IN4030 Introduction to bioinformatics

Exercises - week 1

Molecular Biology, dot plots and alignments

1. Can you identify the 5' UTR, the 3' UTR, the coding sequence, and the encoded protein sequence in the following mRNA sequence?

    ACTTGTCATGGTAACTCCGTCGTACCAGTAGGTCATG

2. The similarity between two sequences can be visualized with a dot plot. The two most important parameters that can be adjusted when making dot plots are the window length and the number of mismatches allowed. In the two dot plots below, the same two sequences are against each other using the same program. Only parameters are changed between the left and right image.
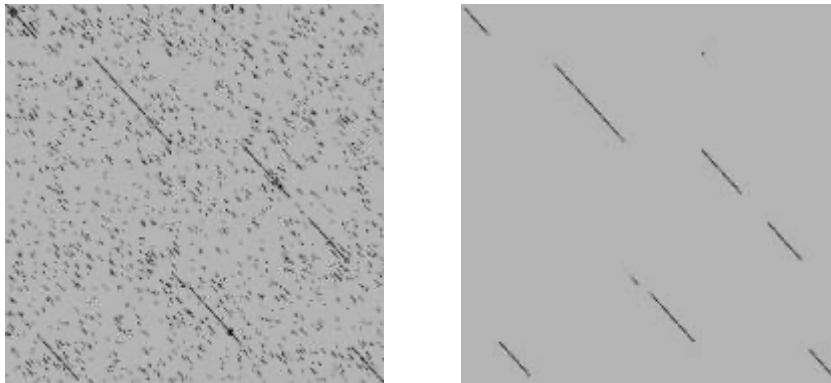
a) In the two plots above – which one has the larger window size (assuming the number of allowed mismatches per window is constant)?

b) Do you think that these sequences are homologous?

c) Draw, by connecting the diagonal lines in the right plot, the best global alignment.

d) Would a global alignment contain all significant similarities between these two sequences?

3. Retrieve the sequence of the human zinc finger transcription factor gene MAZ with accession number NM_002383 from GenBank (https://www.ncbi.nlm.nih.gov/) and save it in FASTA format as a text file. Compare it with itself using Dotlet or dotmatcher. Try different parameters (window size etc). Describe what you see. What are appropriate parameters?

Dotlet JS (javascript) can be found here:

https://dotlet.vital-it.ch/

An alternative is the old Dotlet applet (requires Java):

https://myhits.isb-sib.ch/cgi-bin/dotlet

Another alternative is the Emboss Dotmatcher, here:

https://www.ebi.ac.uk/Tools/seqstats/emboss_dotmatcher/

or other Emboss tools at https://www.ebi.ac.uk/Tools/seqstats/

4. In each of the two cases below, indicate possible steps of single mutational events (substitution, insertion and deletion of single symbols) that would transform the first sequence into the second sequence. How many steps are necessary in each case? What would the global alignment score be if the score for a match is 1, the score for a mismatch is 0 and the penalty for a gap is 1?

a)
CAGGTTGCA
TAGGTCA

b)
PVALGLKEK
PVIGLKDK