

605.620 Foundations of Algorithms

Project 3

Design and implement an algorithm to solve the longest common subsequence problem (chapter 14.4).

S1 = ACCGGTCGACTGCGCGGAAGCCGGCCGAA

S2 = GTCGTTCGGAATGCCGTTGCTCTGTAAA

S3 = ATTGCATTGCATGGGCGCGATGCATTGGTTAATTCCTCG

S4 = CTTGCTTAAATGTGCA

Compare each of the provided strings to each other and to your test cases. Compare by pairs only, finding the LCS of the pair. This will give you six strings. Gaps in matching the substring are allowed.

Here is a related YouTube video you might find interesting “A Dynamic Programming Algorithm - Sequence Alignment” - By Tim Roughgarden <https://youtu.be/xccdfMM6l7c>. Roughgarden has other useful videos as well.

Count individual comparisons to use as a basis for cost efficiency in terms of the string lengths. In addition, do timing of the work involved. You must do **both** timing and counting comparisons.

In addition, implement a brute force method for comparison and run all data sets (yours and mine) against it. Do **both** timing and counting comparisons.

A file with the required input is provided. The required input is used to demonstrate correctness. You should supplement this with your own cases that demonstrate correctness in various standard error situations. You must provide the output for the required input, but you should **not** include it in your graph and table

To demonstrate the asymptotic cost, you will need to create additional input sets with larger strings. You need to collect enough data to have a meaningful comparison of the theoretical efficiency to the observed efficiency. The analysis should include comments about what you learned, what you might do differently next time, justification of your design decision, and issues of efficiency with respect to time and space. Your analysis must include a table and a graph of the asymptotic costs that you observed. Be sure to consider your experiences with the problem and particularly consider the efficiency. Why is it useful and relevant to Bioinformatics? **Before you hand this in, be sure to reread the Programming assignments guidelines.**