Lab Investigation 3: BLAST

Name(s): __________________________ Date: ____________ Period: ________

Essential Question:
How can bioinformatics be used as a tool to determine evolutionary relationships and to better understand genetic diseases?

Learning Objectives:
• To create cladograms that depict evolutionary relationships
• To analyze biological data with a sophisticated bioinformatics online tool
• To use cladograms and bioinformatics tools to ask other questions of your own and to test your ability to apply concepts you know relating to genetics and evolution

BACKGROUND INFORMATION:
Bioinformatics is an interdisciplinary field that combines computer science, statistics, mathematics, and engineering to analyze and interpret biological data, specifically in the field of genomics. Common uses include identifying genes and DNA nucleotide sequences to determine evolutionary relationships among organisms. BLAST stands for Basic Local Alignment Search Tool and it is designed to search nucleotide and protein databases. The sequences must be in a special format to optimize your search. Anyone in the world with a computer can use BLAST.

You will be using BLASTn, which is only used to search for nucleotides. It searches for matches in the sequence between your gene of interest (the gene you upload) and millions of nucleotide sequences in its database. It takes your DNA sequence and breaks it down into “smaller words” made up of 11 nucleotides and then compares them to those in the database. Then it looks for more similar “words” or sequences adjacent to the first one until it “reads” the entire DNA sequence you uploaded and has found the most similar matches, or alignments.

Why is this information important to know? Being able to identify the precise location and sequence of human genes allows 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 your genes are identical or similar, to those found in other species.

You will use the information to construct a cladogram. A cladogram (also called a phylogenetic tree) is a visualization of the evolutionary relatedness of species.

Historically, only physical characteristics were used to create cladograms. Today modern cladistics also relies heavily on genetic similarities and/or differences. For example, the DNA of chimpanzees and humans is 96-98% the same, depending on how it’s calculated. This would place chimpanzees and humans very close together on a cladogram. Humans and fruit flies share about 60% of the same genes. This would place them farther apart on a cladogram.
**Lab Investigation 3: BLAST**

**PROCEDURE:**

Read the paragraphs at the bottom of page S44. This is an enlarged photo of **Figure 3**. This fossil specimen was found in the Liaoning Province in China. It is a newly discovered species.

*(Remember: DNA nucleotide sequences were extracted from soft tissue in this fossil.)*

Your task is to use BLASTn to analyze three genes and determine the place where the fossil species is most likely to be placed on the cladogram on page S45.

**STEP 1:** Using **Figure 4. Fossil Cladogram**, form an initial hypothesis as to where you think the fossil specimen should be placed on the cladogram based on the morphology (form and structure of an organism) you made earlier.

Write your hypothesis to the right of the cladogram below.

Remember to draw where you think the fossil specimen should be on the cladogram in **Figure 4**.

**STEP 2:** Use the gene sequence handout provided by your teacher and choose one of the fossil genes. Highlight and copy your gene of interest.
Lab Investigation 3: BLAST

STEP 3: Upload the gene sequence by doing the following:


b. Paste your gene of interest into the Enter Query Sequence box.
Lab Investigation 3: BLAST

*Be sure to add a job title so you know which gene nucleotide sequence you uploaded.

c. Scroll down to the bottom of the page. Click on the BLAST button.

STEP 4: Be patient. It may take up to 60 seconds to complete the BLASTn search. You’ll see the results in 2 forms – as a diagram, which you see below, and a table.

a. Click on Graphic Summary to see the Table of your BLASTn results.

b. View the **Table of your BLASTn results**: The list of species that appears in this section are those with sequences identical to or most similar, to the gene you uploaded. The sequences become LESS similar are you go down the list.
c. Click on one of the species listed to find out more specific information, including the classification scheme (scientific name), the protein that the sequence codes for, and the sequence of bases that appear to align with your gene of interest. Copy the scientific name of that species and paste it into a google search to find out the common name. See the example below.

d. You must click on the box of the species you want to use to create the “Distance Tree of Results.”
e. Scroll back to the top of the page and click on **Distance tree of results**. Watch what happens!

THIS is a cladogram of your results. This might not mean much to you right now, but it’s a starting point. This cladogram shows the species with similar sequences to the gene you uploaded (gene of interest), according to how closely their matched gene aligns with your gene. (See below.)

Take a closer look at the legend on the right hand-side of the page. The arrow points to your gene.
Lab Investigation 3: BLAST

f. Modify the view of the cladogram by changing the **Max Seq Difference** from 0.75 to 0.15. This will give you a more comprehensive overview of the evolutionary relationships. Use the legend to determine evolutionary relationships. The gene you uploaded is highlighted in **yellow**.

Analyzing Results (page S48):

As you collect information from BLASTn searches for each of the gene files, think about your original hypothesis and whether the data support your original placement of the fossil species on the cladogram on page S45. The more similar genes 2 species have in common, the more recent their common ancestor and the closer the 2 species will be located on a cladogram.

For **each** BLASTn query, consider the three criteria listed on page S48 for **each** of the gene files before answering the following questions.

1. What species in the BLASTn results has the most similar gene sequences to the gene of interest?
   - First species:
   - Second species:
   - Third species:

2. Where is that species located on your cladogram?
   - First species:
   - Second species:
   - Third species:
Lab Investigation 3: BLAST

3. How similar is the gene sequence to your gene of interest?
   First gene:

   Second gene:

   Third gene:

4. Name the species that has the next most similar gene sequence to the gene of interest. Include both the scientific name and the common name.

EVALUATING RESULTS (page S49):
Compare and discuss your cladogram with your classmates. Do everyone’s results agree with the placement of the fossil specimen on your cladogram? If there is disagreement, what is the basis for it?

Identify other evolutionary data that could be collected from the fossil specimen to help properly identify its evolutionary history.

DESIGN AND CONDUCT YOUR OWN INVESTIGATION: Example Procedure
You will use BLAST to find your own genes of interest to investigate. Go to page S50 to read the list of questions and the table with suggested genes for human proteins that you could investigate. Pick one in the table or find a different one. For example:


b. If you want to find one of the genes listed in the table on page S50, you can add the name to your search. The example below shows the results of a search using only “human” as the criteria. You will see a long list of human proteins. Choose a protein.
c. This example shows a search for the gene sequence of HLA-B, an MHC protein. Click on HLA-B. A new page will open.

d. Scroll down the page until you see this. Click on FASTA. (Pronounced as Fast-AY)

e. This is just a snapshot of your results. You need to scroll down to view the entire nucleotide sequence for HLA-B.
f. Copy the entire gene sequence and then go back to the BLASTn home page. Paste it into the “Enter Query Sequence” box, click BLAST and follow the same procedure you used to create a cladogram.

g. Answer the following questions to the best of your ability as you look at a specific gene.

1. What is the function in humans of the protein produced from the gene you chose?

2. Would you expect to find the same protein in other organisms? If so, which ones? Explain why.

3. 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 your answer.

4. If you found the same protein in all organisms you test, what does this suggest about the evolution of this gene in the history of life on earth? Justify your answer with evidence.