Lab 3 Biol 4150/6150 Fall 2021 Pairwise Alignments

Due 1:00 pm Monday, September 20. Paste your responses and results below, rename the document to YourLastName-Lab3 and upload to Canvas Assignments folder.

1. (10 pts) Create and show a **protein** sequence dot plot of any animal calmodulin Refseq sequence entry versus any plant calmodulin Refseq sequence. Adjust parameters (window size, threshold value) until the background is reduced to a suitable level. Paste a screenshot of the final dotplot.

* 1. (5 pts) What would be the effect of having a window size that is too large? Overlapping matches
  2. (5 pts) Do you see multiple diagonals, and what do they indicate?

This relationship is affected by certain sequence features such as frame shifts, direct repeats, and inverted repeats. Frame shifts include insertions, deletions, and mutations. The presence of one of these features, or the presence of multiple features, will cause for multiple lines to be plotted in a various possibility of configurations, depending on the features present in the sequences.

>sp|P62152|CALM\_DROME Calmodulin OS=Drosophila melanogaster OX=7227 GN=Cam PE=1 SV=2

>sp|P59220|CALM7\_ARATH Calmodulin-7 OS=Arabidopsis thaliana OX=3702 GN=CAM7 PE=1 SV=2

Programs for creating dot plots may be found on the web, or you may use **dotmatcher** in EMBOSS (European Molecular Biology Open Software Suite) that can be run either via command line or using Jemboss (Java GUI for Emboss). EMBOSS and Jemboss are freely available for download and installation on your own laptop or desktop PCs.

If using **dotmatcher** in EMBOSS, use the BLOSUM60 scoring matrix (or the closest to it).

2. (10 pts) Create and show a **nucleic acid** sequence dot plot of the plant calmodulin sequence versus the genomic sequence of the Arabidopsis calmodulin gene.

a. (5 pts) What scoring matrix did you use?

b. (5 pts) Where do any gaps in the dot plot occur, and what causes these gaps?

3. (10 pts) Use the MHoffman Excel file (downloadable from the course Canvas site) to align the first seven letters of your name to Sequence 1 (HEAGAWGHEE). If your name contains any of the letters that are not represented in the single-letter code for amino acids, make substitutions as listed below:

B = H

J = A

O = E

Q = W

U = G

X = G

Z = E

a. (5 pts) global – perform the traceback and show the highest scoring alignment – use a nonproportional font such as Courier to show the alignment.

b. (5 pts) local – perform the traceback and show the highest scoring alignment – use a nonproportional font such as Courier to show the alignment.

4. (10 pts) Perform an NCBI **BLASTP** search with the query sequence NP\_001269550 against the nr database, to find homologues in higher plants (angiosperms). Be sure to click on Algorithm Parameters to see what the parameters of the blastp search is.

What is the default scoring matrix?

What is the default word size?

Paste the pairwise alignment view of the highest scoring sequence alignment, showing the bitscore and E-value.

5. (10 pts) Perform a global pairwise alignment between NP\_001269550 against the highest scoring sequence from part 4. The NCBI BLAST home page has a Global Align tool that applies Needleman-Wunsch. Be sure to switch to the Protein tab. You can either enter the accession numbers of the two sequences or their FASTA sequences.

Paste the global alignment of the two sequences below.