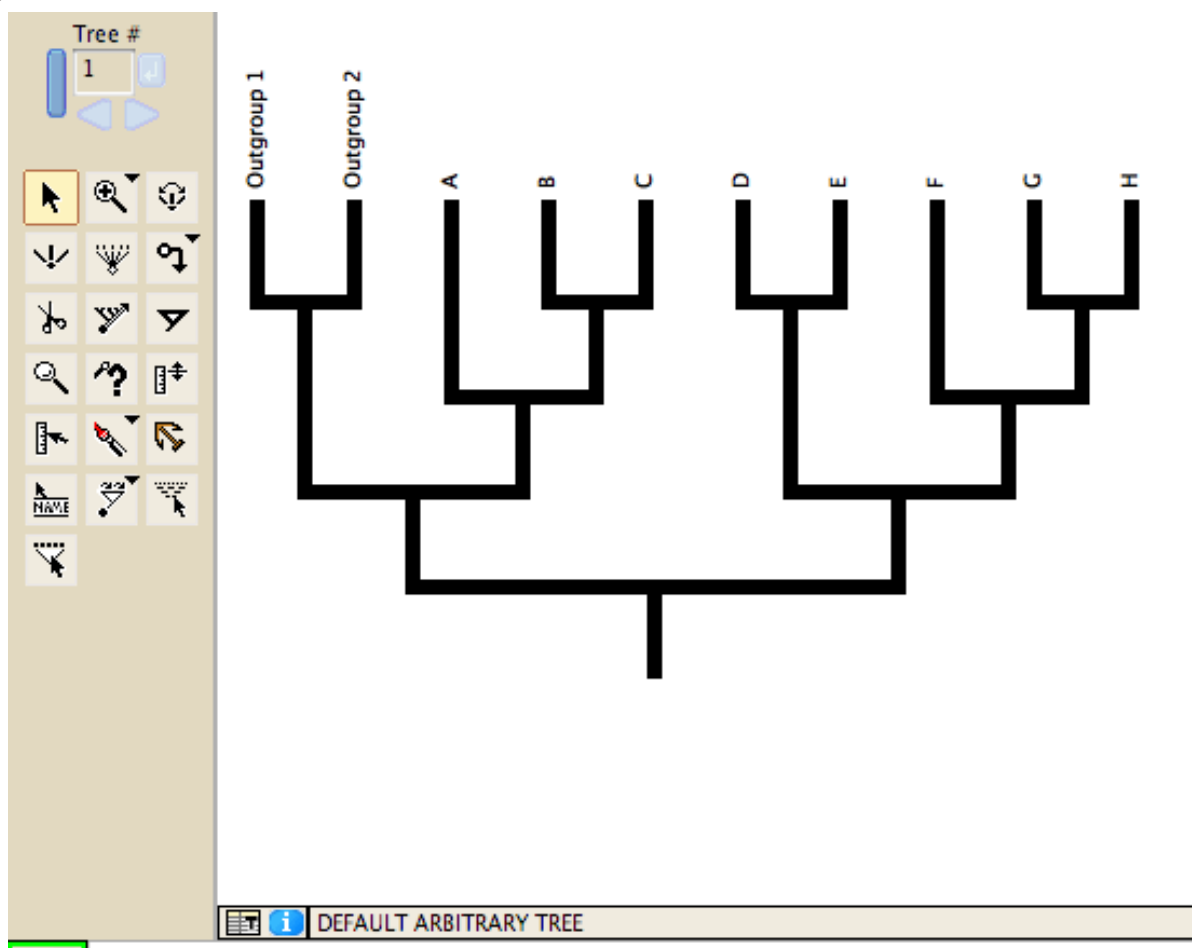
IMPORTANT! There are two spaces for Outgroups – Outgroup 1 and Outgroup 2. Enter the same data in both. (Mesquite does not support a “Make ancestral” command, which is useful to us. Doubling the Outgroup accomplishes something similar. Advanced “Tree thinking”: Any thoughts on why we do this, and why duplicating the Outgroup is similar to making a single outgroup ancestral?)

After entering your data, go to the menu and select File -> Save file.

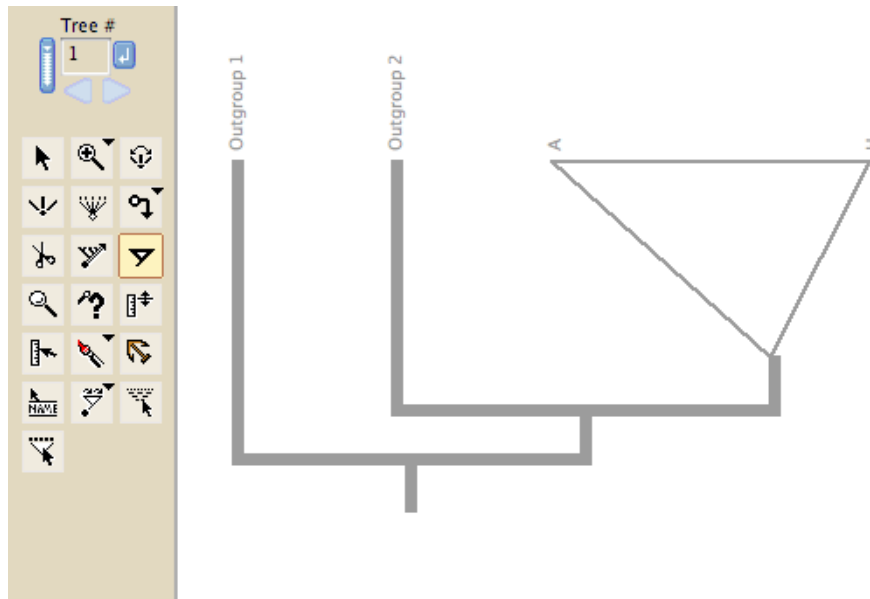
Congratulations – you have now entered your data into a data file. Now it’s time to find the shortest tree. What does the shortest tree represent, and why would we be interested in finding it?

Finding the shortest tree using Mesquite.

On the top menu bar selection Taxa&Trees -> New Tree Window -> With Tree to Edit by Hand. A tree will appear in the window. It will be an arbitrary tree (with no phylogenetic information). Your first task will be to create a tree with the out-groups separated from the in-group. Click on the “Move Branch” tool - the arrow in the upper left of the tree tool panel (indicated below):



Move the branches around until you’ve made the first to splits in the tree the out-groups. The arrangement of the other branches doesn’t matter. e.g.

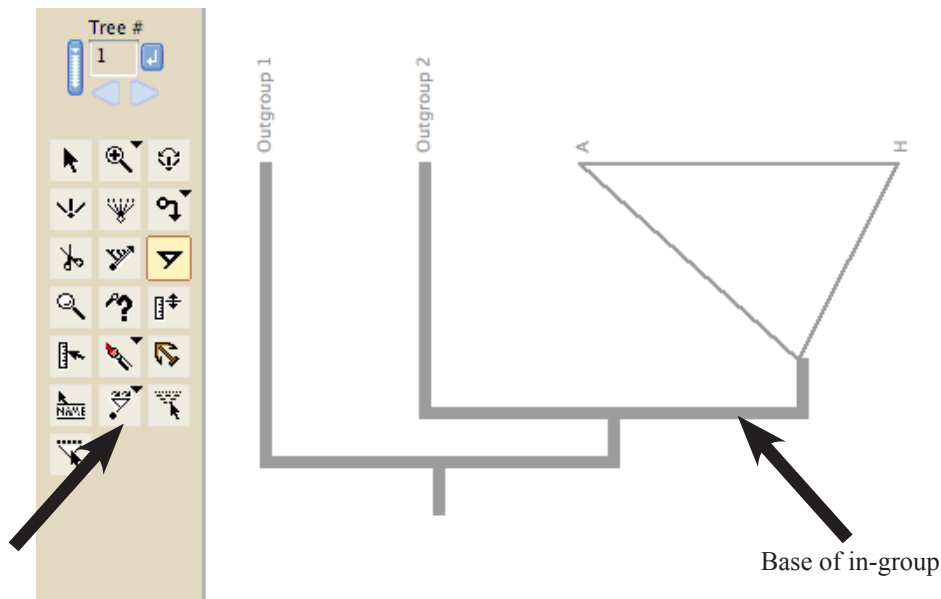


Now, you want to find the arrangement of taxa in your in group that most parsimoniously explains the evolution of the characters in your data matrix. To do this, you will need to have Mesquite compare various arrangements of the tree branches in the in-group. [Note: For a more complicated data set we would need to use a more powerful searching program, but Mesquite is fine here].

To do this, select the “Search within clade tool” from the tools panel (indicated below) and use it to click on the base of the in-group (see arrow – note that the location of the tool may change, but it will still have the same icon).



Click the tool on the base of the in-group:



You may be prompted for some additional options (depending on what the last user did). If windows pop up prompting you to make a selection, choose:

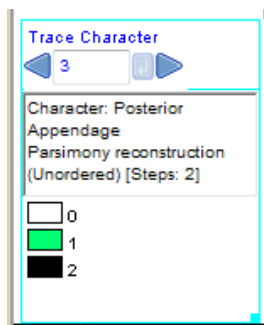
“SPR Rearranger” and/or Statistic to calculate for tree = “Tree length”

Mesquite will now chug through the various possible arrangements of the tree and calculate which arrangement(s) of branches require the fewest changes in the characters (i.e. which tree is the most parsimonious). You should now have a newly arranged and most parsimonious tree!

Analyzing character evolution using Mesquite

Computing the shortest tree is great, but often we are very interested in the evolution of individual characters. Mesquite has a great tool for analyzing character evolution called Trace Character. Go up to “Analysis” on the menu bar and select “Trace Character History.” In the pop-up window that appears select “Parsimony Ancestral States” and click “OK.” You should now see a tree with different shadings or different colors for different branches.

Pay particular attention to the box in the window that looks like this:



For character 1, Lateral Bump in this case, this box tells you what the most parsimonious character state is along each of the branches. For branches that are unshaded the most parsimonious state is 0; for branches that are shaded like the box next to 1, the most parsimonious state is 1; etc. In some cases, multiple states might be equally parsimonious (i.e. the data don't favor one interpretation over another). These branches will be shaded with multiple colors. If you move your mouse over a branch, it will show you the most parsimonious states at the bottom below the shaded boxes.

OK ... so why do we want to know this? A change in color along a branch indicates that the character evolved from one state to another at that point. Cool – we can tell where features change on the tree! If a feature shows up independently in different parts of the tree that indicates convergence. If a feature shared by two taxa is found in their common ancestor that indicates homology (they share this feature because they inherited it from a common ancestor). We can learn a lot from a tree! Let's apply this to some questions about the evolution of Caminalcules.

You can move from character to character by clicking on the little triangles next to the number in the trace character box. Is this trace character feature to answer the following questions about Caminalcules.

Interpreting the cladogram:

Did the posterior sucker evolve from paired hind limbs? The reverse? Neither?

Did hands evolve from paddles? The reverse? Neither?

Did a triangular lateral spot evolve from a circular lateral spot or square lateral spot? Can you tell? What are the possible sequences of evolution?

Are there any examples of convergent evolution among Caminalcules? If so, where?

Are derived (i.e., more recently evolved) characters always more complex than primitive characters? Why or why not? Provide an example.

Clearly, knowing the phylogeny helps to answer these questions. Could you answer these questions in the absence of a phylogeny? How?

Is the pointed head of Caminalcule F a phylogenetically informative character? Why or why not?

Is the absence of posterior appendages in the outgroup phylogenetically informative? Why or why not?

While strolling across the valley in bare feet you accidentally step on Caminalcule H. Remarkably, an infected cut on your foot immediately heals as a result of some antibiotic compound in the fluids of the now departed Caminalcule. You eagerly search for another Caminalcule H to heal the cut on your arm (it's been a rough week) but, unfortunately, none can be found. Which other Caminalcules do you think would be most likely to share the antibiotic properties of Caminalcule H? Explain your reasoning. (There's no "right" answer—just come up with a plausible rationale for choosing another Caminalcule.)

Materials

- Models – 1 set per group
 - Caminalcules
- Mesquite
 - <http://mesquiteproject.org>

Notes on Models

Plastic casts of Caminalcule models are prepared in-house using techniques outlined in Lemke (2011).

Camialcules

The models are based on the creations of Joseph Camin (Sokal, 1983). Though they are imaginary, Camin went through an evolutionary process to derive a set of 29 extant and 48 fossil caminalcules. We originally used images of the caminalcules (see below) to conduct this exercise. When we converted them to 3-D models, we removed the leg joints and turned the stripes into ridges to allow for easier casting. Students could be provided with a sheet containing these images or they could be printed on separate cards for use in the exercise. A wide variety of sets could easily be constructed in this way.

Other Options

This exercise can be easily adapted to other organisms. We have found that it works best when there are actual evo-

lutionary relationships among the specimens, and have not had good luck using items such as paper clips, hardware or sports balls. Look around at collections that you already have available. We have used these procedures with shark teeth, hominid skulls and bivalve shells. Julius and Schoenfuss (2006) report on using a set of mammal skulls. The important thing to keep in mind is that the first set of organisms needs to fit together cleanly as this is the tree that the students will construct by hand. It is also important to pick something with enough easily recognizable characters that everything will sort out cleanly and avoid polytomies.

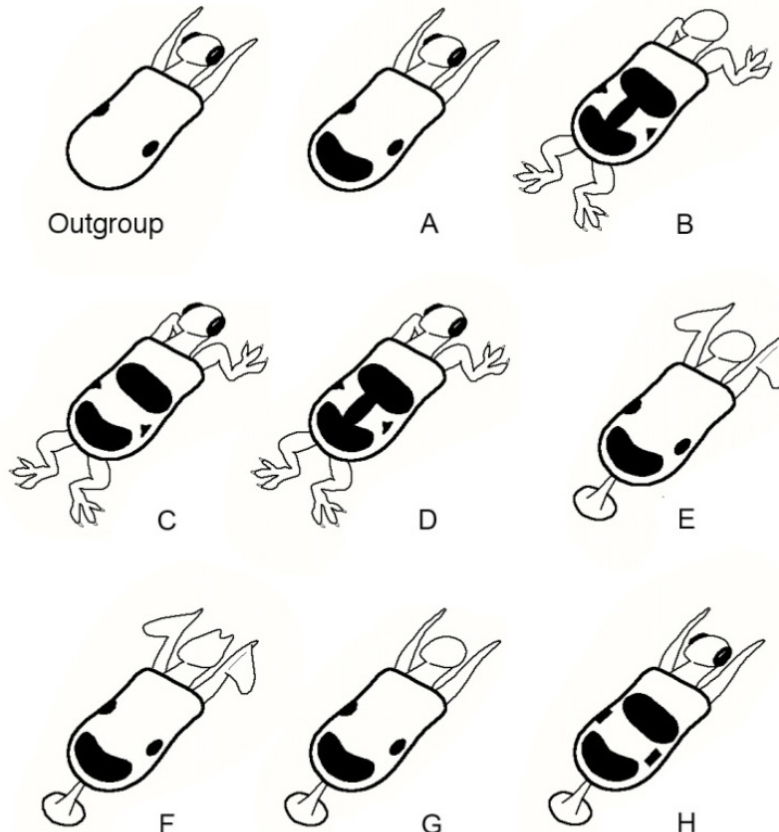
Software

Mesquite

This free software package replaces the use of PAUP and MacClade in earlier versions of this exercise. It is available for Mac, Windows and UNIX platforms and can be downloaded from <http://mesquiteproject.org>

Alternatives

There are a wide variety of phylogenetics programs available. Many of them are designed for specific applications such as DNA sequences. A comprehensive list of software is available from <http://evolution.genetics.washington.edu/phylip/software.html#Windows>



Notes for the Instructor

Teaching Notes

The Big Picture: By the end of lab, the students should understand

- Basic phylogenetic terminology, including
 - Ancestor, Descendent
 - Phylogeny, Cladogram
 - Clade, Outgroup
 - Homology, Analogy, Convergence
 - Derived similarity, Primitive similarity
- The meaning of the “Tree of Life” – i.e. there is a historical relationship among organisms, and this relationship can be represented as a tree
- Why it is important to reconstruct phylogeny
- The value of parsimony as a criterion for choosing among competing hypotheses of relationship - Why choose the shortest tree?

I: Introduction – What is a phylogeny

To get the idea of a phylogeny across to students, I usually draw a line from the bottom towards the top of the board, and say to students “Imagine this is a population of organisms going through time.” The two things that can happen are 1) new characteristics arise (draw a slash across the line); and 2) the populations split into different species. They usually get this, and once I’ve drawn a tree the relationship between the diagram and patterns of relationship are more apparent than if I started with a complete tree.

II: Why bother reconstructing phylogeny?

Emphasize this! Otherwise, phylogeny reconstruction will seem like a pointless and arbitrary exercise. It’s worth mentioning that the field of phylogeny reconstruction (and the related discipline of systematics) has undergone tremendous upheaval in the last 30 years, and this has had tremendous impact on all of evolutionary biology. Reasons for reconstructing phylogenies include:

- Generating and testing ideas about evolutionary processes. Unless we have an understanding of how organisms are related to each other, we can’t begin to understand the factors that caused them to diverge. A silly example: Our ideas about the evolution of human characteristics would certainly be different if we thought our closest living relative was a cricket (as opposed to chimps).
- Identification and control of disease vectors. If we know that a particular insect species carries a disease, we would be well-advised to examine its close relatives as well, since they may share this trait.
- Search for pharmaceuticals. If we find that a plant has medicinal properties, we might also check to see if its close relatives produce the same benefit at lower cost or with fewer side effects. We need a phylogeny to know what the close relatives are.

III: Phylogenetic terms and concepts.

Concepts such as homology are surprisingly slippery and difficult to get across, and it’s not worth getting too bogged down with them in your intro. Once they understand what a phylogeny is, and you’ve described characters and character states, you can try switching to a real phylogeny (e.g. Figure 1 in the lab) to get the basic ideas across. Here are some ideas and examples:

- **Homology** e.g. hair, lactation in mammals; four limbs in tetrapods
- **Analogy, convergence** e.g. forelimbs independently modified into wings in the ancestors of bats and birds; “warm-bloodedness” independently arising in the ancestors of mammals and birds
- **Derived similarity** e.g. crocodiles share several derived similarities with birds (heart structure, parental care, etc.) they got from a recent common ancestor; all mammals share features (hair, lactation) they got from a common ancestor not shared with any other group.
- **Primitive Similarity (plesiomorphy)** e.g. Crocs and lizards share scales and cold-bloodedness, but since this is an ancient feature also found in the ancestor of birds, it doesn’t imply a close relationship between crocs and lizards; some mammals lay eggs, like reptiles and birds, but since egg-laying is an ancient feature this doesn’t imply that egg-laying mammals are more closely related to birds than they are to other mammals. Or, a more extreme example, humans and lizards each have five digits (a very old feature), does this mean that humans are more closely related to lizards than to horses (which have only one digit)?

IV: Outgroup analysis and parsimony – Why choose the shortest tree?

You may wish to save this until after the introduction (once the students have started building trees), but at some point you should mention why the shortest tree is “best.” Primarily, it’s because the most parsimonious tree offers the simplest explanation for how the features found in the species you are comparing could be derived from the features found in the outgroup. We use this reasoning implicitly all the time. For example, we agree that bats are more closely related to other mammals than to birds (despite the presence of wings). Why? We intuitively understand that it is more likely for wings to evolve independently in bats and birds than for hair and lactation and live birth to evolve independently in bats and other mammals. This same logic, applied to all the characters, means we prefer the shortest tree (i.e. the one requiring the fewest changes).

V: The steps of phylogeny reconstruction

Briefly describe the importance of an outgroup, how to create a data matrix, and the role of the computer programs in finding shortest trees.

VI: The exercises

The students will be doing two exercises, each designed to highlight different aspects of phylogeny reconstruction.

- Simple tree – Students use a pre-defined set of characters to build a tree from a set of five Caminalcules. Building the tree by hand gives the students the opportunity to work through the basics of tree construction and interpretation.
- Complex tree - Students will have to develop their own characters, develop a tree using computer software, and *apply* their phylogeny to answer some basic questions about a larger set of Caminalcules.

Logistics

Time

This activity is designed to run in a single, three-hour lab period. Depending on how much background the students have had, the introductory lecture can range from 20-40 minutes. It takes about one hour for the students to complete the exercise for each group of organisms.

Groups

Students should work in groups of two to four. If groups are too large, then there will be lots of downtime.

Caminalcules

For the simple tree, students will likely need guidance in how to build and interpret their trees. For the complex tree, let the students decide on how they want to define characters and character states. They may break them up differently than the sample data shown here. This often results in finding polytomies or different trees than their classmates. This is an excellent opportunity to encourage them to look again at their data and revise it to come up with a more parsimonious solution. It is most important to focus on their interpretation of the tree that they came up with and not that they got the tree exactly “correct.”

Software

Students will be using **Mesquite** to analyze their data. You may want to demonstrate this using the computer projection – this will save you lots of time explaining things later. Instructions for the program are provided in the lab. It is pretty straightforward, but you should still run through it with them. They will be using several of the tools, so prepare to demonstrate the use of these (e.g. moving branches, tracing characters, etc.)

Sample Data and Answers to Questions in the Exercise

Exercise 1: Phylogeny of Caminalcules

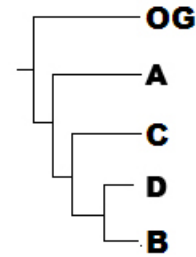
Table 6

Taxon/ Char	PosRdg	Posapp	Eyes	DorRid	Lat- Bum
O'Group	0	0	0	0	0
A	1	0	0	0	0
B	1	1	1	1	1
C	1	1	0	0	1
D	1	1	0	1	1

Table 7

	OG	A	B	C	D
OG		0	0	0	0
A			1	1	1
B				3	4
C					3
D					

Tree



Interpreting the cladogram.

Congratulations! You have constructed... the same trait evolve elsewhere on the tree?.

There are no reversals on this tree.

For each of the characters, diagram... occur on your tree?

___5___

Can you see any ways to rearrange the branches to require fewer changes? *No.*

What do the “nodes” (i.e. the intersections of lines on the tree) represent?

Common ancestors

The simplest way a trait... example from your tree:

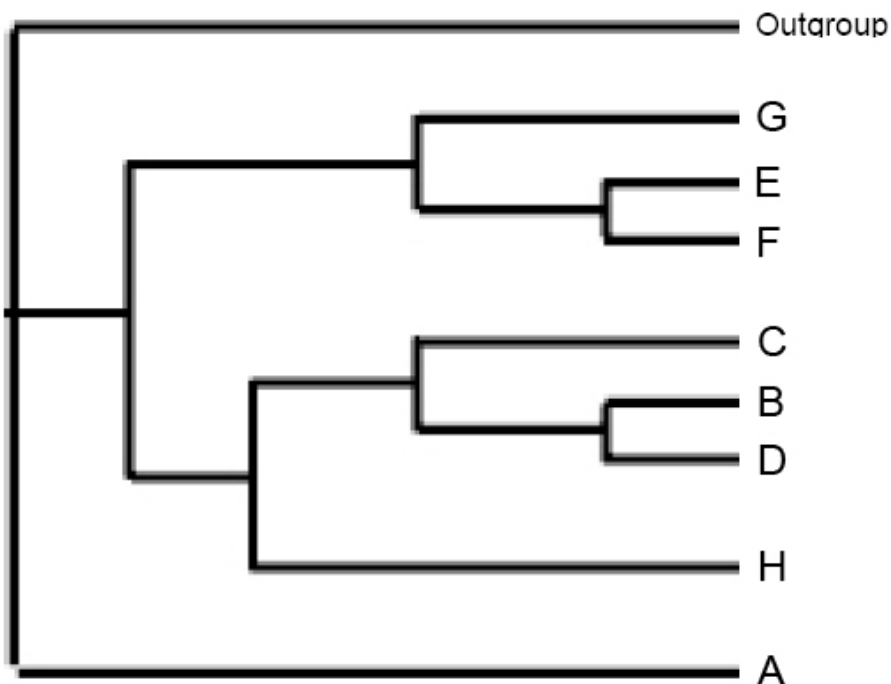
Any of the derived traits work here

Exercise 2: Phylogeny and Evolution of Caminalcules*Characters:*

- Posterior Ridge: (0=absent; 1=present)
- Posterior Appendage: (0=absent; 1=posterior sucker; 2=paired)
- Eyes: (0=present; 1=absent)
- Anterior Appendages: (0=absent; 1=paddles; 2=hands)
- Anterior Ridge: Anterior Bar (0=absent; 1=present)
- Head Shape: (0=round; 1=flat)
- Dorsal Ridge: (0=absent; 1=present)
- Lateral Bump: (0=circular; 1=square; 2=triangular)

Data Matrix

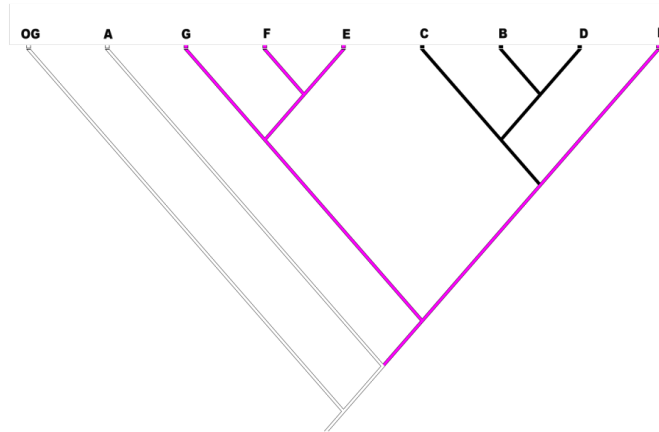
Taxon/ Char	LatBum	PosRdg	Posapp	Eyes	AntApp	AntRid	Head	DorRid
Out-Group	0	0	0	0	0	0	0	0
A	0	1	0	0	0	0	0	0
B	2	1	2	1	2	1	0	1
C	2	1	2	0	2	1	0	0
D	2	1	2	0	2	1	0	1
E	0	1	1	1	1	0	0	0
F	0	1	1	1	1	0	1	0
G	0	1	1	1	0	0	0	0
H	1	1	1	0	0	1	0	0

Shortest Tree

Answers to text questions

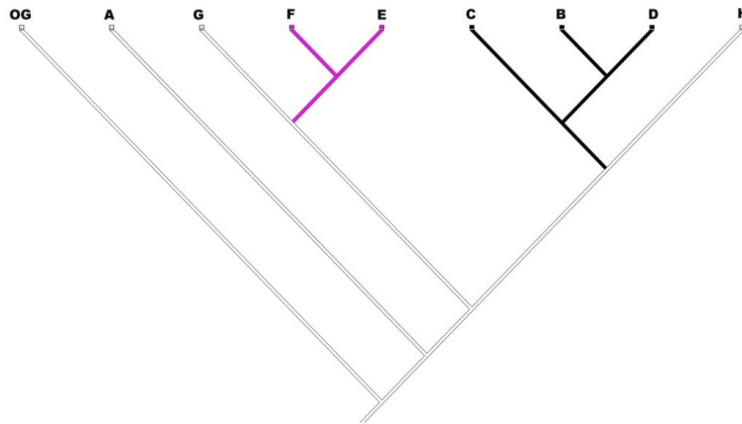
Did the posterior sucker evolve from paired hind limbs? The reverse? Neither?

Paired hind limbs evolved from posterior suckers.



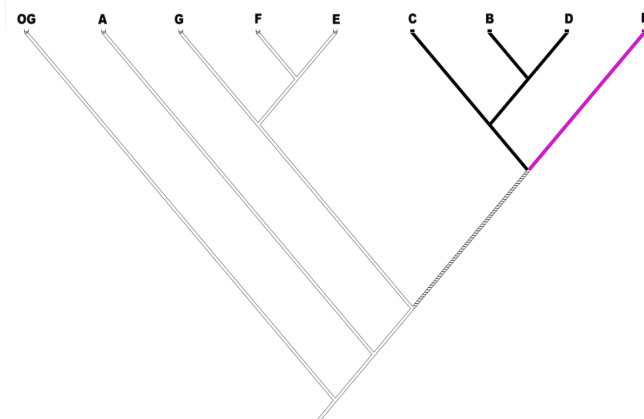
Did hands evolve from paddles? The reverse? Neither?

Neither. Paddles and hands evolved separately.



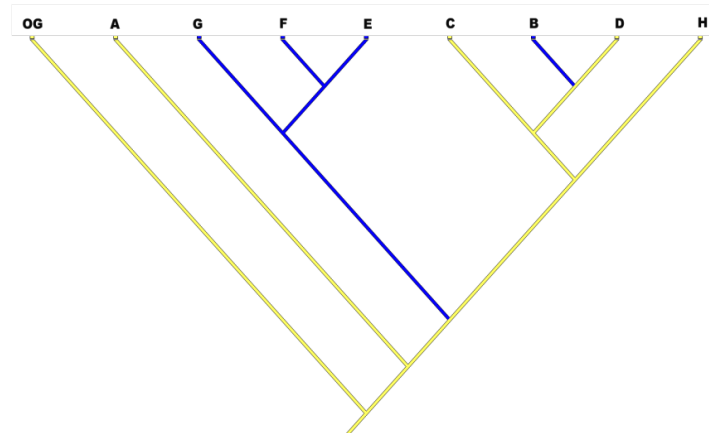
Did a triangular lateral spot evolve from a circular lateral spot or square lateral spot? Can you tell? What are the possible sequences of evolution?

Can't tell. Either circle -> square -> triangle; circle -> triangle -> square; or square and triangle independently from circle. All require two steps.



Are there any examples of convergent evolution among Caminalcules? If so, where?

Yes. The loss of eyes.



Are derived (i.e. more recently evolved) characters always more complex than primitive characters? Why or why not? Provide an example.

No. Absence of eyes is clearly less complex than presence of eyes, and yet eye absence is derived.

Clearly, knowing the phylogeny helps to answer these questions. Could you answer these questions in the absence of a phylogeny? How?

Answers may vary considerably. The main point is for them to recognize the value of the phylogeny. For convergent characters, closer examination of structure or development may indicate independent origin. Likewise, the order of appearance of characters might be revealed by looking at development, or assumed by proposing some directions of change to be more plausible than others.

While strolling across the valley in... caminalcule H? Explain your reasoning. (There's no "right" answer - just come up with a plausible rationale for choosing another caminalcule).

Two plausible arguments are: Any one of B, E, or C, since these are the closest relatives of F, or; D, H, or A which, although more distantly related, share characters that indicate a similar lifestyle (e.g. sucker, no joints) Perhaps this lifestyle promotes the evolution of antibiotic.

Acknowledgements

We would like to thank Brian Coyle for his significant input into the creation and production of the Caminalcule models.

Literature Cited

- Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life*. John Murray, London. [1964. Facsimile edition. Ernst Mayr, ed. Harvard University Press, Belknap Press, Cambridge, Mass.]
- Julius, M. L. 2006. Phylogenetic reconstruction as a broadly applicable teaching tool in the biology classroom. *Journal of College Science Teaching*, 35: 40-45.
- Lemke, H. D. (2011). Simple mold making techniques for the creation of teaching specimens. Pages 389-390, in *Tested Studies for Laboratory Teaching*, Volume 32 (K. McMahon, Editor). Proceedings of the 32nd Conference of the Association for Biology Laboratory Education (ABLE), 445 pages. <http://www.ableweb.org/volumes/vol-32/v32reprint.php?ch=42>
- Maddison, W. P. and D. R. Maddison. 1992. *Mac-Clade 3: Analysis of Phylogeny and Character Evolution*. Sinauer Associates. Sunderland, Massachusetts.
- Sokal, R. R. (1983). A phylogenetic analysis of the Caminalcules. I. The data base. *Systematic Zoology*, 32: 159-184.

About the Authors

Hans is the Lab Coordinator for Principles of Biology II at the University of Maryland. He holds a B.A. in Biology from St. Mary's College of Maryland, an M.S. in Entomology from the University of Maryland, and an M.D.E. in Distance Education from University of Maryland, University College. He currently teaches labs for both semesters of introductory biology and experimental design. Current research focuses on educational outcomes in laboratory and online settings and a survey of tiny Miocene shark and ray fossils found along the Chesapeake Bay.

Jeff is a Senior Lecturer and the Associate Chair and Director of Undergraduate Studies in the Department of Biology, University of Maryland. He holds a Ph.D. from Harvard University, 1993. Jeff was a 2009 National Academies Life Science Education Fellowship, and teaches Ecology, Evolution, Comparative Physiology, and Comparative Anatomy. His current research investigates resource partitioning in surfperches, a family of nearshore fishes common along the west coast of North America. Most of his work is on feeding mechanics and the interface among form, function, and ecology.

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with biology learning and teaching in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit <http://www.ableweb.org/>

Papers published in *Tested Studies for Laboratory Teaching: Proceedings of the Conference of the Association for Biology Laboratory Education* are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board.

Citing This Article

Lemke, H.D., and Jensen, J.J. 2012. Exploring Systematics and Phylogenetic Reconstruction Using Biological Models. *Tested Studies for Laboratory Teaching*, Volume 33 (K. McMahon, Editor). Proceedings of the 33rd Conference of the Association for Biology Laboratory Education (ABLE), 390 pages. <http://www.ableweb.org/volumes/vol-33/?art=9>

Compilation © 2012 by the Association for Biology Laboratory Education, ISBN 1-890444-15-4. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner.

ABLE strongly encourages individuals to use the exercises in this proceedings volume in their teaching program. If this exercise is used solely at one's own institution with no intent for profit, it is excluded from the preceding copyright restriction, unless otherwise noted on the copyright notice of the individual chapter in this volume. Proper credit to this publication must be included in your laboratory outline for each use; a sample citation is given above.