Introduction to Bioinformatics online course: IBT

Practical Assignment

Module name: Sequence alignment theory and applications
Session name: Pairwise Sequence Alignment
Trainer: Jonathan Kayondo, Sonal Henson
Participant: *<write your name here>*
Date: *<write today's date here>*

Pairwise Sequence Alignment

Introduction
Sequence alignment is a complex biological task with several strategies and programs to pick from. The methods usually require intelligent use of several variables in the programs. Dynamic programming helps with finding optimal sequence alignments.

This exercise demonstrates the usefulness of the local alignment in bioinformatics, and also what can go wrong if you are not careful in choosing your parameters. We also look at the dynamic programming applications in the Needleman-Wunsch algorithm.

Tools used in this session

Uniprot - www.uniprot.org
EMBOSS Water – www.ebi.ac.uk/Tools/psa/emboss_water/
EMBOSS Needle - www.ebi.ac.uk/Tools/psa/emboss_needle/
TeachEnG - http://teacheng.illinois.edu/

Please note

- **Hand-in information** If you are formally enrolled in the IBT course, please upload your completed assignment to the Vula 'Assignments' tab. Take note of the final hand-in date for each assignment, which will be indicated on Vula.

Task 1: Local vs Global alignment and choosing the correct parameters.

Task 1: instructions

Introduction to Bioinformatics online course: IBT

The local alignment finds regions of local similarity. It is implemented trough Smith–Waterman (but other also) algorithm. It has many purposes, and just some of them are finding motifs, finding one sequence within another, search sequence databases, etc.

Using the Smith-Waterman algorithm we will make a local pairwise alignment of an unknown sequence to the ACV synthetase of Penicilliumchrysogenum and investigate the effect of adjusting the alignment parameters on the output.

1. Go to www.uniprot.org and download the sequence for ACV synthetase of *Penicillium chrysogenum* (accession number: P26046).
2. Download unknown_sequence.fasta from Vula -> Practical Assignments -> Module 3 Session 2
3. Go to www.ebi.ac.uk/Tools/psa/emboss_water/ to make your alignment.
4. Upload or paste the *P. chrysogenum* sequence in the first box and the unknown sequence in the second box.
5. Click 'Submit' to execute the alignment.
6. Click on 'View Alignment File' and save your alignment output in a text file.
7. Go back to the alignment submission page and repeat step 4.
8. Click on 'More options' under Step 2 and set the Gap Open and Gap Extend to minimum.
9. Submit your alignment.
10. How does your alignment from step 5 and step 9 compare in terms of alignment score, % identity, % similarity, gaps, and the extent (positions) of alignment of the two sequences?

11. Now use the Needleman-Wunsch algorithm for global pairwise alignment to align the two sequences from steps 1 and 2. You can use the EBI tool Needle for that (http://www.ebi.ac.uk/Tools/psa/emboss_needle/)
12. Follow steps 4-6 to submit the alignment.
13. How does your alignment compare with that from step 6?
14. Repeat steps 8 and 9, this time on the Needle program.
15. How does the alignment compare to that from step 12 (using default parameters)?
16. This exercise demonstrates the usefulness of the local alignment in bioinformatics, and also what can go wrong if you are not careful in choosing your parameters.

Task 1: participant's answer
<start typing your answer here>

Introduction to Bioinformatics online course: IBT

Task 2: Dynamic programming

Task 2: instructions

1. Go to TeachEnG (http://teacheng.illinois.edu/) and play the Sequence Alignment game. Click on the 'Play' button on the right of the page. Try all the three levels. Test the effect of adjusting the parameters on the overall score.

2. Similarly, play the Needleman-Wunsch algorithm game. Fill out the dynamic programming matrix. Refer to the Instructional Material for help on how to play the game.

Task 2: participant's answer
<start typing your answer here>