**Bioinformatics Worksheet #2: BLAST Practice – Instructions**

**Procedure**:

1. On the fragments page, there are multiple sequence fragments that you have to identify. BLAST them by going to the blastn, or nucleotide blast, at this link: <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome>. This will take you directly to the BLAST Search Page. Take one of your sequences and enter it into the large box near the top of the page. Then, at the bottom of the page, click the button that says BLAST.
2. The BLAST search may take anywhere from a few seconds to a few minutes, so be patient. Once BLAST is done searching, identify the probable function of the protein by looking at the most similar result (These start below the box filled with red lines). Write down the organism the closest match came from and record the total score, E- value, query cover, and % Identity as well as the length of your protein in both base pairs and amino acids in **Table 1**. This may not always be the first result. If you don’t recognize the scientific name, enter it into google and learn a little more about the organism before continuing.
3. Repeat steps 2 through 4 for each of your sequences until you have identified each of the organisms then answer the question.

**Bioinformatics: Blasting for Answers - Assignment**

**Names of individuals in group:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Table 1: Table of results for 8 largest proteins in operon**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sequence #** | **Max Score** | **Total Score** | **E-value** | **Query Cover** | **% Identity** | **Name of Organism** |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |

1. What organism is responsible for the child’s illness? Explain your evidence and reasoning.
2. Why do the other organisms appear in the metagenomic results? Could the sample have been contaminated? If so, identify which organisms were the contaminants and hypothesize about how they might have gotten there. (Hint: There are two definite and one possible/arguable contaminant.)
3. Do some research on the organism you think is responsible. Likely, there will be a Wikipedia page that will have a summary of information, but also try looking at primary literature in PubMed (http://www.ncbi.nlm.nih.gov/pubmed). Write a brief summary of what you learned and try to identify possible treatments (For example, which antibiotic would work).
4. Each sequence given came from the small subunit of the ribosome called 16s rRNA. This is quite often used to identify prokaryotic species and to construct phylogeny for prokaryotes. Hypothesize why 16S rRNA is used for this.
5. Sometimes when patients take antibiotics to treat a pathogenic bacteria infection, they will feel nausea. Hypothesize about what might cause this, given what you now know about the organisms found in someone’s intestines.
6. Go back to the article by Carl Zimmer that you were provided at the beginning of the exercise. How was the situation for that 14 year old male different from your 15 year old female presented in this exercise? How were the techniques for diagnosing the illnesses in these two patients similar? Why is this medical approach better than techniques that have been used in the past (before this technology was available)?