Lab02

Introduction

We will recreate some of the analysis depicted in Lecture 02: Dotplots, alignment scores, and global alignment using Dynamic Programming.

Document all work during the dry lab for each exercise, defining all your tools and parameters, data output, and interpretation.

From Lab02, there are three Assignments to be submitted to myCourses:
Discussion 2.1
Activity 2.1
Discussion 2.2

Dotplot

Using a Dotplot graphic, identify all the sequence alignments and the differences between the sequences. Make sure to describe and interpret all types of similarities/dissimilarities:

1) Matches

On the dot plot graphic, a match between sequences resembles a diagonal line, representing the continuous match (or repeat).

2) Frame shifts

a. Mutations

Mutations are distinctions between sequences. On the graphic, they are represented by gaps in diagonal lines. They interrupt matches.

b. Insertions

Insertions are parts of one sequence that are missed in the other while the surrounding parts match. In other words, an insertion is a subsequence inserted into a sequence. Graphically, insertions are represented by gaps that lie only on one axis. A little shift towards the other axis indicates a mutation involved.

c. Deletions

A deletion is a subsequence that was deleted from a sequence.

A deletion from sequence A found in sequence B can be considered as an insertion into sequence B and contained in sequence A.

- Low-complexity regions

Redundancy in a particular part of the sequence produces a low-complexity region. On a plot, it is represented as a rectangular area filled with matches.

***Hint**: compare a sequence with itself to easily find inverted repeats or low-complexity regions in it.*

Discussion 2.1
There are many Dotplot programs (java applets, standalone software packages, check-in https://en.wikipedia.org/wiki/Dot_plot_(bioinformatics)). Today, I recommend using NCBI Blast to compare two sequences and visualize the Dotplot for large datasets. There is a Blast 2sequence suite, which is very appropriate for this exercise:
https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=MegaBlast&PROGRAM=blastn&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq
*But it has a limit of 10 Megabases

We will compare two chloroplast genomes (ca. 116 kb and 140 kb). Accession numbers:

NC_014267
NC_014287

Search for those two genomes, save them in fasta format (nucleotide sequence), and use them to run **blast2seq**. Megablast (quicker) or blastn (slower) will work as an appropriate algorithm (Program selection). Leave parameters as default, but check the (?) to know their meaning. If all looks good, "BLAST"!

Let's discuss the output options. There are four tabs:

Descriptions//Graphic Summary//Alignments//Dot Plot

Save the Dotplot image and try to identify (using an image editor/drawing application, like Paint or PowerPoint) what type of similarities/dissimilarities (see above descriptions) are in the plot. E.g., use squares/rectangles/circles (not filled) and name them (1a, 1b, 2a, etc....) for each kind. Then, add a description for each.

What happens if you blast the same genome vs itself? What is the dot blot showing?

Discussion 2.1
Once you finish, submit the report (edited figures and explanations) to myCourses (in Assignments).

Activity 2.1

Protein sequence comparison using Dotplot. Our subject is the "muscle" protein Filamin-A, a large cytoplasmic protein that associates with actin filaments in the cytoskeleton.
However, we would like to see the degree of conservation between species (e.g., humans, mice, and drosophila).
-Download the protein sequences fasta files from

myCourses/Content/Labs/Lab02_Activity2.1.zip

-Set a three-way comparison (3 pairwise alignments with Dotplot)
-For this activity, you can use the online Java applet Dotlet (https://dotlet.vital-it.ch/)
-What type of alignments do you see represented in each plot?
-Adjust **window size** and **Scoring matrix** for better visualization.
-Save Dotplot images along with the parameters used (window size, scoring matrix)
-What happened while using distant/closer protein homologs? Could you interpret such results? ***Hint****: check the protein structure and domain organization of Filamin-A (NCBI, Uniport, Wikipedia)*.

Activity 2.1
Once you finish, submit the report (edited figures and explanations) to myCourses (in Assignments).

Dynamic Programming

Discussion 2.2

Complete a dynamic programming matrix using global alignment between two DNA sequences and calculate the optimal global alignment and the alignment score.

Subject sequence: GACGCAGT
Query sequence: ACCCAT

Scoring system:
Match: 5
Mismatch: -2
Gap: -6

Use the Excel spreadsheet (in myCourses, under Lab02_DynamicProgramming.xlsx) or paper printout to complete your calculations.

Discussion 2.2

Once you finish, submit the report (edited Excel or scanned document) with your solutions to myCourses (in Assignments).