Name _____   Period _____

AP Biology   Date _____

LAB _____. WALRUSES AND WHALES AND SEALS, OH MY!

Walruses and whales are both marine mammals. So are dolphins, seals, and manatee. They all have streamlined bodies, legs reduced to flippers, blubber under the skin and other adaptations for survival in the water. Although mammals evolved on land, these species have returned to the sea. Did they evolve from a single ancestor who returned to the ocean, or were there different return events and parallel evolution? We can't go back in time to observe what happened, but DNA sequences contain evidence about the relationships of living creatures. From these relationships, we can learn about the evolutionary history of marine mammals.

In this lab, we will use sequence information in GenBank (the public repository of all known DNA sequences from many species) and bioinformatics software to test hypotheses about the relationship between aquatic mammals (seals, whales, dolphins, walruses, manatees, and sea otters) and their potential ancestral relationship to land mammals.

We will use a protein that all mammals share: the hemoglobin beta protein. Hemoglobin is a good test molecule since it shows both conservation across species (since it performs the essential function of carrying oxygen in the blood), and variation between species. Species with unique challenges such as holding their breath for long underwater dives, may have evolved changes in their hemoglobin which improved their supply of oxygen. In addition, hemoglobin has been studied by many evolutionary biologists, so sequences are available in GenBank from many different organisms.

PROCEDURE

In this lab, we will be testing hypotheses about the evolutionary ancestry of different marine mammals. To repeat, we are trying to answer the question: Did marine mammals evolve from a single ancestor who returned to the ocean, or were there distinct return events from separate ancestors? As a starting point let's hypothesize that marine mammals have a single common land mammal ancestor.

Part A: First, we will explore the relationship of the marine mammals to each other vs. their evolutionary relationship to land mammals. To do this, we will test whether seals and whales are more closely related to each other than either of them are to representative land mammals: dogs (land carnivores) or cows (land herbivores). This exercise will mainly train you in using the bioinformatics software.

Part B: Second, each student will then develop a cladogram which includes a selection of marine mammals and land mammals which represent the major mammalian orders. You will then use this phylogenetic tree to test our hypothesis that all marine mammals have a single common land mammal ancestor.
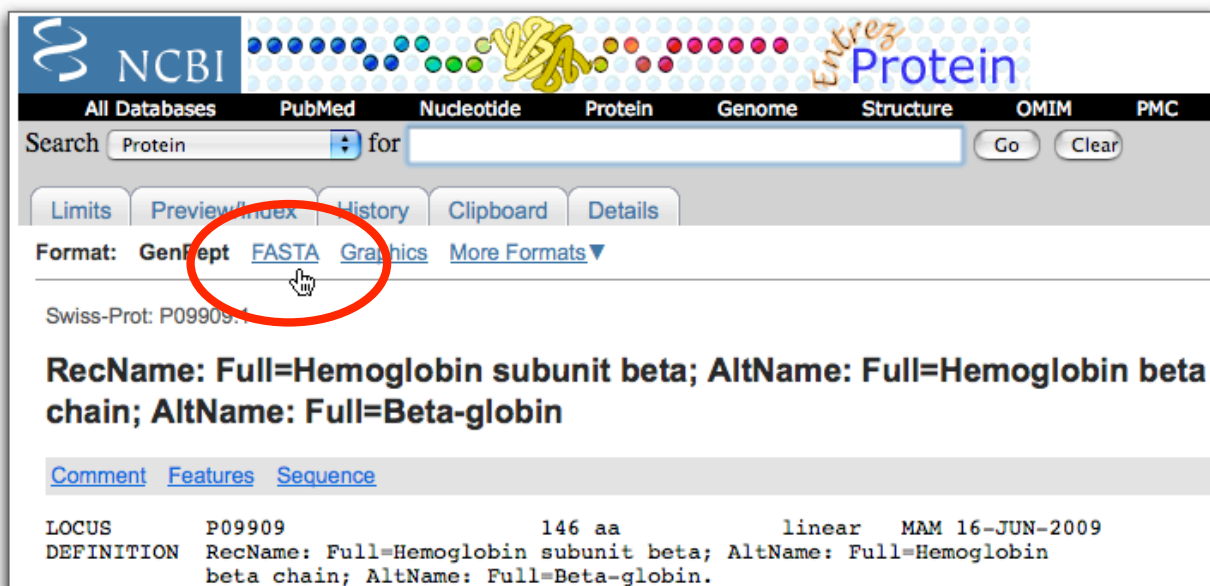
PART A

FINDING AMINO ACID SEQUENCES

1. First we need to get the sequence data for the hemoglobin protein from our marine and land animals: seals, whales, dogs, cows. Go to **GenBank**, a DNA and protein sequence database hosted by the National Center for Biotechnology Information (NCBI) in Maryland at: **http://www.ncbi.nlm.nih.gov/Genbank/** (Note: You can also go to Google and search for "GenBank" and it will always come up as the first link.)

2. One thing you need to know, is that the hemoglobin beta gene and protein is known as "**HBB**" in GenBank. In the "Search" window, select "Protein" from the pull-down menu, because we want to compare the amino acid sequence from each animal for this protein. We need to be specific about the identity of each animal, so we will use the harbor seal, the minke whale (a baleen whale), *Canis familiaris* (dog) and *Bos taurus* (cow). Type in the protein you're looking for and the organism. Push "Go".
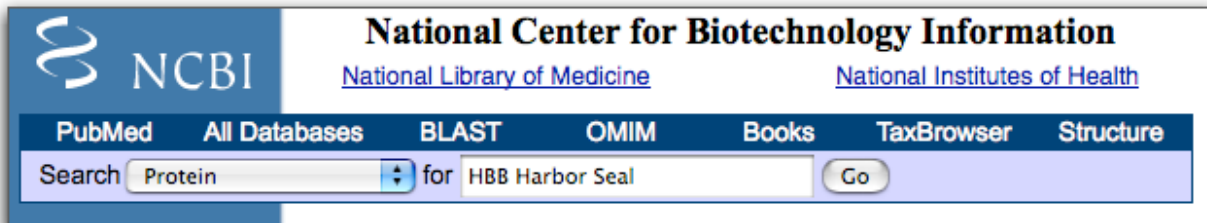
NOTE: This database was developed by humans, so sometimes it is not organized logically. Don't get frustrated; be patient. If your search returns several answers that you have to choose between, look for the listing with a "P0####" number. That will be the correct protein!

3. The search result is a page with a lot of information about the protein from this organism. To see the actual amino sequence for this protein, click on the "FASTA" link near the top of the page.
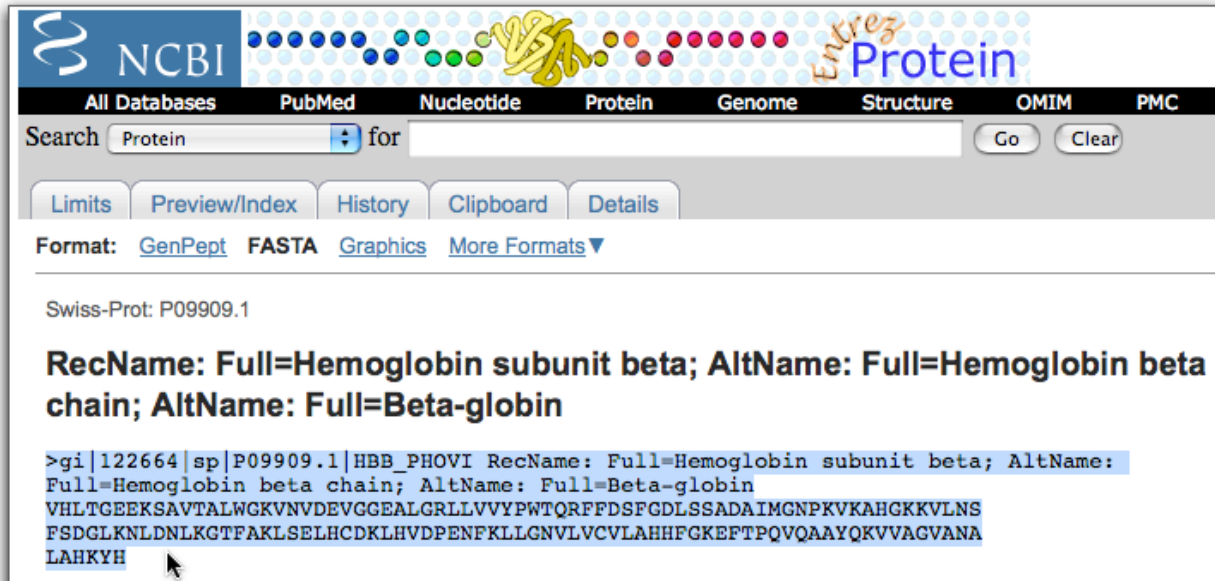
4. The FASTA page presents the amino acid sequence of the protein in a coded format using single letters to represent each of the 20 amino acids (A=alanine, M=methionine, P=proline, etc.) Copy the amino acid sequence. You **must** include the header line, starting from the greater than symbol (>). THIS IS IMPORTANT!

CREATE A FASTA FILE

5. Paste the amino acid sequence into a text file on your computer. Open up NotePad (on a PC) or Microsoft Word (on a Mac). Save as a .txt or "Text only" file. Save it in a logical location on your computer. We will refer to this as your "FASTA text file".

6. Go back to GenBank and collect the amino acid sequences for the HBB gene from minke whale (a baleen whale), *Canis familiaris* (dog), and *Bos taurus* (cow). Paste these amino acid sequences onto separate lines in the same FASTA text file.

7. Once you have all your sequences, we need to edit the file a bit to make our phylogenetic tree read more clearly. Let's look at the **sequence header** at the top of each protein sequence that we have copied into the FASTA text file. The species title that will show up on your tree will be the first line of each set of sequence data following the ">" symbol. So right now, our tree will say things like "gi|122664|sp|P09909.1". That won't make a lot of sense will it! This header can be edited for clarity, **but you MUST preserve the ">" symbol**. You can use the scientific name or the common name to identify your sequence.

For example:
The harbor seal sequence begins like this:

>gi|122664|sp|P09909.1|HBB_PHOVI RecName: Full=Hemoglobin subunit beta

This can be edited to simply say this:

>Harbor_seal

Tip: If you want to use more than one word in your label, like "harbor seal" you must add an underscore "_" between the words (harbor_seal) instead of a space between words. This is the only way that all the words will show up as labels on your tree. Do not use names longer than 30 characters.

8. Scan through your FASTA text file, it is critically important that it is formatted correctly. There must be a "paragraph return" or "hard return" (created by the Enter key) only after your header and only after the complete end of the sequence. Although it may appear that a hard return is already there, it is good practice to add one, because the hidden characters do not always cut and paste correctly.

    For example:

```
>Harbor_seal¶
VHLTGEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFDSFGDLSSADAIMGNPKVKAHGKKVLNSFSDGLKN
LDNLKGTFAKLSELHCDKLHVDPENFKLLGNVLVCVLAHHFGKEFTPQVQAAYQKVVAGVANALAHKYH¶
¶
>Minke_whale¶
VHLTAEEKSAVTALWAKVNVEEVGGEALGRLLVVYPWTQRFFEAFGDLSTADAVMKNPKVKAHGKKVLASFSDGLKH
LDDLKGTFATLSELHCDKLHVDPENFRLLGNVLVIVLARHFGKEFTPELQAAYQKVVAGVANALAHKYH¶
```

ALIGN YOUR SEQUENCES

9. Open **ClustalX**. This is the program that will align you sequences.

10. In the File menu, choose **Load Sequences**.

11. Select your FASTA text file. Your sequences should show up in the ClustalX window. Check to see that they are labeled correctly and that the first few letters in the ClustalX window correspond to the first few amino acids of each sequence. (If not, return to Step 8 and make changes.)
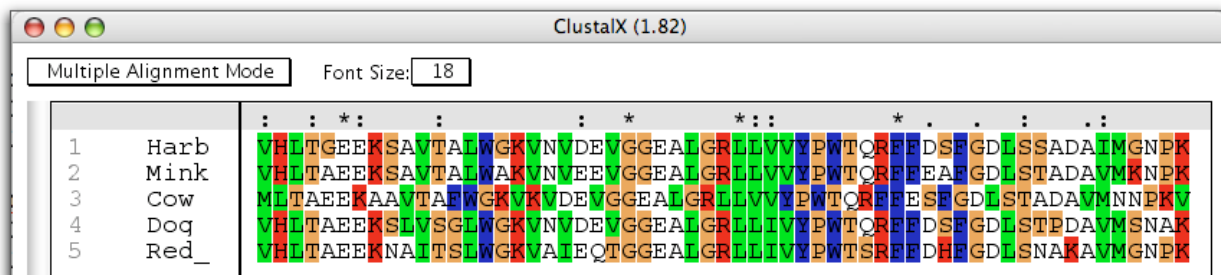
Tip: If your file will not load into ClustalX, or does not load correctly, check for the following common problems:

a. *Your file is in .doc or .rtf format. Look at the extension after the file name. It must end in .txt. Open it in Notepad or Word and save as a **plain text file**.*

b. *You have accidentally deleted the ">" character at the beginning of each sequence header. Simply add ">" back to each sequence header.*

c. *You are missing one or more hard returns at the end of each header and sequence. To fix this, place your cursor at the end of each sequence and header and consciously add a return even if one appears to be there already.*

Or you may have too many hard returns. Make sure there aren't any at the end of lines in the middle of the sequence of amino acids, but only at the end of the complete sequence.

d. *If a.-c. do not help, call your teacher over for assistance.*

12. Take a moment and look at the screen in front of you. Each line is the amino acid sequence of the same protein (hemoglobin beta) in different species. It is interesting to scan along the amino acid sequences and look how they line up — how they are the same in the different species and how they are different. You can see the traces of evolutionary processes here: where amino acids have changed, where they have stayed the same, and where amino acids have been lost. You are looking at the record of evolutionary history!

13. Now we need to better align the sequences to account for these changes and losses of amino acids. To do this go to the Alignment menu and choose **Do Complete Alignment**. (Alignment > Do Complete Alignment).

14. A new window will pop up that provides you with the name and file path of your results.
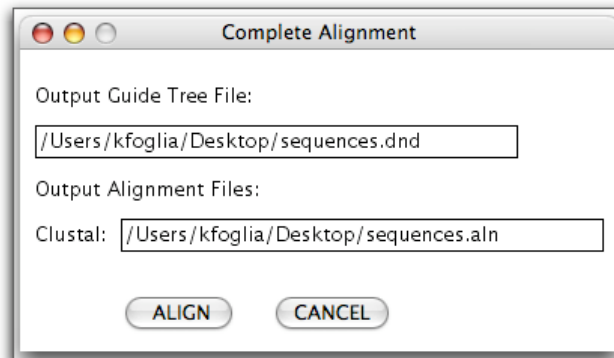
15. Make note of what is in each field so you will be able to find your file, and press **Align**:

    Tip: Your file path will appear differently than that shown above since this was specific for my computer when I was testing the exercise. That's OK. Just make a note of where your own file is for the next step. It should be in a logical place on your computer.

16. This align process has now lined up the amino acid sequences for each of the species vertically. It is now easier to see which parts of the protein are well-conserved (unchanged) and which parts of the protein have experienced mutations. The taller gray bars at the bottom of this chart represent the areas of the amino acid sequence that have been highly conserved through evolutionary time and the shorter gray bars represent the areas of the amino acid sequence that have experienced (and tolerated) genetic changes.

17. Take a screenshot of this alignment chart to use in your lab report (press the "Printscreen" key, typically labeled "PrtScn"). The picture of the screen is now waiting to be pasted into a document. Paste the screenshot into a Word document to use later.

BUILD THE TREE

18. Open **PhyloWin**. This is the program that will build the cladogram. Choose to **Open** and select your **sequences.aln** file using the file browser. This file can be found in the location you made note of in Step 13. You should see a colorful window with all your sequences.
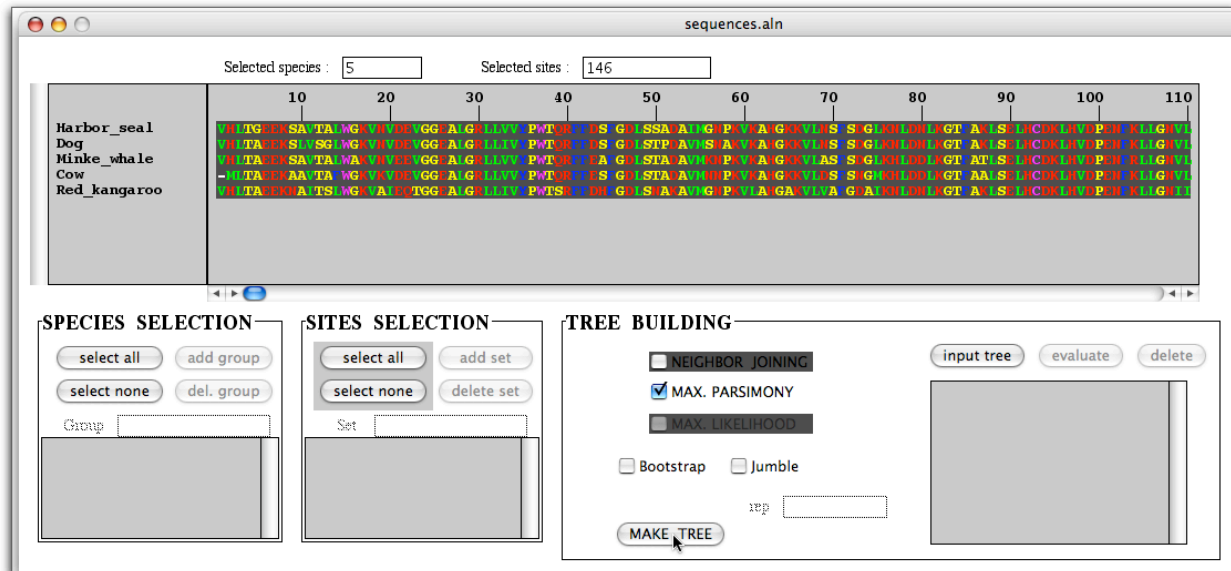
19. In the Tree Building section, check the **Max Parsimony** box. Click the **Make Tree** button. A new window will open with your phylogenetic tree.

20. We need to add one more step to make our tree more accurate. We need to add an "**outgroup**" to the mix of species we are analyzing. An outgroup provides a "root" to the tree by serving as an example of an ancestral state for the traits we are comparing. This clarifies the evolutionary relationships better. So we need to choose a species as an outgroup. We are going to use kangaroo as our outgroup in this investigation since it is a marsupial in contrast to all the other mammals in our study, which are all placentals. Therefore, the kangaroo is selected to be the *most different* organism from the other mammals on your tree. Go back to GenBank and get the amino acid sequence for hemoglobin for "red kangaroo". Paste this sequence into your original FASTA text file. Save it again. Load it and align it in ClustalX. Open the new sequences.aln file in PhyloWin.

21. Select the **new outgroup** button. Click the black box next to the **red kangaroo** to serve as the outgroup. Click the Make Tree button.

22. Unfortunately, you cannot print the tree directly from PhyloWin. Take a screenshot of your tree and then paste it into a Word document to be used in your lab report.

23. Discuss the tree with your class. What conclusions do you come to about the evolutionary relationship amongst seals, whales, dogs, and cows?

   NOTE: _Phylogenetic trees built with this software can only be used to make conclusions about common ancestry. They cannot be used to make conclusions about the timeframe of evolution. The length of branches is **not** a measure of evolutionary time. It is merely an artifact of physically arranging the tree._

PART B

24. Below are the representative marine mammals for your new cladogram:

   - minke whale (baleen whales)
   - dolphin (toothed whales)
   - harbor seal
   - walrus
   - otter
   - manatee

25. In this exercise, we will compare the hemoglobin proteins of these marine mammals to some representatives of the major taxa of land mammals:

   - carnivora: dog, _Canis familiaris_
   - rodentia: rat, _Rattus norvegicus_
   - herbivore: cow, _Bos taurus_
   - primates: human _Homo sapiens_
   - proboscidea: African elephant, _Loxodonta africana_
   - marsupials: red kangaroo, _Macropus rufus_

26. Reminder: We are testing the hypothesis that all marine mammals have a common land mammal ancestor.

27. Once again, use the hemoglobin beta protein sequence to test this hypothesis. You will be building a tree with **all** the marine mammals and **all** the land mammals listed above.

28. Follow the complete process we used in Part A to obtain your amino acid sequences from GenBank, align them in ClustalX, and build the evolutionary tree in PhyloWin. Take a screenshot of your aligned amino acid sequences and your phylogenetic tree and save them in a Word document to be included in your lab write up.

29. Then use your phylogenetic tree to determine the evolutionary relationships amongst marine mammals and the representative land mammals. Determine whether your hypothesis was supported by the molecular data.
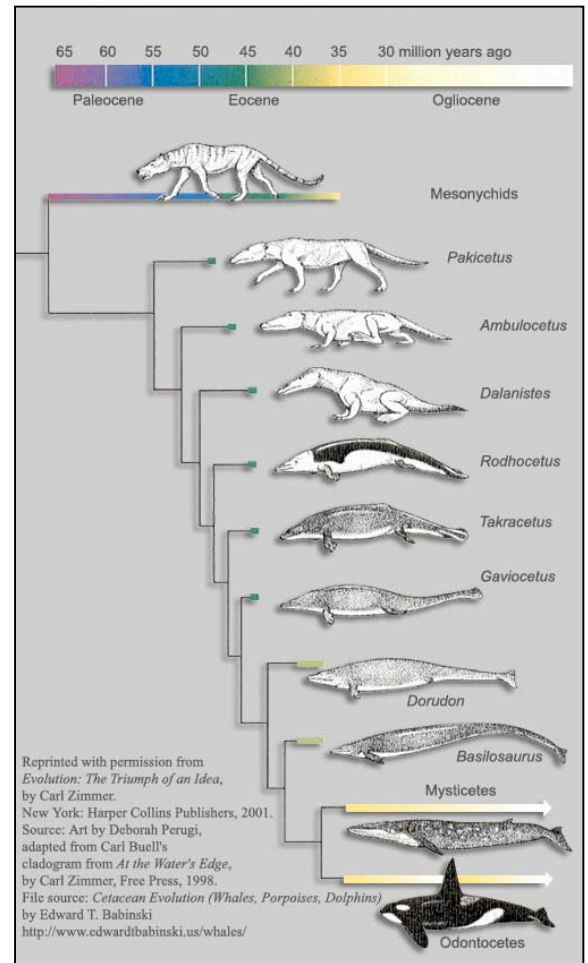
DISCUSSION

You will produce a standard lab report to describe this research. Be sure to include answers to the following questions in your lab report.

1. Did your tree support your hypothesis? Explain.

2. What does this phylogenetic tree structure suggest about the evolutionary history of marine mammals? Go into detail here about what parts of the tree lead you to what conclusions about the evolutionary history of the marine mammals.

3. If marine mammals share common morphological characteristics, what do your conclusions about their evolutionary history imply about these common characteristics?

4. Why did you use the protein sequence from the hemoglobin beta gene?

5. Why do you need to align the sequences (with ClustalX) before inputting them into PhyloWin?

6. What organism served as your outgroup? Why? What function does the outgroup serve?