

## PHYLOGENETICS PROBLEM SET

### Phylogenetics

A **phylogeny** is a representation (you can think of it as a hypothesis) of how different types of organisms are related, and often looks something like a family tree (see Fig. 1 for an example).

How do evolutionary biologists make decisions about the relatedness of different organisms? They study what the organisms (or fossils, in some cases) look like, what their patterns of development are, and (more recently) their DNA sequences.

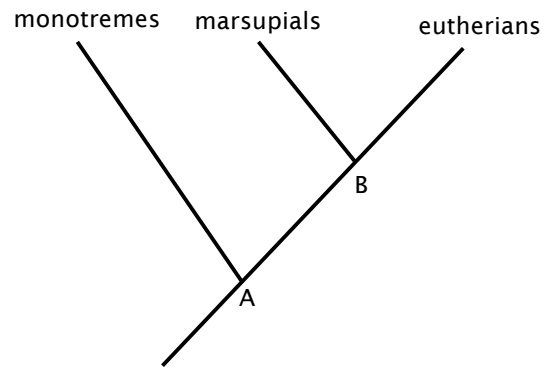
### Phylogenetic Trees

At the end of each branch on a phylogenetic tree is the name of a group of organisms; in Figure 1, each group represents a different type of mammal: monotremes include the platypus and the echidna (they lay eggs), marsupials include kangaroos and opossums (their young develop in pouches), and eutherians include rats and humans (development takes place in the uterus, and involves a well-developed placenta). 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. So, for example, the marsupials and eutherians shared a common ancestor at point B, and the monotremes and the point B ancestor shared a common ancestor at point A. 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.

### Cladistics

One commonly used method for constructing a phylogeny is called **cladistics**. In cladistics, “clade” specifically refers to a group of organisms and their common ancestor(s). In Figure 1, a group including marsupials, eutherians, 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. To do this, it is useful to have an **outgroup**. An outgroup is a group of organisms that is somewhat, but not too closely related to the organisms you want to study;



*Figure 1: Phylogenetic relationships between different types of mammals.*

it is helpful if the outgroup has characteristics (commonly referred to as **characters** in the context of phylogenetics) which are ancestral to those of the groups you're studying. A helpful outgroup to use when assembling the mammalian tree (Fig. 1) would be lizards. Lizards, like mammals, are vertebrates (have a backbone), so they are a better choice for an outgroup than mushrooms or other very distantly related organisms. Compared to mammals as a whole, lizards have more **ancestral** character states (like scales); 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 body covering example, scales are the ancestral character state while hair 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 frogs as an outgroup, scales would be the derived character state, while moist skin would be ancestral.

You will often hear people refer to **ancestral** character states as “primitive” character states. We will stick with the term “ancestral,” because it doesn't have quite the same connotations. It is important to remember that although reptiles have some ancestral characters in the context of looking at mammals, reptiles are not primitive organisms: all organisms have a mix of ancestral and derived characters.

Cladistics creates phylogenies by grouping organisms that have the same derived character states (**shared derived character states**) close to each other on the tree. If all character states just changed once in evolutionary history, then there would be no need for a complex methodology. 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, on a separate branch from baboons, opossums, and triceratops. This is fairly absurd given the other things we know about these organisms and their close relatives.*

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

*For example: neither alligators nor penguins can fly. Geese and hummingbirds both fly. If you created a phylogeny based on this character, using (for example) lizards as the outgroup, you would group geese and hummingbirds together (they share the derived character state of "flight") apart from alligators and penguins (they have the ancestral character state of "flightlessness"). However, based on many other things we know about these organisms (e.g. penguins have feathers and wings), penguins are much more closely related to geese and hummingbirds than they are to alligators. Penguins have actually undergone a reversal at some point in their evolutionary history, from being capable of flight back to the ancestral character state of flightlessness.*

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 which 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 which represent their relationship. With four organisms, there are fifteen possible trees<sup>1</sup>. 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, placental presence, 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.

As you'll see, 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.

<sup>1</sup>The equation to determine the number of possible trees is:  $\prod_{i=3}^s (2i-5)$  where  $s$  is the number of taxa (including the outgroup).  $i=3$   
For our example of 4 taxa plus an outgroup,  $s=5$ , so the calculation is:  $[(2 \times 3)-5] \times [(2 \times 4)-5] \times [(2 \times 5)-5] = 1 \times 3 \times 5 = 15$  possible trees.

## How to Make a Tree

Step #1:  
Choose study organisms.

Mammals:  
monotremes  
marsupials  
eutherians

Step #2:  
Choose outgroup.

lizards

Step #3:  
Choose characters.

eggs  
highly developed placenta

Step #4:  
Score characters for organisms.

organism:	lizard	monotreme	marsupial	eutherian
eggs?	yes	yes	no	no
placenta?	no	no	no	yes

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

eggs: yes = ancestral  
no = derived  
placenta: no = ancestral  
yes = derived

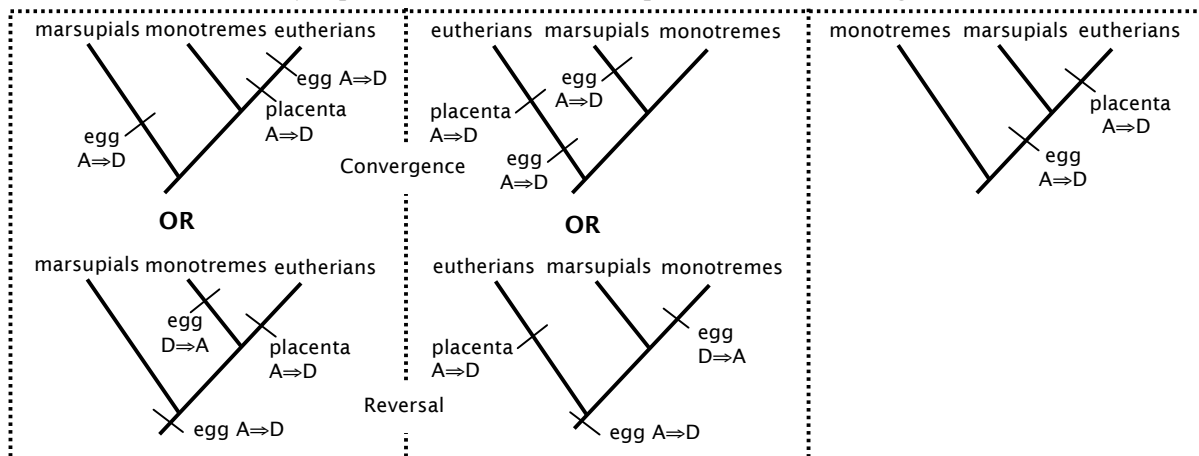
*The outgroup's character state will correspond to the ancestral state.*

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

	egg	placenta
monotreme	A	A
marsupial	D	A
eutherian	D	D

*The outgroup does not appear on this table, because it will not appear on our tree.*

Step #7:  
Use character matrix to analyze possible trees; indicate the presence of state changes with hash marks.



Step #8:  
Determine which tree (or trees) is most parsimonious.

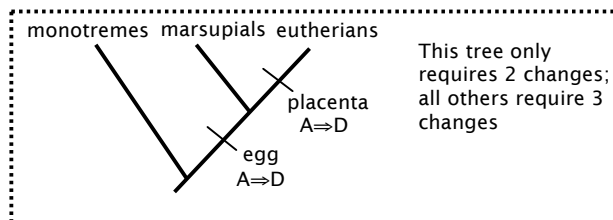


Figure. 2. Diagrammatic representation of the steps involved in reconstructing phylogeny using cladistics, using mammals as an example.

## Problem 1: Dinosaur Phylogeny

This problem is adapted from the fifth edition of *Biology* (Campbell et al., 1999).

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 #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. Pre dentary 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 which 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 which 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

*Allosaur. Pachycep. Parasaur. Stegosaur. Tricerat.*

1 armored plates?	no	no	no	yes	no
2 pre dentary?	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. pre dentary		
3. pubis ext. post.		
4. rostral		
5. skull shelf		
6. symphysis		
7. enamel		

### Step #6: Fill out character matrix.

Character:	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
<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 have 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.**

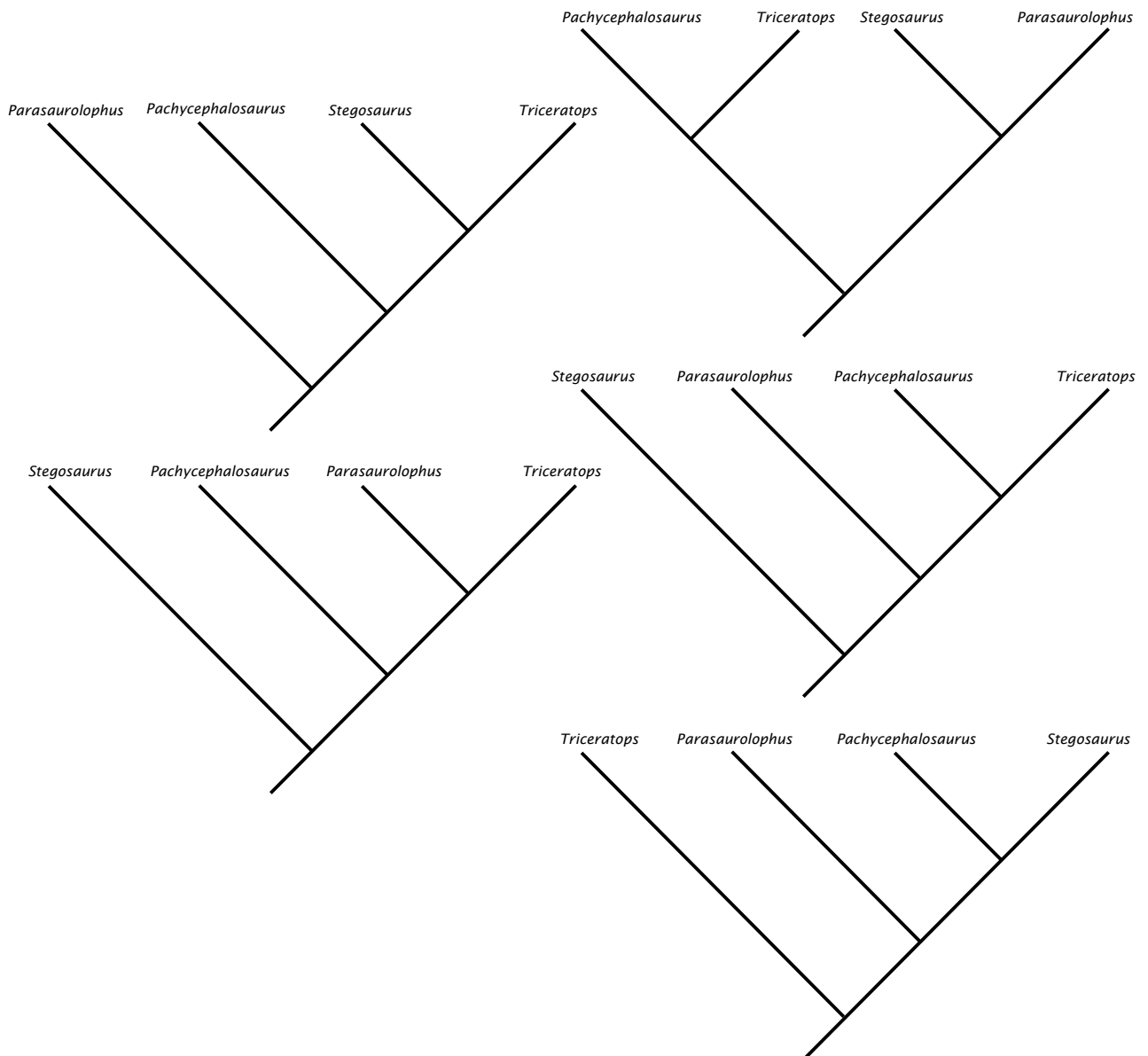
In this problem, rather than making you analyze all 15 possible trees, we have drawn a subset of 5 trees to analyze. You should get the idea from these, but you can always set up the other 10 trees on your own if you like.

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 2 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?



## Problem 2: Primate Phylogeny

In this problem, you'll use data from primate skulls to build a primate phylogeny.

**Step #1**, as we described above, is to choose the organisms you're interested in. We will be studying the primates listed below.

<u>Scientific Genus Name</u>	<u>Common Name</u>
<i>Alouatta</i>	howler monkey
<i>Homo</i>	human
<i>Macaca</i>	macaque
<i>Pan</i>	chimpanzee

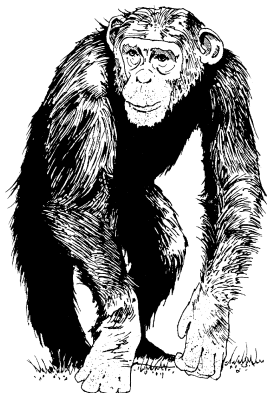
Members of the genus *Alouatta*, the howler monkeys (also called howling monkeys), are found in Central and South America. They roar very loudly, and can be heard up to 2 km away. Their genus name sounds most like a howl of any of the primates we'll work with.



Macaques, 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 fuscata*) basking in hot springs on TV during the 1998 Winter Olympics in Nagano.



*Pan* 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 which peel leaves off branches and stick the branches into termite nests to go "termite fishing."



**Step #2** is to select an outgroup. We have chosen lemurs, of the genus *Lemur*, as our outgroup. Lemurs are among a group of primates which 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.



**Step #3** is to select characters to study. We have selected several characters, primarily on structures associated with primate skulls.

**Step #4** is to score each organism for the characters. We have filled in some of the Data Collection table below for you; you may use the skulls in class and the instructions below to fill out the rest of the table.

- Total number of teeth.** Count the total number of teeth, top and bottom. Record this number on the data sheet. 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.
- 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.
- 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.

**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.

**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.

**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; you will need to indicate which organism corresponds to each number/letter on the diagram.

**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 which 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?)

## References

- Campbell, N.A., J.B. Reece, and L.G. Mitchell. 1999. Biology 5 Ed. Benjamin Cummings, San Francisco.
- Fleagle, J.G. 1999. Primate adaptation and evolution 2 Ed. Academic Press, San Diego.

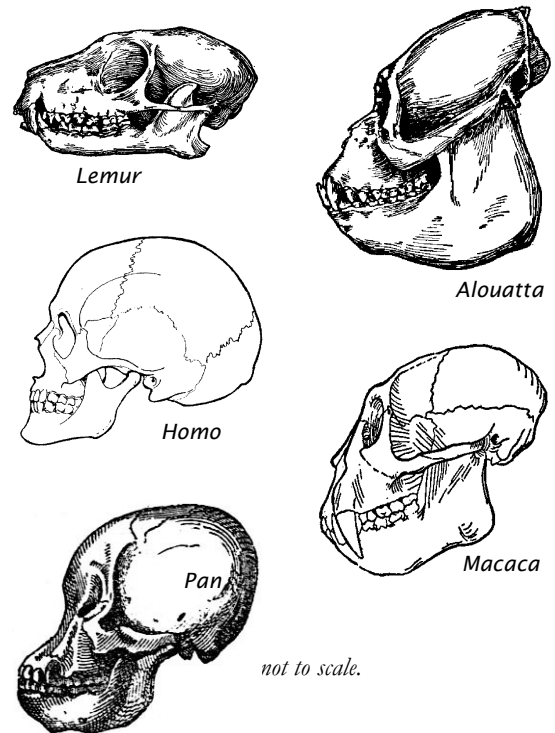


Table for Primate Skull Morphology Data Collection

Measurement	Outgroup: <i>Lemur</i>	<i>Alouatta:</i>	<i>Homo:</i>	<i>Macaca:</i>	<i>Pan:</i>
<b>Dentition</b>					
Total number of teeth (see #4a)	36				
<b>General Cranium Shapes</b>					
Development of ear tube	simple	simple	tube-like	tube-like	tube-like
Post-orbital closure	open	closed	closed	closed	closed
Skull ridge presence (see #4b)	yes				
Two skull ridges (see #4c)	yes				
<b>Size Measurements</b>					
Skull length (cm)	7.5	10	18	11	13.5
Snout length (cm)	2.2	2.6	0.8	1	0.5
Cranial capacity (cm <sup>3</sup> )	14	33	1500	77	460
<b>Size Ratios</b>					
Snout length ÷ skull length	0.29				
Cranial capacity ÷ skull length	1.87				
<b>Other</b>					
Tail?	yes	yes	no	yes	no



### Character Matrix

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; also pay attention to the footnotes below.

	1. Number of Teeth	2. Development of Ear Tube	3. Post-orbital closure	4. Skull ridge presence	5. Two skull ridges	6. Snout length ratio <sup>1</sup>	7. Cranial capacity ratio <sup>2</sup>	8. Tail				
<i>Alouatta</i> :												
<i>Homo</i> :												
<i>Macaca</i> :												
<i>Pan</i> :												

---

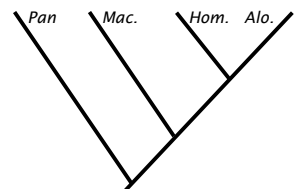
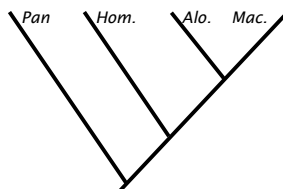
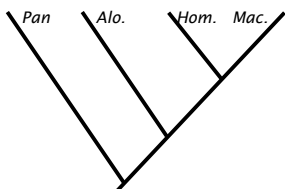
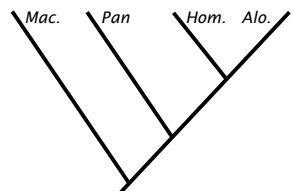
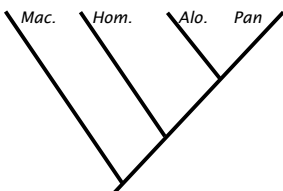
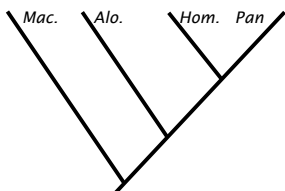
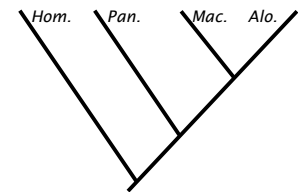
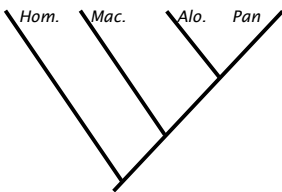
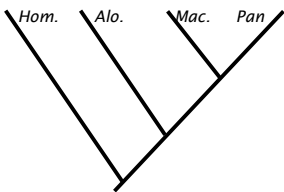
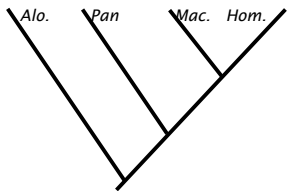
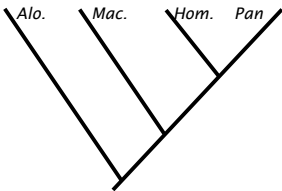
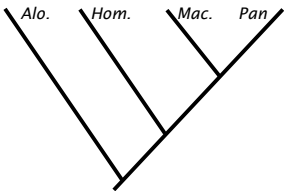
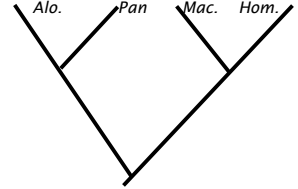
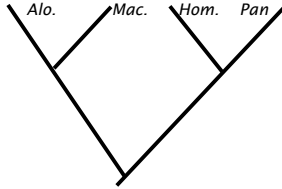
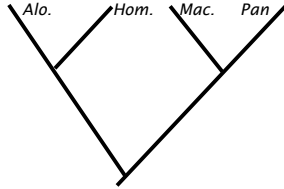
<sup>1</sup> Check each datum to see if it is greater than or less than 0.17; if it is greater than 0.17, like the Lemur, it should be considered ancestral.

<sup>2</sup> 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.

**Notes:**

## Possible Phylogenetic Relationships

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 bash mark on the tree and indicate which character must change (you can use the number associated with each character in the character matrix) and what the change is. Count the number of changes required and write it below each tree.



**Phylogenetics Assignment**

Name: \_\_\_\_\_

*Turn in just pages 11 & 12.*

1. Complete the phylogenies following the instructions at the top of page 11. Circle the most parsimonious tree (or trees, if necessary).
2. Which of the 8 characters were informative in helping you determine the most parsimonious tree?
3. For each of the characters which were not useful in cladistics, explain why.