BLASTing Through the Kingdom of Life- Extra Credit   (Biotech Series)

Background: You were first exposed to the idea of Bioinformatics in chapter one of your textbook (page 6) and you will learn much more about this area of biotechnology your junior year in chapter seven (pages 228-229). This optional extra-credit assignment will introduce you to the use of BLAST, which is an online collection of gene sequences and genomes. Before you start this assignment, you will need to read and summarize the information on pages 228-229 of your textbook. The instructions for this assignment are located online at the following website: http://www.digitalworldbiology.com/BLAST/62000sequences.html#BM2 . You will need to carefully follow the instructions on the above website. You will also need to view the BLAST tutorial that is referenced: http://digitalworldbiology.com/dwb/BLAST .   If you are interested in this extra-credit opportunity, you will need to email me at mpedersen@paulding.k12.ga.us to receive your unique DNA sequence. In the subject line of the email, you will need to type BLAST Extra Credit. Once you receive your DNA sequence, you can continue working through the activity and answering the questions below. When the website refers to "the worksheet", it is referring to the 14 questions below on this worksheet. Do not forget to also include a short summary of pages 228-229 in you textbook. You will be allowed to complete one BLAST assignment each semester for 100 formative points of extra credit.

Questions (1-14): Be sure to include the source of the information along with your answer. In this case, the source will be the database or web page that provided the information.

1. How long is the sequence that was used to search the database? *Hint: This sequence is called the "query" sequence because you used it to ask a question (or query) of the database.*

2. What is the most likely identity of this sequence? What data supports this conclusion? *Hint: Refer to the slide in the BLAST tutorial that discusses the E value.*

3. What organism was the most likely source of the sequence? *Hint: Refer to the BLAST tutorial to find an overview of the GenBank nucleotide record. If more than one organism matches, look at the E values to determine the most likely match.*

4. What is the common name for this organism? *Hint: Refer to the GenBank nucleotide record. It may also help to look at the Taxonomy database. The BLAST tutorial shows where to find this link.*

5. What phylum contains this organism? *Hint: Refer to the taxonomy database. The BLAST tutorial shows how to find the link.*

6. What is the accession number for the best-matching sequence?

7. Is this sequence expressed? How do you know? *Hint: Gene expression includes the processes of transcription (making RNA) and translation (making a protein). Determine if either of these molecules is described in the sequence record.*

8. If your sequence is expressed, where is this gene expressed? *Hint: The sequence record usually indicates the source of the material that was sequenced. In some cases, however, you will probably need to look at other sources of information. Some places to look are the title of the submission, PubMed records (you might need to look at more than one record), the UniGene database, and the Gene database. If your BLAST results show boxes with U's or G's (shown below) those are links to the UniGene and Gene databases respectively. The Expression Profile in the UniGene database is linked to a set of tables that show the tissues where mRNAs have been found and the developmental stage where they were found.*

9. Is there a specific time during development when this gene is expressed? *Hint: See the hint for question 8.*

10. Is anything known about factors that cause your sequence to be expressed? *Hint: The title of the submission is a good place to start. PubMed records and the Entrez Gene database are also helpful; see the hint for question 8.*

11. Estimate the number of sequences with an E value less than 0.01. *Hint: Refer to the blast results.*

12. If possible, give the names of three different organisms with significant E values. If organism is represented, then write down the name of that organism. *Hint: Refer to the*

BLAST tutorial slide on E values for a description.

13. Look at the first matching sequence; determine the length of the alignment and the fraction of nucleotides that match your sequence. Draw a picture to represent the alignment between the two sequences and include the starting and ending map positions for the both sequences.

14. Use GenBank, PubMed, Gene, and UniGene records to find the possible function of the protein that's specified by your DNA sequence. Describe what's known about the role of this protein in the organism that provided the DNA.