

Homework – sequence alignment and distance estimation

Due on Friday, March 29, 5pm.

1 Problem

Consider two sequences: $v = \text{TACGGGTAT}$ and $w = \text{GGACGTACG}$.

1. Find the longest common subsequences for v and w by filling out the dynamic programming table (10 points).
2. Assume the scoring matrix: match is +1, mismatch and indel penalties are -1. Make a global sequence alignment by filling out the dynamic programming table (20 points). Note, I did not talk about this in detail. So I expect you to be able to find some reading on google and to solve this problem.
3. How many subsequences do v and w have respectively? (4 points)
4. Can you come up with a way of computing how many subsequences a sequence of length m has (6 points, please do some derivations instead of giving only a number).

2 Problem

Consider a mini-gene v : 5' CTAAATGAGCAAGTACCGGCTCCCCAT 3'. The 5' and 3' denote the DNA strand directions.

1. Let the transcribed form of v be w (i.e. the mRNA form of v), what is w ? (2 points)
2. Translate w into a protein sequence. (2 points)

3. How long is the mini-protein sequence (i.e. how many amino acids? In reality, proteins are much longer than what we see here). (1 point)

3 Problem

Consider the multiple sequence alignment shown in the figure. It contains the hemoglobin gamma A gene from the human, chimp, and rat.

1. How many amino acid does the hemoglobin gamma A protein have in the human? Is it the same length as the hemoglobin gamma A protein in the rat? Why? (4 points)
2. Compute the p-distance between: (1) the human and chimp hemoglobin gamma A gene, (2) human and rat, and (3) chimp and rat. (6 points)
3. Compute the Jukes Cantor distance between (1) human and chimp, (2) human and rat, and (3) chimp and rat (6 points)
4. Compute the number of synonymous differences for the gene segment from 169 base pairs to 189 base pairs: (1) between human and chimp, (2) between human and rat, and (3) between rat and chimp. (6 points)
Note, for simplicity, you are asked to compute only a small segment of the entire gene. There are many free programs that can compute the Nei-Gojobori distance.
5. Compute the number of nonsynonymous differences for the same segment: (1) between human and chimp, (2) between human and rat, and (3) between rat and chimp. (6 points)
6. Compute the number of synonymous differences per synonymous site for the same segment: (1) between human and chimp, (2) between human and rat, and (3) between rat and chimp. (6 points)
7. Compute the number of nonsynonymous differences per synonymous site for the same segment: (1) between human and chimp, (2) between human and rat, and (3) between rat and chimp. (6 points)

Please turn in your homework outside of my office in Torgensen 2160K on or before the due date. I will have a box for this purpose.