**Investigation 3**

**Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST (Basic Local Alignment Search Tool)**

**Student AP Lab Manual Page S41**

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.

Work through AP Investigation 3 and answer the following questions as you come to them in the lab:

1) Referring to Figure 2 (pS42). Cladogram of Several Animal Species

The cladogram above can be used to answer several questions.

A) Which organisms have lungs?

B) What three structures do all lizards possess?

C) According to the cladogram, which structure — dry skin or hair — evolved first?

2) Draw a cladogram that depicts the evolutionary relationship among humans,

chimpanzees, fruit flies, and mosses?

3) Use the following data to construct a cladogram of the major plant groups:

**Table 1. Characteristics of Major Plant groups**

|  |  |  |  |
| --- | --- | --- | --- |
| **organisms** | **Vascular Tissue** | **Flowers** | **Seeds** |
| Mosses | 0 | 0 | 0 |
| Pine trees | 1 | 0 | 1 |
| Flowering plants | 1 | 1 | 1 |
| Ferns | 1 | 0 | 0 |
| Total | 3 | 1 | 2 |

4) 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, the GAPDH gene in chimpanzees is 99.6% identical to the

gene found in humans, while the protein is identical..

**Table 2. Percentage Similarity Between the GAPDH gene and Protein in humans and other Species**

|  |  |  |
| --- | --- | --- |
| **Species** | **gene Percentage****Similarity** | **Protein Percentage****Similarity** |
| Chimpanzee *(Pan troglodytes)* | 99.6% | 100% |
| Dog *(Canis lupus familiaris)* | 91.3% | 95.2% |
| Fruit fly *(Drosophila melanogaster)* | 72.4% | 76.7% |
| Roundworm *(Caenorhabditis elegans)* | 68.2% | 74.3% |

A) 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.)

B) Draw a cladogram depicting the evolutionary relationships among all five species

(including humans) according to their percentage similarity in the GAPDH gene.

5)

A) Go to and complete ONE of the three online activities:

Online Activities

• “The Evolution of Flight in Birds”

<http://www.ucmp.berkeley.edu/education/explorations/reslab/flight/main.htm>This activity provides a real-world example of how cladograms are used to

understand evolutionary relationships.

• “What did T. rex taste like?”

<http://www.ucmp.berkeley.edu/education/explorations/tours/Trex/index.html>• “Journey into Phylogenetic Systematics”

<http://www.ucmp.berkeley.edu/clad/clad4.html>

B) State WHICH activity you did and summarize in a paragraph what you learned from the activity here:

6) A team of scientists has uncovered the fossil specimen in Figure 3 Fossil Specimen near Liaoning Province, China (page S44). Make some specific observations about the morphology (physical structure) of the fossil, and then record your specific observations here.



7) Scientists were able to extract DNA nucleotides from the tissue of the fossil in Figure 3 and use the information to sequence several genes. Your task is to use BLAST to analyze these genes and determine the most likely placement of the fossil species on Figure 4.



**■■Analyzing Results**

Recall that species with common ancestry will share similar genes. The more similar

genes two species have in common, the more recent their common ancestor and the

closer the two species will be located on a cladogram.

As you collect information from BLAST for **EACH** of the gene files, you should be

thinking about your original hypothesis and whether the data support or cause you to

reject your original placement of the fossil species on the cladogram.

For **EACH** BLAST query, consider the following:

• The higher the score, the closer the alignment.

• The lower the e value, the closer the alignment.

• Sequences with e values less than 1e-04 (1 x 10-4) can be considered related with an

error rate of less than 0.01%.

**COMPLETE 7 A-D for EACH of the three genes you choose AND take a “Screen Shot” of the BLAST result table with Max score values showing, print it, and attach it to the lab to verify your work below.**

7A) What species in the BLAST result has the most similar gene sequence to the gene of

interest? Name it and give the score.

B) Where is that species located on your cladogram? **Circle it and place (B) beside it.**

C) How similar is that gene sequence? Give the score and % (Max Ident.)

**(D on next page)**

D) What species has the next most similar gene sequence to the gene of interest?

**Name it and give the score and % (Max Ident.)**

**COMPLETE 7 A-D for EACH of the three genes, list screen shots and answers for each gene separately.**

**E)** Based on what you have learned from the sequence analysis and what you know from

the structure, decide where the new fossil species belongs on the cladogram with

the other organisms. If necessary, redraw the cladogram you created before.

Evaluating Results

8) Compare and discuss your cladogram with your classmates. Does everyone agree with

the placement of the fossil specimen? If not, what is the basis of the disagreement? *Explain the final basis for the species placement in the cladogram (Figure 4) and be sure to mark this placement clearly on the cladogram in Figure 4.*

**Describe the outcome of your discussion here: (go onto another page as needed).**

9) On the main page of BLAST, click on the link “List All Genomic Databases.”

A) How many genomes are currently available for making comparisons using BLAST?

B) How does this limitation impact the proper analysis of the gene data used in this lab?

C) What other data could be collected from the fossil specimen to help properly identify

its evolutionary history?

Designing and Conducting Your Investigation

Now that you’ve completed this investigation, you should feel more comfortable using

BLAST. The next step is to learn how to find and BLAST your own genes of interest. To

locate a gene, you will go to the Entrez Gene website (http://www.ncbi.nlm.nih.gov/

gene). Once you have found the gene on the website, you can copy the gene sequence

and input it into a BLAST query.

Follow the directions for the “human actin” (ACTA1 actin) gene search.

10) Name what genome(s) you compared the “human actin” gene sequence to, whether you chose “highly similar” or “somewhat similar” gene sequences to compare, and what you found out after completing the BLAST:

11) Below is a list of some gene suggestions you could investigate using BLAST.

|  |  |
| --- | --- |
| **suggested genes to explore** | **families or genes studied Previously** |
| ATP synthase | Enzymes |
| Catalase | Parts of ribosomes |
| GAPDH | Protein channels |
| Keratin |  |
| Myosin |  |
| Pax1 |  |
| Ubiquitin |  |

Pick ONE or name another HUMAN gene of interest and use that.

**GENE NAME:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

For your particular gene, answer the following questions:

A) What is the function in humans of the protein produced from that gene?

B) Would you expect to find the same protein in other organisms? If so, which ones or which types?

Follow the BLAST procedure done for “human actin” on S49 and S50

Now scroll down the screen to below the graphical output, to examine the table

with the species names and BLAST scores of these 100 hits.

The species with the gene most closely-related to your chosen gene are at the top of the

table and have the highest scores.

**\*\*\*Take a “Screen Shot” of the BLAST result table with Max score values showing, print it, and attach it to the lab to verify your work below.**

C) Name the top five species most closely related to humans for your gene of choice. Give the Genus and species and look up the common name using a Google search.

1)

2)

3)

4)

 5)

What do you think this result means in terms of evolution?

If you were studying drugs that impacted the activity of this protein in humans, which species might be the best model organism on which to run pre-human drug trials? Explain.

D) Is it possible to find the same gene in two different kinds of organisms but not find the protein that is produced from that gene? Explain.

E) If you found the same gene in all organisms you could possibly test, what does this suggest about the evolution of this gene in the history of life on earth?

F) Does the use of DNA sequences in the study of evolutionary relationships mean that other characteristics are unimportant in such studies? Explain your answer.