**Lab 3 Biol 4150/6150 Fall 2021 Pairwise Alignments**

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.

* **I performed the dot plot on EMBOSS with the EBLOSUM 60 scoring matrix with Solanum tuberosum (Potato) and Drosophila melanogaster. The results can be interpreted with the plot below ( I got the best results with the default threshold and window size and did try adjusting it to other values to view changes)**

A picture containing diagram

Description automatically generated

* 1. (5 pts) What would be the effect of having a window size that is too large?
* **The effect of having too large window size is overlapping matches. Additionally, it could also result in missing of detection of repeating sequences and gaps if present in the sequences** 
  1. (5 pts) Do you see multiple diagonals, and what do they indicate?
* **Multiple diagonals indicate that the 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.**

1. (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 picture containing table

Description automatically generated

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

* **DNAfull**
  1. (5 pts) Where do any gaps in the dot plot occur, and what causes these gaps?
* **Gaps occur in case of missing data in the sequences due to improper sequencing. Gaps can be caused due to insertion, deletions and mutations.**

**This would result in different views of the dot plot according to the type of anomaly observed.**

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.

**After performing global alignment, the match that I obtained with my name is illustrated below:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| H | E | A | G | A | W | G | H | E | E |
| \* | \* | A | S | H | I | \* | K | A | R |

**The highest scoring alignment was -16 and illustrated in red as:**

Text

Description automatically generated

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

**After performing local alignment, the best match that I obtained with my name is illustrated below:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| H | E | A | G | A | W | G | H | E | E |
|  |  |  |  |  | A | S | H |  |  |

**The highest scoring alignment was +7 and illustrated in red as:**

Calendar

Description automatically generated with low confidence

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?

* **BLOSUM62**

What is the default word size?

* **6**

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

Table

Description automatically generated

Graphical user interface, application

Description automatically generated with medium confidence

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.

Please find the results from the global pairwise alignment as well as the

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated with medium confidence