AP Bio BLAST Lab - Searching for Fossil Genes   Name _____

Introduction:

You are a manager of a new animal food supply company, You need to find out if Vitamin C needs to be included in new animal foods designed for dogs, cows, cats, mice, and guinea pigs. Based on your research on the GULO gene, you will be able to determine if you need to include Vitamin C in these foods.

Most mammals, such as mice, can produce their own vitamin C and therefore do not need a source of vitamin C in their diet. The GULO gene codes for an enzyme, L-gulonolactone oxidase, involved in vitamin C synthesis. The GULO gene is present in mice and most other mammals, but is either missing, or is nonfunctional, in some mammals. These animals cannot make their own vitamin C and must have this vitamin present in their diet. You will see if various mammals, including the primates, such as humans and chimps, have a functional GULO gene. A functional gene is able to produce a functional protein, in this case the GULO enzyme.

In some mammals, the GULO gene is an example of a pseudogene. Pseudogenes are vestigial genes. That is, they were once functional in an ancestral species, but since they were no longer needed they accumulated mutations until they became non-functional. In many cases they evolve to the point where a protein can no longer be produced at all. Pseudogenes represent molecular evidence for evolution. As fossils are the remains of extinct organisms, pseudogenes are the remains of extinct genes.

Procedure:

The U. S. Government maintains a set of databases called GenBank, which contains nucleotide and amino acid sequences for those genes and proteins whose sequences have been determined. For your research we will use a computer program called **BLAST**. BLAST is able to search the GenBank databases. If you input a nucleotide or amino acid sequence into BLAST, it will search for any known genes or proteins that are similar to the one that you entered. You will use BLAST to determine if the genomes of cows, pigs, humans and chimps contain functional GULO genes, or if they contain vestigial GULO pseudogenes, which do not result in production of the GULO enzyme.

Part I. Is the GULO gene present in various mammals?
A. Is the GULO gene present in various mammals?
1. Go to the class website and open the file, "**Mouse GULO Gene**".
2. Copy the nucleotide sequence for the mouse GULO gene.
3. Go to www.ncbi.nlm.nih.gov.
4. From the menu at the top right hand side of the page, select **BLAST**.
5. From the BLAST menu, select "**nucleotide BLAST**". It is under "**Basic BLAST**".
6. Copy the mouse GULO sequence into the box under "**Enter Query Sequence**".
7. Scroll down until you see "**Database**".
8. Check the "**Others**" box. Change the database setting in the box to "**nucleotide collection**".
9. In the "**Organism**" window, type in "**cow**". When you see cow appear in the box, select it.
10. Scroll to the bottom and find the "**BLAST**" button. Above the "**BLAST**" button, you will see the "**optimize for**" box. Select optimize for "**somewhat similar sequences (BLASTn)**"
11. Scroll back down and hit the "**BLAST**" button.
12. When **BLAST** is done with its search, you can scroll down and see a colorized diagram indicating the degree of similarity of the BLAST hits to your mouse GULO nucleotide sequence. Red and pink/purple mean a good match, while green, blue and black indicate a poor match. If the colored line spans the entire length of the window, then the "hit" sequence matches the inquiry sequence along its entire length. We want to see a high quality match along a majority of the inquiry sequence.
13. Below the colorized diagram is a "hit list" of your results. This shows the quality of matching as an E-value. An E-value is the chance that the matchup may be due to a random matching of a sequence of bases. The smaller the E-value, the more confidence you can have in your matching. A good match should have a low E- value (red or pink line) and an alignment along a large segment of the sequence.
14. Note your result in the chart below.
15. Now start again and do BLAST searches for pig (*Sus scrufa*), human (*Homo sapiens*) and chimpanzee (*Pan troglodytes*) GULO genes. Record your data in the chart.

Mammal GULO genes

Species	Present or Absent?	% of gene that matches	Comments
Mouse (*Mus musculus*)	Present	100%	Good match over entire length of gene.
Dog (*Canis familiaris*)			
Cow (*Bos Taurus*)			
Pig (*Sus scrufa*)			
Guinea Pig (Cavia porcellus)			
Human (*Homo sapiens*)			

PART II. Does the human GULO gene produce a functional protein?
We will now use protein **BLAST** to search for GULO proteins in cows, pigs, humans and chimpanzees.
1. Copy the mouse GULO protein from the **"Mouse GULO Protein"** text file.
2. Go to www.ncbi.nlm.nih.gov.
3. Select **BLAST** from the menu at the top of the page.
4. From the BLAST menu, select "**protein BLAST**". It is under "**Basic BLAST**".
5. Copy the mouse GULO sequence into the box under "**Enter Query Sequence**".
6. Scroll down until you see "**Organism**".
7. In the "**Organism"** window, type in "**cow**". When you see cow appear in the box, select it.
8. Scroll to the bottom and select the "**BLAST"** button.
9. When BLAST is done with its search, you can scroll down and see a chart of your results. Note your result in the chart below.
10. Now start again and do BLAST searches for pig (*Sus scrufa*), human (*Homo sapiens*) and the other GULO genes.
11. Look for proteins with the same name (, L-gulonolactone oxidase). If the GULO protein is not present, other, more distantly related proteins may come up. They will have a much lower score and a higher E-value. Note that the E-value represents the chance that the result is due a random matching of some amino acid sequences from both proteins. An E-value of 0 means a statistically perfect match. A good E-value should be much lower than e^{-4}.
12. Record your data in the chart.

Functional GULO Proteins

Species	Present or Absent?	E value		Species	Present or Absent?	E value
Mouse (*Mus musculus*)	Present	0.0		Cow (Bos Taurus)		
Dog (*Canis familiaris*)				Guinea Pig (*Cavia porcellus*)		
Pig (*Sus scrufa*)				Human (*Homo sapiens*)		

Critical Thinking

1. Why do you think that primates (monkeys, apes and humans) have lost the ability to produce vitamin C? (Hint: think about the diet of early primates).

2. Explain why the GULO gene in humans may be considered vestigial.

3. Explain how the presence of the GULO gene in humans provides evidence for evolution.

4. The new pet food company you work for is designing healthy foods for dogs, pigs, cows, mice and guinea pigs. Write a memo to the product development division to make recommendations regarding the addition of vitamin C in animal foods.

PRACTICE!
Designing and Conducting Your 0wn Blast Investigation

Now that you've completed a BLAST 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.

Example Procedure
Your starting question may be: What is the function of actin in humans? Do other organisms have actin? If so, which ones?
1. Go to the Entrez Gene website (*http://www.ncbi.nlm.nih.gov/gene*) and search for "human actin."
2. Click on the first link that appears and scroll down to the section "NCBI Reference Sequences."
3. Under "mRNA and Proteins," click on the first file name. It will be named "NM001100.3" or something similar. These standardized numbers make cataloging sequence files easier. Do not worry about the file number for now.
4. Just below the gene title click on "FASTA." This is the name for a particular format for displaying sequences.
5. The nucleotide sequence displayed is that of the actin gene in humans.
6. Copy the entire gene sequence, and then go to the BLAST homepage (*http://blast.ncbi.nlm.nih.gov/Blast.cgi*).
7. Click on "nucleotide blast" under the Basic BLAST menu.
8. Paste the sequence into the box where it says "Enter Query Sequence."
9. Give the query a title in the box provided if you plan on saving it for later.
10. Under "Choose Search Set," select whether you want to search the human genome only, mouse genome only, or all genomes available.
11. Under "Program Selection," choose whether or not you want highly similar sequences or somewhat similar sequences. Choosing somewhat similar sequences will provide you with more results.
12. Click BLAST.

Below is a list of some gene suggestions you could investigate using BLAST.
- **ATP synthase**
- **Catalase**
- **GAPDH**
- **Keratin**
- **Myosin**
- **Pax1**
- **Ubiquitin**

Lab Report
Your write-up for this lab will consist of a summary of your research and findings, and should demonstrate a sophisticated level of scientific literacy. It should include an introduction (5 points), a description of the methods you used (5 points), data (5 points), an analysis of your results (5 points), and a conclusion (10 points). Your report will also be evaluated based on organization and AP Biology-level writing (10 points).

You will want to use additional internet resources, being sure to cite the sources you use.

There is no required minimum length for your report, but it must be produced as a Word document, which you can then print out and submit as a hard copy, or it may be submitted electronically.

Consider including graphics as part of your report, either those that you come across in your research or create yourself.

You can use the following as guiding questions:
- What is the function in humans of the protein produced from that gene?
- Would you expect to find the same protein in other organisms? If so, which ones?
- 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?
- If you found the same gene in all organisms you test, what does this suggest about the evolution of this gene in the history of life on earth?
- Does the use of DNA sequences in the study of evolutionary relationships mean that other characteristics are unimportant in such studies?