Investigating Evolutionary Questions
Using Online Molecular Databases

Lesson Background and Overview   [Student Information Handout]

How does an evolutionary biologist decide how closely related two different species are? The simplest way is to compare the physical features of the species (their "morphologies"). This method is very similar to comparing two people to determine how closely related they are. We generally expect that brothers and sisters will look more similar to each other than two cousins might. If you make a family tree, you find that brothers and sisters share a common parent, but you must look harder at the tree to find which ancestor the two cousins share. Cousins do not share the same parents; rather, they share some of the same grandparents. In other words, the common ancestor of two brothers is more recent (their parents) than the common ancestor of two cousins (their grandparents), and in an evolutionary sense, this is why we say that two brothers are more closely related than two cousins.

Similarly, evolutionary biologists might compare salamanders and frogs and salamanders and fish. More physical features are shared between frogs and salamanders than between frogs and fish, and an evolutionary biologist might use this information to infer that frogs and salamanders had a more recent common ancestor than did frogs and fish.

This methodology certainly has problems. Two very similar looking people are not necessarily related, and two species that have similar features also may *not* be closely related. Comparing morphology can also be difficult if it is hard to find sufficient morphological characteristics to compare. Imagine that you were responsible for determining which two of three salamander species were most closely related. What physical features would you compare? When you ran out of physical features, is there anything else you could compare?

Many biologists turn next to comparing genes and proteins. Genes and proteins are not necessarily better than morphological features except in the sense that differences in morphology can be a result of environmental conditions rather than genetics, and differences in genes are definitely genetic. Also, there are sometimes more molecules to compare than physical features.

In the three following exercises, you will use data in a public protein database of gene products (proteins) to evaluate evolutionary relationships. In all three exercises, you will work with the hemoglobin beta chain. You will obtain your data from a public online database that contains the amino-acid sequences of proteins coded for by many genes for many different organisms. Hemoglobin, the molecule that carries oxygen in our bloodstream, is composed of four subunits. In adult hemoglobin, two of these subunits are identical and coded for by the alpha hemogobin gene. The other two are identical and coded for by the beta-hemoglobin genes. The hemoglobin genes are worthy of study themselves, but today we will just use the protein sequences as a set of traits to compare among species.

Materials for each student or pair of students:
 1) Background and Overview package
 2) Packet of figures for reference
 4) Copies of two articles for reference
 3) Packet of tables for data collection
 5) One computer with access to the Internet and with a word-processing program and an Internet browser
     running simultaneously, and students prepared with basic internet and word-processing skills.

PART I: "Are Bats Birds?"
Calvin, from the "Calvin and Hobbes" cartoon series, failed his first school report "Bats are Bugs". He is trying to repair his report and has changed the title to "Bats are Birds". Help convince Calvin that just because bats have wings, they are not really birds, but in fact, bats share more features with mammals than with birds.

Procedure Part I: (Answer questions in spaces provided on Data Table sheets)
1.1) What morphological features do bats share with mammals? With birds? Fill in **Table 1**.
1.2) Answer question 1.2 and 1.3 on Data Sheet.

2. Generate a **distance matrix** for the beta-hemoglobin chain for two bird species, two bat species, and two non-bat mammal species into a word processing worksheet. Follow the steps below to do this:

Step 1: Begin by going to the ExPASy server at http://ca.expasy.org (see **Figure A**).

Step 2: In the "Databases" box, find "SWISS-PROT and TrEMBL", and click on it. This will fetch the SWISS-PROT Protein Knowledgebase (**Figure B**). Scan down to the "Access to SWISS-PROT and TrEMBL" box, and click on "Full text search". This will display the "Full text search" page (**Figure C**). In the future, you may want to go here directly, by typing in http://au.expasy.org/cgi-bin/sprot-search-ful

Step 3: Under "Enter search keywords:" type the phrase **hemoglobin beta** and mouse-click on "**submit**." The computer will retrieve many entries and display them (**Figure D**).

Step 4: Use the right-hand scroll bar to scroll through the names of the many entries. Find one for either a bird, a bat, or some other mammal. When you find one, check to make sure that it is the hemoglobin *beta* chain (preferably without a number after it) and *not* the alpha or gamma or other hemoglobin subunit. If the sequence is for the beta chain and it is for an appropriate species, click on it and the computer will retrieve the sequence.

Step 5: The next screen contains a lot of information. The protein sequence is near the bottom of the information sheet in the "Sequence information" section (for an example, see **Figure E**). Using the right-hand scroll bar, find the amino-acid sequence. The amino acids are indicated with their single-letter symbols (see **Figure F** for their full names), and every 10th amino acid is marked with its position.

Step 6: At the lower right of the sequence (see **Figure E**), click on "**FASTA** format."   This will simply provide you with the condensed sequence for that species, along with the species identification.

Here is a sample of that species information and its beta hemoglobin sequence:
>sp|P02140|HBB_CARAU HEMOGLOBIN BETA CHAIN - Carassius auratus (Goldfish).
VEWTDAERSAIIGLWGKLNPDELGPQALARCLIVYPWTQRYFATFGNLSS
PAAIMGNPKVAAHGRTVMGGLERAIKNMDNIKATYAPLSVMHSEKLHVDP
DNFRLLADCITVCAAMKFGPSGFNADVQEAWQKFLSVVVSALCRQYH

Step 7: Use your mouse to **select** and **copy** the information, go to a blank sheet in your word processor, and **paste** the information into that word-processing sheet.

Step 8: Repeat the above steps until your word-processing sheet contains the FASTA formatted sequences for **six** different species: two bird species, two bat species, and two non-bat mammal species. Write the **names** of the species you chose into **Table 2**.

Step 9: **Save** your word processing sheet but do not close it.

Step 10: To align the sequences and determine how similar they are, go to an internet alignment program, e.g., "LALIGN" at: http://fasta.bioch.virginia.edu/fasta/lalign.htm (**Figure G**).
(*Other online alignment programs exist:* http://www.expasy.ch/tools/sim-prot.html; change the defaults to user entered sequence or, to make it even easier, just put in the sequence code given in Swis-prot or http://www.ch.embnet.org/software/LALIGN_form.html where you should change the default to *global.*)

Step 11: **Select**, **copy** and **paste** one sequence from your word processing sheet into the first sequence box and another sequence into the second sequence box as shown in **Figure G**. For simplicity's sake, just copy the protein sequence and not any of the identification information. However, make sure you keep track of which two species' data you have entered.

Step 12: Click on "**Align Sequences**." (You may need to wait a short time initially for this to be done.)

Step 13: The computer will return a set of information including the "percent identity in the 146 aa overlap" (**Figure H**). Record that piece of information in the **Table 3** grid. This value is essentially the percent of amino acids that are similar. If all the amino acids were the same, the percent would be 100%. Not only does LALIGN give you the percent similarity, it also shows you the actual allignment of the two sequences. Identical amino acids are marked with two dots between them (:). If there is one dot, the change in amino acid is conservative (both amino acids have similar properties and charge), and if there are no dots, then the two amino acids have different biochemical properties.

Step 14: A **distance matrix** is a table that shows all the pairwise comparisons between species. Continue to make all pairwise comparisons until **Table 3** is filled. For each comparison, use the **percent identity** for the overlap of all the 146 amino acids.

Step 15: Use **Table 3** to answer the 6 questions listed below that table

PART II: "Whale Feet?"

Help a graduate student figure out what the feet of the ancestors of whales were like! She knows that whales are mammals and that, based on earlier studies, they very likely evolved from terrestrial four-legged hooved mammals (ungulates) which, over many generations and millions of years, gave rise to animals fully adapted to life in the sea: whales. She heard that whales could have evolved either from perissodactyl (odd-toed) or from artiodactyl (even-toed) ungulates, and that these two ungulate groups have distinct differences in their feet. Unfortunately, modern whales do not have feet so she can not examine the morphology of those whales to figure out which is the more likely scenario. Help her use hemoglobin beta chain sequence data to evaluate these two hypotheses.

Procedure Part II: (Answer questions in spaces provided on data table sheets)

Examine some specimens or pictures of whales, fish, and perissodactyl (odd-toed) and artiodactyl (even-toed) mammals, and take note of the morphological differences of these. **Repeat the process used in Part I**, but this time construct a **distance matrix** for one whale, one fish, two perissodactyl mammals, and two artiodactyl mammals. Describe the morphological distinctions between perissodactyls and artiodactyls. What are some examples of perissodactyl ungulates? What are some examples of artiodactyl ungulates?
Fill in and use **Tables 4 and 5** and answer the questions on your Data Sheets (4.1 - 5.3b).

The graduate student's advisor suggested that she examine published paleontological and molecular studies to see what other researchers have concluded about the closest relatives of whales. Examine the papers referenced below (or others suggested by your teacher), and determine whether your conclusion is the same as their conclusions regarding the feet of the ancestors of whales. Return to Question 5.4 on your Data Sheets.

Gingerich, P.D., Smith, B. H., & Simons, E.L. (1990). *Science* **249**: 154-157.
Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I., & Okada, N. (1997). *Nature* **388**: 666-670.

Part III: "Reptiles With Feathers?"

Some phylogenetic systematists (scientists who work to make the classification of organisms match their evolutionary history) complain that the vertebrate class Reptilia is improper because it should include birds. In technical terms, the vertebrate class Reptilia is *paraphyletic* because it contains some but not all of the species that arose from the most recent common ancestor to this group. Just how similar are reptiles and birds in terms of the beta-hemoglobin chain? Should birds be considered a type of reptile? You will evaluate this question in this exercise using a BLAST (Best Local Alignment Search Tool) search.

Procedure Part III: (Answer questions in the spaces provided on the Data Table Sheets)
Step 1: As in the previous two exercises, start by going directly to the SWISS-PROT database:
http://au.expasy.org/cgi-bin/sprot-search-ful

Step 2: Find a beta-hemoglobin chain for any type of crocodile. You can do this just as in the previous two exercises or you can type in **hemoglobin beta** for the search keywords or you can type in **crocodile** instead**.**

Step 3: A BLAST (Best Local Alignment Search Tool) search takes a particular sequence and then locates the most similar sequences in the entire database. A BLAST search will result in a list of sequences with the first sequence being most close to the one entered and the last sequences being least similar. The easiest way is for us to do a BLAST search is using links within SWISS-PROT as follows. Once you have found and clicked on a crocodile entry for the beta-hemoglobin chain, you will enter the screen with the amino-acid sequence as you did before (**Figure E**). This time, do *not* mouse click on FASTA format as in the previous exercises. *Instead*, look below the sequence at the very bottom of the web page (you will need to scroll down the page), and mouse click on either "Direct BLAST submission at EMBnet-CH and CSCS (Switzerland)" or "Direct BLAST submission at NCBI (Bethesda, USA)." On the screen that appears next, you will see the crocodile sequence pasted into the BLAST search screen. All you need to do is mouse click on "Run Blast" at the bottom of the page. The default options are appropriate for our search. It may take a few minutes for the results to appear.

Step 4: The next screen will have a list of sequences in order of similarity. Mouse click on the first ten sequences to determine to what species they belong. List those species in **Table 6** beginning with the first most similar species that is not a crocodile, then answer the 5 questions (6.1-6.5) asked following the data table.

PART IV: "What in the World...?" (optional part, if time allows or assigned by teacher)
What in the world is a spectacled caiman? A red gurnard? A northern gundi? A Gray francolin? A slender loris? A European suslik? Any other animal in Prot-Swiss that you do not know!?

Procedure Part IV: (Answer questions in spaces provided on data table sheets)

Obtain the beta-hemoglobin sequence for one of the animals listed above (or any other animal you do not know). Use a BLAST search to determine which is most closely related to one of the animals above or any other animal in the database that you do not know, then answer the questions 1-4 for Part IV.