Peyton Somerville

COMP. 4290

2/7/21

Homework 2 Report

1. a) ATGTTAT

   ATCGTAG

   The sequences are already aligned.

   b) - - - - -ATGAATGCGATTTCGGGTGGCC -

   TTGGCAGGACATGAAGTTCGATACGGAA

2. I tested some other gap penalty values, like 10, 15, 50, 100 and 500, and they all resulted in the same sequence alignment. So I think the gap penalty has no impact on the alignment result.

My approach:

I used C++ to make this program. I originally implemented this program using a 2D array to represent a matrix. But the problem with the array was it could not have variable size values, so I decided to use a vector or vectors instead. I start by setting the size of the matrix based on the lengths of the two DNA sequence strings. Then I fill up the first row and column of the matrix based on the gap penalty, and fill the rest of the matrix with 0s to start. Then I implemented the described algorithm using the formula $D(i,j) = \max[\ D(i-1,j) + g,\ D(i-1,\ j-1) + S(a_i,\ b_i),\ D(i,j-1)+g]$. I made a function to find the score based on the two letters being compared, which is added to the D(i-1,j-1). I utilized C++'s max algorithm to find the maximum of the three values in the formula. This formula is in a double for loop, so the rest of the matrix is filled. The optimal alignment is calculated based on the full matrix.