Biology 211
Lab 1 - Sequence Data Analysis

Objectives:

- Analyze a small sequence from the ComGen project.
- Begin to become familiar with sequence analysis using BLAST.

Before lab:

- Since it is the first day of class, no pre-reading is necessary.
- For background and tutorials on how to do your BLAST analysis and what this all means, view the Lab 9 Bioinformatics.

In lab:

- Analyze your sequence data to determine the genetic sequence.
- Work with the BLAST analysis to determine what the best matches to your sequence are.
- Answer the questions below.

BLAST stands for Basic Local Alignment Search Tool. It allows you to compare your sequence to all known sequences and find other sequences that are the same or similar (we call that alignment). An online computer program does the comparison and gives you the results you see here. The higher the alignment score, the greater the degree to which the sequences match. Determine your alignment score on the color-coded diagram. If you have good alignment, determine what genetic sequence your stretch of DNA matches on the table that follows your alignment diagram.

You will need to heavily utilize the BLAST handbook, accessible here:
http://www.ncbi.nlm.nih.gov/books/NBK21097/

In your background:

- What are the different types of BLAST? When would you use each? Which one(s) are we using here?
- What is E-value? What is considered a "good" E-value score? Why?
- Why can close nucleotide alignments between two separate organisms be used to predict function? For example, we are sequencing *Pseudomonas fluorescens*, yet your top hits may be from *Pseudomonas brassicacearum*. Why can we use the *brassicacearum* data to infer function in *fluorescens*?

In your discussion:

- What is the most likely identity of your sequence match and what organism(s) is it from?
- What is the most likely function of your sequence? How do you know?
- Generate a hypothetical phylogenetic tree of *Pseudomonas* based only on your sequence.