BIOL 3001 GENETICS LABORATORY

NAME:

# LABORATORY 8; EXERCISE 2. INTRODUCTION TO DNA SEQUENCING

**Purpose –** In this exercise, you will practice the major steps involved in nucleotide sequencing. You will choose one of the unknown DNA sequences provided to analyze. Once this sequence has been successfully read and scored, you will use some of the bioinformatics tools available in GenBank to analyze your unknown DNA sequence. Specifically, instead of using the query searches you did in Exercise 1 and 2, you will use the BLAST tool to search your unknown sequence against a database in GenBank. You will gain more experience with reading DNA sequences and you will learn to use GenBank as a major tool to answer bioinformatics questions during this laboratory exercise.

## Starting the Exercise –

1. You will begin this exercise by choosing one of the UNKNOWN TRAINING SEQUENCES.
2. Once you have chosen your sequence, be sure to write down all the information you are given regarding your sequence in the box below.
3. You are now ready to determine the DNA sequence for your Unknown Training Set. Below is a quick reminder of steps to electrophoresis. You should be familiar with steps from Laboratory 1. I am not asking you to perform the electrophoresis. The sequence you chose above will be the results of your electrophoresis.
    - Choose DNA samples to prepare.
    - Load DNA samples to appropriate wells of plate.
    - Heat DNA samples for a minimum 5 minutes. (*WHY?*)
    - Pour electrophoresis gel and place the comb in the gel.
    - Load the gel with your DNA samples.
    - Connect the power source and run the gel.
4. Read your gel for the correct DNA sequence (remember, you should start at the bottom of the gel and read a <u>minimum</u> of 100-200 bases. Enter the sequence of your unknown below.

5. Enter some basic information about your sequence below.

Total number of nucleotides that you read. ☐

| | | | |
|---|---|---|---|
| Number of Adenines in your sequence. | ☐ | % of total nucleotides | ☐ |
| Number of Guanines in your sequence. | ☐ | % of total nucleotides | ☐ |
| Number of Thymines in your sequence. | ☐ | % of total nucleotides | ☐ |
| Number of Cytosines in your sequence. | ☐ | % of total nucleotides | ☐ |

How many purines are present in your sequence? ☐

How many pyrimidines are present in your sequence? ☐

How many phosphodiester bonds would you expect to find in your sequence? ☐

Was the DNA sequence that you read single-stranded or double-stranded? Fully explain your answer below. Be sure your explanation includes what you have learned about DNA structure up to this point.

If your DNA sequence was actually double stranded, what percentage of the nucleotides in the entire sequence would be Cytosines? ☐ Adenines? ☐

9. Enter at least 100 bases from your Unknown sequence into the BioEdit program. Follow the Quick Start guide below and save the information in FASTA format.

- Open BioEdit.
- Go to “Sequence” and open “New Sequence”.
- Enter your DNA sequence into the window box that opens. Be sure to give your sequence a name!
- Select “Apply and Close”.
- Use your pointer to select the NAME of your sequence and highlight it by clicking.
- Go to “Edit” and “Copy Sequence to clipboard (FASTA format)”.
- Go to “File” and “Save” your sequence.

10. You are now ready to determine the identity of your Unknown DNA sequence. Open the GenBank website and click on the “GenBank” drop down menu at the top left of the page. Select “BLAST” from your choices.

11. You should notice you have a number of choice of different BLAST alignments to choose from. You are looking at nucleic acids (i.e. DNA) so you should choose the "Nucleotide BLAST" option.
12. This will open your data entry portal for a Standard Nucleotide BLAST. Right click and Paste your sequence in FASTA format into the "Enter Query Sequence" box. You will know you are in FASTA format if the first line of entry looks something like : ">{your sequence name}" and your actual sequence data starts on line two.
13. Take some time to examine all the options you have for searches on this page. You can choose different databases, you can choose different organisms. You can use the same Boolean operators to select search parameters if so you choose.

    For our purposes, go with all the default entries, mainly be sure the database chosen is the "Nucleotide Collection (nr/nt)" and click on the blue "BLAST" at the bottom of the page. Depending upon traffic at the site, you may immediately get an answer or you may get an updated status every few seconds before your Results are given.
14. Using information from your BLAST search, answer the following questions.

    What gene or gene region did your Unknown most likely come from? (Note: It is possible that your DNA fragment will be matched to a segment of DNA from an entire chromosome. If this occurs, you will not be able to determine the specific gene region directly from the BLAST output. Ask your instructor for some assistance.)

    What was the GenBank Accession number for the sequence most closely matched to your Unknown?

    Accession number –

    In Biology, most organisms are known by two types of names. One is the unofficial Common Name (Red Fish, Highland Rose, Bengal Tiger) and the other more formal Latinized Scientific Name that consists of at least two parts, Genus and species. Based on your GenBank output, what are the common AND scientific names of the organism that most closely matches your Unknown sequence?

    Common Name –

    Genus - Species –

Carefully look at the list of sequences indicated in GenBank that share some similarity to your Unknown. Write a short summary of what you observe. For example, are all of these similar sequences from the same Genus? Are they all one kind of organism, like are they are all mammals? All vertebrates? Does the unknown cross over different kingdoms: plants vs. animals? Which do you think is the most unusual match and why?