

BACKGROUND

The main goal of this problem set is to give you some first-hand experience in phylogenetic analysis. An additional goal is to make some of the concepts we discuss in lecture more concrete by forcing you to familiarize yourself with a set of organisms and their relationships to one another. Before you begin, get it straight in your mind that there may not be one best answer to the questions we ask, although there are wrong answers. We are primarily interested in having you learn the process of how to do systematics. In short, you are presented with a set of organisms that look different, and are faced with solving the puzzle of who is related to whom. By using different types of information to solve this puzzle we hope that you will learn how evolution shapes different traits in different ways.

This problem set involves **Two Steps**: 1) A problem involving vertebrate evolution and 2) A problem with Cetacean (whale) phylogeny where you have to **collect your own data set** from web pages, and from literature on reserve in the Science Library (Bio 48 Whale Data), and then use this data set in two pieces of Macintosh software. You will compare the results from your own data set to those from a molecular data set of mitochondrial DNA (mtDNA) sequence. You will do the vertebrate evolution problem and the first part of the whale phylogeny problem at one sitting. You will then go do research on whales and return for further analysis.

The two different pieces of Macintosh software work together (i.e., can read each others files). One program (**PAUP** for Phylogenetic Analysis Using Parsimony) finds the shortest evolutionary tree based on the data set provided. The other program (**MacClade**) allows you to move branches of the tree around with the mouse and trace the evolution of characters on different trees. This allows you to test different hypotheses about the evolution of characters and play "what if..." (e.g., "what happens to the statistics of the evolutionary tree if we force these species to be more closely related?").

Using the Macintosh Cluster Servers. 1) To use the Cluster programs you need activate the CLUSTER.CLASSES and the CLUSTER.MAC servers on the desktop of the Mac you are using. Instructions on how to do this can be found at <http://www.brown.edu/Facilities/CIS/Doc/> (follow the Using a Mac in the Clusters link). The **MacClade** and **Paup** software are in the CLUSTER.MAC server, listed alphabetically as **MacClade 3.01** and **Paup 3.0s** (**NOTE: if PAUP is not seen under CLUSTER.MAC it can be found in the BI0048 folder in the CLUSTER.CLASSES server**). For Part 1 below you will use the Vertebrates file in the MacClade folder. For Part 2 below you will use the **Whale.mtDNA.Bio48** data set is in the BI0048 folder in the CLUSTER.CLASSES server.

NOTE: THERE IS A LIMIT TO THE NUMBER OF PEOPLE (37) THAT CAN USE MacClade SIMULTANEOUSLY. PLAN AHEAD AND/OR WORK DURING NON-PEAK HOURS
THERE IS NO LIMIT TO THE NUMBER OF PEOPLE USING PAUP.

Part 1. Vertebrate Evolution

- Open the **MacClade 3.01** folder, then open the **MacClade Examples** folder, then double click on the **Vertebrates** file. A Data matrix comes up with **taxa** in the left column, **characters** along the top of each column and **character states** in each cell. Inspect the entire data set using the scroll bar at the bottom. If the terms don't make any sense, consult your local medical dictionary or morphology text.
- Pull down the **Display** menu and select on **Go To Tree Window**; a tree will be displayed. Pull down the Σ menu and select **Tree Changes**; repeat with **Consistency Index**. These will be displayed in a box at the bottom of the screen; find them. (Tree changes are the number of character state changes that are imposed by the current structure of the tree. Consistency Index is a measure of how "good" the tree is at

evolving these characters: it is a ratio of the number of characters in the data set to the length of the current tree displayed. A tree with no reversals (or convergence) would have a C.I. of 1.0; a tree with twice as many steps as needed would have a C.I. of 0.5).

- Pull down the **Trace** menu and select **Trace Character**. This will highlight in color each branch on which the alternative character **states** have evolved, and put a character box in the lower right which describes the character being displayed. This box also has a scroll bar so you can examine each character in turn. You will see on the current tree that the amnion has evolved three separate times. **Is this parsimonious?** Use the scroll bar in the character box to examine the pattern of character evolution for each character on the current tree (if you want to look at the data matrix again, select Data Editor under the Display menu). Make note of which taxa appear to evolve character states independently on this tree; this will help in the next step. Move the scroll bar back so amnion is showing in the character box.
- Now you will make the evolution of the characters more parsimonious by rearranging the branches on the tree. Use the mouse pointer to click on branches and drag them to new locations on the tree (point to a branch, click the mouse button, hold it down while you move the pointer to a new branch position and release the button). Make the amnion evolve once on the tree. Make a note of the new C.I. and Treelength. Is the new tree parsimonious with respect to other characters? (use scroll bar again).

Problem 1: Find the shortest tree (Treelength = 19; yes, 19! There is more than one tree of length 19, can you find them??). Use the patterns of character evolution as a guide to finding the shortest tree.

Question 1: i) Are reptiles a natural group (monophyletic)? **ii)** What term do we use to describe this kind of “group”? **iii)** The two fish taxa (rayfined and lungfishes) are not sister taxa, hence are not a natural (monophyletic) group. But the tree length is the same when they are monophyletic vs. when displayed on separate branches (i.e., paraphyletic). Explain why. Hint: Look at the data set (under the Display menu, select ‘Go to Data Editor’), and examine the character states for the characters in these two kinds of fishes to provide a answer. **iv)** Which character is the least parsimonious in vertebrate evolution?

Part 2 Whale Phylogeny.

- Open the **Paup** folder in the CLUSTER.MAC folder (if it is not there, go to the BI0048 FOLDER in the CLUSTER.CLASSES server, and drag the PAUP folder to the desktop of the Mac you are using). Double click on the **PAUP 3.0s** application. After it launches, open the **Whale.mtDNA.Bio48** file (pull down the File menu and select Open; you will need to go to the CLUSTER.CLASSES folder to access the **Whale.mtDNA.Bio48** file). DO NOT change any of the text or the program will not run. (If you do by mistake, just quit without saving anything and re-open the file).
- Pull down the **File** menu and select **Execute " Whale.mtDNA.Bio48 "**. Click **YES** if prompted with questions. The screen should say Processing of file "whale.mtDNA.Bio48" completed.
- Pull down the **Data** menu and select **Define Outgroup**. A dialogue box will come up; click on **Camel** then click on **>To Outgroup>** and the Camel will be moved over to the outgroup box. This tells PAUP that the camel is known to lie outside the group of taxa in the rest of the data set, and helps to root the tree. Click **OK**.
- Pull down the **Search** menu and select **Heuristic**, then click on search. The program will read the data set that appeared in the first window (sequences of the cytochrome b gene in whale mitochondrial DNAs) and will find the shortest tree. When the computer has finished searching (beep or flash) note the number of trees saved and number of rearrangements tried, then click **Close**.
- Pull down the **Trees** menu and select **Describe Trees**. A dialogue box comes up; click on the **Rooting** button. A new dialogue box comes up; select three buttons: **Outgroup rooting**, then **Make Ingroup Monophyletic** and then **Make Outgroup Monophyletic Sister Group to Ingroup**. What do these

mean? Click **OK**. It goes back to the Tree Description Option box. In the boxes shown, click on **Cladogram** and **Phylogram**. These will lead to the display of your phylogeny as a simple cladogram, and as a phylogram where the branch lengths are proportional to the number of changes along each branch, akin to an evolutionary distance. If there is more than one tree listed in the box in the upper left (only one now, but might be >1 with your own data set), select the ones you want. Click **OK**. Trees are displayed with simple statistics.

- To save your tree, pull down the **Trees** menu and select **Save Trees** at the bottom. Save the tree to your desktop or Zip disk. Now you know how to use PAUP. Play around with the various menus and options.

Curious students do this: One interesting variation is the **Bootstrap** (under the Search menu). This randomly throws out characters from your data set, and builds the tree on the reduced data set. After the tree is built, the discarded characters are restored to the data set, and a different random set of characters are sampled. This process is repeated as many times as you specify in the "Number of Replications" box that come up when you run the Bootstrap. When the bootstrap is finished, click on the Close button, and a tree will be scroll onto the window. The numbers on the branches represent the percent of bootstrap replications where all members of the group included in that branch were clustered together. This is a sort of significance test for the reliability of the cluster. For example, a bootstrap value of 65 means that 35% of the bootstrapped data sets placed one member of the group at a different place in the tree. High numbers (70-90) give you confidence that the group in question is "supported" by the data. Check it out. For starters, try only 25 replications; BE FOREWARNED: it can take a long time to run when many people are on the network, but if the runs go quickly, try 100 replications. **Draw the bootstrap tree in your notes, with the bootstrap values**; the simple Save command only saves the tree shape.

Question 2: Try each non-cetacean outgroup taxon (camel, cow or hippo) as the single assigned outgroup (see **Define Outgroup** above) and answer: **can you prove which species (camel, cow or hippo) is the sister taxon to the Cetacea?** Briefly describe a comparison between specific Tree Length and Consistency Index (C.I.) statistics that support your answer.

You are done with the first session.

Collect Cetacean Data. On the Problem set link of the Other Bio 48 Web page (<http://biomed.brown.edu/Courses/BIO48/problems.html>) is a preliminary data set and some links to various sources of information about whales, Cetaceans in general, and their relationship to other mammals. There are also some information packet at the Honors Reserve desk at the Science Library, and in the Walter Hall Conference room in the Bio 48 Box. Use these materials (and other references if you know some good ones) to find **characters** and **character states** for each of the species in question. You can use morphology, color, ecology, geographic range, etc. Characters **must have two or more character states** and they should be **phylogenetically informative** (see lectures on Systematics and Phylogenetic Inference). **Avoid using more than three character states** per character, if possible. Some characters are listed as ranges of values (e.g. Body Length: x - y...). For data of these types, break them up into discrete values that can be coded as **discrete character states**. Try to use a good combination of characters that you think are biologically meaningful (morphology, ecology, geography). Find a **minimum of 5 additional informative characters** beyond the 10 characters provided on the web page. What you are trying to do is compare the tree produced from the mitochondrial DNA data to a tree produced from your own data so that you can study the evolution of characters.

Enter your data set into MacClade A MacClade file has been set up with some of these data already added. Go to the BI0048 folder, and locate the file **Whale.Characters**. Double click on it to launch MacClade. You will see a data editor window as before with Whale species in the left column, and characters across the top. Only two are provided now. **You need to code the available whale data as discrete characters (see above).** Now pull down the **Display** menu and select **State names...** You will be presented with a box for naming your characters at the top and the character states in the boxes listed vertically. This tells the program what to recognize when you type your data into the data editor (if confused, go back to the Vertebrate data set from Part 1 and look at the data window). The scroll bar at the top of the box allows you to move to the next character when you have named all the character states for a character. It will scroll only as far as the number of columns you defined with the mouse in the data editor; if you need more, go back to the data editor and move the right edge out.

Type your characters and character states into the data editor (**remember to code continuous vales as discrete states!**). The little ruler in the upper left corner of the Data Editor allows you to widen or narrow the columns to fit words you use in naming characters. As you enter the data, **Save** the file: pull down the **File** menu and select **Save**. **Note:** you cannot save files on the Cluster server, so you must save the file on the **Desktop** or **Hard Disk (or a Zip disk)**. Click on the rectangular box just above the list of folders in the dialogue box (box with scroll bar on right edge) and choose Desktop (or Hard Disk). When you can save the file, the **Save** button at the lower right will be active. Name your file in the box provided (don't name it "Whales"; try Lastname.data or Ilovewhales, or WhyDoWeHaveToDoThis.data, or something original). When it is successfully saved, close the file (under the File menu), but leave MacClade running

Open your file under PAUP Launch PAUP. You copied this from the BI0048 folder onto your desk top in the first session. Do this again if PAUP is not on your desktop. After PAUP opens, pull down the **File** menu and select **Open**. A dialogue box will appear; click on the rectangular bar above this scroll box and locate your data file from MacClade). Your new whale data file should appear in the window; double click on it to open it. Now go through the steps described above for running this data set under PAUP and finding a tree (set Camel to the outgroup). Because your data set may have a lot of parallel or convergent (=homoplasiuous') changes, PAUP may find several trees that are the same length. If the program produced many trees, your data set is not very informative and you may need better character data. Examine the tree (or first few trees) using the Describe trees... Rooting procedure described above. Study the tree(s) and statistics.

Analyze character evolution. Now go back to **MacClade** (click on the icon in the upper right corner) and activate your data file. Pull down the **Display** menu and select **(Go To) Tree Window...** Using the mouse, **build the tree that PAUP generated from your data set**. Select Treelength and C.I. under the Σ menus and Trace Character under the **Trace** menu (see above, Part 1). Scroll through the characters as you did with the Vertebrates problem and examine the patterns of character evolution. Now, use the mouse to rearrange the branches of the morphology-based tree so that it matches the **tree that PAUP generated from the mitochondrial DNA data set**. You may need to open the mtDNA tree or re-run the data set under PAUP to get this tree. Compare patterns of character evolution (i.e. **Trace** characters): which characters are **most** and **least** parsimonious in character state changes in the two different tree topologies?.

WRITE-UP: You should hand in the following 1) Write short but coherent and meaningful answers to the questions below, 2) Print out your whale data set, 3) Print out the shortest tree(s) that PAUP found from your data set including C.I. and Treelength, 4) Print out one informative tree from MacClade that shows how different characters you found evolve on the tree based on the mitochondrial DNA and 5)

provide a written analysis of important observations about the analysis of character evolution. All written answers **should not exceed 1 1/2 pages ! Do not print out anything until you have your final versions .**

Don't forget Questions 1 and 2 above

Question 3: **i)** How many different trees did PAUP find from your data set, and how do your trees compare to those of the mtDNA data set? **ii)** Which character(s) show the most disagreement between the morphological and mtDNA trees? (i.e., are parsimonious in one topology, but not in another? (support for this should be provided in the trees printed from Paup and MacClade). **iii)** How does the comparison between the morphological and molecular trees provide evidence for adaptive evolution?

Question 4: In the whales, is the order Odontoceti (toothed whales) a natural group? How do the traditional groups of Odontoceti and Mysticeti relate to the concepts of grade and clade?

Question 5: Most cetaceans have lost their hind limbs, but vestigial pelvic girdles are found in some of the baleen whales. What does this fact say about the selection pressures on pelvic girdles in the three lineages (porpoises+dolphins), baleen whales, and sperm whales.

Question 6: What is the difference between: "whales are descended from hippos", and "whales are sister taxon to hippos"? Can DNA studies from living organisms distinguish between these statements? Explain the role fossils could play in addressing these questions.