

DETERMINING EVOLUTIONARY RELATIONSHIPS

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

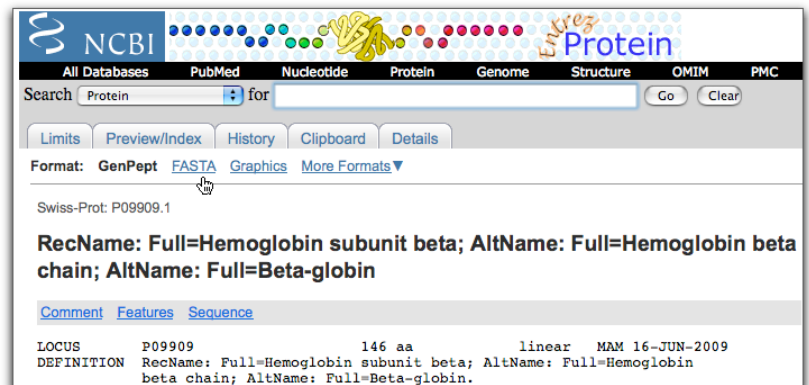
FINDING AMINO ACID SEQUENCES

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: www.ncbi.nlm.nih.gov/genbank/ (Note: You can also Google search "GenBank" and it will come up as the first hit.)

2. 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.

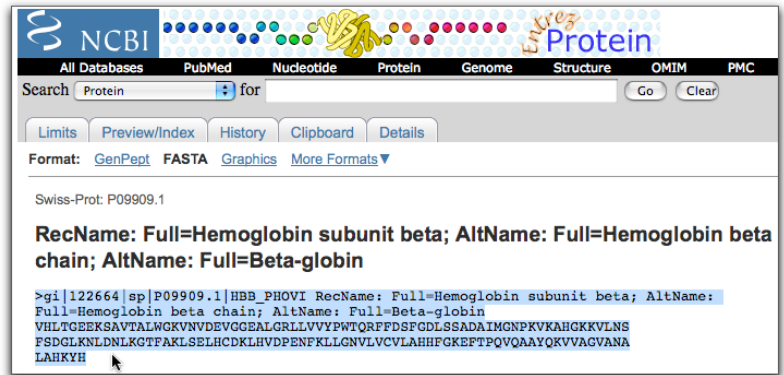


The screenshot shows the NCBI GenBank protein search results page. The search criteria are: Protein, for [search term]. The format options are: GenPept, FASTA, Graphics, More Formats. The protein name is: RecName: Full=Hemoglobin subunit beta; AltName: Full=Hemoglobin beta chain; AltName: Full=Beta-globin. The accession number is: Swiss-Prot: P09909.1. The FASTA link is highlighted.

LOCUS	P09909	146 aa	linear	MAM 16-JUN-2009
DEFINITION	RecName: Full=Hemoglobin subunit beta; AltName: Full=Hemoglobin beta chain; AltName: Full=Beta-globin.			

DETERMINING EVOLUTIONARY RELATIONSHIPS

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.) Highlight and copy the entire amino acid sequence and the wording preceding it. You **must** include the header line, starting from the greater than symbol (>). THIS IS IMPORTANT!



The screenshot shows the NCBI Entrez Protein search results for the protein Hemoglobin subunit beta. The search term "Protein" is entered in the search bar. The results show the Swiss-Prot accession number P09909.1. The record name is "Full=Hemoglobin subunit beta; AltName: Full=Hemoglobin beta chain; AltName: Full=Beta-globin". The amino acid sequence is displayed below the header, starting with the greater-than symbol (>).

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; AltName: Full=Beta-globin; AltName: Full=Hemoglobin beta chain
```

This can be edited to simply say this:

```
>Harbor_seal
```

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 (see next page for example).

DETERMINING EVOLUTIONARY RELATIONSHIPS

For example:

```
>Harbor_seal
VHLTGE EKSAVTALWGKVNVEVGGEALGRLLVVPWTQRFFDSFGDLSSADAIMGNPKVKAHGKKVLNLSFSDGLKN
LDNLKGTFAKLSLHCDKLHVDPENFKLLGNVLVLCVLAHFGKEFTPQVQAAYQKVVAGVANALAHKYH

>Minke_whale
VHLTAE EKSAVTALWAKVNVVEVGGEALGRLLVVPWTQRFFEAFGDLSTADAVMKNPKVKAHGKKVLASFS DGLKH
LDDLKGTFAKLSLHCDKLHVDPENFRLLGNVLVIVLARHFGKEFTPQLQAAYQKVVAGVANALAHKYH
```

ALIGN YOUR SEQUENCES

9. Open the website **Clustal Omega**. This is the program that will align your amino acid sequences. <http://www.ebi.ac.uk/Tools/msa/clustalo/>

10. Paste in your set of amino acid sequences for your sample organisms and hit "submit". It may take a few seconds to process.

Note: Leave the sequences formatted exactly as you copied them from GenBank. DO NOT remove line breaks to make it one continuous sequence in the input box.

11. You will now see various tabs at the top of the page. Click on "Result Summary" then click on "Percent Identity Matrix" at the bottom. This table shows you the percent similarities between the species' hemoglobin sequence, and therefore their **evolutionary relationship**. The more amino acids they have in common, the more closely related the species.

To help you read this data, imagine there is a top row listing the organisms in the order as presented on the left. This means if you match up each row and column, you will see the percentage of how similar their amino acid sequences are. E.g., a cow and whale share 86.9% of the same hemoglobin beta amino acid sequence.

	Cow	Whale	Seal	Dog	
1: Cow	100.00	86.90	84.14	82.07	Imagine this row is here
2: Minke_whale	86.90	100.00	88.36	86.30	
3: Harbor_seal	84.14	88.36	100.00	93.15	
4: Dog	82.07	86.30	93.15	100.00	

12. Click on "Show Colors" to create a color-coded grid comparing the amino acids of each organism.

13. 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!

DETERMINING EVOLUTIONARY RELATIONSHIPS

CLUSTAL O(1.2.1) multiple sequence alignment

```
Cow      -MLTAEKKAAVTAFWGVVDEVGGEALGRLLVYPWTQRFESFGDLSTADAVMNNPKV
Minke_whale VHLTAEKSAVTALWAKVNVDEVGGEALGRLLVYPWTQRFEEAFGDLSTADAVMKNPKV
Harbor_seal VHLTGEKSAVTALWGVNVDEVGGEALGRLLVYPWTQRFDFSGDLSSADAIMGNPKV
Dog       VHLTAEKSLVSGLVGVNVDEVGGEALGRLLVYPWTQRFDFSGDLSTPDVMSNAKV
          **.*:** *:.:*.**:*:*****:*****:*****: **:* **
```

14. Now we need to better align the sequences to account for these changes and losses of amino acids. To do this, click on the "Phylogenetic Tree" tab. You will analyze the phylogenetic tree to see if it supports or refutes your hypothesis.

PART B

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

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

16. 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*

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

18. 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.

19. Follow the complete process we used in Part A to obtain your amino acid sequences from GenBank, align them in Clustal Omega, and build the evolutionary (phylogenetic) tree. Alternatively, your teacher may already have the amino acid sequences for Part B saved for you, in which case bypass the GenBank step and go straight to Clustal Omega.

20. 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.