

Homework #3

Computer Science for Bioinformatics (140.636)

Implement the Smith-Waterman local alignment algorithm and output the alignments that have the top two scores. Use these sequences

S = GGAGTGAGGG

T = TGCGGAGTGAGGGGAGCAGTTGGGGAGTG

Solution strategy:

1. Don't write something from scratch. Just modifying the Needleman-Wunsch example that was described in class.
2. Paste in the above sequences instead of the ones in `nw.pl` and run the code to perform a global (Needleman-Wunsch) alignment. This will just verify that you have a working version of `nw.pl` and that it works on your sequences.
3. Next, modify the script to perform the Smith-Waterman algorithm instead of Needleman-Wunsch. As a first cut output just one alignment. Note that there might be two alignments with the maximum score. At this point, keep your life simple and just output one of them. You might want to write a traceback function that takes the starting cell and does the work.
4. Finally, rank all the cells in the scoring matrix by their scores and apply your traceback routine to the all the cells that have the top 2 scores. Note: If several alignments have the same scores, you might have more than 2 alignments to output.