BIN506 Assignment 3

Due Date: 25 March, 23:59

Late Submission Policy: After the due time, 10 points of deduction will be applied for each extra hour.

Submit your answers in one pdf file please.

1) You will see 5 sequences (sequence_#.fasta) attached to this assignment. Use blastn to align them to 'sequence_DB.fasta' and answer the following questions. Add screenshots to your assignment.

   a) Align sequence_1 to sequence_DB, comment on the scores and the dot-plot.

   b) Align sequence_2 to sequence_DB, comment on the scores and the dot-plot.

   c) Align sequence_3 to sequence_DB, comment on the scores and the dot-plot.

   d) Align sequence_4 to sequence_DB, comment on the scores and the dot-plot.

   - Important: Please show or give information about the event you see on the plots, i.e. if you see a deletion, show it on the plot, write down how many events are there, etc. Provide details.

   - sequence_DB is the reference sequence here. You will be comparing all other sequences to it.

2) Use the **msa.txt** file attached to the assignment to answer the following questions.

   a) Can you figure out the origins of these sequences from the identification line?

   b) Use **T-Coffee** and **Clustal-O** to perform MSA for the given sequences. Compare the results.

      - What are the main differences in the alignment results? Tip: Look at the aligned regions in both results. Compare the length of gaps and aligned segments. (Show some examples with screenshots)

      - Why do you think these differences occur? (No need to explain the algorithmic differences, just try to explain the main difference between the two tools).

      - Compare the phylogenetic trees provided by the tools (Ignore the numbers, just look at the clusters).

3) Answer the following questions about Clustal-O.

   a) What type of MSA method does Clustal-O use?

   b) What is the substitution matrix used in Clustal-O?

   c) What are the default gap opening and extension penalties for Clustal-O?

   d) What do the (*), (:) and (.) symbols at the alignment positions indicate? What does it mean if there are no symbols?

Bonus Question: Have you used ChatGPT or any other Large Language Model to assist with these questions? If so, please share the prompts you used (or better yet, a link to the full conversation). Using AI tools is completely acceptable, as long as you understand what you're asking and can critically evaluate the responses you receive.