

**Walruses, Whales and Hippos, Oh My:
Using Bioinformatics to Teach Cladistics and Evolution**

Kim B. Foglia
Biology Educator
Division Avenue High School
120 Division Ave.
Levittown, NY 11756
kim@ExploreBiology.com
(516) 520-8350

Dr. Stuart M. Brown
Associate Professor
Center for Health Informatics and Bioinformatics
NYU School of Medicine
550 First Ave
New York, NY 10016
stuart.brown@med.nyu.edu
(212) 263-7689

Walruses, Whales and Hippos, Oh My: Using Bioinformatics to Teach Cladistics and Evolution

In introductory biology courses, molecular evidence is presented as one of the key lines of evidence in support of evolutionary theory. Textbooks standardly employ diagrams which illustrate the relatedness of a range of species by comparing a homologous gene such as hemoglobin or cytochrome C. However, students are rarely offered the hands-on experience of building such a phylogeny from these data and then analyzing the tree.

The use of molecular sequence information to build phylogenetic trees and their subsequent analysis is now seen as a core concept in introductory biology and, in fact, was the topic of a 2009 essay question on the AP Biology exam. The use of software tools for the analysis of molecular biology data such as DNA and protein sequences lies within the increasingly important field of bioinformatics. We have developed a laboratory exercise that can be conducted in an Introductory or AP biology course (with access to computers or a computer lab), which introduces students to the methods of bioinformatics, allows them to explore the wealth of resources available in public online molecular biology databases (GenBank), and illustrates the basic principles of phylogenetics and cladistics. This exercise engages students by investigating the taxonomy of marine mammals through the proposed question: *Did all modern marine mammals evolve from a single common land mammal ancestor which returned to the ocean, or were there different return events and parallel evolution?*

In this way, the laboratory offers reinforcement of the evolutionary concepts of homology and convergent evolution. By approaching the exercise as an hypothesis-driven inquiry activity in evolutionary relationships, using state-of-the-art software tools, this laboratory supports the acquisition of content knowledge as well as the development of the scientific process skills of prediction, generating and testing hypotheses, gathering empirical data, analyzing data, and inference. Using the same tools as research biologists and gathering actual molecular data (rather than mock abridged sequences) gives students a feel for current phylogenetic research, promotes content retention, and offers practice in problem-solving.

The use of DNA and protein sequence information has come to dominate modern taxonomy, and these methods are employed in many different fields ranging from microbiology, epidemiology, conservation biology, and anthropology. While the computational tools used by experts in these fields may be complex, the essential aspects of building phylogenetic trees from sequences can be accomplished with two programs that are free, fairly easy to use, and available in both Mac and PC formats: ClustalX (Larkin et al, 2007) and PhyloWin (Galtier et al, 1996). It is advisable for the instructor to download and install these programs in advance on the computers that will be used by the students to reduce the likelihood of unexpected technological problems and so that class time is spent on the phylogenetics work rather than on the technology.

As a special note to high school teachers, one of the authors was successful in getting this software installed on the computers in her public school district, once the district's Information Technology (IT) staff was assured that (1) the software is state-of-the-art professional, scientific software, so there was no worry about receiving viruses or malware in the download, (2) the software is a free download, so it would not cost the district, and (3) the software is not resource-intensive, so neither program would tax either the network or the individual computers.

It is the authors' experience that many teachers do not feel confident with technology, in general, and bioinformatics tools, in specific. This lab is designed to make these tools easier to use and less threatening. A second agenda of this exercise is to show teachers that bioinformatics can be used to teach many aspects of the biology curriculum rather than remaining isolated as a footnote within the biotechnology unit. This lesson uses bioinformatics to reinforce evolution, taxonomy, cladistics, and molecular genetics.

This lab is designed to be undertaken across two lab days. On the first day, students are led by the teacher through an initial exercise of gathering protein sequence data on a small group of marine mammals and land mammals to learn how to use the GenBank Web site and the bioinformatics software. On the second day, to reinforce the lessons learned and to promote a more student-driven inquiry, students are given the opportunity to explore the evolutionary question of the origin of marine mammals on their own. In this part of the lab, students collect a wider range of sequence data and develop a more robust phylogenetic tree. This tree is then analyzed to make conclusions about the evolutionary history and relatedness of the marine mammals to the representative land mammals. Students are expected to write a lab report about their research from the second day to discuss their procedures, findings and conclusions.

BACKGROUND KNOWLEDGE AND TECHNICAL TIPS FOR TEACHERS

- Adult hemoglobin is a protein which is assembled from four subunits, consisting of two α and two β subunits, non-covalently bound (Bellelli et al, 2006; Schechter, 2008). In this exercise, the sequence of the beta subunits is compared amongst different mammalian species.
- Instead of typing in the long URL for GenBank, students can also search for “GenBank” at Google and it will always come up as the first link.
- GenBank is an actively growing database (Benson et al, 2000) and is therefore an evolving entity. Its interface is always changing, so the screenshots included in this article may be different than what you see when you visit the Web site. Do not be daunted by the change. It is merely the appearance and organization of the page. All the core functionality is still there and usually in a similar location.
- The GenBank database is designed to serve many different purposes, this lab uses only one aspect of it. For this exercise, we want to find protein sequences in different organisms. This is made easier if teachers know GenBank's conventions. All entries into GenBank are given “accession numbers” as ID numbers. Complete protein sequences generally have accession numbers in the format of “P#####”. If a search for the hemoglobin sequence of an organism returns several answers, look for the listing with this formatting.
- The GenBank accession numbers for the hemoglobin beta protein sequences of the marine mammals in this laboratory are: walrus (P68046), minke whale (P18984), dolphin (P18990), harbor seal (P09909), and manatee (P07415).
- The GenBank accession numbers for the hemoglobin beta protein sequences of the land mammals in this laboratory are: dog (P60524), rat (P02091), cow (P02070), human (P68871), hippopotamus (P19016), and red kangaroo (P02107).
- Genes and proteins are also listed in GenBank by their gene symbol. The hemoglobin beta gene and its protein are known by the symbol “HBB”.
- Whales are divided into two taxonomic groups, the baleen whales and the toothed whales (Fordyce & Barnes, 1994). In this exercise, the minke whale represents the baleen whales and the dolphin represents the toothed whales.
- When students save their files on networked computers, we have found it easiest to manage if the teacher creates a central folder, such as “Cladistics Lab”, for all students to use. Then, if

the teacher needs to access an individual student's file, they are all in one place. It is best if each student adds their last name to the file name as an identifier, like "sequences-Smith.txt".

- On a Macintosh computer, if you have any difficulty loading the FASTA sequence files into ClustalX, then move your sequence.txt files onto the Desktop first then load them into ClustalX again. We have had no such issue on PC computers.

LABORATORY EXERCISE

BACKGROUND INFORMATION FOR STUDENTS

Walruses, whales, dolphins, seals, and manatees are all marine mammals. 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? It is not possible to go back in time to observe what happened, but DNA and protein sequences contain evidence about the evolutionary history of organisms and the relationships between living creatures. Once we collect and analyze DNA or protein sequences of marine and land mammals, maybe the data will reveal the evolutionary history of marine mammals.

This lab uses 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 (dog, rat, cow, human, hippopotamus).

The analysis uses 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 (Bellelli et al, 2006). In addition, hemoglobin is also an easy test protein to use because it has been studied by many biologists, so sequences are available in GenBank from many different organisms.

The goal of this lab is to test hypotheses about the evolutionary ancestry of different marine mammals: Did marine mammals evolve from a single ancestor which returned to the ocean, or were there distinct return events from separate ancestors? A useful starting hypothesis is that all modern marine mammals have a single common land mammal ancestor.

Day 1 is a small-scale investigation in which students will explore the evolutionary relationships of two marine mammals and two land mammals. The exercise tests 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 provide training in using the bioinformatics software that will analyze the sequences and develop the phylogenetic tree.

Day 2 is a larger analysis, in which each student will develop a cladogram that includes a wide selection of marine mammals and land mammals, which represent the major mammalian orders. This phylogenetic tree is then used to test the hypothesis that all marine mammals have a single common land mammal ancestor.

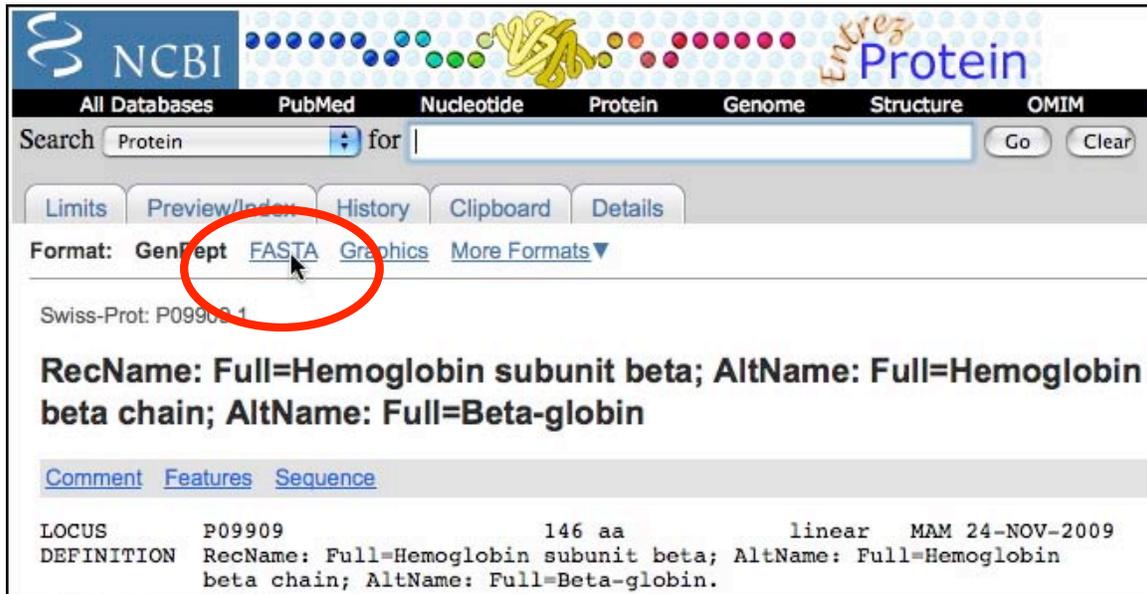
PROCEDURE: DAY 1

FINDING AMINO ACID SEQUENCES

1. In a web browser, go to GenBank, a DNA and protein sequence database (Benson et al, 2000) hosted by the National Center for Biotechnology Information (NCBI) in Maryland at: <http://www.ncbi.nlm.nih.gov/Genbank/>.
2. To measure the relatedness of species, students will compare the amino acid sequence of the hemoglobin protein from marine and land animals. There are a number of different hemoglobin proteins in mammals. This exercise will use the hemoglobin beta gene and its protein, which is known by the gene name “HBB”. In the “Search” window at the top of the NCBI home page, select “Protein” from the pull-down menu (Figure 1). To search for a protein in an organism, it is necessary to include the specific protein and the specific organism in the search field. In this part of the exercise, students will collect the amino acid sequences from the hemoglobin beta protein for the following animals: harbor seal, minke whale (a baleen whale), *Canis familiaris* (dog) and *Bos taurus* (cow). In the search field, type in “HBB” for the protein and then the organism name (like “HBB harbor seal” as shown in Figure 1), then click “Search”. In some cases, a Latin name for a species may yield a more precise query than a common name.



Figure 1. NCBI search tool.



NCBI Protein

All Databases PubMed Nucleotide Protein Genome Structure OMIM

Search Protein for | Go Clear

Limits Preview/lookup History Clipboard Details

Format: GenBank **FASTA** Graphics More Formats

Swiss-Prot: P09909.1

RecName: Full=Hemoglobin subunit beta; AltName: Full=Hemoglobin beta chain; AltName: Full=Beta-globin

[Comment](#) [Features](#) [Sequence](#)

LOCUS P09909 146 aa linear MAM 24-NOV-2009

DEFINITION RecName: Full=Hemoglobin subunit beta; AltName: Full=Hemoglobin beta chain; AltName: Full=Beta-globin.

Figure 2. NCBI search result for hemoglobin beta protein in harbor seal. The link to the FASTA format page for the amino acid sequence is highlighted.

- The search result returns a page containing a lot of information about the hemoglobin beta protein from this organism. To see the actual amino acid sequence for this protein, click on the “FASTA” link near the top of the page (Figure 2).
- 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.) To prepare the data for the tree-building software, copy the complete amino acid sequence. Include the whole header line, starting with the greater than symbol (>). *That detail is important!* (Figure 3)

The screenshot shows the NCBI Entrez Protein search results for Hemoglobin subunit beta. The page includes the NCBI logo, navigation tabs for All Databases, PubMed, Nucleotide, Protein, Genome, Structure, and OMIM. A search bar contains the word "Protein" and a search button. Below the search bar are buttons for Limits, Preview/Index, History, Clipboard, and Details. The format is set to FASTA. The protein information is displayed as follows:

Swiss-Prot: P09909.1

RecName: Full=Hemoglobin subunit beta; AltName: Full=Hemoglobin beta chain; AltName: Full=Beta-globin

```
>gi|122664|sp|P09909.1|HBB_PHOVI RecName: Full=Hemoglobin subunit beta; AltName: Full=Hemoglobin beta chain; AltName: Full=Beta-globin
VHLTGEKSAVTALWGKVNVDVVGGEALGRLLVVYPWTQRFFDSFGDLSSADAIMGNPKVKAHGKKVLNS
FSDGLKRLDNLKGTFAKLSLHCDKLVDPENFKLLGNVLVLCVLAHHFGKFTPPQVQAAAYQKVVAGVANA
LAHKYH
```

Figure 3. FASTA format page. The amino acid sequence of the hemoglobin beta protein in harbor seal.

CREATE A FASTA SEQUENCE FILE

- Paste the amino acid sequence and its header 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 so that you can easily find it again and add your last name to the file name as an easy identifier (such as sequence-Smith.txt). This is will be referred to as the “sequence text file”.
- Go back to GenBank and collect the amino acid sequences for the HBB protein from minke whale, *Canis familiaris* (dog), and *Bos taurus* (cow). Paste these amino acid sequences onto separate lines in the *same* sequence text file.
- Once all these sequences are saved in the sequence text file, it is useful to edit the file so that the phylogenetic tree will read more clearly. Look at the *sequence header* on the first line of each protein sequence in the sequence text file. The species name that will show up in the final phylogenetic tree will be the first word following the “>” symbol. So right now, the tree will be labeled with phrases like “gi|122664|sp|P09909.1”. Instead, a researcher would want the tree to list the species name for clarity. Edit this header by changing it to the scientific name or the common name of the organism, *but remember the “>” symbol must be*

preserved. However, do not mistakenly insert a blank space after the “>” symbol in this process. You are limited to 30 characters (Figure 4).

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 a species 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 the tree.

8. Scan through the sequence text file; it is critically important that it is formatted correctly. There must be a “hard return” (created by the Enter key) only after the header and only after the complete end of the sequence (Figure 4). 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.

```
>Harbor_seal
VHLTGE EKSAVTALWGKVNVEVGGEALGRLLVYYPWTQRFFDSFGDLSSADAI MGNPKVKAHGKVKVLSFSDGLKNLDN
LKGTFAKLSELHCDKLHVDPENFKLLGNVLCVLAHHPGKEFTPQVQAAYQKVVAGVANALAHKYH

>Minke_whale
VHLTAE EKSAVTALWAKVNVVEVGGEALGRLLVYYPWTQRFFEAFGDLSTADAVMKNPKVKAHGKVKLASFSGLKHLDD
LKGTFA TLSELHCDKLHVDPENFRLLGNVLVIVLARHFGKEFTPPELQAAYQKVVAGVANALAHKYH

>Cow
MLTAE EKA AVTAFWGKVKVDEVGGEALGRLLVYYPWTQRFFESFGDLSTADAVMNNPKVKAHGKVKVLD SFSNGMKHLDDL
KGTFAALSELHCDKLHVDPENFKLLGNVLVVV LARNFGKEFTPVLQADFQKVVAGVANALAHRYH
```

Figure 4. The amino acid sequence for the hemoglobin beta protein of several mammalian species formatted in the sequence.txt file.

ALIGNING THE SEQUENCES

9. Open ClustalX. This is the program that will align the amino acid sequences to each other.
10. In the File menu, choose “Load Sequences”.
11. Select your sequence text file. Your sequences should show up in the ClustalX window (Figure 5). 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, use the troubleshooting tips listed below.)

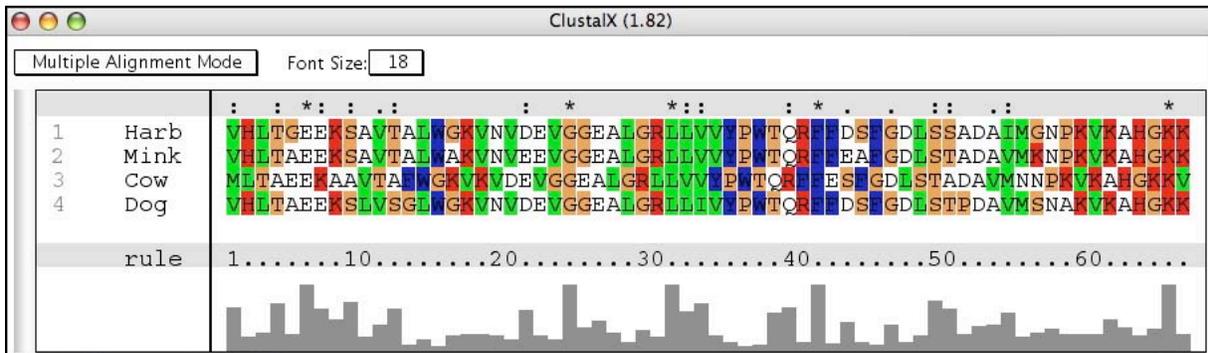


Figure 5. The sequence alignment window of ClustalX. These sequences have been loaded into the software, but have not yet been aligned.

Troubleshooting:
 If your file will not load into ClustalX, or does not load correctly, check for the following common issues:

- 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 or sequence. To fix this, place your cursor at the end of each sequence or header and add a return even if one appears to be there already.

12. The sequences need to be aligned to account for changes, additions, and losses of amino acids in the proteins from different species. To do this, go to the Alignment menu and choose “Do Complete Alignment”. (Alignment > Do Complete Alignment).
13. A new window pops up that provides the name and file path of the alignment results (Figure 6). This saves two files: a .dnd file and a .aln file. Each field shows the student the path to where the file is saved. It should be in the same folder as the sequence text file. Make note of the location, in case it is different. Press “Align” (Figure 6).

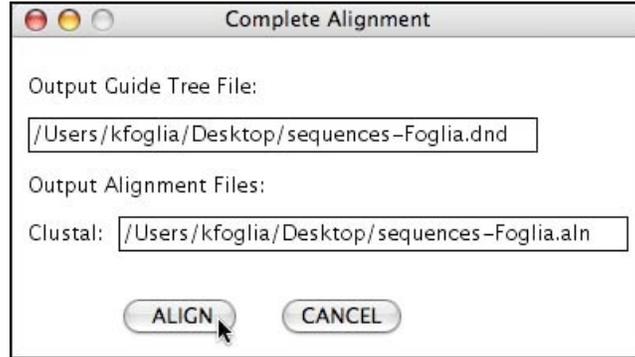


Figure 6. The Complete Alignment window in ClustalX showing the path where the alignment files will be saved.

14. This process has now vertically aligned the sequences. Remember each row is the amino acid sequence of the same protein in different species. It is interesting to observe how the sequences line up—how they are the same in the different species and how they are different. One can see the traces of molecular processes here: where amino acids have changed, where they have stayed the same, and where amino acids have been lost. This is a record of evolutionary history!

An interpretation of the alignment is also illustrated in the bar graph at the bottom of the alignment image (Figure 5). The taller grey bars represent the areas of the amino acid sequence that have been highly conserved through evolutionary time and the shorter grey bars represent the areas of the amino acid sequence that have experienced genetic changes.

Consider why some changes to the amino acid sequence were tolerated by the protein and why some areas of the protein remain unchanged over millions of years. Suggest molecular events that could have changed the amino acid sequence of this protein.

15. Take a screenshot of this alignment chart, in case it is needed in a student lab report.

Tip: To take a screenshot on PC: press the “Print Screen” key, typically labeled “PrtScn” and then Paste into a Word document. On Mac: pressing Command+Shift+3 will take a screenshot of the whole screen; pressing Command+Shift+4 will turn your cursor into a cross-hair so that you can click and drag to the exact dimension of your preferred shot. The screenshot is saved to the Desktop and can be inserted into a Word document.

BUILD THE TREE

- Open PhyloWin, the program that will build the cladogram. Choose to “Open” and select the “sequences.aln” file (which was created by ClustalX) using the file browser. The sequences should now be visible in the PhyloWin window (Figure 7).

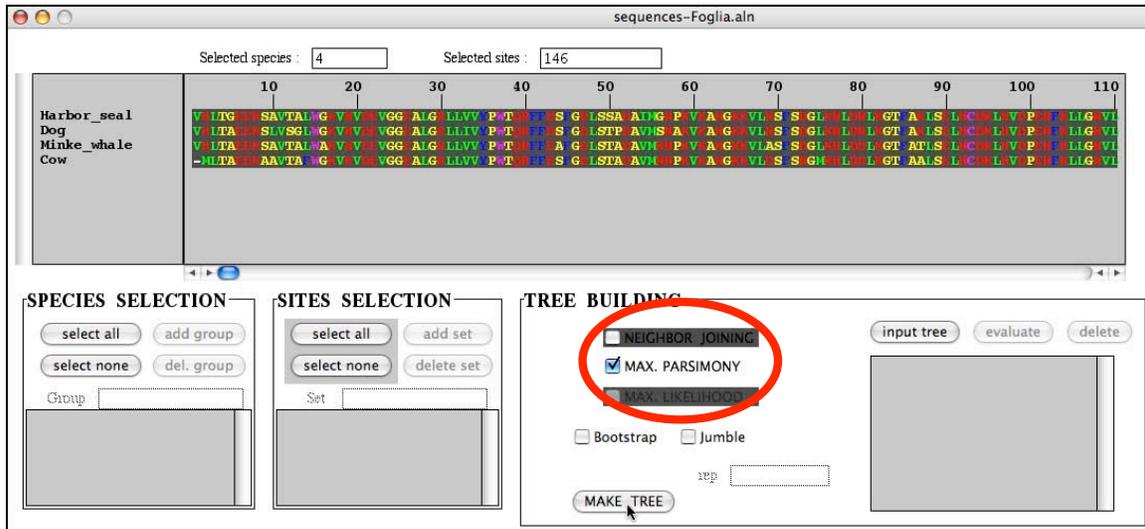


Figure 7. The PhyloWin window for building phylogenetic trees. Maximum parsimony option is selected.

- In the Tree Building section, check the “Max Parsimony” box. Click the “Make Tree” button (Figure 7). A new window will open with your phylogenetic tree.
- There is one additional step which will make the tree more accurate. It is necessary to add an *outgroup* to the mix of species being analyzed. An outgroup provides a *root* to the phylogenetic tree by serving as an example of an *ancestral state* for the traits we are comparing. This clarifies the evolutionary relationships. So a species needs to be chosen to serve as an outgroup. Kangaroo is a good choice as an outgroup in this investigation, since it is a marsupial mammal in contrast to all the placental mammals in our study. Therefore, the kangaroo is selected to be the *most different* organism from the other mammals on the tree. Go back to GenBank and get the amino acid sequence for hemoglobin for *red kangaroo*. Paste this sequence into the original sequence text file. Save it again. Then load it and align it in ClustalX. And once again, open the new “sequences.aln” file in PhyloWin and make the tree.
- Select the “new outgroup” button. Click the black box on the red kangaroo branch to select it as the outgroup (Figure 8). Then select “show tree” to remove the black boxes.

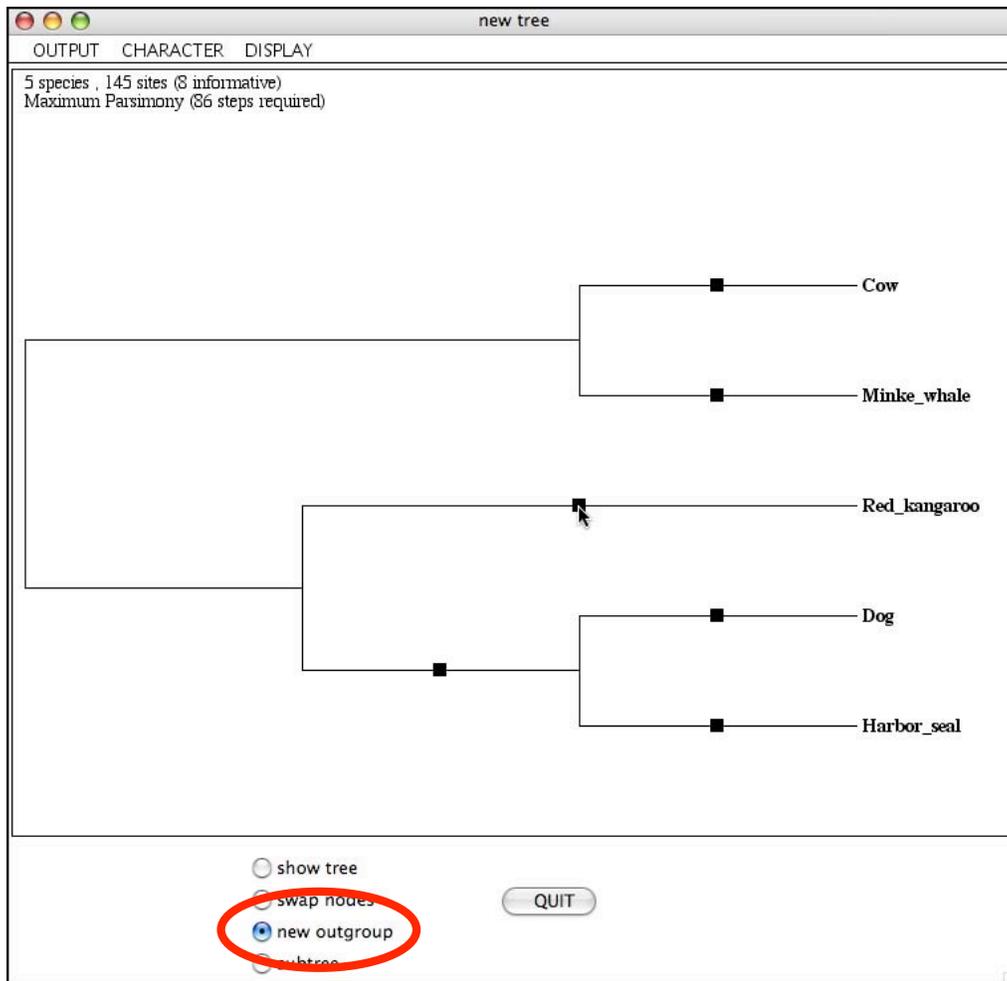


Figure 8. Phylogenetic tree built by PhyloWin. The outgroup is being selected.

20. Unfortunately, the phylogenetic tree cannot be printed directly from PhyloWin. Take a screenshot of this alignment chart, in case it is needed in a student lab report.
21. Study the phylogenetic tree and discuss with other students. What conclusions can be made about the evolutionary relationships 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.

PROCEDURE: DAY 2

Now that students understand how to use the GenBank database and the alignment and tree-building software, it is time for a broader analysis of the origin of marine mammals. Each student will develop a cladogram that includes a wide selection of marine mammals and land mammals which represent the major mammalian orders in the sea and on land. Students will then use this

highlights the traces of molecular processes: where amino acids have changed, where they have stayed the same, and where amino acids have been lost. An interpretation of the alignment is also illustrated in the bar graph at the bottom of the alignment image. The taller grey bars at the bottom of the chart represent the regions of the amino acid sequence that have been highly conserved through evolutionary time and the shorter grey bars represent the regions of the amino acid sequence that have experienced genetic changes. The highly conserved amino acids are assumed to be regions that are critical to the protein's oxygen-transport function and therefore cannot tolerate mutations whereas the sections that have accrued genetic changes were able to tolerate those changes and are therefore assumed to be non-critical regions of the protein's structure. Note the loss of the initial amino acid in the hemoglobin beta sequence of cow. This indicates that there was a deletion of an initial triplet codon in the ancestry of this land mammal. The final alignment from Day 2 will shed more light on this molecular record. Scan along the complete sequence to see other changes. You cannot over-emphasize to students that this alignment is a record of evolutionary history.

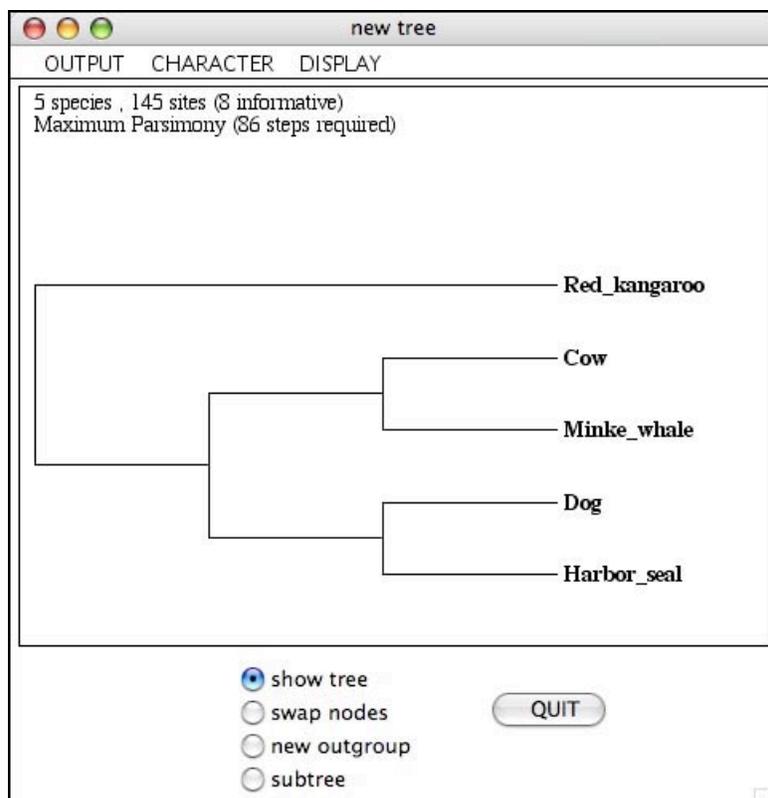


Figure 10. Phylogenetic tree built by PhyloWin showing evolutionary relationships from Day One exercise.

The tree built from these protein data (Figure 10) gives the teacher an opportunity to lead the class in an interactive analysis of its implications. This discussion lays the foundation for students so they can analyze their expanded tree from the research on Day Two. The branching pattern on this cladogram is already suggesting that marine mammals arose from more than one line of land mammals. Harbor seal is more closely related to dog, the representative of land carnivores, than to minke whale. Minke whale is more closely related to cow, the representative of the land herbivores, than to harbor seal. Therefore, the molecular evidence of just these few species leads one to the conclusion that these marine mammals do not share a recent common ancestor and that

they are the products of separate evolutionary origins—distinct “returns to the sea”—and the morphological similarities between them were shaped through convergent evolution. Be sure to review the need for and value of the outgroup, red kangaroo. The segue to Day Two is the suggestion that the class needs to collect more data on more species to clarify the relationships further.

When students complete the Day Two exercise they will have built the sequence alignment and phylogenetic tree seen in Figures 11 and 12.

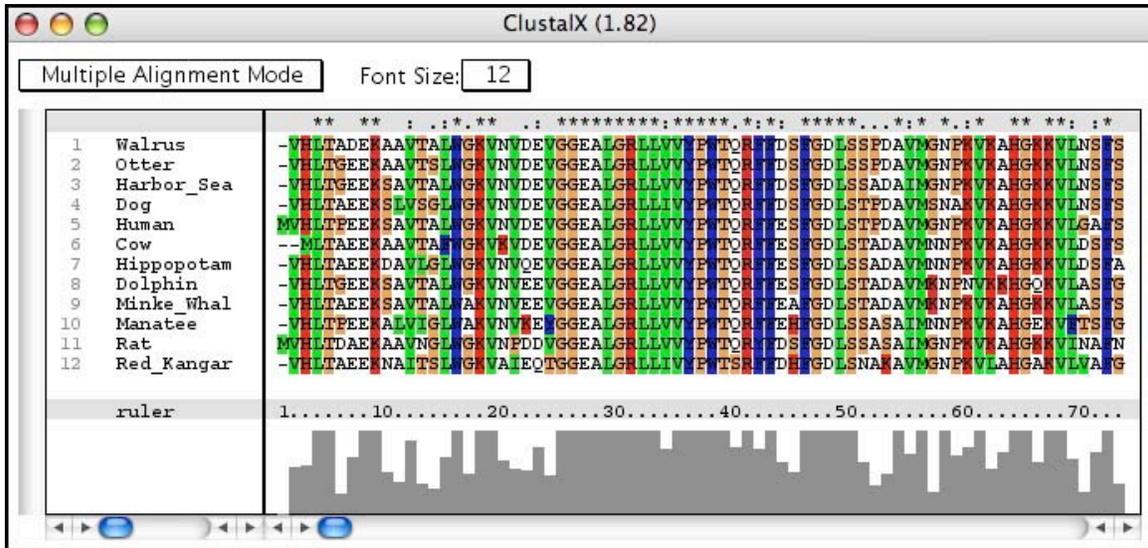


Figure 11. ClustalX window showing aligned HBB sequences from Day Two exercise.

Just as in the sequence alignment from Day One, there is a lot of evolutionary history displayed in this protein comparison. What we can now see—which was not evident in the first alignment—is that there has actually been a loss of the initial methionine at the start of the polypeptide chain in all the mammals except for human and rat. This indicates that there was a deletion of a triplet codon from the gene for this protein in the ancestry of the other mammals. And we can also now see that cow has actually experienced the deletion of two triplet codons at the start of its hemoglobin beta gene. This is a great opportunity to reiterate the connection between changes in a gene to changes in a protein and emphasize that mutations are the raw materials of evolution.

The phylogenetic tree built from this amino acid alignment (Figure 12) has a wealth of information in it and can support the following conclusions about the evolutionary history of the marine mammals.

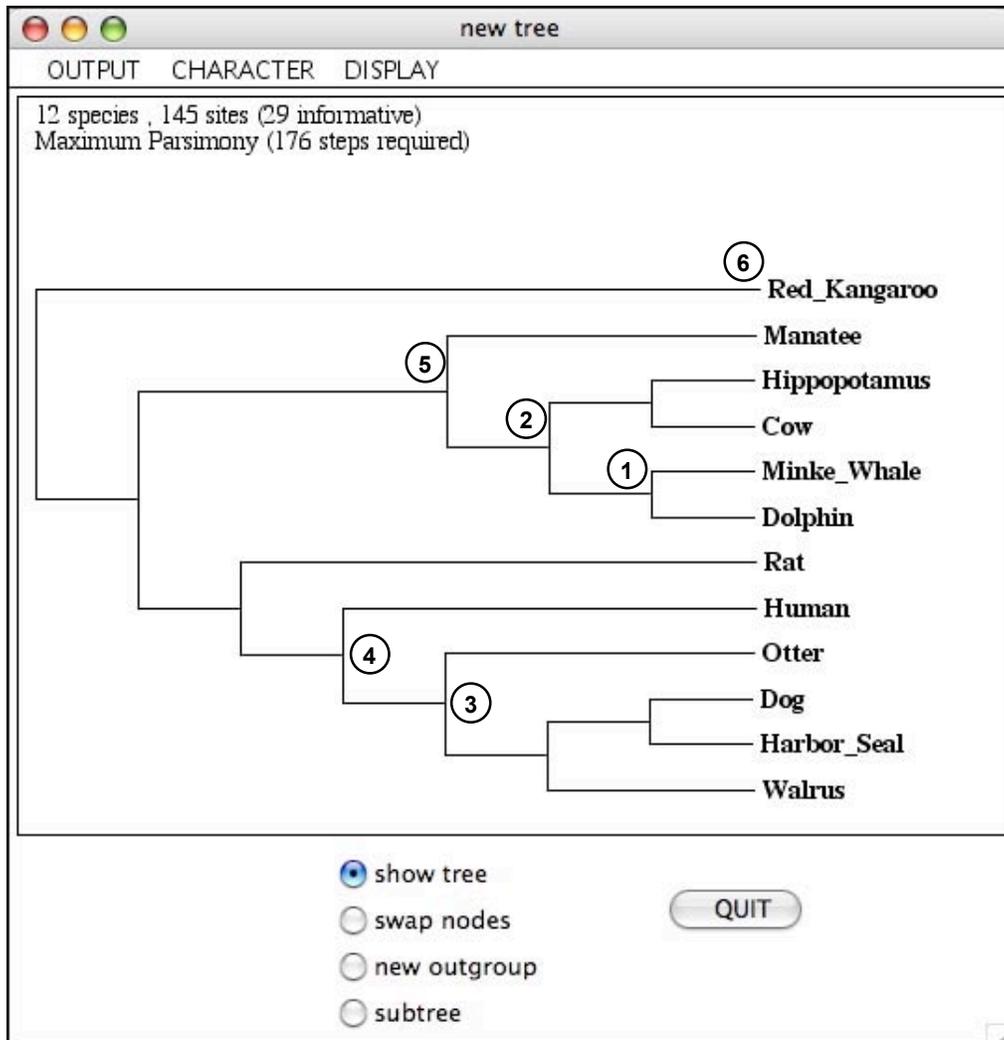


Figure 12. Phylogenetic tree built by PhyloWin showing evolutionary relationships from Day Two exercise. The numbered sites are discussed in the text.

Site 1 signifies that the whales, both toothed and baleen, come from a common ancestor. In addition, Site 2 indicates that the whales share a common ancestor with the land mammal group that includes cow and hippopotamus (Artiodactyla)—the even-toed ungulates (land herbivores). This suggests several conclusions about the physiological and morphological adaptations of these mammals. It would seem that having teeth is the ancestral condition and that the baleen of the minke whale is a derived characteristic. These four species share a common ancestor, yet cow and hippopotamus are herbivores whereas the whales are carnivores. This offers an opportunity to discuss how rapidly diet, feeding behavior, nutritional requirements, and the associated anatomical structures can change over relatively short periods of evolutionary time, just like Darwin's finches.

Site 3 signifies that walrus, otter and seal have a common ancestor and they share this ancestor with dog and other members of the order Carnivora. However, these marine mammals have a distinct lineage from the whales and dolphins. This branch of the cladogram clearly suggests that there were separate evolutionary events that produced marine mammals. Site 4 indicates that human and rat are more closely related to the carnivore group of the walrus, otter, seal and dog than the whale and Artiodactyla group.

Site 5 shows manatee on its own branch, sharing no common ancestor with any representative marine or land mammal. This would suggest that manatee ancestors returned to the sea in a separate evolutionary event and that the ancestors and land mammal relatives of the manatee are now extinct.

Please remember that the relative lengths of branches on this cladogram cannot be used to suggest a rate of evolution for these species. The PhyloWin analysis of the aligned sequences does not take time into consideration, so no conclusions can be made about the relative timing of each branch point as compared to any other.

The cladogram offers an answer to the research hypothesis: *Marine mammals evolved from a single common ancestor—a land mammal who returned to the ocean and diversified there.* It appears that the marine mammals represent at least three distinct return events: the whales (both toothed and baleen), the walrus/otter/seal group, and the manatee. Their ancestors are represented by a range of modern-day mammals from different orders. Consequently the hypothesis in this research was not supported and the common morphological traits of the marine mammals are examples of convergent evolution.

However, the cladogram analyzed in this research was based on the amino acid sequence of a single protein chain. This research would be just a beginning of the taxonomic analysis. In order to develop a more robust understanding of the evolutionary relationships between marine and land mammals, more research would need to be done using the analysis of other proteins and genes as well as evaluating fossil data.

Systematics is often seen by students as a static science. This exercise can be concluded by discussing with students that systematics is actually a very active and evolving field, especially in the light of new molecular data and new fossil discoveries. The evolutionary relationships of present-day and extinct organisms—including marine mammals—is still being hotly debated. Some systematists place whales within the Artiodactyla closely related to hippopotamus and others keep whales as a distinct taxon separate from the Artiodactyla although sharing a common ancestor just as this exercise concluded (Spaulding et al, 2009).

REFERENCES

- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., & Wheeler, D.L. (2000) GenBank. *Nucleic Acids Research*, 28, 15-18.
- Bellelli, A., Brunori, M., Miele, A.E., Panetta, G. & Vallone, B. (2006). The allosteric properties of hemoglobin: insights from natural and site directed mutants. *Current Protein Peptide Science* 7, 17-45.
- Fordyce, R.E. & Barnes, L.G. (1994). The evolutionary history of whales and dolphins, *Annual Review of Earth Planetary Science*, 22.

- Galtier, N., Gouy, M. & Gautier, C. (1996) SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. *Computer Applications in the Biosciences*, 12, 543-548.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H.*, Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G. (2007) ClustalW and ClustalX version 2. *Bioinformatics*, 23, 2947-2948.
- Schechter, A.N. (2008). Hemoglobin research and the origins of molecular medicine. *Blood*, 112, 3927-3938.
- Spaulding, M., O'Leary, M.A. & Gatesy, J. (2009). Relationships of Cetacea (Artiodactyla) among mammals: increased taxon sampling alters interpretations of key fossils and character evolution. *PLoS One*, 4(9):e7062.