

Phylogeny Construction: Primate Skulls and Protein Sequences

Sarah Deel

Carleton College, Biology Department, One North College St., Northfield MN 55057 USA
(sdeel@carleton.edu)

In this introductory lab, students generate a primate phylogeny using skull reproductions to collect data including number of teeth and cranial capacity. They determine whether traits are similar to the outgroup organism (ancestral) or different (derived). They then find the most parsimonious tree for four organisms, determining which of the possible trees requires the fewest changes from ancestral to derived character states. As a student follow-up to the lab, protein sequences are used from the same primates to determine relationships based on sequence similarities. This assignment makes use of free, online resources to collect sequences, generate alignments, and construct phylogenetic trees.

Keywords: phylogenetics, primate, skull, sequences

Introduction

This lab is designed to introduce introductory biology students to the construction of phylogenetic trees using cladistics. No prior background is assumed.

In order for students to prepare themselves for the lab, a pre-lab practice problem is provided in the lab manual, with keys posted on our course management system so they can check their work. A short, multiple-choice pre-lab assignment is also posted, which they are expected to complete and hand in at the beginning of their lab session.

During the lab, students collect morphological data from reproduction primate skulls and use these data to determine which of the possible phylogenetic trees is/are most likely to represent the phylogenetic relationships between the organisms. A class discussion based on data from all the lab groups follows. This takes approximately 3-4 hours.

After lab, students generate phylogenetic trees for the same set of primates using sequence data and free, web-based sequence alignment and tree-making software. They turn in an assignment summarizing this work.

Student Outline

Primate Phylogenetics: Reconstructing Evolutionary History Using Morphological Characteristics

Project Goals

We want you to understand what a phylogeny is, and how phylogenies are constructed. You will learn how to construct a phylogeny based on morphological data using the techniques of cladistics. In the process of constructing a phylogeny, you will gain experience collecting data. You will explore the evolutionary relationships among living primates, including humans.

Additional Reading

There is no additional reading for this week's lab; you will want to be sure and read this section of the lab manual carefully, and complete the (ungraded) dinosaur practice problem (see below). Please bring a calculator and a pencil with you to lab this week.

Phylogenetics

Phylogenetics is the study of evolutionary history; evolutionary biologists try to piece together a picture of how different types of organisms are related to one another. A **phylogeny** is a representation of this relatedness, and often looks like a branching tree (see Fig. 1 for a simple example). Many evolutionary questions can be answered once a phylogeny is well established. Are birds really a lineage of dinosaurs that did not become extinct? How many times did flight originate in the insects? Did the ancestor of all eukaryotes have mitochondria? These are just a few examples of questions that are cropping up in evolution journals currently.

How do evolutionary biologists make decisions about the relatedness of different organisms? The earliest biologists grouped organisms together based on how similar they looked, and this use of morphological (form-related) characteristics is still valuable today. The number and type of characteristics used by biologists to reconstruct phylogenies have grown as technology has made more information available. In Darwin's time, phylogenies were based on "macro" characteristics. An example might be something as simple and quantitative as the number of teeth an organism has, or as complex and qualitative as the shape of a particular bone in the skull. Fossils may be used to collect this type of evidence as well (though fossils would not be helpful, for example, if you were looking at something like the color of an organism). Light microscopes made the inclusion of tissue and cellular characteristics possible, and electron microscopes brought in subcellular characteristics.

Biologists have also looked at similarities between the development of different organisms for evidence of relatedness. One example is the appearance of a tail early in human development, supporting a link between humans and other vertebrates not apparent by looking at only adults. Most recently, the ability to sequence DNA and proteins has led to an immense new source of data evolutionary biologists can use to compare organisms.

This week in lab, we'll be using morphological characteristics to investigate the relationships between several different primate species; after lab, you'll continue this investigation using genetic sequences for a homework assignment.

Phylogenetic Trees

To illustrate historical relationships, evolutionary biologists use diagrams to connect organisms based on their relatedness; these are commonly referred to as "phylogenetic trees," "phylogenies," or just "trees." The diagrams are called trees because they often begin with a single "trunk" and break off into multiple branches. Trees can look very different depending on what an evolutionary biologist is trying to show. The trees you'll be working with today will look something like that in Figure 1.

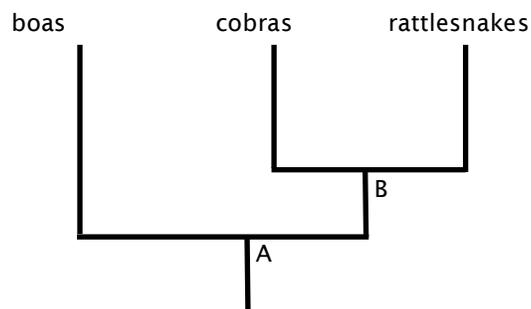


Figure 1. Phylogenetic relationships between different types of snakes.

At the end of each branch is the name of a group of organisms; in this case, each group represents a different type of snake. In a tree like this, there is an assumption that each branch point represents some (now extinct) ancestor which was common to the groups of organisms at both ends of the branches. As you move down the branches of the tree, you are moving backward in time. So, for example, the cobras and the rattlesnakes shared a common ancestor at point B. This point of divergence occurred between 30 and 40 million years ago (Wiens et al., 2006). The boas (like *Boa constrictor*), the point B ancestor, the cobras, and the rattlesnakes shared an earlier common ancestor at point A (approximately 95 million years ago (Wiens et al., 2006)). At each of those branch points, **speciation** occurred: a new species was formed which was different from the species below it on the tree. Over time, both species arising from a branch point evolve so that they are different from their common ancestor. While none of the snakes have limbs, boas have a pelvis and also vestigial hind legs. Boas suffocate their prey by constriction, squeezing tightly enough to prevent the rib cage's normal expansion. Rattlesnakes and cobras both use venom, delivered via fangs, to subdue their prey, but boas have no venom or fangs. Therefore the ancestors of cobras and rattlesnakes, at some point along the line between A and B in Figure 1, evolved fangs for venom delivery.

Cladistics

There are several different methods to go about constructing a tree. One commonly used method is called **cladistics**, also called “phylogenetic systematics.” The term “cladogram” refers to a tree generated with this method. The word **clade** comes from a Greek word which means “branch;” basically, a clade is one branch of a phylogenetic tree. In cladistics, “clade” specifically refers to a group of organisms and their common ancestor(s). In Figure 1, a group including the cobras, the rattlesnakes, and the ancestor at B forms a clade. The entire tree can also be considered a clade.

Cladistics is based on looking at changes between ancestors and modern species. You can think of each change as occurring at some point along each branch of the tree, and causing the modern species to look different from their ancestors. To study these changes and determine evolutionary relationships, it is useful to have an outgroup. An **outgroup** is a group of organisms that is closely related to the organisms you want to study, but is definitely not a part of your study subjects. The outgroup should share a common ancestor with the organisms you are studying. Choosing an outgroup requires knowledge of more general phylogenetic relationships; you can use this general information to help you determine the evolutionary relationships among a more limited, closely related set of organisms.

For example, a helpful outgroup to use when assembling the snake tree (Fig. 1) would be lizards. Lizards, like snakes, are vertebrates (have a backbone), and share a relatively recent common ancestor compared to other organisms. Snakes diverged from lizards approximately 165 million years ago; this general relationship is known from previous phylogenetic studies (Wiens et al., 2006). Lizards are a better choice for an outgroup than earthworms or other very distantly related organisms. If you tried to use earthworms as an outgroup, you'd lose the chance to make comparisons in the shape, number, or size of bones, since earthworms are invertebrates.

Many of the outgroup's characteristics (commonly referred to as **characters** in the context of phylogenetics) should be ancestral to those of the groups you're studying. While lizards are an appropriate outgroup for studying snake phylogeny, if you were trying to construct a lizard phylogeny, you would not want to use snakes as the outgroup. We have strong evidence that the ancestors of snakes had limbs like lizards: the boas have a pelvis and vestigial hind limbs; also, a fossil snake has been discovered with a pelvis and fully formed hind limbs (Apesteguia and Zaher, 2006).

Compared to snakes as a whole, lizards have more **ancestral** character states (like legs); in contrast to the term ancestral, we use the term **derived** to refer to character states that differ from the ancestral forms. In the case of this appendage example, legs are the ancestral character state while leglessness is a derived character state. Keep in mind that context is vitally important to how these two terms (ancestral and derived) are used. If we were studying lizards instead of mammals, and using fish as an outgroup, legs would be the derived character state, while leglessness would be ancestral.

Cladistics creates phylogenies by grouping organisms with **shared derived character states** close to each other on the tree. The outgroup is necessary in cladistics because it is assumed to have the ancestral state for each character: knowing this, you can then determine for a given character if one of the organisms you're interested in has the ancestral state (i.e. looks the same as the outgroup) or the derived state (i.e. looks different from the outgroup in some way).

If all character states just changed once in evolutionary history, then there would be no need for a complex methodology: we could just group organisms together based on similarities in their derived states. Unfortunately, there are two types of events which can muck up reconstruction of a true phylogeny. They are:

convergence: the same derived character state evolving twice, independently

For example: primates, marsupials, and dinosaurs all include bipedal members. If you use bipedalism as a derived character state (walking on all fours being ancestral), you might incorrectly group these organisms (humans, kangaroos, and Tyrannosaurus rex) together on a phylogenetic tree. This is absurd given the other things we know about these organisms.

reversal: possession of the ancestral character state by an organism whose ancestor actually had the derived character state

For example: the lizards most closely related to snakes are venomous. The lack of venom in boas is a reversal; the production of venom was present in the ancestors of these organisms, but this capability was lost in the boas. If you created a phylogeny for lizards and snakes based only on this trait, you might group the boas with other non-venomous lizards, separately from the venomous lizards and snakes. Based on other things we know about the boas and other snakes, this would be inaccurate.

Since convergence and reversals can obscure the true phylogeny, we need to be able to correct for them. We do this by taking many characters into account when we are developing a phylogeny, and by avoiding characters that we know include reversals or convergence. Even though these confounding events occur, in cladistics we make the assumption that they are relatively uncommon.

Given that convergence and reversals, although rare, will still occur, how do we find the true phylogeny for a group of organisms? If we are comparing three different types of organisms, there are three possible trees that represent their relationship. With four organisms, there are fifteen possible trees. How do we know which is most likely to be correct? In cladistic analysis, we use the concept of **parsimony** to select the best tree. Parsimony is, generally speaking, the notion that a simpler explanation for something is more likely to be correct than a more complicated explanation. In cladistics, parsimony is the idea that the simplest tree, requiring the fewest changes in character state (ancestral to derived or derived to ancestral) is most likely to be the correct tree. The convergence example above is a good one; a tree which placed humans, kangaroos, and *T. rex* close together would require many, many changes in other character states (e.g., the character states of endothermy, hair, and lack of an egg would each have to have evolved more than once, for many different large groups of organisms). Such a tree would not be very parsimonious. Conversely, a tree placing humans with other placental mammals, kangaroos with other marsupials, and *T. rex* with other dinosaurs would be much more parsimonious, even though the “bipedal” character state would have to arise multiple times.

Selecting the most parsimonious tree literally involves counting the number of changes required for each possible tree, and identifying the tree with the fewest changes.

How to Make a Tree

We'll describe the steps you might use to make the snake tree (Fig.1); in Figure 3 a diagrammatic representation of this process is presented. You should go through the same basic process with a dinosaur example, below, before coming to lab this week. In lab, we'll follow this process once again to generate a tree for primate phylogeny.

1. Select the group of organisms you're studying.

We chose snakes for this first example, dinosaurs for a practice problem, and a variety of primates for the lab exercise. Decisions about which organisms to look at can be driven by many different sorts of questions.

2. Select an outgroup for your group of interest.

For the example snake tree, we'll use lizards as the outgroup. Selection of an outgroup requires some knowledge about how the organisms you're interested in are related to the broader tree of life.

3. Select which characters you'll use to determine relatedness.

We've chosen characters that should help sort out the groups of organisms as much as possible. Generally, you look for characters that have a clear difference between the ancestral and derived states, and characters that show some variation within the group of interest. In the snake example, we have only selected two characters; generally, you would use hundreds of characters to sort out relationships, in order to help minimize the effects of reversals and convergence.

4. Score each character for each organism.

This just means to decide what the character is for each of your organisms, including the outgroup. In our snake example, it is just the answer to a yes or no question. In some cases, this might be a quantitative character, like “length of jawbone.”

5. Using the outgroup, decide what the ancestral and derived states are for each character.

Generally, the character state of the outgroup will correspond to the ancestral character state.

6. Fill out a character table (also called a character matrix) for the group of interest. Use an “A” to designate the ancestral state and a “D” to designate the derived state.

Here we're just trying to get everything into a binary system; this allows us to count up differences between organisms. It may seem obvious in the snake example, where we are just using "yes" or "no" answers, but keep in mind that "yes" can refer to an ancestral state for some characters or to a derived state for other characters. Also, for quantitative traits, like length of jawbone, you may be able to decide on ancestral and derived categories to sort out your results (e.g., <20 mm is ancestral and >20 mm is derived). Notice that the outgroup does not appear in this table, because it will not appear on our tree.

7. Fill in possible trees with character matrix data; indicate the presence of state changes with hash marks. Use the total number of changes as an indication of parsimony.

For our example, there are only three possible trees: one with boas and rattlesnakes most closely related, one with boas and cobras most closely related, and one with rattlesnakes and cobras most closely related. Draw a hash mark across the branch whenever there is a change in state necessary (Fig. 2). Label each hash mark with the change required there. Note that for the first two possible trees in the diagram, there are different ways to draw in the changes; the top diagrams are drawn with a convergence, and the bottom two diagrams are drawn with a reversal. There is no way to know beforehand which is correct, and in this case, the total number of changes for these trees is the same. (If the number of changes differed, you would choose the most parsimonious option).

8. Determine which tree is most parsimonious by counting the number of changes required to make each tree.

Count up the total number of changes required in each tree and write that below the tree. The smallest number indicates the most parsimonious tree. In the snake example, Possible Tree 3 requires only 3 changes, making it more parsimonious than the other two possible trees, which each require 4 changes.

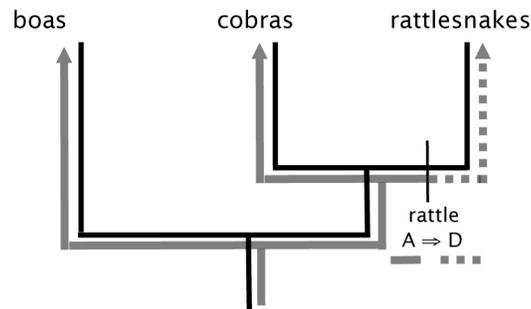


Figure 2. Example of use of a hash mark. Placement in this example indicates that there has been a change from the ancestral state (A) lacking a rattle on the tail (solid arrows) to the derived character state (D), the presence of a rattle (dotted arrow).

Practice Problem: Dinosaur Phylogeny

This problem is adapted from Campbell's Biology, 5th Edition (Campbell et al., 1999). Work through the problem before coming to lab; hints and answers are posted. We've done steps 1-4 for you below; complete steps 5-8 to find the most parsimonious phylogeny.

Step #1: Choose study organisms:

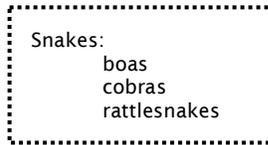
Pachycephalosaurus
Parasaurolophus
Stegosaurus
Triceratops

Step #2: Choose outgroup:

Allosaurus

All dinosaurs are in one of two groups: the Saurischians (saur=lizard, ischia=hip joints, so "lizard-hipped;" their pelvic bone structure is similar to that of lizards) or the Ornithischians (ornith=bird, ischia=hip joints, so "bird-hipped;" their pelvic bone structure is similar to that of birds). Our study organisms are all Ornithischians, and the outgroup *Allosaurus* is a Saurischian dinosaur. To see images of these dinosaurs, try using Google's Image search.

Step #1:
Choose study organisms.



Step #2:
Choose outgroup.



Step #3:
Choose characters.



Step #4:
Score characters for organisms.

| organism: | lizards | boas | cobras | rattlesnakes |
|-----------|---------|------|--------|--------------|
| pelvis? | yes | yes | no | no |
| fangs? | no | no | yes | yes |
| rattle? | no | no | no | yes |

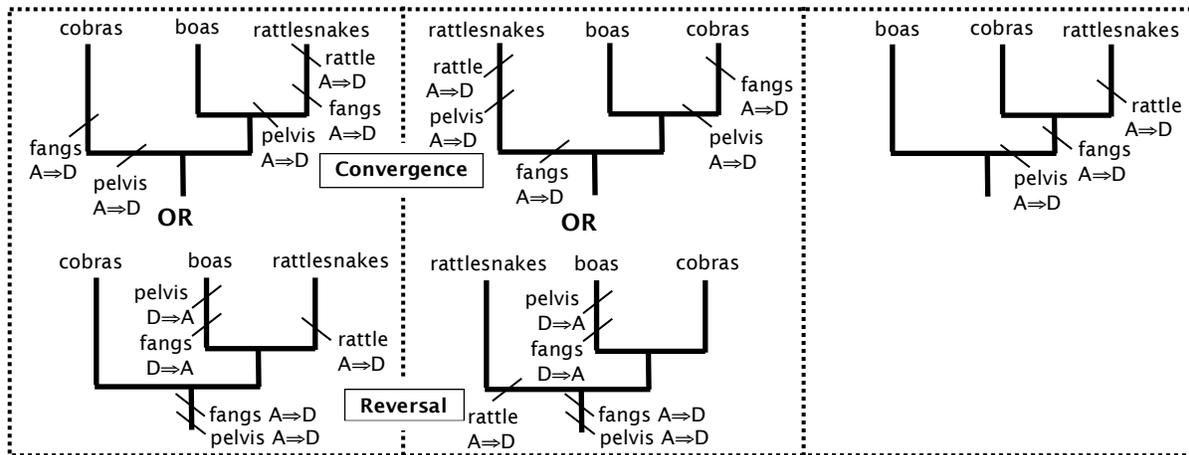
Step #5:
Determine which character states are ancestral and which are derived using outgroup data.

| | | |
|---------|-----------------|---------------|
| pelvis: | yes = ancestral | no = derived |
| fangs: | no = ancestral | yes = derived |
| rattle: | no = ancestral | yes = derived |

Step #6:
Fill out character matrix.
A=ancestral, D=derived

| | pelvis | fangs | rattle |
|--------------|--------|-------|--------|
| boas | A | A | A |
| cobras | D | D | A |
| rattlesnakes | D | D | D |

Step #7:
Use character matrix to analyze possible trees.



Step #8:
Determine which tree is most parsimonious.

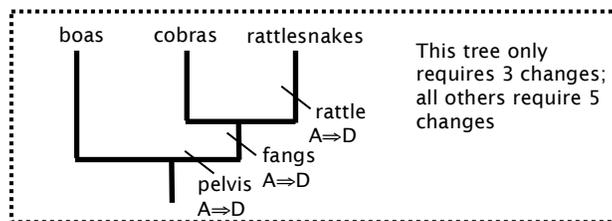


Figure 3. Diagrammatic representation of the steps involved in reconstructing phylogeny using cladistics, using snakes as an example. See the section on How to Make a Tree for more information.

Step #3: Choose characters:

1. Armored plates along back

One hypothesis for the plates is that they functioned in thermoregulation; other suggestions are that they were used in defense or as a display.

2. Prementary bone

This is an extra bone at the front of the lower jaw in some dinosaurs, which did not contain teeth but supported the lower half of a beak

3. Pubis bone extends posteriorly

The pubis bone is found in the pelvis, and in some dinosaurs it has an extension which points backwards; one hypothesis is that this allows for a long digestive tract.

4. Rostral bone

This is a bone at the front of the snout that forms the upper part of the beak; this can serve as a base for a horn on the snout.

5. Shelf at base of skull

Some dinosaurs have an enlarged area at the base of their skull; this can be bulbous or shelf-like.

6. Pubic symphysis

Some dinosaurs have pubis bones that join in the center of the pelvis (one bone from the right and the other from the left); this tight joint is called a symphysis.

7. Unequal enamel layer on teeth

On upper teeth, there is more enamel on the outer surface of the tooth than on the inner surface of the tooth; on lower teeth, the reverse is true.

Step #4: Score characters for organisms

| | <i>Allosaur.</i> | <i>Pachycep.</i> | <i>Parasaur.</i> | <i>Stegosaur.</i> | <i>Tricerat.</i> |
|-----------------------------------|------------------|------------------|------------------|-------------------|------------------|
| 1 armored plates? | no | no | no | yes | no |
| 2 prementary? | no | yes | yes | yes | yes |
| 3 pubis bone extends posteriorly? | no | yes | yes | yes | yes |
| 4 rostral bone? | no | no | no | no | yes |
| 5 shelf at base of skull? | no | yes | no | no | yes |
| 6 pubic symphysis? | yes | no | yes | yes | no |
| 7 unequal enamel layer on teeth? | no | yes | yes | no | yes |

Step #5: Determine which character states are ancestral and which are derived using outgroup data.

| | Ancestral | Derived |
|---------------------|-----------|---------|
| 1. armored plates | no | yes |
| 2. prementary | | |
| 3. pubis ext. post. | | |
| 4. rostral | | |
| 5. skull shelf | | |
| 6. symphysis | | |
| 7. enamel | | |

Step #6: Fill out character matrix.

| Character: | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------------------|---|---|---|---|---|---|---|
| <i>Pachycep.</i> | A | | | | | | |
| <i>Parasaur.</i> | A | | | | | | |
| <i>Stegosaur.</i> | D | | | | | | |
| <i>Tricerat.</i> | A | | | | | | |

A=ancestral, D=derived

You'll need to apply the information from Step #5 to your data from Step #4. We have filled in the chart for Character 1, armored plates, for you. As you can see in Step #6, three of the four organisms share the ancestral state of no armored plates; the only organism with armored plates (the derived character state) is Stegosaurus.

Step #7: Use character matrix to analyze possible trees.

In this example, rather than making you analyze all 15 possible trees, we have drawn a subset of 5 trees to analyze (Fig. 4). You should get the idea from these, but you can always set up the other 10 trees on your own if you like.

For each tree, you will need to determine where state changes must have occurred. In some cases, you will need to decide whether a convergence or a reversal is necessary; choose the option which puts fewer changes on the tree. (If the number of changes is identical, it doesn't matter which you choose.)

Review the trees in Figure 3 if you have questions. This process of analyzing trees may take a little time, but the more you can get figured out now, the more comfortable you will be with the procedure during lab.

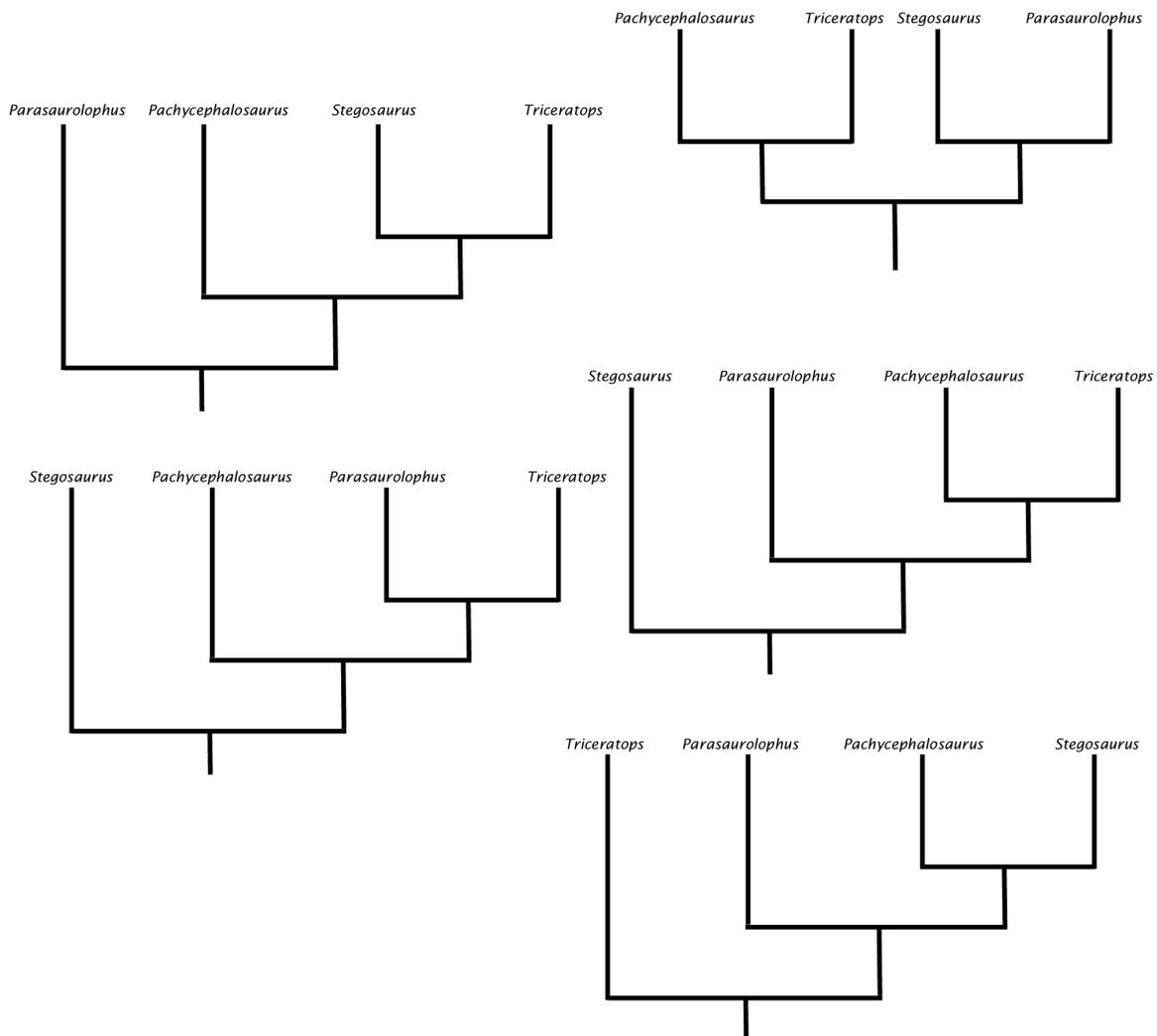


Figure 4. Five possible trees to analyze.

Step #8: Determine which tree is most parsimonious.

After you have counted up the changes required for each tree, find the tree with the smallest number. Congratulations! If you haven't already, step back and look at how much each character contributed to your selection of that tree. Were there characters that had the same effect on all the trees? What are the implications of characters like those? Which characters were the most important in determining your phylogeny?

Jot down any questions you have about this procedure and bring them to lab. You can ask your lab instructor or TA about them, or just discuss them in your lab group.

Determining Primate Phylogeny

This week in lab, you'll be studying primate phylogeny. We want you to gain experience observing and recording differences in skull morphology for a number of different primates, then use the data you've collected to select the most parsimonious tree.

Step #1, as we described above, is to choose the organisms you're interested in. We chose primates because they are well-studied, with several well-established phylogenetic groups as well as a few more controversial branches. Many people find it inherently interesting to study human skulls in comparison with other primate skulls; there is the hope that morphological studies like this can help provide a framework for studying fossils which look similar to modern humans. We will be studying the primates listed below.

| <u>Scientific Genus Name</u> | <u>Common Name</u> |
|------------------------------|--------------------|
| <i>Alouatta</i> | howler monkey |
| <i>Gorilla</i> | gorilla |
| <i>Homo</i> | human |
| <i>Lagothrix</i> | woolly monkey |
| <i>Macaca</i> | macaque |
| <i>Pan</i> | chimpanzee |
| <i>Pongo</i> | Orangutan |

Members of the genus *Alouatta* (Fig. 5), the howler monkeys (also called howling monkeys), are found in Central and South America. They roar very loudly, and can be heard up to 2 km away. Check out <www.naturesongs.com/howler5.wav> to listen to one. Their genus name sounds most like a howl of any of the primates we'll work with this week.

Gorillas (Fig. 6) have one of the easiest genus names to remember: *Gorilla*. Gorillas are the largest living primates; males can weigh up to 200 kg (441 lb) (Fleagle, 1999). These endangered animals are found in central Africa. Famous gorillas include Koko, who was taught sign language (you might remember a children's book called *Koko's Kitten*). Dian Fossey's studies of gorillas were popularized in the movie *Gorillas in the Mist*. In Disney's animated *Tarzan*, Rosie O'Donnell, Glenn Close, and Lance Henriksen were all voices of gorillas.

Woolly monkeys (Fig. 7) belong to the genus *Lagothrix*. They are found in northern South America. They are named for their thick fur, which ranges in color from brown to grey to black depending on the species. They use their prehensile tails to help with locomotion through the trees. If you happened to catch the 1964 sci-fi flick *Robinson Crusoe on Mars*, the monkey stranded with the astronaut is a woolly monkey (apparently a male woolly monkey played this female role, and so had to wear fur trousers for the part).

Macaques (Fig. 8), in the genus *Macaca*, are more widely distributed in their range than any other primate genus except *Homo*; they are found from southeastern Asia to northern Africa. You might have seen Japanese snow monkeys (*Macaca fus-cata*) basking in hot springs on TV during the 1998 Winter Olympics in Nagano.

Pan (Fig. 9) is the chimpanzee genus (note the genus name is contained within the common name); chimps are found in western and central Africa. Jane Goodall documented the first use of tools by non-humans when she discovered chimps that peel leaves off branches and stick the branches into termite nests to go "termite fishing."

Orangutans (Fig. 10) belong to the genus *Pongo* (to remember this one, note that both common and genus names contain "ng"). They are large primates that live in Asia. Orangs had starring roles in those cinematic greats *Every Which Way But Loose* and *Dunston Checks In*, and you might have seen a PBS special with Julia Roberts on orangutans.

Step #2 is to select an outgroup. We have chosen lemurs, of the genus *Lemur* (Fig. 11), as our outgroup. Lemurs are among a group of primates that are considered to have a relatively high number of ancestral characters compared to other primates. Lemurs are found only on the island of Madagascar, off the eastern coast of Africa. There are fourteen genera of primates (25% of the primate genera living today) on this island, which is the size of California and Oregon combined (Fleagle, 1999). You might have seen lemurs in the movie *Fierce Creatures*. *Zoboomafoo*, the PBS kids' show, stars a sifaka, which is a close relative of the *Lemur* genus.



Figure 5. Howler monkey, *Alouatta*.



Figure 6. Gorilla, *Gorilla*.



Figure 7. Woolly monkey, *Lagothrix*.



Figure 8. Macaque, *Macaca*

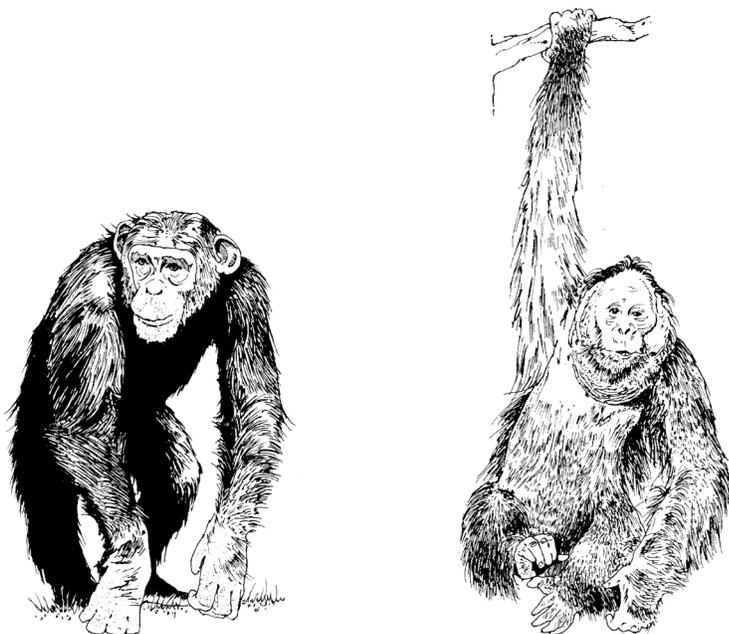


Figure 9. Chimpanzee, *Pan*.

Figure 10. Orangutan, *Pongo*.



Figure 11. Lemur, *Lemur*.

(Image Sources: Figures 5-11 all licensed from Clipart.com (www.clipart.com))

Step #3 is to select characters to study. You will record data on several characters in lab today (see Table 1). You will focus primarily on structures associated with primate skulls for several reasons. There are many bones that make up the skull (Fig. 12); each one of these may be shaped differently in different organisms, and the way the bones join together can also vary. Many fossil primate skulls have been carefully analyzed for evidence of relatedness to modern humans, so there is a lot of existing data on primate skull morphology. Teeth in particular are often well preserved (in comparison with other fossils), so they have also been well-studied. If you glance through the list of characters in Table 1, you can get an idea of the different types of data you'll be recording. There are obviously many more skull characters you could look at; we selected this subset based on what we know about major, defining differences between different types of primates and how easy it is to observe the characters in these skulls. You may notice additional patterns in your observations today; feel free to make other measurements and record your data for use in developing your phylogeny.

Step #4 is to score each organism for the characters. We'll give you more specific directions for particular measurements below. As an overview, you should know that each lab group (of two or three students) will score only four of the six primates we're studying in the lab; this will keep the number of possible trees manageable. Each group will take measurements of the Lemur skull in addition to the four primates you'll be assigned in lab. At the end of the lab, your whole lab section will work as a group to come up with the most likely phylogeny for all six of the primates.

Step #5 is to determine whether each character state is ancestral or derived for each organism. For the characters we have chosen, the *Lemur* has the ancestral state. Check with your lab instructor or TA if you have any questions about a particular character.

Step #6 is to fill out a character matrix based on the categories you came up with in Step #5 and the data you collected in Step #4. Fill out Table 2 similar to the character matrix you completed for the dinosaur problem. Your lab instructor may ask your lab section to compare data at this point, and resolve any inconsistencies between groups.

Step #7 is to mark the changes in character state necessitated by arranging your four assigned primates onto different phylogenetic trees. We have drawn the fifteen possible trees for you on Figure 16; you will need to indicate which organism corresponds to each number/letter on the diagram. You may decide to divide up the trees among the members of your lab group, and each analyze a different subset of the trees. If you do this, just make sure you are all using the same numbering scheme.

Step #8 is to determine the most parsimonious tree(s). Count the number of changes required for each tree and write this number at the base of the tree. Indicate somehow which tree is the most parsimonious. If more than one tree is most parsimonious, indicate all these trees. If there are some trees that are close to the most parsimonious, try to determine which characters are responsible for the extra changes and think about how confident you are that those trees are incorrect. (What is the likelihood that one of your not-most-parsimonious trees actually represents the true phylogeny of these organisms?)

Discuss your findings within your group, and be prepared to informally share your results and your reasoning with the rest of the lab section.

References

- Apesteguia, S. and H. Zaher. 2006. A Cretaceous terrestrial snake with robust hindlimbs and a sacrum. *Nature*, 440: 1037-1040.
- Campbell, N. A., J. B. Reece, and L. G. Mitchell. 1999. *Biology*. Fifth edition. Benjamin-Cummings, Menlo Park, California, 1175 pages.
- Fleagle, J.G. 1999. *Primate adaptation and evolution*, 2nd edition. Academic Press, San Diego, 596 pages.
- Weins, J. J., M. C. Brandley, T. W. Reeder. 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution*, 60: 123-141.

Practical Information for the Lab

First of all, you need to get some general familiarity with the skulls you are working with. Be aware that some of these skulls are real, and many are excellent (expensive!) reproductions. Be very careful with the skulls. Hold them over the foam pads on the lab benches. Do NOT use pen or pencil to point to features of the skull; stray marks can be almost impossible to remove. We have blunt probes in the lab that you can use while examining and discussing the skulls.

1. Pick up a human skull and gently turn it around and over to get a sense of it. Compare it to the human skull in Figure 12.
2. Find a **suture** somewhere on the skull (Fig. 12).

There are 22 bones in a human skull. A suture is a place where two different bones have come together and formed a very tight joint. Some of these joints will actually fuse so tightly that you will not be able to find a seam. Since most of the non-human skulls in the lab are reproductions, you won't be able to find all the sutures that would be present on a real skull.

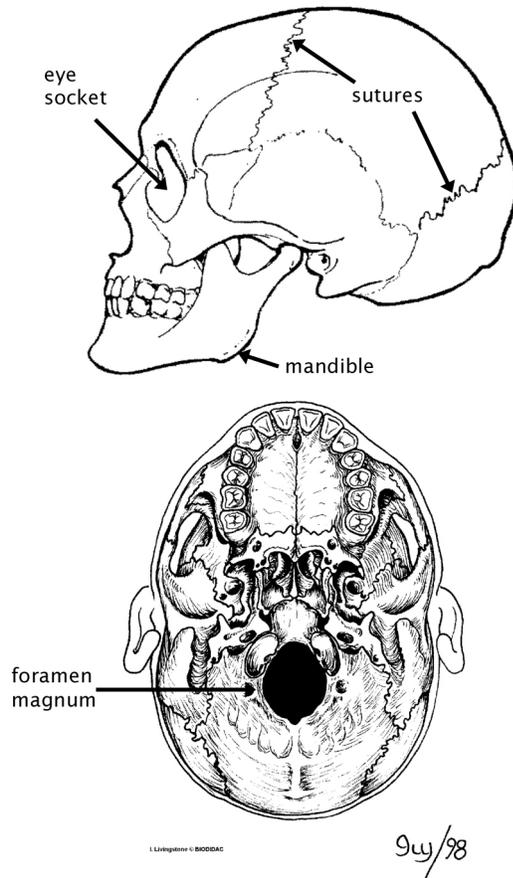


Figure 12. Human skull, side view and view from below. The mandible is not present in the view from below. Apparently the rear-most molars are not present on this individual. (Sources: upper image downloaded from www.clipart.com; lower image by Ivy Livingstone, downloaded from biodidac.bio.uottawa.ca.)

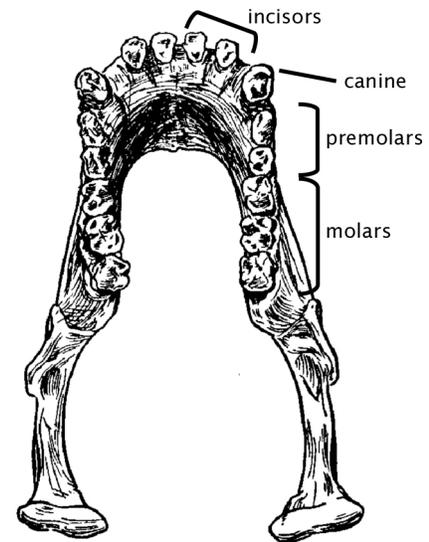


Figure 13. Chimpanzee mandible, view from above. (Source: www.clipart.com)

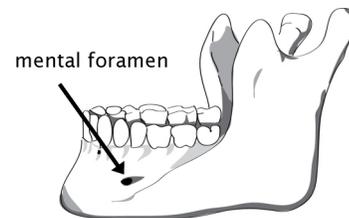


Figure 14. Human mandible showing location of mental foramen. (Source: Public Domain image under Creative Commons CC0 1.0 Universal Public Domain license, downloaded from commons.wikipedia.org.)

- Practice fitting the mandible (jaw bone) into the proper place in the skull (Fig. 12, Fig. 13). Note how the upper and lower teeth meet.
- Find the **foramen magnum** (foramen: hole, magnum: large) in the human skull (Fig 12).
The foramen magnum is where the spinal cord meets the skull.
- Orient the **eye sockets** forward.

It can be a little misleading to look at the skull sitting on the lab bench. The best way to get an idea of how the skull looked on the animal is to hold it so that the eye sockets (they are called orbits) are facing directly forward. If you look directly at the front of the skull, and angle it so that looking into the sockets gets a little creepy, you've probably got it right. It will be important to do this for some of the data you're collecting today.

Next, it is time to start recording data. See the sheet at your lab bench for information on which primates your group is using. Coordinate efficient data collection with the group sitting across from you at the lab bench. Use the descriptions below to help you fill out Table 1.

6. Dentition

- a. **Total number of teeth.** Count the total number of teeth, top and bottom. Record this number on Table 1. Be aware that the lemur skull has several teeth modified to form a comb structure; you should still be able to count all of the individual teeth in the comb.

7. General cranium shapes

- a. **Post-orbital closure.** When you look into the front of the skull, is the area behind the eye socket mostly closed and forming a pocket the eyeball fits into, or is the area more open, with much less bone protecting the back of the eye? Write “open” or “closed” on the data sheet.
- b. **Skull ridge presence.** Are there one or more skull ridges on the top of the skull? (They may be small.) Write “yes” or “no” on the data sheet.
- c. **Two skull ridges.** Are there (at least) two entirely separate skull ridges on top of the skull, which do **not** meet? (Look carefully; they may appear to merge but not actually do so.) Write “yes” or “no” on the data sheet.
- d. **Shape of mandible.** Remove the cranium from the mandible (lower jaw bone) and set the mandible on the lab bench with the teeth pointing up. When you look at the angle of the row of teeth compared to the bench top, are they roughly parallel (a shallow rear mandible), or do the teeth slope down toward the front (because of a deep rear mandible)? Write “deep” or “shallow” on the data sheet.
- e. **Mental foramen location.** Locate the mental foramen (plural: foramina) on the mandible (Fig. 14). Note that on some of the skull reproductions (particularly the Lemur), the foramina can be difficult to find. In real skulls, these are holes for nerves to pass through. Counting from the center incisor (Fig. 13) on the same side as the foramen, is the foramen located beneath the fourth tooth back, or the fifth? Write “4” or “5” on your data sheet.

8. Size measurements.

- a. **Cranial capacity.** We will measure cranial capacity by carefully filling each skull with small, round mung beans (the source of bean sprouts), and then measuring the volume taken up by the beans. You will need to be extremely careful during this procedure, both to avoid harming the skulls and to avoid spilling the beans. **Do not use the real human skulls;** borrow a replica from another bench if you are collecting data on humans. Work over the lab bench, and use funnels as needed to fill the skull through the foramen magnum. For some of the larger skulls, you may need to plug the holes in the eye sockets with cotton to prevent the beans from spilling out.

You should fill the cranial cavity completely, with two exceptions. The orangutan (*Pongo*) and gorilla (*Gorilla*) skulls should only be filled so the back part of the skull is filled, not the snout. We are actually interested in the amount of space taken up by the brain, and the orangutan and gorilla brains do not go forward as far as the snout.

After you have filled the cranium with beans, start dumping them out into an empty plastic box on your lab bench. When you have emptied the skull, pour the beans from the box into a graduated cylinder. (You may empty a skull directly into a graduated cylinder if you prefer.) You’ll probably want to use a funnel again. Choose a graduated cylinder of a size appropriate to the size of your skull. If the cylinder is not big enough, fill it to the top line, record the volume, return those beans to their original container, and refill the cylinder. Keep track of the number of milliliters (mL) of beans it took to fill the skull.

A milliliter is equivalent to a cubic centimeter, so you should report your data as cm³. Believe it or not, the use of seeds to measure cranial capacity has been used historically by biologists to determine cranial capacity.

- b. **Skull length.** You will use calipers to measure the length of the skull from front to back. Find the large, curved, hinged calipers at your bench. Place one point of the caliper at the top, center tip of the nasal opening (on you, this would be the bridge of your nose, where the bone stops and cartilage begins). The caliper hinge should be above the center of the skull. Place the other point of the caliper at the center back of the skull, at approximately the same height as the front point of the caliper while the skull is held in the anatomical position with eye sockets oriented forward (Fig. 15). For skulls with a large ridge, you may need to use a pencil or ruler to help determine the proper horizontal distance.

Once you have the calipers adjusted properly, carefully remove the skull (without changing the angle of the calipers) and use a ruler to measure the distance between the points of the calipers. Record this distance, in centimeters, on the data sheet.

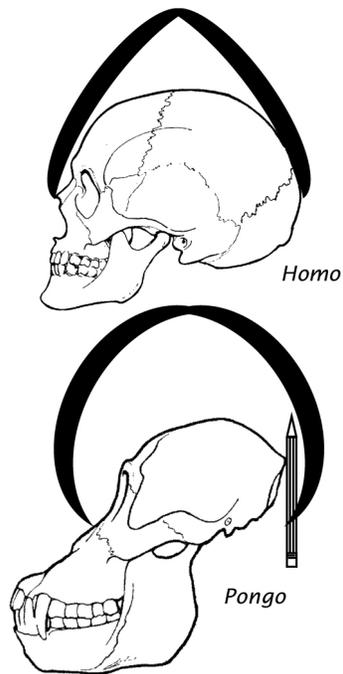


Figure 15. Proper measurement of total skull length. (Source: www.clipart.com)

9. Size ratios

We won't actually use the raw measurements we just made when we try to determine ancestral versus derived traits, because body size becomes a factor. We will use the total skull length to very roughly correct for differences in body size. There are more mathematically complex ways to do this, but using skull length will give us a crude idea.

- a. **Cranial capacity:skull length.** Divide the cranial capacity by the skull length and record the value on the data sheet.

10. Other

- a. **Tail.** Based on what you know about these organisms, the illustrations in the lab manual, and the pictures in lab, does the organism have a tail? Write "yes" or "no" on the data sheet.

11. Continue with Steps #5-8 as described above. Refer back to the dinosaur practice problem and the example snake problem as necessary.

Table 1. Primate Skull Morphology Data Collection. Be sure to number your primates consistently within your lab group. See above for complete descriptions of these characters.

| Measurement | Outgroup: <i>Lemur catta</i> | Your Primate I: | Your Primate II: | Your Primate III: | Your Primate IV: |
|-------------------------------------|---------------------------------|-----------------|------------------|-------------------|------------------|
| Dentition | | | | | |
| Total number of teeth | | | | | |
| General Cranium Shapes | | | | | |
| Post-orbital closure | | | | | |
| Skull ridge presence | | | | | |
| Two skull ridges | | | | | |
| Shape of mandible | | | | | |
| Mental foramen location | | | | | |
| Size Measurements | | | | | |
| Cranial capacity (cm ³) | | | | | |
| Skull length (cm) | | | | | |
| Size Ratio | | | | | |
| Cranial capacity : skull length | | | | | |
| Other | | | | | |
| Tail | | | | | |

Table 2. Character Matrix. Make sure all members of your group agree on the data collected in Table 1. Write an “A” (for “ancestral”) in each box which shares the Lemur data for that character; write a “D” (for “derived”) in each box which is different from the Lemur data. (See Steps #5 and #6 for instructions.) For character #7, check each datum to see if it is greater than or less than 12; if it is less than 12, like the Lemur, it should be considered ancestral.

| | 1. Number of Teeth | 2. Post-orbital closure | 3. Skull ridge presence | 4. Two skull ridges | 5. Shape of mandible | 6. Mental foramen location | 7. Cranial capacity ratio | 8. Tail |
|-------------------|--------------------|-------------------------|-------------------------|---------------------|----------------------|----------------------------|---------------------------|---------|
| Your Primate I: | | | | | | | | |
| Your Primate II: | | | | | | | | |
| Your Primate III: | | | | | | | | |
| Your Primate IV: | | | | | | | | |

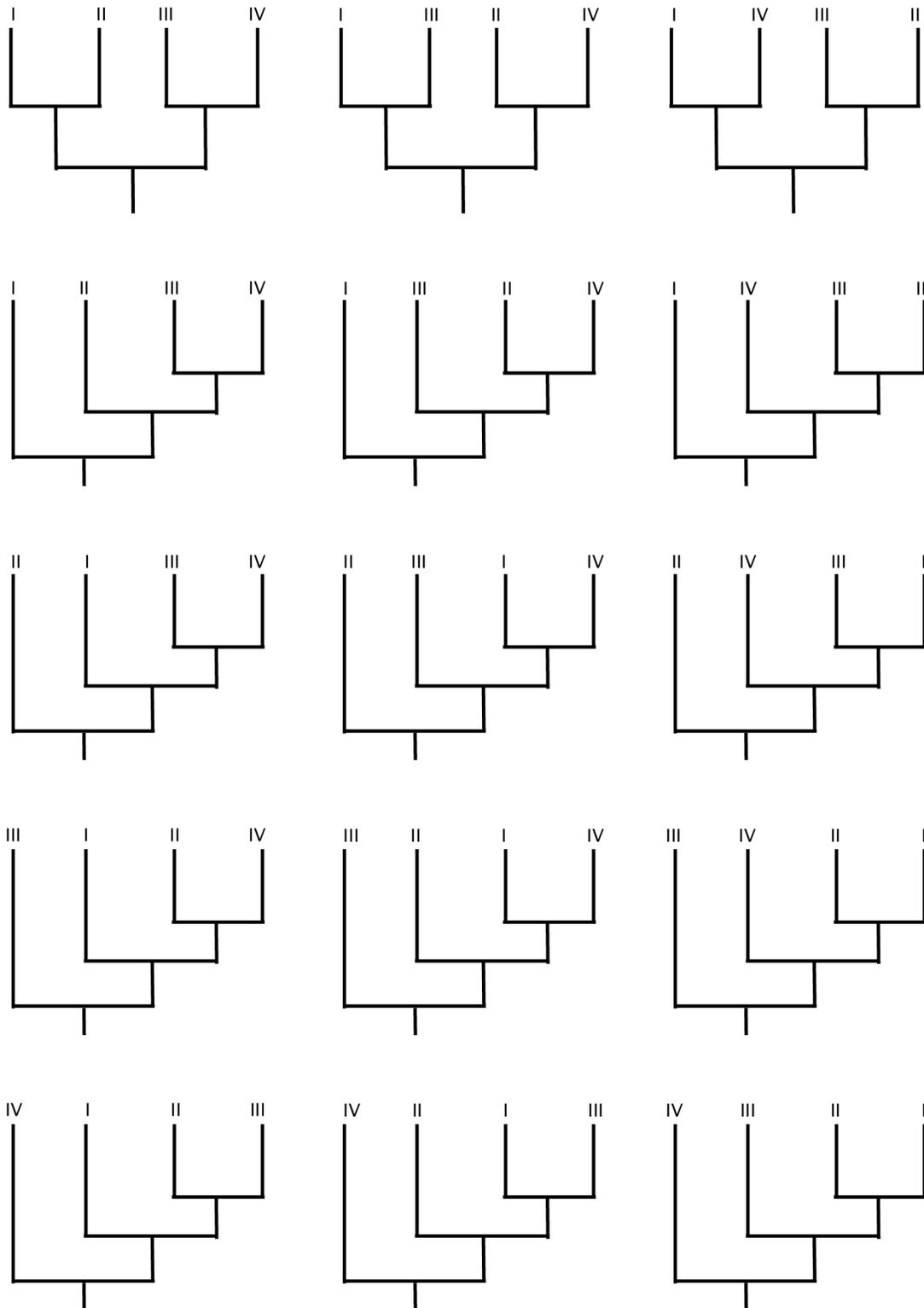


Figure 16. Possible Phylogenetic Relationships. Before working on these trees, check in with your lab instructor. When you're all set, see the instructions above for Step #7. Each time there is a change from A to D or D to A required based on your character matrix and the arrangement of primates on the tree, put a hash mark on the tree and indicate which character must change and what the change is. Count the number of changes required and write it below each tree. Prepare for class discussion as described above for Step #8.

Primate Phylogenetics Assignment: Using Amino Acid Sequences to Determine Phylogeny

Introduction

In this assignment, you'll continue to look at the evolutionary relationships between primates. Based on morphology alone, it is possible to determine the probable relationship between some of the genera, and less easy to discern the relationship between others (including *Homo*). Many people are interested in finding out which genera are most closely related to humans. As molecular data have become available, the picture may be getting clearer.

Why do we use molecular data (such as genetic sequences) to determine relationships? An organism's DNA is what is passed from one generation to the next, and evolution in the morphology of an organism occurs because of changes in the sequence of its DNA. These changes in DNA are called **mutations**. Looking at the differences and similarities of DNA sequences from different organisms is one way to get detailed information about their evolutionary relationships.

We can also use the amino acid sequences of proteins to develop phylogenies. Proteins are three dimensional molecules, but they are made up of a long chain of amino acids. The sequence of amino acids is determined by the sequence of bases in the DNA coding for the protein. A mutation in DNA may cause a change in the sequence of amino acids. Today we'll be comparing amino acid sequences for the same protein in different organisms to develop a phylogenetic tree.

What is known about primate phylogeny?

The lemurs are the members of a group called "**prosimians**" (pro: before, simia: ape), which are thought to have relatively more ancestral characters when compared with other primates (this is why we chose Lemur as our outgroup the first week of lab). Prosimians are found in Africa and Asia. The other primates we studied are all considered "**anthropoids**" (members of the suborder Anthroidea) (anthro: human, oid: like); they share relatively derived characters when compared with the prosimians. The anthropoids are commonly divided into three groups: the "New World monkeys," "Old World monkeys," and the "hominoids" (Fig. 17). The **New World monkeys** are those found in Central America and South America, and include *Alouatta* and *Lagothrix*. The **Old World monkeys** and the hominoids share some derived characters not found in the New World monkeys. Old World monkeys, which include members of the genus *Macaca*, are found in Africa and Asia. The **hominoids**, including *Gorilla*, *Homo*, *Pan*, and *Pongo*, share additional derived characters not found in Old World monkeys; the non-human hominoids (often referred to as **apes**) are found in Africa (e.g., *Gorilla* & *Pan*) and Asia (e.g., *Pongo*). The classification we are using is from a primate text by Fleagle (1999).

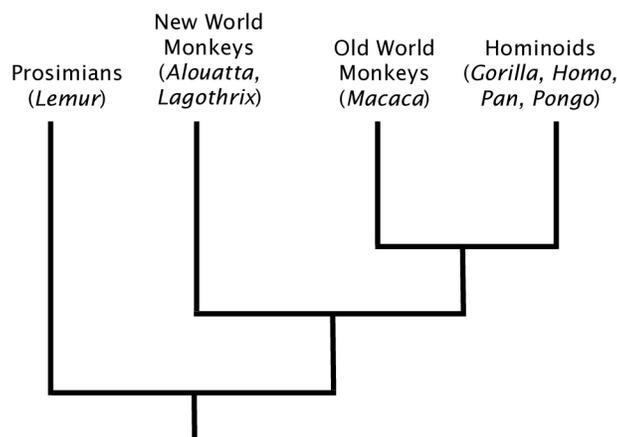


Figure 17. Phylogenetic relationships among primates.

Overview of Goals for This Assignment

One of your tasks will be to select a primate from a genus we did not study in lab and attempt to determine the group of primates (prosimians, New World monkeys, Old World monkeys, or hominoids) it is most closely related to. This will give you some practice comparing amino acid sequences using the amazing resources now available online. Another task is for you to use molecular data to determine the most likely relationship among the hominoids we looked at in lab (*Gorilla*, *Homo*, *Pan*, and *Pongo*).

We want you to understand the process of making a phylogeny using sequence data; our techniques will be relatively simple, but sometimes it is easy to lose sight of the big picture and get bogged down with when to click where. Try to be mindful of the process you're following. You will find fewer instructions as you progress through the assignment; we want you to get to a point where you could explore additional questions on your own if you want, rather than being tied to a particular sequence of mouse clicks.

Making Trees Using Sequence Data

Before using the computer, you should get a general sense of what this process will be like. After you have identified the organisms you're interested in, and determined what sequences to compare, **the first step** in making a tree is to find those sequences (we'll be doing this online). As an example, we downloaded the amino acid sequence for the same protein from a kangaroo and a wallaby (Fig. 18). The protein is sperm protamine P1, a structural protein that tightly binds to DNA in sperm and prevents new protein from being made using that DNA. We chose it because it is short, variable and has been sequenced for many organisms. Each amino acid has been given a one-letter abbreviation; Table 3 shows a key to the abbreviations.

Macropus rufus, red kangaroo:
 ARYRHSRSRSRSRYRRRRRRRSRYRSQRRRYRGRRRRRSRRGRRRGYSRRRYSRRRRRY

Macropus eugenii, tammar wallaby:
 ARYRHSRSRSRSRYRRRRRRRSRYRSRRRRSRGRRRRRSRRGRRRRRGYSRRRYSRRRRRY

Figure 18. Amino acid sequence of sperm protamine P1 in two marsupial species.

Table 3. Standard one-letter abbreviations for the amino acids.

| <u>Abbreviation</u> | <u>Amino Acid</u> |
|---------------------|-------------------|
| A | alanine |
| C | cysteine |
| D | aspartic acid |
| E | glutamic acid |
| F | phenylalanine |
| G | glycine |
| H | histidine |
| I | isoleucine |
| K | lysine |
| L | leucine |
| M | methionine |
| N | asparagine |
| P | proline |
| Q | glutamine |
| R | arginine |
| S | serine |
| T | threonine |
| V | valine |
| W | tryptophan |
| Y | tyrosine |

The **second step**, once you have downloaded your sequences, is to **align** the sequences. In addition to the possibility of a mutation occurring which changes one amino acid to another, it is possible for a mutation to insert an amino acid into a sequence or delete an amino acid from a sequence. In aligning the sequences, we try to factor in these types of mutations before we do an amino acid-by-amino acid comparison.

Figure 19 shows the same sequences in a format that is slightly easier to view. The numbers, written in columns above the

letters, can help you refer to particular amino acid positions by number. Note that the two sequences differ in length by two amino acids. If you were to compare these sequences as written in Figure 19, there would be many differences near the end of the molecule solely because the sequences are not properly aligned. Can you find the spots where the alignment needs to be adjusted by inserting spaces in the *M. rufus* sequence? (Try this on your own before reading further.)

| | | | | | | | |
|--------------------|------------|------------|------------|------------|------------|------------|---|
| | 1 | 1111111112 | 2222222223 | 3333333334 | 4444444445 | 5555555556 | 6 |
| | 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1 |
| <i>M. rufus:</i> | ARYRHSRSRS | RSRYRRRRR | RSYRSQRRR | YRGRRRRSR | RGRRRGYSR | RYSRRRRY | |
| <i>M. eugenii:</i> | ARYRHSRSRS | RSRYRRRRR | RSYRSRRR | SRGRRRRSR | RGRRRGYSR | RRYSRRRRR | Y |

Figure 19. Unaligned amino acid sequence of sperm protamine P1 in two marsupial species, with spacing and numbering.

Figure 20 shows the alignment we got using an alignment program available online. This program inserted a space, marked with a dash, at position 46 and at position 60. Note that we do not know for sure the locations; the first difference could just as likely be at position 43, 44, or 45. It will not change the results of our comparison, however.

| | | | | | | | |
|--------------------|------------|------------|------------|------------|------------|------------|---|
| | 1 | 1111111112 | 2222222223 | 3333333334 | 4444444445 | 5555555556 | 6 |
| | 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1 |
| <i>M. rufus:</i> | ARYRHSRSRS | RSRYRRRRR | RSYRSQRRR | YRGRRRRSR | RGRRR-GYSR | RRYSRRRRR- | Y |
| <i>M. eugenii:</i> | ARYRHSRSRS | RSRYRRRRR | RSYRSRRR | SRGRRRRSR | RGRRRGYSR | RRYSRRRRR | Y |

Figure 20. Aligned amino acid sequence of sperm protamine P1 in two marsupial species.

The **third step** is to determine the differences between the sequences. Looking at the alignment in Figure 20, how many differences do you see? Do not count the missing bases at positions 46 and 60 as differences.

Write the number of differences here: _____

So, now that you have determined the number of differences between the sequences, you can calculate the **proportional difference** between the two protein sequences. Divide the number of differences by the total number of amino acids in the shorter sequence. $\frac{\text{number of differences}}{\text{total number of amino acids}} = \text{proportional difference}$

You should get a result of 0.034. You can multiply that by 100% to think of this relationship as a 3.4% difference between the sequences.

In the tree-building technique we'll be using today, these proportional differences are calculated for each pair of species to be included on the tree. For example, Figure 21 shows an alignment for sperm protamine P1 for several mammalian species, and Figure 22 shows a table (called a distance matrix) with all the pairwise comparisons. Check Figure 22 to see how *M. rufus* and *M. eugenii* compare; is the number the same as our calculation above? Note that the table is mirrored diagonally through the middle; each species pair is listed twice.

| | | | | | | | |
|--------------------|------------|------------|------------|------------|------------|------------|----------|
| | 1 | 1111111112 | 2222222223 | 3333333334 | 4444444445 | 5555555556 | 66666666 |
| | 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 | 12345678 |
| <i>M. rufus:</i> | ARYRHSRSRS | RSRYRRRRR | ----RSYR | SQRRYRGR | RRRSRRGRR | -GYSRRYSR | RRRR-Y-- |
| <i>M. eugenii:</i> | ARYRHSRSRS | RSRYRRRRR | ----RSYR | SRRRSRGR | RRRSRRGRR | RGYSRRYSR | RRRRY-- |
| echidna | ARFRPSRSRS | RSLYRRRRS | RRQRSRRGR | QTGPRKITR | GRGKSR | RGRSMRSSR | RRRRRRN |
| platypus | ARFRPSRSRS | RSLYRRRRS | ----RRGGR | QTRSRKLSR- | SRRGRSRRR | KGRSRRSSR | RSRRN- |
| rat | ARYRCCRSKS | RSRCRRR-RR | ----RC--R | RRRRCCRRR | RR--RCCRRR | RSY-----T | FRCKRY-- |
| human | ARYRCCRSQS | RSRYRQ-RQ | ----RS--R | RRRRSCQTR | RRAMRCCRPR | --Y-----R | PRCRRH-- |

Figure 21. Aligned amino acid sequence of sperm protamine P1 in six mammalian species.

| | <i>M. rufus</i> | <i>M. eugenii</i> | echidna | platypus | rat | human |
|-------------------|-----------------|-------------------|---------|----------|-------|-------|
| <i>M. rufus</i> | 0.000 | 0.034 | 0.441 | 0.397 | 0.367 | 0.440 |
| <i>M. eugenii</i> | 0.034 | 0.000 | 0.426 | 0.400 | 0.340 | 0.400 |
| echidna | 0.441 | 0.426 | 0.000 | 0.200 | 0.540 | 0.600 |
| platypus | 0.397 | 0.400 | 0.200 | 0.000 | 0.531 | 0.571 |
| rat | 0.367 | 0.340 | 0.540 | 0.531 | 0.000 | 0.292 |
| human | 0.440 | 0.400 | 0.600 | 0.571 | 0.292 | 0.000 |

Figure 22. Distance matrix for sperm protamine P1 in six mammalian species. Each number represents the proportional difference between two species' sequences.

The **fourth step** is to use the information about how similar/different each pair of species is to construct a phylogenetic tree. This is not the same technique we used in lab. Determining ancestral and derived forms is relatively difficult for changes in amino acids or base pairs; problems with reversals and convergence are more common and complicate matters. So we will be using (or, more accurately, our computer program will be using) the Neighbor Joining method (Saitou and Nei, 1987), which places two species as far apart from each other on the tree as the difference in the distance matrix. So, two very similar species, like the wallaby and kangaroo, will appear close to each other on the tree.

The most striking difference between the trees we used in lab and the trees that you'll make today is in their overall appearance, and is due to the different method by which they are constructed. The trees you'll make for this assignment are **unrooted trees**; we aren't taking into account anything that is known about ancestral or derived characters, so we won't really be drawing in points of divergence that represent ancestors. We also are not limited to branches with two limbs only; three limbs may come off at the same point if distances are identical.

Figure 23 shows the results of using the difference information from Figure 22 and the Neighbor Joining method to generate a tree. Which organisms cluster together? How does this tree compare to Figure 24, which is the generally accepted tree for the major groups of mammals? (Check your textbook if you're not sure which organisms belong to which group.) We may not be able to glean as much information about evolutionary history from an unrooted tree like Figure 23, but we can use this quick method to get a good idea of relationships between organisms.

To predict where the root might be for the tree in Figure 23, we can include an outgroup on the tree; the point where the outgroup branches off represents a likely spot for the root of the tree. There are no sperm protamine P1 sequences for non-mammalian vertebrates, but sequences are available for *Octopus vulgaris* (the common octopus) and *Murex brandaris* (the purple dye murex, which is a type of snail). Compare the tree in Figure 25 to the other two trees; does the outgroup root the tree in Figure 25 in a way that is consistent with the tree in Figure 24? You might try thinking of the outgroup branch as the "trunk" of the tree; as you move away from the outgroup along the trunk, what group branches off first? Second?

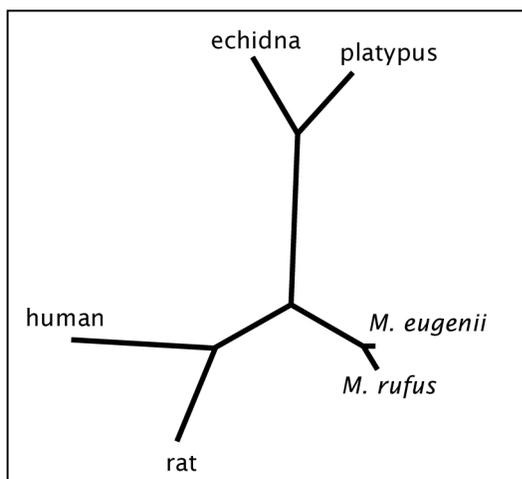


Figure 23. Tree of mammalian relationships based on sperm protamine P1 sequences.

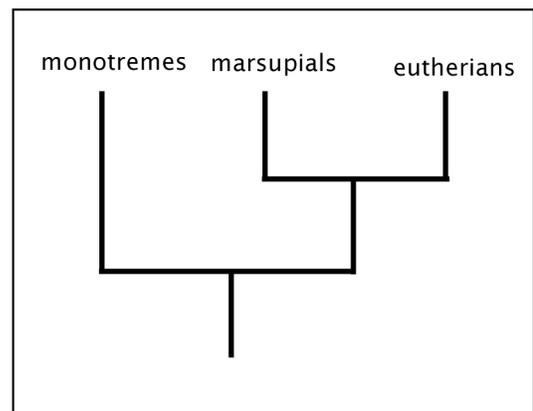


Figure 24. Tree of mammalian relationships based on morphological characters.

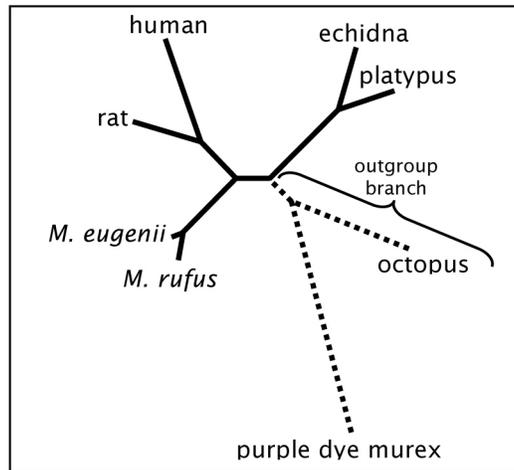


Figure 25. Tree of mammalian relationships based on sperm protamine P1 sequences, rooted with two invertebrates.

Using SWAMI to Construct Phylogenies

Here is an outline of your assignment:

- Part I: Tutorial: make an initial primate tree using cytochrome c oxidase protein sequences.
- Part II: Choose a different primate and make a new tree that includes this primate as well as the primates from Part I.
- Part III: Make your own hominoid tree using sequences from a new protein.

There is an Answer Sheet below that will be due at the beginning of lab next week.

Part I: Tutorial Tree Using Cytochrome C Oxidase Polypeptide II

You should work individually on this assignment; attend help sessions or contact your lab instructor with questions.

1. Get into SWAMI: The Next Generation Biology Workbench.

a. Type in the URL: <http://www.ngbw.org>

b. You will need to register to use SWAMI, but they do not require very much information, and they will not fill your email box. Click the “Use the Workbench now” button and then the “Create an account” button on the next page. Follow the online instructions to register. Register with a username and password you will remember; this will allow you to save information about your sequences on their web server and access them from any computer with an internet connection. *You will only be asked to provide your name, email address, a username, and a password.*

2. Learn how to get around in SWAMI.

a. You should be in a tab labeled “My Workbench.” Click on the text in the “**Create New Folder**” button to begin.

The buttons in SWAMI are a little tricky: you must click on the text within the button for them to work properly; if you click on the outside edge of the button, it will blink convincingly, but nothing else will happen.

Folders are used to organize data sets. You will likely want to store all your sequences in one folder; but you may decide to create a new folder for Part III below.

b. Give your folder a clear name, like “Primate Phylogeny.” You may add a description if you like, but it must be less than 100 characters.

c. In the left-hand column, you’ll notice that your folder contains two sections: “Data” and “Tasks.” The “Data” section is where you will store sequences. The “Tasks” section allows you to tell the software what to do with your sequences.

3. Find and save protein sequences (First Step).

- a. Navigate to the “Data” page for your new folder (link on left side of page).
- b. Click the “Search for Data” button.
- c. Try typing “cytochrome c oxidase” (without the quotes) into the top box. Set the Entity Type field to “Protein.” Set the Data Type field to “Sequence.” In the Data Sets field, select “Swiss-Prot,” which indicates you want to search this European database of protein sequences. Click the “Submit Search” button to begin the search.

Cytochrome c oxidase is actually a complex of several proteins working together; it is found in mitochondria, where it is an important component of the electron transport chain. The presence of cytochrome c oxidase and other components of the electron transport chain allows energy to be generated during the transfer of electrons from one molecule to another.

GenBank is a database maintained in the US; we’re using Swiss-Prot because their sequence-naming conventions are a little easier to work with.

- d. It may take a moment for the results to appear on the “Results Page” tab. How many results come back? SWAMI will only return a maximum of 500 sequences for a given search. Scroll through the list to get an idea of the different organisms available. Clearly, we need to limit our search. *Cytochrome c oxidase is one of the most commonly used proteins in sequence comparisons, because it is present in all eukaryotes and has changed little enough over evolutionary time that even distantly related organisms can be compared.*
- e. Click the “Basic Search” tab to go back to the search page. In the search box this time type “cytochrome AND c AND oxidase AND II AND primates” without the quotes. Be sure to capitalize the word “and.” You can leave the other settings as they were in step #3c. Click the “Submit Search” button. *You may have noticed that there were many different polypeptide sequences in the previous list; this is because cytochrome c oxidase is actually many proteins (each with its own sequence) working together. We are limiting our search in two ways: first, we are looking for only one cytochrome c oxidase sequence, “polypeptide II;” second, we are only searching for primate sequences.*
- f. Do you get a more manageable list of sequences this time? You can view all the results on one page by changing from “Show 20 results on each page” to “Show 100 results on each page.” Look carefully at the “Alternative ID” column, and notice that though many of the names say “COX2,” indicating we have the correct sequences for subunit II, some of the names are for “COX3,” which represents subunit 3.
- g. Look for sequences of cytochrome c oxidase polypeptide II from the following organisms in the genera we looked at in lab (be sure to choose the correct species):

Alouatta palliata
Gorilla gorilla gorilla
Homo sapiens
Lagothrix lagotricha
Lemur variegates
Macaca fascicularis
Pan troglodytes
Pongo abelii

Click the checkbox beside each of these; check marks should appear when they are selected. Be sure to only select Alternative ID names that begin with COX2. *The sequences you select will be available for aligning and tree-building. You’ll notice that some of these organisms have a third name (after genus and species); this is a subspecies name, and often denotes a geographic region where that organism came from. You can use the “find” feature in your web browser to help you find the correct sequences; look for it under the “Edit” menu.*

- h. Once you have selected all the sequences you are interested in, click the “Save Results” button at the top or bottom of the screen. Click the “Data” link in the left-hand column to see the list of your saved sequences.
- i. Double-check to make sure all eight of your sequences have an Alternative ID that begins with “COX2.”
 - *If you are missing a sequence or have made an error, you can return to your search results by clicking the “Search for Data” button at the bottom of the Data page, and then clicking the “Search Results” tab.*
 - *If you have saved a sequence in error, you can click the checkbox next to it to select it and then click the “Delete Selected” button at the bottom of the page.*

4. Look at a sequence.
 - a. On the Data page, click on the Label for one of your saved sequences. *It doesn't matter which sequence you select.*
 - b. Note the information available on the page: the first few lines give some information about the source and name of the sequence. These are followed by information about references related to the sequence. Further down the page you may see lines that include a description of the function of the protein. If you keep scrolling down to the bottom of the page, you will find the length of the protein (on the line starting with "SQ"), a number followed by "AA," which stands for "amino acids." The letters on the lines below that represent the sequence of amino acids. Each letter represents an amino acid; see Table 3 for a key.
 - c. Click the "Return to Data List" button to get back to the page with your data on it.
5. Align your protein sequences (Second Step).
 - a. Go to the Toolkit tab at the top of the page.
 - b. Go to the Protein Sequence Tools tab.
 - c. Click the ClustalW_P tool.
 - d. Type a brief description into the box (like "Part I Alignment") and click the Set Description button. *If you do not click the Set Description button now, you will have to re-type it in later.*
 - e. Click the Select Input Data button. This will take you to your list of saved sequences. Next, click the "Select all" checkbox, followed by the "Add Selected to Task" button. *If you have extra sequences in your data, which you do not want to include in the alignment, then click the checkboxes for the sequences you want individually rather than clicking the "Select all" checkbox.*
 - f. You should have been returned to the Task Summary tab. Click "Save and Run Task" to create the alignment. Click on the "Refresh Tasks" button at the top of the page every few seconds until your task's button in the "Action" column reads "View Output."
 - g. Click "View Output" button. This will show you a list of the files associated with the task; click the "View File" button for one of the grey lines, with the filename "outfile.aln." *If you don't see a line with this filename, check back over your work and make sure you performed all the steps correctly. Contact your lab instructor if the problem persists.*

Here's a key to interpret your alignment:

| <u>Genus/Species</u> | <u>Primary ID</u> | <u>Alt ID</u> |
|--------------------------------|-------------------|---------------|
| <i>Alouatta palliata</i> | P98024 | ALOPA |
| <i>Gorilla gorilla gorilla</i> | P26456 | GORGO |
| <i>Homo sapiens</i> | P00403 | HUMAN |
| <i>Lagothrix lagotricha</i> | P98036 | LAGLA |
| <i>Lemur variegates</i> | P98047 | LEMVA |
| <i>Macaca fascicularis</i> | P11948 | MACFA |
| <i>Pan troglodytes</i> | P50690 | PANTR |
| <i>Pongo abelii</i> | P92693 | PONAB |

As in the sperm protamine P1 example above, each row represents a sequence from a different organism. The name of each sequence is on the left, and starts with "ngbw|" followed by the Primary ID number from the table above. Because this sequence is quite long, it is scrolled onto multiple lines. For example, the human sequence (P00403) is four and a half lines long, and it appears at the top of the group of eight sequences in each of those lines. Dashes represent missing bases in the alignment. The bottom row below each group of sequences is a string of symbols that are associated with the alignment at each position in the amino acid chain. The different symbols can be used to get an idea of which areas of the protein are highly variable and which stay the same across all the genera you are looking at. People use the term "conserved" to refer to the parts of the sequence that do not vary, or vary little, between organisms.

CLUSTALW Alignment Symbol Key

Asterisk: the amino acid at this position is identical in all the species whose sequences you've aligned

Colon: the amino acids in this position are quite similar to one another in their chemical structure

Period: the amino acids in this position are somewhat similar to one another in their structure

No symbol: the amino acids at this position have quite different structures from one another

- h. Save this alignment by clicking the “Save to Current Folder” button at the top of the page. Choose a more informative name, like “Part I Alignment.aln,” and choose the following settings:

Entity Type: Protein

Data Type: Sequence Alignment

Data Format: Clustal

Finally, click the “Save” button.

6. View the proportional differences between each pair of organisms (distance matrix) (Third Step).
- Go to the Toolkit tab at the top of the page.
 - Go to the Phylogeny/Alignment Tools tab.
 - Click on the ClustalW_Dist tool.
 - Type a brief description into the box (like “Part I Distance Matrix”) and click the Set Description button.
 - Click the Select Input Data button. This will take you to your list of saved sequences. Click the checkbox next to the .aln file you just saved, followed by the “Add Selected to Task” button. *If you saved your .aln file but do not see it, try clicking on the “All Data” tab.*
 - You should have been returned to the Task Summary tab. Click “Save and Run Task” to create the distance matrix. Click on the “Refresh Tasks” button at the top of the page until your task’s button in the “Action” column reads “View Output.”
 - Click the “View Output” button. This will show you a list of the files associated with the task; click the “View File” button for the grey line, with the filename “infile.dst.” *This page shows you a table similar to the one in Figure 22 above, but the sequence names have not been repeated across the top (they are in the same order left to right as up and down).*
 - Take a minute to figure out how the table is set up; which two genera are the most similar to each other (have the lowest difference)? Which are the most different? Be sure to use the table above to convert the Primary ID information to genus names. Write the names and differences on the Answer Sheet below.
7. Make a phylogenetic tree using your data (Fourth Step).
- Click on “Tasks” in the left hand column.
 - Click the “View Output” button for the original alignment you performed, with the ClustalW_P tool. *You will need to use a file generated by this tool to construct your tree.*
 - Click the “View File” button in the “newtree” row. Click the “Save to Current Folder” button and rename the file something logical, like “For Part I Tree.dnd.” Click the “Save” button to save the .dnd file; you can leave the fields set to “Unknown.”
 - Go to the Toolkit tab, click to the Phylogeny/Alignment Tools tab if it is not already selected, and click on the Drawtree tool.
 - Set a description to something like “Tree for Part I” and click on the “Set Description” button.
 - Click the “Select Input Data” button. Click in the checkbox next to your tree (.dnd) file, and click “Add Selected to Task.”
 - Click the “15 Parameters Set” button and click the “Advanced Parameters” link. Change the “Angle of Labels” from M: Middle to A: Along. Click the “Save Parameters” button.
 - Click the “Save and Run Task” button.
 - Click the “Refresh Tasks” button until the “View Status” button becomes a “View Output” button for the Drawtree tool.

- j. To view the tree, once the task is complete, click the “View Output” button. Click the “Download File” button in the “psfile” row. Choose to open the file with Preview on a Mac. On a PC, save the file locally, then go to <http://view.samurajdata.se/> in your browser and upload the file to view it. You’re looking at a phylogenetic tree showing similarities in the sequence of the cytochrome c oxidase subunit II protein. *This tree is unrooted (we aren’t using an outgroup), and so is based only on the proportional difference between each organism. Remember that longer lines represent more differences between organisms.*
- k. Are the species arranged as you would have predicted? Which organisms are grouped together? Draw your tree and write your answers on the Answer Sheet. *Use the key above to determine which label corresponds to each organism.*
- l. Assuming that Lemur is still our outgroup, complete the tree on the Answer Sheet in the same geometric format we used the first week in lab. *See Figures 24 & 25 for help with this redrawing. You will be taking a tree that looks something like that in Figure 25 (variable branch lengths, random patterns) and making it look more like Figure 25 (a more ordered arrangement emphasizing evolutionary relationships rather than differences between organisms). It might help to start with the outgroup and follow the line from the outgroup to the other branches of the tree; the group which branches off first will be the lowest branch point on the tree. If you think of following the outgroup-branch like a trunk, watching where other organisms branch off it, that might help. Remember the trees can be rotated or pivoted at any branch point and it doesn’t change the phylogeny.*
- m. How does your tree compare to the trees we generated in lab? How does your tree compare to the tree in Figure 17? Write your answers on the Answer Sheet.
- n. What can you say about the relationship of the hominoids (gorillas, humans chimps, and oranges), based on these data? Which of the fifteen possible cladistic trees below are supported by this tree? Write the name of the protein (or an abbreviation, like “cyt c ox II”) beside each tree on the Answer Sheet which is consistent with the tree on your screen. *Keep in mind that rotation can occur at branch points of trees, and that the cladistic-style tree diagrams in the lab manual do not reflect difference as distance (the length of these branches is irrelevant).*
- o. Why might this tree not give you the definitive answer to questions about the evolutionary relationship of the hominoids? *Ponder this question; we’ll return to it in our discussion of the results from Part III.*

Part II: Determine the Relationship of an Unknown Primate Species Using Cytochrome C Oxidase Polypeptide II

8. Search for and save a cytochrome c oxidase polypeptide II sequence from some other primate genus.
- a. Return to SWAMI in your web browser and go to the Data page by clicking on the “Data” link in the left-hand column and click the “Search for Data” button near the bottom of the page. Repeat the search for “cytochrome AND c AND oxidase AND II AND primates” using the Swiss-Prot database as before. (Check the “Results Page” tab first: your earlier results may still be available.) *Refer back to the instructions in #3 as needed.*
- b. From the list of results, choose one cytochrome c oxidase polypeptide II sequence that is from a new primate genus (one we did not look at in lab). Select and save the new sequence. *The point of this is to make a new tree including this new primate, and to use this tree to predict whether the primate you’ve chosen is a prosimian (like Lemur), a New World monkey (like Alouatta and Lagothrix), an Old World monkey (like Macaca), or a hominoid (like Gorilla, Homo, Pan, and Pongo). You can choose any primate you are interested in. Some of the primates will be simpler to place than others. Going through this process will increase your comfort in SWAMI and cement your understanding of the procedures you’re following.*
- c. On the page with the list of your data, you will see the Primary ID and Alternative ID listed for your new organism. Write its scientific name and ID numbers below:

Genus & species: _____ Primary ID: _____ Alternative ID: _____

You may need to click on the label for your new organism to see the scientific name. The line with the scientific name will start with the code “OS.”

9. Make a new alignment including the new primate sequence.
 - a. Use the ClustalW_P tool to make an alignment that includes the new primate as well as the original eight sequences you saved earlier. *Refer back to #5 if you have questions. This tool is found in the Protein Sequence Tools tab.*
 - b. Take a look at the resulting alignment, then save it with a descriptive name (like “Part II Alignment.aln”).
10. View the new distance matrix.
 - a. Use the ClustalW_Dist tool to create a distance matrix for your new alignment (not the old one!). View the resulting distance matrix file. *Refer back to step #6 if you have questions.*
 - b. Based on the distance matrix, which organism is most closely related to your new primate? Write your answer on the Answer Sheet.
 - c. Based on the distance matrix, and your answer above, which group of primates do you predict your new primate belongs to? See Answer Sheet.
11. Make a new tree including all eight primates.
 - a. View the .dnd file from your alignment in step #9 and “Save to Current Folder.” *Refer back to step #7 for help.*
 - b. Use the Drawtree tool on the Phylogeny/Alignment Tools tab to create a new tree including your new primate. *Refer back to step #7 for help.*
 - c. Does the tree agree with your prediction in step #10? See Answer Sheet.

Part III: Use a Different Protein Sequence to Make a Hominoid Tree

The rest of the assignment will focus on trying to tease apart the relationships among the hominoid genera we looked at in lab.

12. You should choose ONE protein at random from the list below:

- 40S ribosomal protein S4 Y isoform 1
- Abnormal spindle like microcephaly associated protein
- ATP synthase protein 8
- Galactoside 2 alpha L fucosyltransferase 2
- ribonuclease like protein 9 precursor
- C5a anaphylatoxin chemotactic receptor
- cytochrome c oxidase subunit IV isoform 1
- fMet Leu Phe receptor
- FMLP related receptor I

Write your protein name on the Answer Sheet. *You might want to create a new SWAMI folder at this point; you can do this in the “My Workbench” tab.*

13. Search for and import sequences of your protein.

Use the Swiss-Prot database, and limit your search by including the name of the protein and the word “primates.” Connect all words with capitalized “AND.” All your sequences should have the same protein abbreviation (e.g. “COX2” for the protein sequences from Parts I and II) for part of the Alternative ID of their Swiss-Prot record.

If your search returns no results, check your spelling (including I vs. 1) carefully and retry the search.

Our goal is to develop trees for the hominoids, so you’ll want to find at least one sequence from each of those four genera. You need to include at least one other primate sequence as an outgroup. Use one of our genera (Macaca, Alouatta, Lagothrix, or Lemur) if these are available. (Hylobates is a gibbon genus, and is an acceptable outgroup if the others are not available.) When possible, avoid using sequences labeled “fragment.” You may include more than one species from each primate genus or additional genera if you like.

14. Keep a record of your organisms and ID names here:

| Genus & species: | Primary ID: | Alternative ID: |
|----------------------|-------------|-----------------|
| <i>Gorilla</i> _____ | _____ | _____ |
| <i>Homo</i> _____ | _____ | _____ |
| <i>Pan</i> _____ | _____ | _____ |
| <i>Pongo</i> _____ | _____ | _____ |
| _____ | _____ | _____ |
| _____ | _____ | _____ |
| _____ | _____ | _____ |
| _____ | _____ | _____ |
| _____ | _____ | _____ |
| _____ | _____ | _____ |

15. Align your sequences using ClustalW_P. Save the alignment (.aln) file.

16. Save the .dnd file from your alignment and make a tree using the Drawtree tool. Refer back to step #7 for help.

Sketch this tree on your Answer Sheet, using genus names (include species if you have multiple species from a single genus).

17. Using the possible trees on the Answer Sheet, determine which are supported by your protein sequence data. Write the name of your protein by each tree on the Answer Sheet which is consistent with the tree on your screen.

18. Assess your data.

- a. Which primates did you use to build your tree? Which served as an outgroup? Record your answers on the Answer Sheet.
- b. How long were your sequences? Record your answer(s) on the Answer Sheet. Remember you can look at a sequence by clicking its Label on the Data page. Don't count them all—the number you're looking for is right before "AA," which stands for "amino acids."
- c. Try to get an idea of the variation among your sequences; use the alignment results and the ClustalW_Dist tool to try and decide how much confidence you have in your tree. Get a sense of how many amino acid differences are affecting your tree. Why might this matter? Record your answers on the Answer Sheet.

If you are curious & have extra time...

- Do a Google image search for any of the primates you used in your analysis to see what they actually look like.
- Check out "PRIMATES: The Taxonomy and General Characteristics of Prosimians, Monkeys, Apes, and Humans." It's a great site with not-too-much-text and lots of images; it will allow you to quickly reinforce what you've learned about primate phylogenetics. <http://anthro.palomar.edu/primate/>

References

Fleagle, J.G. 1999. *Primate adaptation and evolution*. Second edition. Academic Press, San Diego, 596 pages.
 Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4: 406-425.

Materials

The equipment and supplies are minimal, beyond the large one-time investment of the skull reproductions.

For eight lab groups of two to four students per group, assuming two groups can share each skull:

- 3 *Alouatta palliata* (howler monkey) skulls (Skulls Unlimited BC-017, \$168 US)
- 4 *Gorilla gorilla* (gorilla) skulls (Skulls Unlimited BC-001, \$270 US)
- 2 *Homo sapiens* (human) skulls (Skulls Unlimited SP-01, \$99 US)
- 4 *Lagothrix lagotricha* (woolly monkey) skulls (Skulls Unlimited TQ-245, \$92 US)
- 4 to eight *Lemur catta* (ring-tailed lemur) skulls (Skulls Unlimited BC-087, \$119 US) (we use one skull/lab group (eight), but groups could share(four))
- 3 *Macaca mulatta* (rhesus monkey) skulls (Skulls Unlimited BC-137, \$160 US)
- 4 *Pan troglodytes* (chimpanzee) skulls (Skulls Unlimited BC-003, \$218 US)
- 4 *Pongo abelii* (orangutan) skulls (Skulls Unlimited BC-002S, \$245 US)

For each lab group of two to four students:

- large foam pad for setting skulls on (optional)
- approximately 2 Lof mung beans (these are bean sprout beans, and will last many years if they stay dry)
- 1 caliper for measuring skull length (we use 12-inch aluminum calipers like those from Bailey Ceramic Supply C-090-138, \$6.50 US)
- cotton balls (approximately 10)
- 12-inch ruler, metric
- blunt probe
- large plastic beaker or spouted cup (3-L capacity, for holding mung beans)
- 1000-mL graduated cylinder (plastic preferred, for measuring mung beans)
- 250-mL graduated cylinder (plastic preferred)
- 25-mL graduated cylinder (plastic preferred)
- 2 to 3 funnels with a variety of sized openings, at least 1 cm at the small end (small end needs to fit into foramen magnum of skull)
- large, flat plastic bin or box (for dumping beans out of skulls easily)

Caution:

SWAMI has been increasingly difficult to use over the last year. The site has become uncomfortably slow, or gone down while students were using it. The site managers have been quick to reboot, and are hoping to upgrade shortly. The author is working to find alternative programs as well; contact her for more information.

Notes for the Instructor

Big Picture Information

This week in lab, students have several things to accomplish:

- collect morphological data on skull anatomy
- find most parsimonious tree for four primates
- discuss (as a lab) the most parsimonious tree for seven primates

The primary goal for the lab section is to develop a phylogeny for seven primate species, using an eighth species as an outgroup. In the process, we want students to be able to interpret phylogenetic trees and explain how cladistics is used to create trees (especially that organisms are grouped together based on shared, derived characters).

Students should work in groups of 2-4, but for data collection, a pair of groups (4-8 students) should work together and coordinate their efforts. We typically try to have eight groups of students, even if there are several groups of 2, so we can collect more data. Each lab group will be finding the most parsimonious tree for a different subset of four primate species. The four species for each lab group should be listed at each bench (Fig. 28). Lab groups will need to share two of their skulls across the lab bench. Encourage students at a bench not to duplicate work with the people sitting across from them.

Because of all the background reading for the lab, we typically do not provide much of a verbal intro for the lab activities, other than to map them out. Students should have worked through the dinosaur practice problem before coming to lab; you might go around and see if they have questions about the dinosaur trees as they start working on the primate trees (though you may not have time to do this). Students should follow the instructions in the lab manual to collect data about the skulls, then figure out what is derived and what is ancestral, and finally find the most parsimonious tree for their genera.

One suggestion for running the lab smoothly is to have the students stop once they have determined “ancestral” and “derived” character states for each of their organisms, collect data from the entire lab on the board or an overhead sheet, and check to see if everyone is in agreement. Not everyone will get to this point at the same time, but it makes for a good snack-break time. If there is disagreement, the students need to go back and reevaluate their data, and then come to some consensus. This provides a great opportunity for communication between different groups in the lab section. If there is a problematic character, as long as there is some consensus, it is okay if the class’ results for this trait disagree with the key below.

Once there is a consensus, students can go ahead and start determining the most parsimonious tree for their four genera. This is the tough part for some students. It is helpful if lab instructors go around and really engage students in conversation once they’ve started; if they are struggling, you

might start by helping them “eliminate” some of their traits from consideration—those that are all ancestral or all derived, and those which only have one organism with a derived trait will never have an effect on the tree. Do this by asking the students questions and letting them work it through. If some group members seem to be really struggling, you might sit down with them and make sure they understand Figure 3 in the lab manual—do they get what the hash marks represent? If a whole group has questions, and one person gets things figured out while talking to you, have that person explain it to their lab partners—it will help solidify that student’s understanding and give other students in the group another way of getting at the information. If the dynamics of the group seem tenuous, stay there while the student explains.

After all the groups have completed their 15 trees, the instructor should run a class discussion with the goal of figuring out what the most likely tree is for all 7 genera. I suggest having all the groups draw their most parsimonious tree(s) on the board. Then, I usually ask people to identify which are most closely related, and how we might arrange the organisms on a single tree. The main patterns will probably be quite obvious. Groups will need to be able to explain and defend their trees. Do not expect to come up with a single, neat tree; you may want to talk about (or have students talk about) what data you have confidence in and what data you don’t, and why. Typically, students arrive at a consensus tree that corresponds to the accepted phylogeny for these organisms, with the exception that they can’t really tease apart gorillas, orangutans, chimps, and humans. That’s fine—their homework assignment, based on genetic sequences, will focus on these organisms.

Information to Share With Students

- Make sure students know which genera they are focusing on, and that they can share skulls (and data collection responsibilities!) with the group they are paired with.
- Remind students not to mark on skulls.
- If you have real human skulls available, tell students to be particularly careful with them.
- Suggest students use pencil to fill out their data tables.
- In Table 3 I’ve listed the size measurements I recorded on single skulls. You can use my data as a general guide, and check for ballpark agreement, but students may come to slightly different conclusions based on their own data. Again, this is fine, and illustrates why multiple specimens are so important.
- For cranial capacity, we’ll have several sizes of funnels and graduated cylinders; tell them to choose the combination that works best for each skull. We’ll also have large plastic containers for dumping skulls into, rather than trying to pour them directly into funnels (student choice).
- Warn students to be careful to clean up spilled mung beans, which are likely to be hazardous on the floor. (Think of walking on ball bearings...)

- In the discussion, the status of *Alouatta*, *Lagothrix*, and *Macaca* will most likely not be in question. The other four genera are pretty impossible to tease apart, though. It is FINE to let this remain unsettled: discerning the relationship among these four is what the students will do for their post-lab assignment (using molecular data). So this can be a puzzle: it’s okay.

Common Student Misconceptions About Phylogenies

Students often have some difficulty understanding how to interpret phylogenies. The use of square bracket phylogenies in these handouts is intentional, since it has been shown that they are easier to understand than v-shaped trees when learning about phylogenies (Novick et al., 2011).

One common point of confusion for students is the idea that moving along lines on the tree represents moving through time (moving “up” a line on the tree is moving from more ancient to more recent evolutionary history). I help students work through this by talking with them about the tree in Figure 2 of the student handout. There is a similar figure in the key for the dinosaur practice problem.

Another major misconception students often have is that the order of branch endpoints on a phylogeny is meaningful. In reality, the same phylogeny can be drawn in a number of ways, and the important point is which lines share the most recent common ancestor. I often teach this point by talking about mobiles; if you invert a phylogeny and imagine hanging it up like a mobile, the branches can swing around freely but the relationship being described remains the same. Often this point of confusion arises when students are working on placing hash marks on their trees and claim that there should be more than 15. I typically ask them which tree is “missing” and then explain how that tree is actually already represented by another tree already on the sheet.

The third misconception is more subtle, but is common even among some of my best students. Students often try to say that organisms connected on a tree by fewer “nodes” or branch points are more closely related than those that are connected by more branch points. The key here is for

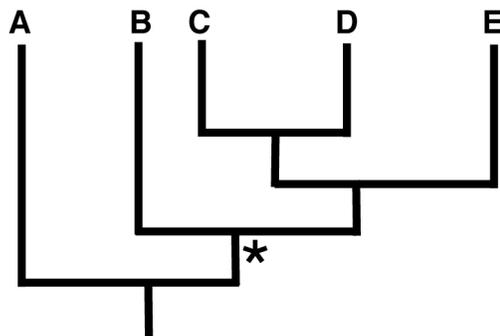


Figure 26. Phylogenetic tree illustrating common student misconception.

them to realize that the location of the Most Recent Common Ancestor is the relevant piece of information, not how many branches are on a tree. For example, in the tree below (Fig. 26), organisms B and C share a Most Recent Common Ancestor at the star. A is more distantly related to both organisms, because the common ancestor for A with B or C occurs earlier in the tree (below the star). Students will often claim that A and B are more closely related to each other than A and C, because only two “nodes” separate A and B, while there are four nodes separating A and C. I think it will be possible to help correct this misconception by providing students with a problem utilizing a tree with abstract or unfamiliar organism names (or just letters like in the tree below) and then showing them the same tree with familiar groups of organisms, so they are encouraged to confront the misconception and correct it.

For more comprehensive information about student misconceptions around reading phylogenies, I highly recommend the very readable (and Open Access!) paper “Understanding Evolutionary Trees” (Gregory, 2008).

Primate Background for Instructors (don’t share with students before lab)

The genus we will be using as an outgroup is *Lemur*, a strepsirrhine primate. Primates can be separated into two

main groups, strepsirrhines and haplorhines. All the primates that we are trying to develop a phylogeny for are haplorhines. I’ve drawn a simple version of the currently accepted tree (including lemurs) below (Fig. 27); you can find a recently published, more complete primate tree in Springer et al. (2012).

Alouatta and *Lagothrix* are new world monkeys (Platyrrhini), a large group of primates found in Central and South America. All the primates in this hemisphere belong to this group (strepsirrhinians and catarrhines are only found in Africa, Asia, and Europe). *Macaca* is a representative of another large group of primates, the old world monkeys (Cercopithecoidea). The apes are actually a very small group of organisms; in addition to the four genera we are looking at in lab (assuming you include *Homo* as an “ape”), there is one other extant ape genus, *Hylobates* (the gibbons). (Gibbons are sometimes called “lesser apes,” you might have heard oranges, gorillas, and chimps referred to as the “great apes.”) Gibbons and oranges are found in Asia; gorillas and chimps are found in Africa. I probably will not use the term “ape” in lab; I use “hominoid” instead, because it is easier for people to feel comfortable applying that term to humans.

Here’s the arrangement of skulls the student groups will use in lab (Fig. 28). Skulls in boldface will need to be shared across the bench. Extra skulls will be on the side bench.

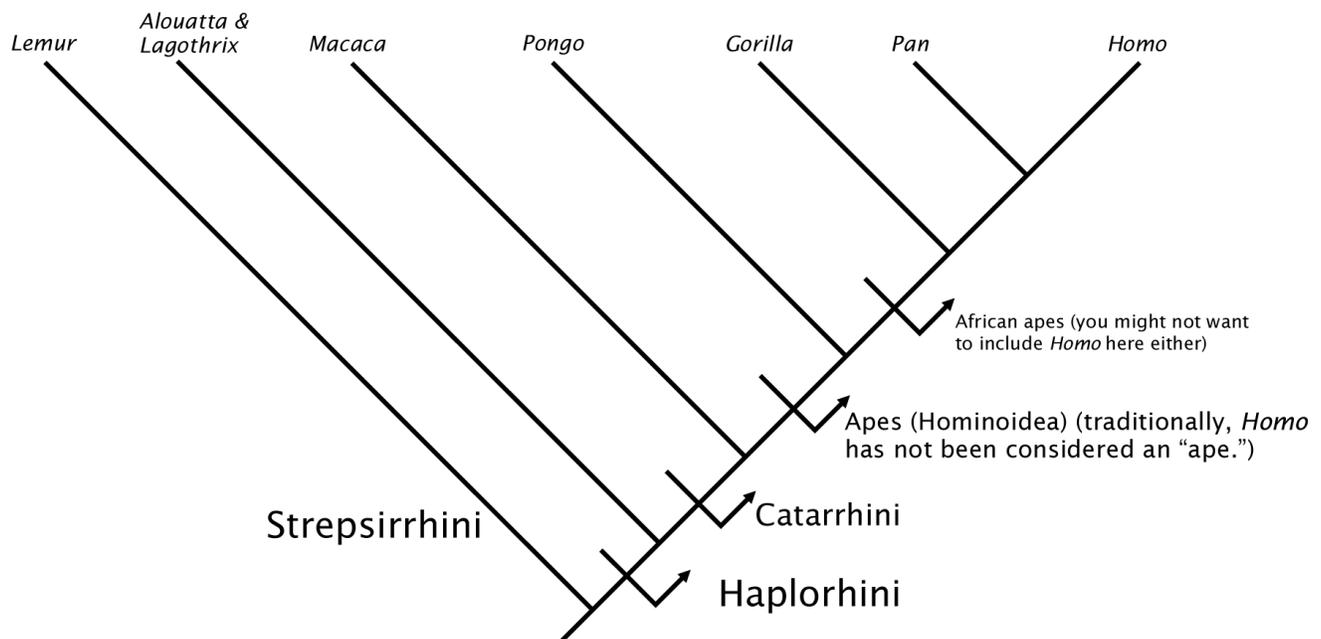


Figure 27. Major groups of primates, based on Perelman et al. (2011).

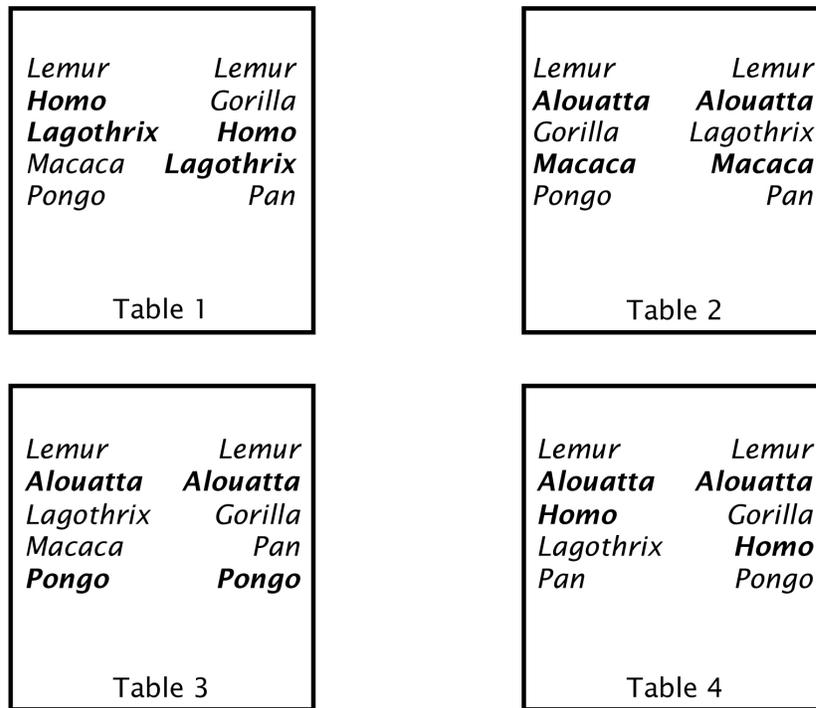


Figure 28. Diagram of skull arrangement for lab.

Table 3. Key for Primate Skull Morphology Data Collection.

| Measurement | Out-group: <i>Lemur</i> | <i>Alouatta</i> (howler) | <i>Gorilla</i> (gorilla) | <i>Homo</i> (human) | <i>Lagothrix</i> (woolly monkey) | <i>Macaca</i> (macaque) | <i>Pan</i> (chimp) | <i>Pongo</i> (orangutan) |
|-------------------------------------|----------------------------|-----------------------------|-----------------------------|------------------------|-------------------------------------|-------------------------------|-----------------------|-----------------------------|
| Dentition | | | | | | | | |
| Total number of teeth | 36 | 36 | 32 | 32 | 36 | 32 | 32 | 32 |
| General Cranium Shapes | | | | | | | | |
| Post-orbital closure | open | closed | closed | closed | closed | closed | closed | closed |
| Skull ridge presence | yes | yes | yes | no | yes | yes | yes | yes |
| Two skull ridges | yes | yes | no | no | yes | yes (but tough; okay if “no”) | no | no |
| Shape of mandible | shallow | deep | shallow | shallow | deep | shallow | shallow | shallow |
| Mental foramen location | 5 | 4 | 5 | 5 | 4 | 5 | 5 | 5 |
| Size Measurements | | | | | | | | |
| Cranial capacity (cm ³) | 14 | 33 | 591 | ~1500? | 88 | 77 | 459 | 310 |
| Skull length (cm) | 7.5 | 10 | 19.2 | 18 | 9.1 | 11 | 13.5 | 14.5 |
| Size Ratios | | | | | | | | |
| Cranial capacity: skull length | 1.87 | 3.3 | 30.8 | ~83+ | 9.67 | 7 | 34 | 21.4 |
| Other | | | | | | | | |
| Tail | yes | yes | no | no | yes | yes | no | no |

Table 4. Key for Character Matrix.

| | 1. Number of Teeth | 2. Post-orbital closure | 3. Skull ridge presence | 4. Two skull ridges | 5. Shape of mandible | 6. Mental foramen location | 7. Cranial capacity ratio | 8. Tail |
|--|--------------------|-------------------------|-------------------------|---------------------|----------------------|----------------------------|---------------------------|---------|
| <i>Alouatta</i> howler monkey | A | D | A | A | D | D | A | A |
| <i>Gorilla</i> gorilla | D | D | A | D | A | A | D | D |
| <i>Homo</i> human | D | D | D | D | A | A | D | D |
| <i>Lagothrix</i> woolly monkey | A | D | A | A | D | D | A | A |
| <i>Macaca</i> macaque | D | D | A | A (D ok) | A | A | A | A |
| <i>Pan</i> chimp | D | D | A | D | A | A | D | D |
| <i>Pongo</i> orangutan | D | D | A | D | A | A | D | D |

Acknowledgements

The premise for this lab was based on a “Phylogenetic Systematics” lab taught at Duke University. We gratefully acknowledge their idea of using primate skulls and genetic information to teach students about phylogeny.

I appreciate the support and encouragement I have received from members of the Biology Department at Carleton College. I would especially like to thank Mark McKone for his perceptive feedback on drafts and his help developing appropriate examples.

Pre-Lab Assignment questions 1 and 2 were modified (only slightly) from the Basic Tree Thinking Assessment in the supporting online material from Baum et al. (2005). This assessment is a great resource for simple questions to find out where students are confused about phylogenies.

The phylogenetic trees in the student handouts are displayed as vertical lines connected by horizontal lines; the choice of this arrangement over V-shaped trees was intentional, and based on research by Novick et al. (2011) on how students understand these diagrams.

Literature Cited

- Apestequia, S., and H. Zaher. 2006. A Cretaceous terrestrial snake with robust hindlimbs and a sacrum. *Nature*, 440: 1037-1040.
- Baum, D. A., S. De Witt Smith, and S. S. S. Donovan. 2005. The tree-thinking challenge. *Science*, 310: 979-980.
- Campbell, N. A., J. B. Reece, and L.G. Mitchell. 1999. *Biology*. Fifth edition. Benjamin-Cummings, Menlo Park, California, 1175 pages.
- Fleagle, J.G. 1999. *Primate adaptation and evolution. Second edition*. Academic Press, San Diego, 596 pages.
- Gregory, T. R. 2008. Understanding Evolutionary Trees. *Evolution: Education and Outreach*, 1: 121-137.
- Novick, L. R., C. K. Shade, and K. M. Catley. 2011. Linear versus branching depictions of evolutionary history: Implications for diagram design. *Topics in Cognitive Science*, 3: 536-559.
- Perelman, P., W. E. Johnson, C. Roos, H. N. Seuánez, J. E. Horvath, M. A. M. Moreira, B. Kessing, J. Pontius, M. Roelke, Y. Rumpler, M. P. C. Schneider, A. Silva, S. J. O’Brien, and J. Pecon-Slattery. 2011. A molecular phylogeny of living primates. *PLoS Genetics*, 7: e1001342.
- Weins, J. J., M. C. Brandley, and T. W. Reeder. 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution*, 60: 123-141.

APPENDIX A

Supplier Addresses

Bailey Ceramic Supply

P.O. Box 1577
62 Tenbroeck Ave.
Kingston, NY 12402
Phone: 1-800-431-6067

Web: <http://www.baileypottery.com/>

Skulls Unlimited

10313 South Sunnyslane
Oklahoma City, OK 73160
Phone: 1-800-659-7585

Web: <http://www.skullsunlimited.com/>

APPENDIX B

Key for Practice Problem

Dinosaur Problem: Hints and an Example Tree

If you are having difficulty with the dinosaur practice problem, read through these pages, especially the example tree we filled out.

To fill out the table in Step #5, look at the data in Step #4: if the character state for *Allosaurus*, the outgroup, is “no,” then that state is ancestral (and “yes” is derived). Remember that this is by definition: we chose *Allosaurus* as the outgroup because we know that it has relatively ancestral character states in comparison with these other dinosaurs. Basically, we are using *Allosaurus* as a tool in this step.

If your answers in Step #5 are correct, this character matrix in Step #6 should be fairly straightforward to complete. Note that the genera are now listed in a column, *Allosaurus* is missing, and we begin referring to the characters by number for shorthand. Before going on to Step #7, or if you are having trouble with Step #7, we suggest you check your answers to Steps #5 and #6 below. Once you are sure your answers for #5 & #6 are correct, you’ll need to be very methodical about placing hash marks on your trees. Remember that in Step #7 of the dinosaur problem, we are only asking you to evaluate 5 of the 15 possible trees that describe the relationship between these dinosaurs. This example tree (Fig. B1) is merely a sixth tree from the 15 possible:

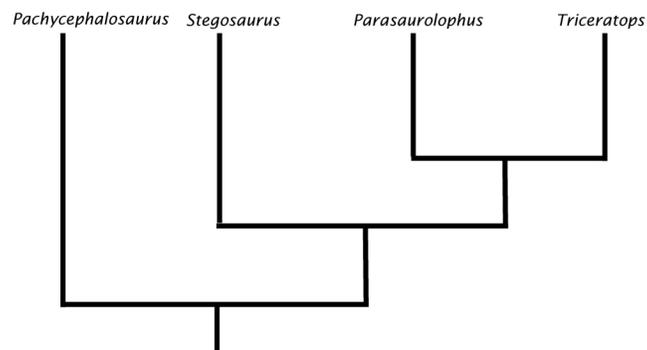


Figure B1. Example tree to practice hash mark placement.

To begin placing hash marks on the tree, start with one character and look at the pattern in Step #6. For example, character 1 is ancestral for *Pachycephalosaurus*, *Parasaurolophus*, and *Triceratops*, but derived for *Stegosaurus*. This means you will need to place a hash mark such that only *Stegosaurus* is derived. The hash mark represents an evolutionary change from the ancestral to the derived state (i.e. this is where armored plates arose, for character 1). On the example tree (Fig. B2), you would put the hash mark here:

We’ve added solid grey arrows to represent the ancestral state, and after the hash mark, the dotted arrow represents the derived state.

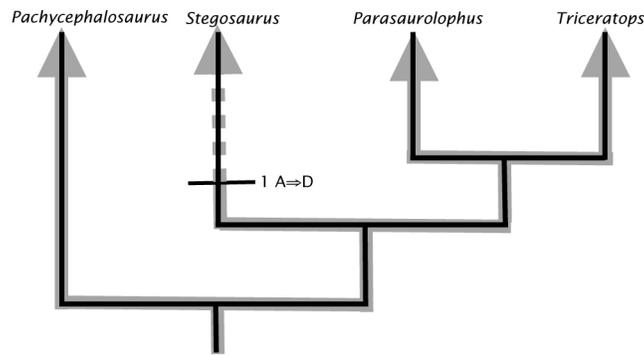
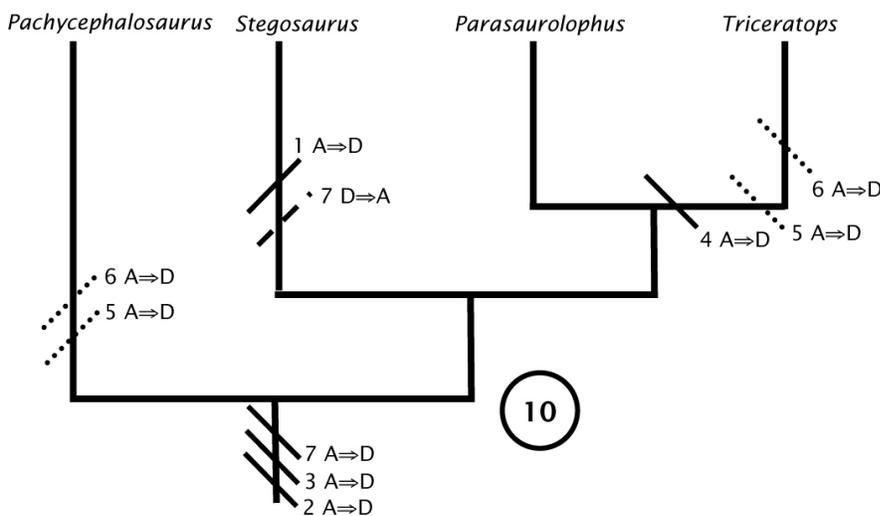


Figure B2. Example tree showing first hash mark placement and “flow” of traits up the tree.

It does not matter where along this branch you place the hash mark. You should realize that placing the mark on any other branch will increase the number of changes required (it will require reversals). Always choose placement such that it invokes the fewest changes on the tree. In some cases, there may be multiple ways to handle hash mark placement with the same number of changes: choose your favorite.

Continue looking at each character individually and placing hash marks on the tree; repeat this process for each of your trees. You may find that some trees require you to draw multiple “A ⇒ D” hash marks (these represent convergence), and at other times you will need to make a “D ⇒ A” change (this represents a reversal).

We completed the example tree (Fig. B3) so you can see what it looks like:



A solid hash mark represents a single change from ancestral to derived.
 A longer-dashed hash mark represents a reversal from derived to ancestral.
 A shorter-dashed hash mark represents a convergence: multiple changes from ancestral to derived.
 The total number of changes required in the tree is in circled.

Figure B3. Example tree completed in the most parsimonious way.

You do not need to code your hash marks with dotted lines; we just wanted to emphasize what is occurring in the tree. Note that this is drawn with a reversal in character 7 and two converging characters: 5 and 6. The reversal could be diagrammed equally effectively as a convergence, with no effect on the total number of changes required. The convergences could be drawn as reversals, but this adds changes to the tree. Below (Fig. B4) is the same tree with those adjustments:

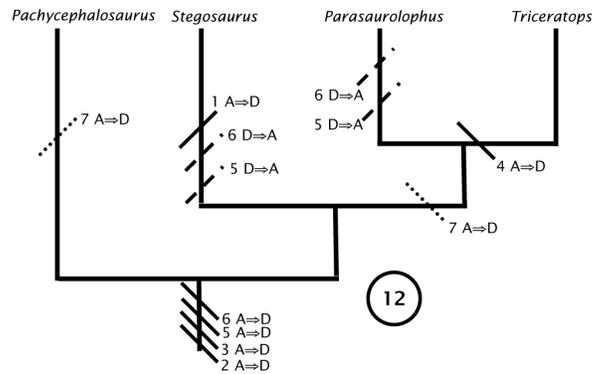


Figure B4. Example tree completed with reversals in place of convergences, less parsimonious in this case.

Dinosaur Practice Problem: Key for Steps 5 and 6

Step #5: Determine which character states are ancestral and which are derived using outgroup data.

| | Ancestral | Derived |
|---------------------|-----------|---------|
| 1. armored plates | no | yes |
| 2. preentary | no | yes |
| 3. pubis ext. post. | no | yes |
| 4. rostral | no | yes |
| 5. skull shelf | no | yes |
| 6. symphysis | yes | no |
| 7. enamel | no | yes |

Step #6: Fill out character matrix.

| Character: | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------------------|---|---|---|---|---|---|---|
| <i>Pachycep.</i> | A | D | D | A | D | D | D |
| <i>Parasaur.</i> | A | D | D | A | A | A | D |
| <i>Stegosaur.</i> | D | D | D | A | A | A | A |
| <i>Tricerat.</i> | A | D | D | D | D | D | D |

A=ancestral, D=derived

Dinosaur Practice Problem: Key for Steps 7 and 8

Step #7: Use character matrix to analyze possible trees.

In this example, rather than making you analyze all 15 possible trees, we have drawn a subset of 5 trees to analyze. You should get the idea from these, but you can always set up the other 10 trees on your own if you like.

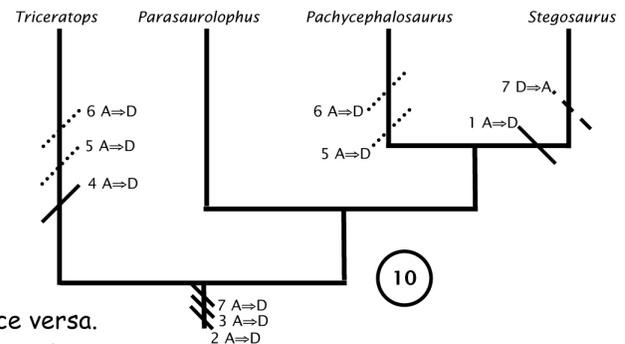
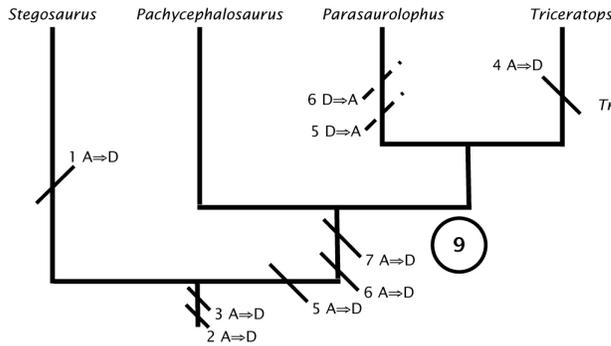
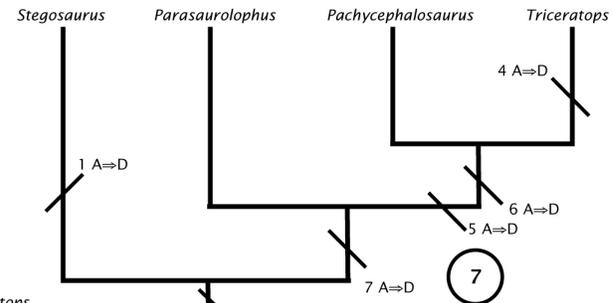
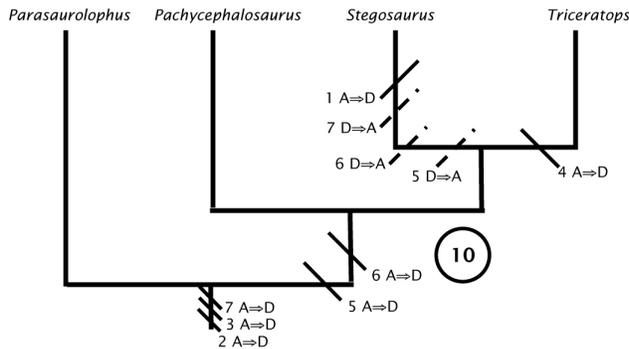
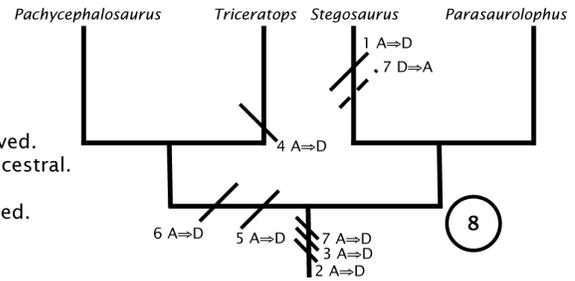
For each tree, you will need to determine where state changes must have occurred. In some cases, you will need to decide whether a convergence or a reversal is necessary; choose the option which puts fewer changes on the tree. (If the number of changes is identical, it doesn't matter which you choose.)

Review the trees in Figure 3 if you have questions.

Step #8: Determine which tree is most parsimonious.

After you have counted up the changes required for each tree, find the tree with the smallest number. Congratulations! If you haven't already, step back and look at how much each character contributed to your selection of that tree. Were there characters that had the same effect on all the trees? What are the implications of characters like those? Which characters were the most important in determining your phylogeny?

A solid hash mark represents a single change from ancestral to derived.
 A longer-dashed hash mark represents a reversal from derived to ancestral.
 A shorter-dashed hash mark represents a convergence:
 multiple changes from ancestral to derived.
 The total number of changes required in the tree is in circled.



You could redraw the reversals as convergence, and vice versa.
 You might try this for practice; it may or may not affect the
 total number of changes in the tree.

Figure B5. Completed trees from the dinosaur practice problem

The most parsimonious tree is the one shown in Fig. B6. It requires only 7 changes, and no reversals or convergence.

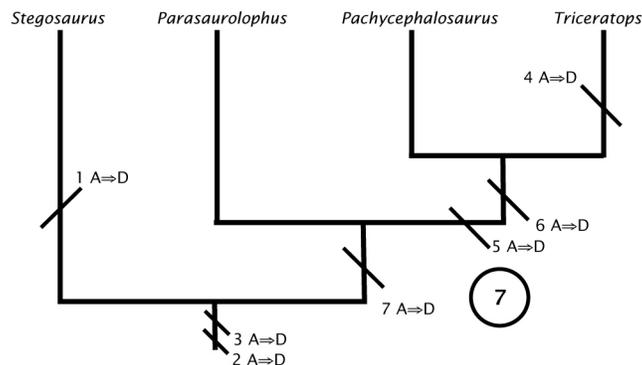


Figure B6. Most parsimonious tree from the dinosaur practice problem.

You should quickly notice that including characters 2 and 3 on your trees did not affect the phylogeny. If you look back at the character matrix in Step 6, you'll see that these characters are derived for all four organisms: this is why they are not useful.

Characters 1 and 4 were also not helpful in creating a phylogeny; since only one organism has the derived character state for each of these characters, it does not help us group organisms with shared derived characters. If you look over your trees, you'll see that characters 1 and 4 are always out on the ends of branches, and so do not affect the overall tree.

Characters 5 and 6 have the same pattern of derived and ancestral states across these organisms. This means they do not provide us with different information from each other, but they do effectively support the same tree. Overall, characters 5, 6, and 7 were the only characters that helped us distinguish between the different possible trees. Here's a revised version of this tree with only the distinguishing changes marked:

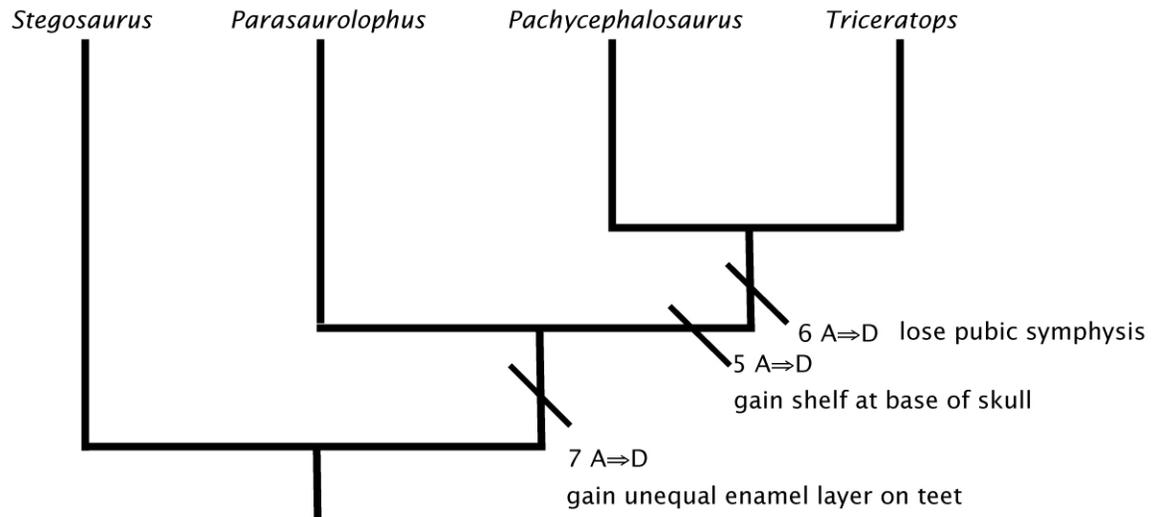


Figure B7. Revised tree from dinosaur practice problem with only characters relevant to cladistics labeled.

APPENDIX C

Phylogeny Construction Pre-Lab Assignment

This assignment is due at the beginning of your lab session. You will want to read this week's lab and work through the problems in it before completing this assignment. We encourage you to work through the problems in the lab manual with other students in the course, but you should complete this assignment independently. You may use your lab manual and any additional posted resources to help you.

1. In the phylogenetic tree in Fig. C1, which of the five marks corresponds to the most recent common ancestor of Lepidoptera and Diptera? Write the correct letter in the blank: _____

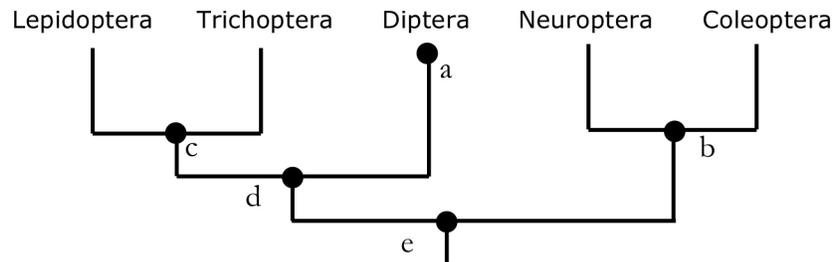


Figure C1. Pre-Lab tree for #1.

2. In the phylogeny in Fig. C2, assume that the ancestor was an herb (not a tree) without leaves or seeds.

Based on the phylogeny and assuming that all evolutionary changes in these traits are shown, which of the organisms is a tree (not an herb) and lacks true leaves? Circle the correct answer. (Just FYI, though most of the trees you are familiar with have both leaves and seeds, it is possible for trees to lack them.)

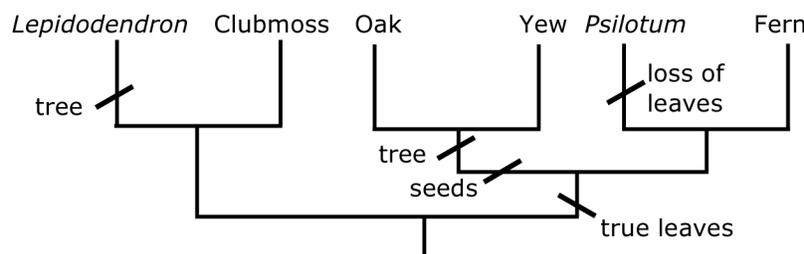


Figure C2. Pre-lab tree for #2.

- a) *Lepidodendron*
- b) Clubmoss
- c) Oak
- d) *Psilotum*
- e) Fern

3. In the dinosaur practice problem in your lab manual, which of the numbered traits were **not** useful in determining evolutionary relationships? Circle all that apply; there may be more than one correct answer.

- a) trait #1 b) trait #2 c) trait #3 d) trait #4 e) trait #5 f) trait #6 g) trait #7

APPENDIX D

Phylogeny Construction Pre-Lab Assignment Key

Note: Questions 1 and 2 were modified (only slightly) from the Basic Tree Thinking Assessment in the supporting online material from Baum et al. (2005).

1. In the phylogenetic tree in Fig. C1, which of the five marks corresponds to the most recent common ancestor of Lepidoptera and Diptera? Write the correct letter in the blank:

Answer: d (3 points)

2. In the phylogeny in Fig. C2, assume that the ancestor was an herb (not a tree) without leaves or seeds.

Based on the phylogeny and assuming that all evolutionary changes in these traits are shown, which of the organisms is a tree (not an herb) and lacks true leaves? Circle the correct answer. (Just FYI, though most of the trees you are familiar with have both leaves and seeds, it is possible for trees to lack them.)

Answer: a) *Lepidodendron* (3 points)

3. In the dinosaur practice problem in your lab manual, which of the numbered traits were **not** useful in determining evolutionary relationships? Circle all that apply; there may be more than one correct answer.

Answers: a) trait #1 b) trait #2 c) trait #3 d) trait #4 (4 points, -½ point for each letter circled/not circled incorrectly)

APPENDIX E

Student Answer Sheet for Primate Phylogenetics Assignment

Phylogenetics Assignment Answer Sheet

Part I: Tutorial Tree Using Cytochrome C Oxidase Polypeptide II

6. h. Two most similar genera: _____ & _____

Numerical difference from distance matrix: _____

Two most different genera: _____ & _____

Numerical difference from distance matrix: _____

7. k. Draw your tree as it appears on the computer screen in the space below. Replace the labels with genus names.

Using the tree as evidence, write a convincing argument that the differences listed on the distance matrix are reflected in the construction of this tree. (Hint: Think about measuring the line lengths.)

7. l. Draw a more ordered tree below which conveys the data from SWAMI but takes the form of the trees we made in lab (i.e. branch points are important and indicate common ancestors, but branch lengths are not meaningful). Use the genus name of each primate (since Lemur is the outgroup, it will not appear on the tree).

7. m. How does your tree in 7.l. compare to the trees we generated in lab?

How does your tree in 7.l. compare to the tree in Figure 17 in this lab assignment?

7. n. Write the name of the protein (or an abbreviation, like “cyt c ox II”) beside each tree in Figure E1 which is consistent with the tree on your screen.

Part II: Determine the Relationship of an Unknown Primate Species Using Cytochrome C Oxidase Polypeptide II

10. b. _____ (genus and species of your new, chosen primate)

is most closely related to: _____

Proportional distance between these organisms? _____ (# from distance matrix)

10. c. Based on the distance matrix, and your answer above, which group of primates (refer to Fig. 1 as needed) do you predict your new primate belongs to? Circle one:

prosimians New World monkeys Old World monkeys hominoids

11. c. Does your tree agree with your prediction?

Part III: Use a Different Protein Sequence to Make a Hominoid Tree

12. Which protein did you choose to investigate? _____

16. Draw your tree as it appears on the computer screen in the space below. Be sure to include genus names, not ID numbers. Include species names if you have multiple members of the same genus.

17. Write the name of your protein (or a clear abbreviation) beside each tree in figure E1 that is consistent with the tree on your screen.

18. a. Which primates did you use to build your tree (list genus & species names below)? Which served as an outgroup (indicate with an asterisk)?

b. Were all your sequences the same length? If "yes," length: _____ amino acids

If "no," shortest length: _____ longest length _____

c. Would you expect to have more confidence in a tree based on sequences with:

more amino acids in length OR fewer amino acids in length? (circle one)

highly similar sequences OR several differences between sequences? (circle one)

Explain your reasoning.

Given your circled answers, assess the protein sequence you chose and describe how much confidence you have in the tree(s) you indicated in #17.

Fifteen Possible Trees for Four Species of Primates

Indicate which tree(s) are best supported by data from each of the two proteins you studied (cytochrome c oxidase subunit II and another protein of your choosing). See #7n & #14.

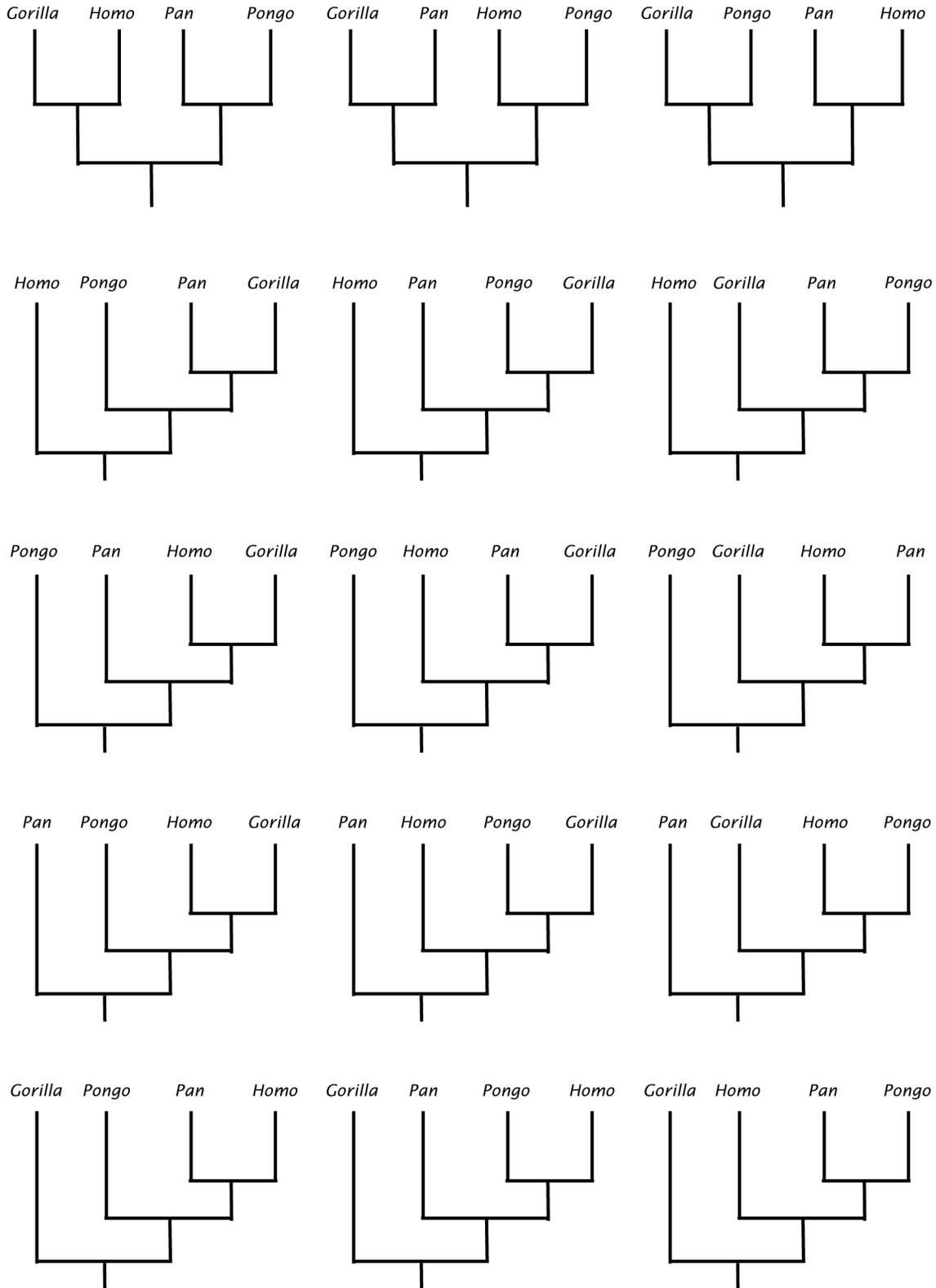


Figure E1. Fifteen possible trees for four species of primates. Part of Phylogenetics Assignment

APPENDIX F

Phylogenetics Assignment Key

The first letter of a genus name should be capitalized, species names should be all lower case, and all should be italicized (underlining ok if hand written). Make comments as needed so students realize this is important.

Part I: Tutorial Tree Using Cytochrome C Oxidase Polypeptide II

6. h. (2 points) Two most similar genera: *Homo* & *Pan*
 (1 point) Numerical difference from distance matrix: 0.022
 (2 points) Two most different genera: *Lagothrix* & *Macaca*
 (1 point) Numerical difference from distance matrix: 0.348

7. k. (3 points) Draw your tree as it appears on the computer screen in the space below. Replace the labels with genus names.

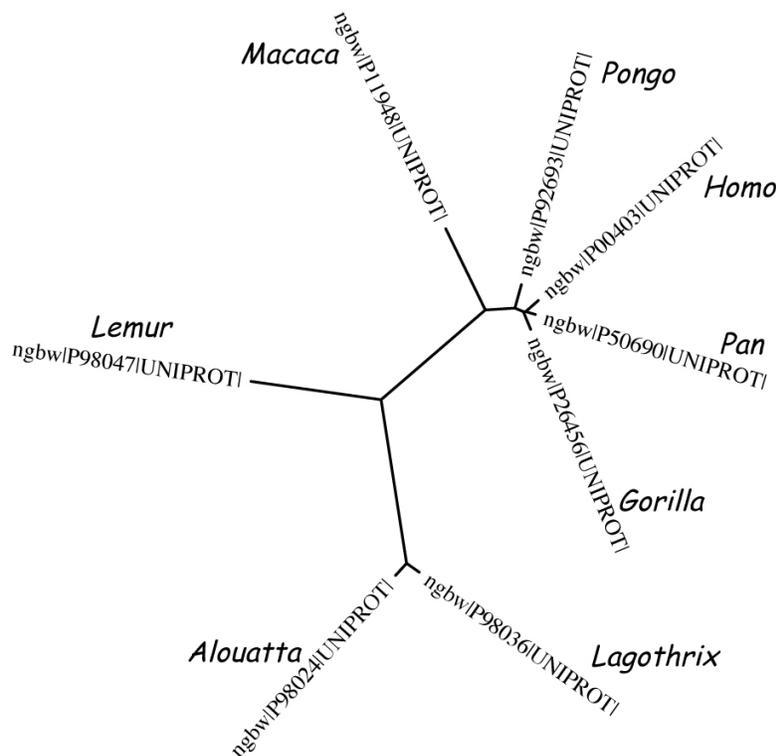


Figure F1. Figure for Phylogeny Assignment key #7.k.

(2 points) Using the tree as evidence, write a convincing argument that the differences listed on the distance matrix are reflected in the construction of this tree. (Hint: Think about measuring the line lengths.)

The lines connecting *Pan* & *Homo* are very short, and these two genera are located next to each other on the tree. The lines connecting *Lagothrix* and *Macaca* are relatively long, which is consistent with the large difference on the distance matrix.

7. l. (5 points) Draw a more ordered tree below which conveys the data from SWAMI but takes the form of the trees we made in lab (i.e. branch points are important and indicate common ancestors, but branch lengths are not meaningful). Use the genus name of each primate (since Lemur is the outgroup, it will not appear on the tree).

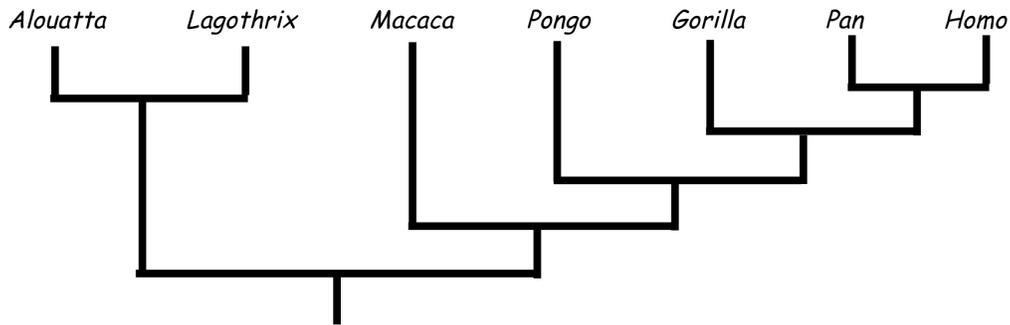


Figure F2. Figure for Phylogeny Assignment key #7.1.

7. m. (2 points) How does your tree in 7.1. compare to the trees we generated in lab?

Overall it is the same, except that the relationship between *Pongo*, *Gorilla*, *Pan*, & *Homo* is more clear.

(2 points) How does your tree in 7.1. compare to the tree in Figure 1 in this lab assignment?

Again, it is the same, except that the relationship between *Pongo*, *Gorilla*, *Pan*, & *Homo* is more clear, and Lemur has not been included on the tree.

7. n. (2 points) Write the name of the protein (or an abbreviation, like “cyt c ox II”) beside each tree below which is consistent with the tree on your screen.

The tree that reads “*Pongo-Gorilla-Homo-Pan*” from left to right should be indicated.

Part II: Determine the Relationship of an Unknown Primate Species Using Cytochrome C Oxidase Polypeptide II

10. b. (1 point) __ Will vary: see below __ (genus and species of your new, chosen primate)

Note: new primate should not be from one of the genera already in the tree, even if a new species.

(2 points) is most closely related to: ____ Will vary: see below _____

(2 points) Proportional distance between these organisms? Will vary: see below (# from dist. matrix)

| Possible Primate Group | Most Closely Related Genus | Distance | Predicted |
|--|------------------------------------|-----------------|-------------------|
| <i>Aotus nigriceps</i> (Q7IC90) | <i>Alouatta</i> & <i>Lagothrix</i> | 0.022 | New World Monkeys |
| <i>Brachyteles hypoxanthus</i> (Q9GCE7) | <i>Alouatta</i> & <i>Lagothrix</i> | 0.078 | New World Monkeys |
| <i>Callimico goeldii</i> (P50661) | <i>Alouatta</i> & <i>Lagothrix</i> | 0.042 | New World Monkeys |
| <i>Cercocebus galeritus</i> (P98020) | <i>Macaca</i> | 0.062 | Old World Monkeys |
| <i>Cercopithecus aethiops</i> (P26455) | <i>Macaca</i> | 0.070 | Old World Monkeys |
| <i>Cheirogaleus medius</i> (P98027) | <i>Lemur</i> | 0.062 | Prosimians |
| <i>Daubentonia madagascariensis</i> (P98032) | <i>Lemur</i> | 0.145 | Prosimians |
| <i>Eulemur macaco</i> (P98033) | <i>Lemur</i> | 0.062 | Prosimians |
| <i>Galago senegalensis</i> (P50688) | <i>Lemur</i> | 0.128 | Prosimians |
| <i>Haplemur griseus</i> (P98034) | <i>Lemur</i> | 0.062 | Prosimians |
| <i>Hylobates lar</i> (Q96127) | <i>Pongo</i> | 0.053 | Hominoids |
| <i>Hylobates syndactylus</i> (P25312) | <i>Pongo</i> | 0.053 | Hominoids |
| <i>Mandrillus leucophaeus</i> (P98037) | <i>Macaca</i> | 0.075 | Old World Monkeys |
| <i>Microcebus tavaratra</i> (Q7IKU8) | <i>Lemur</i> | 0.075 | Prosimians |
| <i>Nycticebus coucang</i> (P98039) | <i>Lemur</i> | 0.123 | Prosimians |
| <i>Papio anubis</i> (P68297) | <i>Macaca</i> | 0.066 | Old World Monkeys |
| <i>Papio hamadryas</i> (P68298) | <i>Macaca</i> | 0.066 | Old World Monkeys |
| <i>Propithecus tattersalli</i> (P98042) | <i>Lemur</i> | 0.079 | Prosimians |
| <i>Tarsius bancanus</i> (P98043) | <i>Lemur</i> | 0.141 | Prosimians |
| <i>Tarsius syrichta</i> (P98046) | <i>Lemur</i> | 0.159 | Prosimians |
| <i>Theropithecus gelada</i> (P98044) | <i>Macaca</i> | 0.070 | Old World Monkeys |

10. c. (2 points) Based on the distance matrix, and your answer above, which group of primates do you predict your new primate belongs to? Circle one: Will vary: see above

prosimians

New World monkeys

Old World monkeys

hominoids

11. c. (2 points) Does your tree agree with your prediction?

Trees should agree with predictions.

Part III: Use a Different Protein Sequence to Make a Hominoid Tree

12. (1 point) Which protein did you choose to investigate? _____ See attached list _____

16. (3 points) Draw your tree as it appears on the computer screen in the space below. Be sure to include genus names, not ID numbers. Include species names if you have multiple members of the same genus.

May vary depending on primates & protein chosen. Should support the correct tree, though.

17. (2 points) Write the name of your protein (or a clear abbreviation) beside each tree below which is consistent with the tree on your screen.

The tree that reads “*Pongo-Gorilla-Homo-Pan*” from left to right should be indicated.

18. a. (3 points) Which primates did you use to build your tree (list genus & species names below)? Which served as an outgroup (indicate with an asterisk)?

May vary for some proteins. They should have the four hominoid genera plus one primate outside the hominoids represented.

18. b. (2 points) Were all your sequences the same length? If “yes,” length: ___ See attached list. ___

If “no,” shortest length: _____ longest length _____

18. c. Would you expect to have more confidence in a tree based on sequences with:

(1 point) more amino acids in length OR fewer amino acids in length? (circle one)

(1 point) highly similar sequences OR several differences between sequences? (circle one)

(3 points) Explain your reasoning.

They should explain that short sequences could have high proportional differences from just a few amino acid differences, and that longer sequences usually mean more data. The more data points, the stronger the evidence for a pattern. Similarly, if the number of differences between sequences is very small, the trees will be based on very little data. Thirty differences in amino acids is much more convincing than three, because the patterns indicated are less likely to be random.

(3 points) Given your circled answers, assess the protein sequence you chose and describe how much confidence you have in the tree you indicated in #24.

They should be thoughtful; if their sequence was very short and their tree choice based only on a few amino acid differences, they should have less confidence in their tree. See attached list. Grade this question generously.

Notes about Proteins & Trees

40S ribosomal protein S4 Y isoform 1

Length: 263 amino acids

Notes: The tree is a bit tricky to interpret; *Gorilla* may not have its own branch, only a bend in the line. Only 17 out of the 263 amino acids show any differences. *Pongo* is the most obviously different from the other three hominoids (based on 7 amino acids), but the grouping of *Pan & Homo* together apart from *Gorilla* is only based on one amino acid.

Abnormal spindle like microcephaly associated protein

Length: *Macaca mulatta* 3479 aa; *Colobus, Hylobates, Homo, & Pan* 3477 aa; *Gorilla & Macaca fascicularis* 3476 aa *Pongo* 3471 aa

Notes: Tree should be very clear. Quite a long sequence, with well over 100 differences in amino acids. A good sequence to use for tree-building.

ATP synthase protein 8

Length: *Alouatta* 66 aa; all others 68 aa

Notes: Tree should be clear. Very short protein, however there is a considerable amount of variability; only 17 amino acids out of the 66 are conserved across all primates studied.

Galactoside 2 alpha L fucosyltransferase 2

Length: 343 aa

Notes: Tree should be clear. Only 20 differences in amino acids; only two of these group *Pan, Homo, & Gorilla* together; only one groups *Pan & Homo* together.

ribonuclease like protein 9 precursor

Length: *Cebus, Gorilla, Homo, and Pan* 205 aa; *Macaca assamensis* 203 aa; *Saimiri* 206 aa; all others 204 aa

Notes: Tree should be clear. Good length with quite a bit of variability across all the primates the sequence is available for.

C5a anaphylatoxin chemotactic receptor

Length: *Homo* 350 aa, all others 340 aa

Notes: Tree a bit confusing but readable. Only a small amount of variability; 20 aa difference in whole sequence, and only 2 of these group *Pan & Homo* together.

cytochrome c oxidase subunit IV isoform 1

Length: *Cebus, Saimiri ustus* 55 aa; *Mandrillus, Presbytis* 99 aa; *Saimiri sciureus* 124 aa; *Perodicticus* 137 aa; *Aotus, Gorilla, Hylobates, Pan, Pithecia, Pongo, Theropithecus* 144 aa; *Homo* 169 aa

Notes: Tree readable. Short protein, particularly for some fragments; only one amino acid groups *Pan & Homo* together. Not a lot of confidence in this tree.

fMet Leu Phe receptor

Length: *Gorilla, Macaca, Pan, Pongo* 346 aa; *Homo* 350 aa

Notes: Tree is clear. Not a lot of differences in sequence, but okay; about 37 of the aa differ. 7 can be used to group *Pan & Homo* together, but often this is because *Gorilla* has diverged; only 1 of these is truly a shared derived character.

FMLP related receptor I

Length: *Homo* 351 aa; all others 348 aa

Notes: Tree clear. Again, few differences; only 21 aa; only 3 of these group *Pan & Homo* (and two of these are the same aa as *Macaca*).

About the Author

Sarah Deel received her B.A. in Biology from Grinnell College in 1991, and her M.S. in Zoology from Oregon State University in 1996. She is a Senior Lecturer in Biology at Carleton College, where she coordinates and teaches introductory labs for “BIOL 125 Genes, Evolution, and Development.” She is interested in methods of teaching introductory biology, particularly those which encourage traditionally underrepresented groups, such as women and minorities, to continue in biology.

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 teaching biology 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: Peer-Reviewed 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

Deel, S. 2013. Phylogeny Construction: Primate Skulls and Protein Sequences. Pages 76-123, in *Tested Studies for Laboratory Teaching*, Volume 34 (K. McMahon, Editor). Proceedings of the 34th Conference of the Association for Biology Laboratory Education (ABLE), 499 pages. <http://www.ableweb.org/volumes/vol-34/?art=4>

Compilation © 2013 by the Association for Biology Laboratory Education, ISBN 1-890444-16-2. 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.