

Bioinformatics computing
CSE40532/60532
Homework #2

Reading:

1. Skim journal article (esp. BLAST) handouts (week of 9/9)
2. Read Chapter 4 and handouts on Viterbi algorithm (week of 9/16)
3. Read Chapter 2 and Glimmer paper (week of 9/23)

Problems: (due 9/23)

1. Download the two genes mentioned in Problem #1 in Section 3.9 (X79493 and AY707088) from fruit fly and human, respectively.
2. Write a program called “nwalgn” to compute the global alignment score of the two sequences above using the following parameters: +2 for a match, -1 for a mismatch, and -2 for a gap. Perform the traceback and display the alignment in a user-friendly format. Please place the source code in your dropbox, and tell the TA how to compile and run it in an accompanying write up. (10 points)
3. Compute the local alignment between X79493 and AY707088 using a modified program “swalgn”. Perform the traceback and display the alignment in a user-friendly format. Submit your code with instructions on how to run and compile. Discuss in your writeup how the local alignment compares to the global alignment from #2 (8 points).
4. BLAST the two sequences from #1 using the bl2seq utility at NCBI BLAST. Summarize the BLAST alignment in your writeup (alignment length, identity, e-value, etc.) and discuss how it confirms or refutes your result from #3 (2 points).
5. Compute the end-gap free alignment of the two mitochondrial sequences from HW #1: NC_001807 and AF254446 using only linear space. Report the score using the scheme used for problem #2 above (+2 match, -1 mismatch, -1 for a gap) in your writeup (5 points) Required CSE60532: Perform the traceback using linear space and submit your code to the TA also (5 additional points).
6. Download the “mystery” sequence from the course website. Report its BLAST hits using blastn and blastx at the NCBI website in your writeup (1 point each).
7. Download the sample sequences from the course website. Write a small program/pipeline that will compute the end-gap free alignment score of all pairs of sequences in the single input file. Output a single line of output with the following information for each pair: SeqA SeqB score. Submit the code with instructions in your dropbox (8 points).