NAME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Period:\_\_\_\_\_\_\_\_ Date:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Investigation 3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST**

*How can bioinformatics be used as a tool to determine evolutionary*

*relationships and to better understand genetic diseases?*

**Learning Objectives:**

* Create cladograms that depict evolutionary relationships.
* Analyze biological data with a sophisticated bioinformatics online tool.

**Background:**

Between 1990–2003, scientists working on an international research project known as the Human Genome Project were able to identify and map the 20,000–25,000 genes that define a human being. The project also successfully mapped the genomes of other species, including the fruit fly, mouse, and *Escherichia coli*. The location and complete sequence of the genes in each of these species are available for anyone in the world to access via the Internet.

Why is this information important? Being able to identify the precise location and sequence of human genes will allow us to better understand genetic diseases. In addition, learning about the sequence of genes in other species helps us understand evolutionary relationships among organisms. Many of our genes are identical or similar to those found in other species.

Suppose you identify a single gene that is responsible for a particular disease in fruit flies. Is that same gene found in humans? Does it cause a similar disease? It would take you nearly 10 years to read through the entire human genome to try to locate the same sequence of bases as that in fruit flies. This definitely isn’t practical, so a sophisticated technological method is needed.

Bioinformatics is a field that combines statistics, mathematical modeling, and computer science to analyze biological data. Using bioinformatics methods, entire genomes can be quickly compared in order to detect genetic similarities and differences. An extremely powerful bioinformatics tool is BLAST, which stands for Basic Local Alignment Search Tool. Using BLAST, you can input a gene sequence of interest and search entire genomic libraries for identical or similar sequences in a matter of seconds.

In this laboratory investigation, you will use BLAST to compare several genes, and then use the information to construct a *cladogram*. A cladogram (also called a phylogenetic tree) is a visualization of the evolutionary relatedness of species. Figure 1 is a simple cladogram.

****

Note that the cladogram is treelike, with the endpoints of each branch representing a specific species. The closer two species are located to each other, the more recently they share a common ancestor. For example, *Selaginella* (spikemoss) and *Isoetes* (quillwort) share a more recent common ancestor than the common ancestor that is shared by all three organisms.

Figure 2 includes additional details, such as the evolution of particular physical structures called shared derived characters. Note that the placement of the derived characters corresponds to when (in a general, not a specific, sense) that character evolved; every species above the character label possesses that structure. For example, tigers and gorillas have hair, but lampreys, sharks, salamanders, and lizards do not have hair.



The cladogram above can be used to answer several questions. Which organisms have lungs? What three structures do all lizards possess? According to the cladogram, which structure — dry skin or hair — evolved first?

Historically, only physical structures were used to create cladograms; however, modern-day cladistics relies heavily on genetic evidence as well. Chimpanzees and humans share 95%+ of their DNA, which would place them closely together on a cladogram. Humans and fruit flies share approximately 60% of their DNA, which would place them farther apart on a cladogram.

Can you draw a cladogram that depicts the evolutionary relationship among humans, chimpanzees, fruit flies, and mosses?

**PRE-LAB CHECKPOINT –** *Answer the following questions using complete sentences.*

1. Use the cladogram in Figure 2 to answer the following:
	1. Which organisms have lungs?
	2. What three structures do all lizards possess?
	3. According to the cladogram, which structure – dry skin or hair – evolved first? How do you know?
2. Use the following data to construct a cladogram of the major plant groups:



1. GAPDH (glyceraldehyde 3-phosphate dehydrogenase) is an enzyme that catalyzes the sixth step in glycolysis, an important reaction that produces molecules used in cellular respiration. The following data table shows the percentage similarity of this gene and the protein it expresses in humans versus other species. For example, according to the table, GAPDH gene in chimpanzees is 99.6% identical to the gene found in humans, while the protein is identical.



* 1. Why is the percentage similarity in the gene always lower than the percentage similarity in the protein for each of the species? (Hint: Recall how a gene is expressed to produce a protein.)
	2. Draw a cladogram depicting the evolutionary relationships among all five species (including humans) according to their percentage similarity in the GAPDH gene.

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

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 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:

1. Go to the <http://www.uniprot.org> website.
2. In the “Query” box, type **giant panda** **alpha hemoglobin** and hit “search”.
3. Scroll down the giant panda protein sequences that have been entered in this database until you see **HBA AILME (P18970)** alpha hemoglobin. 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 code** letters 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.
4. Scroll down to the bottom of the protein information page to the section titled **Sequences**. Find the dropdown menu link that says, “**BLAST**” and click **Go**.
5. This link will paste the alpha hemoglobin sequence of the giant panda onto a search screen and find out how similar the protein chain is to other organisms.
6. Write the common names of the ten species that have the highest **identity** scores and the **% identity** match. Ignore any comparisons to other giant pandas. You may need to sort the Identity column to put them in order.
7. Answer the “Reflection Questions” below for this protein and then return to the search screen by clicking the “search” tap at the top of the page.
8. Compare the protein sequences for beta hemoglobin, delta hemoglobin, BDNF and cytochrome b (these are all the protein sequences that have registered sequences for other types of bears in this database) by repeating steps 2-6.
9. Choose an organism whose evolutionary relationships are unknown to you (ex: platypus, rabbit, armadillo, camel, etc.). Make a hypothesis about the organism’s closest relatives. Conduct a BLAST search of the protein sequences similar to this organism’s to test your hypothesis.

Data Chart for Activity A: Giant Panda

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Alpha hemoglobin | Beta hemoglobin | Delta hemoglobin | BDNF | Cytochrome b |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Your organism: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Hypothesis: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Data Chart for Activity A: *your animal*: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Alpha hemoglobin | Beta hemoglobin | Delta hemoglobin | BDNF | Cytochrome b |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Reflection 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. Look up the order and family of the organisms that have proteins similar to the giant panda. What order do most of the organisms in your chart belong to? Order: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
5. What family do most of the organisms in your data chart belong to? Family: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
6. What surprised you about the evolutionary relationships you discovered about the organism you choose to research?

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

Introduction:

 In this first activity, you will compare birds, reptiles, and mammals – all of which produce four 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 the 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 mat have learned when you were growing up.

**Hypothesis**:

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

Alligators/Crocodiles (1)

Lizards (2)

Turtles (3)

Birds (4)

Mammals (5)

Procedures:

1. Open a blank Word document. Type the following words down the left side of the page: **bird, turtle, lizard, alligator, and mammal**. Leave this document open.
2. Retun to the [www.uniprot.org](http://www.uniprot.org) main page.
3. In the “Query” box, type **hemoglobin beta chain** and hit “search”.
4. Scroll down through the list of organisms and choose one bird, turtle, lizard, alligator, or mammal under these protein descriptions and perform steps 5-8. Make sure the protein you choose is a beta chain!
5. Write down the common name of the sample organism chosen below each category name in the first row of the Data Chart.
6. Click on the blue protein code to pull up the information on the protein sequence.
7. Scroll down to the bottom of the protein information page to find the sequence under the heading “**sequences**” (the list of capital letters is a code for the sequences of amino acids that comprise this protein chain; there should be 146 amino acids in the beta chain for each entry you select). Click on “**FASTA**” to open the sequence into a form you can copy.
8. Select and copy **only** the protein sequence and paste it into your Word document under the appropriate label.
9. When you have pasted a beta hemoglobin sequence into your Word document for each taxon, open another internet window and go to the <http://fasta.bioch.virginia.edu/fasta/lalign.htm> website to perform a protein chain similarity comparison of these sequences.
10. Copy and paste only the amino acid sequence of the first two protein sequences that you want to compare into the alignment window at the website. The first sequence can be entered into the top box and the second sequence can be pasted into the bottom box.
11. Press the “align sequences” button and read the results. There will be a similarity percentage halfway down the screen. Create a data chart in your word document and record this number in the data chart for the two organisms you compared.
12. After you have recorded the similarity percentage, return to the LALIGN-Local Alignments” page. Repeat steps 10-12 until all of the organisms have been compared to one another. Record your results in the data table.
13. Print this data chart and turn it in with your lab.

Reflection Questions for Activity B:

1. Which organisms 2 organisms are most closely related according to comparisons of beta hemoglobin? Which two are least closely related?
2. How did these results relate to your hypothesis? Using your knowledge of evolution, explain the relationships you identified in question 1.
3. How is using biochemical evidence different from other methods of identifying evolutionary relationships? Include data from both Activity A and B to support your answer.