Name(s)	Project Number
Christian Liu	
	31682

Project Title

Effect of Gap Penalty on Global Alignment Accuracy

Abstract

Objectives/Goals

The Needleman-Wunsch global alignment algorithm uses a scoring system consisting of a match, mismatch, and gap penalty to compare two large sequences of data. This research aims to examine effects of the gap penalty on sequence alignments when the match score is 1 and the mismatch score is 0.

Methods/Materials

Sequence-modifying, global alignment, and alignment scoring programs were written in Java and used to compare alignments. The sequence-modifying program controlled how similar sequences were by altering 50, 200, 350, 500, and 600 characters in an original 626-character long sequence. These altered sequences were then globally aligned to the original DNA sequence using 0, -1, -2, -3, and -5 as gap penalties. Finally the scoring program judged the accuracy of the alignment.

Results

At a gap penalty of 0, there were gaps throughout the alignments and alignment scores were low, but at a gap penalty of -5, the algorithm avoided gaps for all alignments. The ideal gap penalty was -2.

Conclusions/Discussion

At very small gap penalties, scores are low because the algorithm doesn't penalize for using many gaps. However, the algorithm quickly stops aligning as penalties grow, so the most effective penalties are small values still large enough to discourage unnecessary gaps.

Summary Statement

Five gap penalties were used to align sequences against an original and model the effect of gap penalties on the accuracy of the global alignment algorithm.

Help Received

My advisor Dr. James Li reviewed the programs I wrote and helped me correct errors.