

COMS4761: ASSIGNMENT 2.1: ezLife

Introduction

Dr. No – a brilliant but vicious individual -- has embarked upon a secret project to replace all living organisms with a new form of life, ezLife. ezLife is planned to accelerate evolution by orders of magnitude to form new species, far more fit than current life. It is estimated that first generation ezLife is already evolving at a rate faster by factor of 10^3 – also called k-PAMs (killo PAMs) -- over natural evolution, with an ultimate target of reaching M-PAMs (10^6) acceleration or even G-PAMs (10^9) acceleration, where ezLife will accomplish in one year the equivalent of 1 billion years of evolution, rapidly replacing current forms of life.



Only sketchy information is available on ezLife. It uses two new complementary nucleic acids, denoted $\{0,1\}$, which form stable ezRNA molecules, not requiring DNA. These ezRNA molecules are translated to ezProtein molecules formed from 4 new amino acids, denoted $\{a,b,c,d\}$. The genetic code consists of the translations: $00 \rightarrow a$; $01 \rightarrow b$; $10 \rightarrow c$; $11 \rightarrow d$; ezRNA sequences of $\{0,1\}$ use a translation start sequence: 11111111, corresponding to dddd.¹

In what follows, you are asked to analyze ezLife data obtained by intelligence sources. The term “compute an algorithm” in the questions below means showing the intermediate results of applying the steps of the algorithm, whether you compute it manually or code it.

¹ It may not be uninteresting to ask whether efficient sustainable living species may be developed with only 2 nucleic acids and 4 AAs; if so, what are the advantages of having 4 nucleic acids and 20 AAs? Similarly, can molecular evolution be accelerated; how; what are the natural bounds on evolutionary rates.

Problem 1: Alignment Of ezLife molecules

You are asked to align two ezGene sequences: X=0010010111 and Y=0110101100, using a fixed gap penalty $g=2$ and the scoring matrix

	0	1
0	2	-1
1	-1	2

- A. Compute optimal global alignments using the Needleman-Wunsch algorithm.
 - (i) Show the alignment matrix, score values and backtracking direction arrows as the algorithm completes: (a) initialization, (b) filling the left-upper 4x4 sub-matrix; and (c) filling the entire alignment matrix.
 - (ii) Show the various optimal paths for backtracking. Compute 3 different optimal alignments.
- B. Compute optimal local alignment using the Smith-Waterman algorithm. Show the computational steps as in (i) above. How many optimal local alignments are there? What are they?.
- C. Compute local alignment using dot-matrix heuristic as follows.
 - (i) Show a dot matrix and draw all exact-match-diagonals of length ≥ 2 .
 - (ii) Derive optimal non-gapped local alignment by extending exact-match-diagonals to maximize the score. Compare these local alignments with the results of B.
 - (iii) Use the DP algorithm for a band near the exact-match-diagonals to compute optimal, possibly gapped, local alignment.
 - (iv) How will the results of (ii) change if the scoring matrix penalizes mismatch with a score of -10 instead of -1?