



Alexandria University
Faculty of Engineering
Computer and Systems Engineering Department

Bioinformatics

Master of Computer Sc.

Fall 2020

Assignment 2

Yousef Mohamed Zook



Problem 1. Profile alignment

In multiple sequence alignment, we often have to align a profile to a sequence. Use the average Sum of Pairs score to align the given sequence and profile. Your answer should include the filled in matrix given below, arrows for the traceback procedure and the final alignment.

The sum of pairs score, $S(m_i)$ for the i^{th} column is defined as

$$S(m_i) = \sum_{k < l} s(m_i^k, m_i^l)$$

where k and l iterate over all rows of the alignment. Use the following scoring criteria $s(x, -) = 0$, $s(-, -) = 0$ where $-$ denotes a gap. $s(x, y) = 2$ when x matches y , and $s(x, y) = 1$ when x and y don't match.

Solution

	-	A	G	G	T
	-	A	-	G	T
-	0	0	0	0	0
A	0	4	1	1	1
G	0	1	6	6	3
A	0	4	6	7	7

A G G T
A - G T
A G A -



Problem 2. Read the paper

Evaluating the accuracy and efficiency of multiple sequence alignment methods.

a) When comparing MSA algorithms, what are the main factors that should be considered?

1. Different parameter settings may improve their performance.
2. Size of indels, and alignments
3. the used dataset - simulated vs BALi-BASE benchmark

b) How to measure the accuracy of a given MSA algorithm?

1. Generate a simulated tree
2. Construct of simulated alignments
3. Construct test alignments
4. Measure a score, for example, sum of pair score SPS

c) Select two algorithms mentioned in the paper. Compare and contrast them by giving the most distinguishing characteristics for each.

Benchmark Alignments

	avg accuracy				Time consumed (*10 ³ s)
	overall	indel size	sequence length	deletion	
SATe	.31	.25	.35	.4	80.475
ProbCons	.36	.2	.5	.5	425.8

1. SATe is 529.10% faster than ProbCons
2. ProbCons has a better accuracy than SATe overall.