

# Computational Molecular Biology 2019 - Homework One

Colm Seale

February 2019

## Question One

Given sequence S of length N and sequence T of length M, how many possible alignments are there? Give an expression in N and M.

Given a sequences of length N and M, we construct a Needleman-Wunsch grid with sides  $n+1$  and  $m+1$  when we include the gap, as show in the completed example in Figure 1. Each path from the top left of the grid to the bottom right taking right, down or down-right diagonal steps provides a valid solution to the problem.

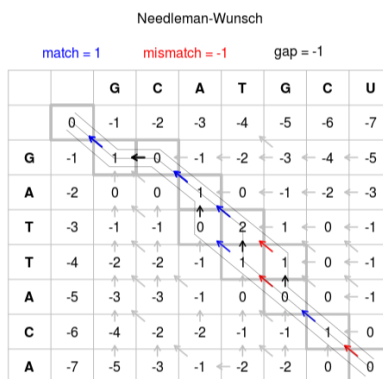


Figure 1: Needleman-Wunsch

This problem can be described using **Delannoy numbers**. Therefore, we can describe the expression for the number of possible alignments in terms of  $n$  and  $m$  as:

$$\sum_{k=0}^{\min(m,n)+1} \binom{n+1}{k} \binom{m+1}{k} 2^k$$

## Question Two

Results were obtained on a random DNA sequence 100000 nucleotides long, generated from a script. The running time was 4129.23 seconds / 68.8205 minutes. The best global alignment score was 76046. The code and test sequences are included.