**Phylogenetic Analysis Using Molecular Data**

In the following activities, you will use molecular databases of genetic information compiled by scientists from all over the world.

When a scientist uses DNA sequencing to resolve the base sequence of a strand of DNA, they can submit the resulting information to an international database called GenBank, so that all of the other scientists in the world can use the data for their own work. If a scientist works with protein sequences (a summary of DNA that has been translated), they can also submit their data to Swiss-Prot, an international information bank that offers universal access.

You will practice using both of these databases over the next few days, comparing strands of DNA or polypeptides to analyze the phylogenetic relationships between taxa.

**Activity A: Comparison of Taxa Using Protein Sequences**

**Purpose:**

- To learn how to use a protein sequence database to identify similar species.
- To understand the difference between monophyletic and polyphyletic taxonomic groups.

**Introduction:**

Protein and DNA sequences have many uses in biology. One of which is to help scientists determine which organisms are most closely related to each other. In this activity we will analyze protein sequences to examine the relationships between several species that are commonly called bears.

**Hypothesis:**

Name 6-8 species of bears that you hypothesize to be most closely related to the giant panda bear:

________________________________________________________________________

________________________________________________________________________

Using the species you named in the question above, draw a tree that shows your hypothesis of how these taxa might be related:
Procedure:

2. Click on the link that says “List of UniprotKB/Swiss-prot (reviewed) entries”
3. In the “query” box at the top of the page type in giant panda, you will receive over 1400 gene results that have been sequenced and identified. Scroll through the first few pages looking at the protein names. Find one that contains terms that you recognize, list it and click on the entry number to figure out what its function is:

   ____________________________________________________________
   ____________________________________________________________

4. In the “query” box at the top of the page type in giant panda, hemoglobin
5. Scroll down until you come across the “Hemoglobin sub unit a” entry.
6. Record the accession # (to the left of the entry name) here ________________.
   The accession # is an organizational system for the protein database. This oxygen carrier protein is homologous among all mammals and so is an excellent choice for comparing organisms that have an unknown evolutionary relationship. Click on the blue accession numbers for alpha hemoglobin to see a full description of the protein and the particular amino acid sequence for the giant panda. Take a few minutes to explore what types of information are available in this database.
   a. What is the genus and species of the giant panda _______________________
   b. What is the length of the amino acid sequence for this particular hemoglobin alpha chain? _______________________
   c. What is a ligand? What position in the amino acid sequence are the two iron ligands in this alpha hemoglobin chain? _______________________

7. Scroll up to the top of the protein information page and click on the tab that says “BLAST”.
8. This will paste the amino acid sequence into a new box. Click on the “Blast” button to the right of the sequence box. This will run a search to determine how similar the protein chain is to other organisms.
9. Write the common names and the similarity scores (you have to click on the green bar) of the ten species in data table that have the highest similarity scores (under the columns titled “Identity”=% and “Score”=number of shared amino acids in sequence). Record both the identity and the score in your data table next to the animals name.
10. Hit the back button twice and repeat steps 6-9 for the other proteins in the data table.
### Data Table A

<table>
<thead>
<tr>
<th>Alpha Hemoglobin</th>
<th>Beta Hemoglobin</th>
<th>Delta Hemoglobin</th>
<th>BDNF</th>
<th>Cytochrome b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Name/Identity %/Positives %</td>
<td>Common Name/Identity %/Positives %</td>
<td>Common Name/Identity %/Positives %</td>
<td>Common Name/Identity %/Positives %</td>
<td>Common Name/Identity %/Positives %</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Follow Up Questions for Activity A:

1. How does the list of most similar species for alpha hemoglobin compare to your hypothesis of the giant panda’s closest relatives?

2. How does the list of most similar species for beta hemoglobin, delta hemoglobin and cytochrome b compare to your hypothesis of the giant panda’s closest relatives?

3. How does the list of most similar species for BDNF compare to your hypothesis of the giant panda’s closest relatives?
4. Using your textbook, write the definition for each of the following terms:

**Monophyletic:**

**Paraphyletic:**

**Polyphyletic:**

5. Are the species that are commonly called bears (panda, grizzly, Asiatic, Malayan, polar and black bears) a monophyletic group? Explain why or why not, using appropriate scientific terminology.

**Activity B: Comparison of Taxa Using Beta Hemoglobin Protein Sequences**

**Purpose:**

- to learn how to use a protein sequence database to make distance matrices.
- to understand the differences between monophyletic and paraphyletic taxonomic groups.

**Introduction:**

In this first activity, you will compare birds, reptiles and mammals – all of which produce for amniotic (extraembryotic) membranes: the amnion, the yolk sac, the allantois, and the chorion. The presence of amniotic membranes holds these taxa together as a group, but what separates them into three classes that you learned as a child? You will use genetic information to see how the evolution of protein sequences points to a different relationship between these taxa than the one you may have learned when you were growing up.

**Hypothesis:**

Draw a tree diagram that shows your hypothesis of how these taxa are related (use the numbers beside each taxonomic group instead of writing in the taxon name):

- Aligators/Crocodiles-1
- Lizards-2
- Turtles-3
- Birds-4
- Mammals-5
Procedure:

1. Go to the following website http://www.expasy.org/sprot/
2. Click on the link that says “List of UniprotKB/Swiss-prot (reviewed) entries”
3. In the “query” box at the top of the page type in beta hemoglobin, black vulture and hit enter.
4. The list should generate 3 or 4 proteins. Click on the blue protein code to pull up the information on the hemoglobin subunit beta. There should be 146 amino acids in the beta chain for each entry you select.
5. Scroll down to the bottom of the protein information page to find the sequence under the heading “Sequence information” (the list of capital letters is a code for the sequences of amino acids that comprise this protein chain. Look just above the sequence and click on “FASTA format” to open the sequence in a form you can copy.
6. Select and copy only the protein sequence and paste it below name of the organism (just as I have done below for the rest of the sequences).
7. Repeat steps 3-6 for the remaining organisms in the list below.

Black Vulture Protein Sequence Chain
Painted Turtle Protein Sequence Chain
VHWTADEKQLITSLGKVNVEEGSEALARLLIVIVPWTQFSTFGNLSSNAEAILHNPHV
HAHGGKVLTSPEAVKLDHIIQTFATLSKLHCERKLHVDPENFKLLGNVLLIVLASHFTK
EFTPAQAAWQLVSAVAHALALGYH

Common Iguana Protein Sequence Chain
American Alligator Protein Sequence Chain
VHWTAEEKQLITCLGWGDVTGADLGMLVMYWTPRTFFADFGNLSSATACGNFKV
RAHGGKVLTAIPAIGNLDNIIKDTFAKLSELHCDLKLHVDPENFKLGNVLLIVLASHGYK
DFTPAHAAYQQLVNVVAAHALARH

Pallid Bat Protein Sequence Chain
VHLTADEKSAVTGGLGKVNVEEGGEALRLVYLIVPWTQFSTFGDLSSAGAVMGANKV
KAHGGKVLNASDGKLNLKGTFAKLSHCLDLHLVDPEFLLLGNVLMIVLARHGFK

8. When you have pasted a beta hemoglobin sequence into your word document for each taxon, open another internet window and go to the following website, http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi to perform a protein chain similarity comparison of these sequences.
9. Click on the button that says “Compare Sequences”, copy and paste only the amino acid sequence (THE CAPITAL LETTERS) of the first two protein sequences that you want to compare into the alignment window at this website. The first sequence can be entered into the top box and the second sequence can be pasted into the bottom box.
10. Press the button that says “Compare Sequences” and read the results of the phentic comparison it generates.
11. There will be similarity percentages about halfway down the screen. Record this number in the data chart below for the two organisms you compared.
12. After you have recorded the similarity percentage, press the back button to return to the “LALIGN – Local Alignments” page. Repeat steps 10-12 until all the organisms have been compared to one another and their similarity percentages have been entered into the data chart below.

Data Chart for Activity B:

<table>
<thead>
<tr>
<th></th>
<th>Bird (Black Vulture)</th>
<th>Turtle (Painted Turtle)</th>
<th>Iguana (Common Iguana)</th>
<th>Alligator (American Alligator)</th>
<th>Mammal (Pallid Bat)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bird (Black Vulture)</td>
<td>100 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turtle (Painted Turtle)</td>
<td>100 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iguana (Common Iguana)</td>
<td></td>
<td>100 %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alligator (American Alligator)</td>
<td></td>
<td></td>
<td>100 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mammal (Pallid Bat)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 %</td>
</tr>
</tbody>
</table>

Follow Up Questions

1. Which organisms are most closely related according to comparisons of beta hemoglobin?

2. Think about the definitions of monophyletic, paraphyletic, and polyphyletic. Are birds a monophyletic group? Explain why or why not, using appropriate scientific terminology.

Activity C: Comparision of Taxa Using Cytochrome b DNA Sequences.

Purpose:

- To learn how to use a DNA sequence database.
- To learn how to use a phylogenetic comparison website
- To create a phylogenetic tree using molecular analysis tools.

Introduction:
Now that you have learned how to use a protein sequence database, you will practice using a DNA database to generate hypothesis trees for the same organism that were in your morphological and ecological characteristics data set.

In order to begin, you must first decide which part of the genome would be a good comparison tool for these taxa. This activity is a much more specific comparison, using a selected region of DNA and a select group of taxa. You must choose a region of DNA common to all taxa for which there is adequate, known information. The region must be one that is fairly stable so that the sequences can be lined up with one another along homologous sections, but not so highly conserved that there have not been any mutations (remember, if there are only shared ancestral characteristics then we cannot tease out difference between taxa). **The cytochrome b gene, which codes for a carrier molecule used in the electron transport chain of the mitochondria, is a region of DNA that meets these requirements.** All eukaryotes have cytochrome b genes from a homologous evolutionary origin, so that the molecular data sequences of the mammals that were in your morphological and ecological data set can be compared using phylogenetic analysis.

**Hypothesis:**

Look at the list of organisms below in the data table and fill in the tree in terms of which organisms you think go where (the closer they are on the tree the more closely related the organisms are).
<table>
<thead>
<tr>
<th>Organism</th>
<th>GenBank Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bear</td>
<td>X82307</td>
</tr>
<tr>
<td>Seal</td>
<td>X82309</td>
</tr>
<tr>
<td>Cat</td>
<td>X82296</td>
</tr>
<tr>
<td>Opossum</td>
<td>U34665</td>
</tr>
<tr>
<td>Deer</td>
<td>X56291</td>
</tr>
<tr>
<td>Muskrat</td>
<td>AF119277</td>
</tr>
<tr>
<td>Beaver</td>
<td>AJ389529</td>
</tr>
<tr>
<td>Raccoon</td>
<td>X94930</td>
</tr>
<tr>
<td>Horse</td>
<td>D82932</td>
</tr>
<tr>
<td>Squirrel</td>
<td>AB126252</td>
</tr>
<tr>
<td>Dog</td>
<td>X94920</td>
</tr>
<tr>
<td>Pig</td>
<td>Z50079</td>
</tr>
<tr>
<td>Cow</td>
<td>AY952966</td>
</tr>
<tr>
<td>Rabbit</td>
<td>AJ279427</td>
</tr>
<tr>
<td>Mole</td>
<td>AB185152</td>
</tr>
<tr>
<td>Skunk</td>
<td>X94927</td>
</tr>
<tr>
<td>Otter</td>
<td>AF057124</td>
</tr>
<tr>
<td>Panda</td>
<td>X94919</td>
</tr>
</tbody>
</table>


2. To find cytochrome b sequences for the organisms that you studied in your morphological comparison, perform step a or b below:
   a. To find a cytochrome sequence, simply choose “nucleotide” from the pull-down menu that says “search” and type the accession code (from the chart above) into the box that says “for”. Press the “Go” button and the DNA sequence should appear. The actual DNA sequence will be at the bottom of the page under the title ORIGIN.

3. I will do the rest for you…

Follow Up Question for Activity C:

1. Record the tree that resulted from the comparison of the organism’s molecular data in the space below.

2. What is the difference between phentic comparison and phylogenetic comparison?