**Biol/Gen 462 Fall 2019 Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Lab Answer Sheet**

**Send via Slack to Jessica by 5PM, Tuesday, 29 October 2019**

**Part I. NCBI Sequence Databases and Tools (17 pts)**

1. **Dyskeratosis congenita can be caused by mutations in a wide variety of genes. Identify three such genes. (3 pt)**
2. **What is the full name of the TINF2 gene? (1 pt)**
3. **On which chromosome, and where on that chromosome, is the TERC gene located? (1 pt)**
4. What are the RefSeq accession numbers for the mRNA and protein sequences for isoform 1 of the TERT gene? (2 pt)
5. **What is the GI number for this protein? (This is a unique identifier for every record in GenBank in NCBI.) (1 pt)**
6. **Paste the protein sequence for NP\_937983.2 in FASTA format below. (1 pt)**
7. **Name any 4 links that are available for this record. (1 pt)**
8. **Identify the conserved domains (concise) in this protein. (1 pt)**
9. **Identify (by its 4-letter PDB ID) at least 1 related structure for this protein. (1 pt)**

10. Write down the best 5 hits for your query sequence and the information for each one (including e-value). From these hits, write what gene your query sequence is. Click on one of the hits to go to the detailed flat file to see more details about the sequence. (2pts)

11. Write down the top 5 hits and their information with tblastx. Are they the same results as what you got for blastn? What about the e-values; why are they different? Can you guess what species this sequence may have come from? Why? Choose which algorithm you think gives the best hit when searching for the identity of an unknown sequence through BLAST (blastn or tblastx). (3pts)

**Part II. Biological Processes in the Community Ecology of a Sloth (8 pts)**

12. Which sequences come from the sloth, the algae growing on the sloth, and the microbes in the gut of the sloth? To figure this out, take each sequence and BLAST it. Because we are working on a community of organisms that are likely NOT represented in this genetic database we are looking for evolutionarily similar hits. As sequences evolve slower at the protein level then nucleotide level (i.e. many nucleotide mutations are “sense” mutations that don’t change the protein) it would be best to use the *tblastx* that will translate our sequence into the 6 possible protein sequences and compare it to the translated version of the nucleotide database.

**What are the top hits?** Hint – if you hit whole genomes in your blast search, you can scroll down to the alignments to see the specific region or gene in the genome that matches your query.

**Fill out the table. (3pts)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sequence** | **Species/organism** | **Gene** | **e-value** |
| **1** | | | |
| **2** | | | |
| **3** | | | |
| **4** | | | |

**13. Which sequences do you think belonged to the (1pt)**

Sloth:

Gut Microbes:

Algae:

**14. What are the biological processes that each of these genes are involved in? (2pts)**

**To investigate use GO Ontology. Type in the gene name (select the gene) and hit Search**

Hint – may have to put a “\_” between the letters and numbers in a gene name. For example if you found gene ABC123, type in ABC\_123

|  |
| --- |
| **Sequence, Gene Biological process** |
| **1** |
| **2** |
| **3** |
| **4** |

**15. Write a few sentences interpreting what is happening in this little community at the gene expression level inferred by the biological processes. (2pts)**