Rec DNA Lab,
1 June 2005-Sequence analysis

For the first part of this exercise we will look at the raw data given by the ABI sequencer. The sequences are located on webCT in a folder labeled "sequences" on Dr. Courcelle's website (http://web.pdx.edu/~justc/courses/RecombDNALab.html).

Drag these files to the desktop and double click to open them. This will start the program called 4peaks. This program will allow you to see the electropherogram of the dye peaks, and allow to edit your sequences and to fix miscalls the machine made. There is a range of results in the sequencing runs, a few good, a few poor, and some in between. Look at your own sequence and note the results. Did you get a good read? If not, what may have happened?

Edit your sequence if possible, if there is nothing in your sequence, choose another sequence that looks good. By clicking on the helix symbol in the lower right hand side of 4peaks you can open a text window showing just the sequence. Single clicks on letters will allow you to shade the text, which you can use to mark the beginning and end of the usable part of the sequence. You can then edit changes in the sequence by double clicking on the letters you wish to change.

When you feel you have edited the sequence as well as you can, save the text to the desktop in word. Repeat this with at least two other sequences. These text files can then be cut and pasted into the online search engines to compare them to the public databases. (apple C cuts things, apple V pastes things)

For the second part of the lab, we will BLAST our gene sequences against the public database and see if we can find some similarities to previously sequenced genes.

The hyperlink below will bring you to the NCBI Blast search page.
http://www.ncbi.nlm.nih.gov/BLAST/

Try several of these different search algorithms including blastn, blastx, and megablast, blasts of environmental samples such as the saragasso sea, and searches by domain architecture.

In most cases you should get a list of results that each look something like this:

```
1_17300 gi|48446|emb|X58791.1|VHLUXAB  100.00  635    0    0    1    635    28
```

The code "X58791.1" is used as the identifying ID for the sequence that was entered into the database. Use codes in this column to look up individual sequences. You can cut and paste the link into the search engine in the following link:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide
or click on the blue hyperlink in the blast results

Try several different Blast algorithms, including protein and nucleotide searches. Also try to align two of your edited sequences (bl2seq). How close are these sequences to each other?

Questions to answer:
1. Were most of your blast hits from bioluminescent bacteria? What were some of the other organisms? What were some of the other genes? Why do you think these results showed up?

2. E values give the probability that a matching sequence matches because of chance and not because of true similarity (E values of 0 means there's no chance that the results are random). Based on looking at your blast results, what range of e values would you trust to be relevant?

3. What is the difference between BLASTn and BLASTx ? Which is better at identifying biologically relevant sequences (eg Genes rather than non-coding regins)?

4. Do different organisms match when you blast against protein database and the nucleotide database? Why is this? If organisms are the same does the E value change?

5. Using this method of amplification and sequencing from primers designed from conserved reigns of bioluminescent Vibrio species; do you think we are getting an accurate representation of the bioluminescent diversity present in the ocean? Why or why not? What are some of the factors that may lead to biases in this type of experiment?

6. What were the results of your sequencing reaction. Explain.