

CSC 110
Class Activity #9b

Name: ___ Joey Koumjian ___

Group #: _____

Optimal Sequence Alignment

Recall that sequence alignment involves determining the most likely relationship between a pair of gene sequences by aligning them to reflect possible evolutionary changes. From a computer science point of view, finding the optimal alignment between a pair of sequences is a string manipulation problem.

1) Given the following two sequences:

Example 1: GGAGTATA and GCGTTA

List all of the possible ways to align them with NO GAPS

Here is the first one:

Alignment 1	Seq 1	G	G	A	G	T	A	T	A	Total Score
	Seq 2	G	C	G	T	T	A			
	Score	1	0	0	0	1	1	0	0	

Alignment 2

GGAGTATA

GCGTTA

01011000

Score = 3

Alignment 3

GGAGTATA

GCGTTA

00000011

Score = 2

2) For each of the alignments listed above, give a score and circle the optimal alignment using the following: MATCH = 1 MISMATCH = 0

Alignments 1 and 2 are the optimal alignments

3) Now suppose you have the following two sequences:

Example 2: CTGAAGTCAGAG and CCTCAAGCGCCTGTG

Fill in the tables below:

	Example 1	Example 2
Number of possible simple alignments	3	4
Number of comparisons per alignment	8	15

Total number of comparisons	24	60
-----------------------------	----	----

Express number of possible alignments in terms of n (length of longer sequence) and m (length of shorter sequence)	$(n - m) + 1$
Number of comparisons per alignment in terms of n and m	n
Total number of comparisons in terms of n and m	$[(n - m) + 1] * n$
What is the order of magnitude for simple alignment?	$O(n^2)$

TA Signature: _____

- 4) Now let's consider alignments where you can insert gaps in the shorter of the two sequences to make the sequences equal length. Using the sequences from Example 1 above, list all possible alignments and score each alignment using the following:

MATCH = 1

MISMATCH = 0

GAP = -1

Here is the first one:

Alignment 1	Seq 1	G	G	A	G	T	A	T	A	Total Score
	Seq 2	-	-	G	C	G	T	T	A	
	Score	-1	-1	0	0	0	0	1	1	

HINT: Start by considering all the ways to add one gap in the sequence, and then add one more gap to each of those (removing duplicates).

Alignments with one:

GGAGTATA

- 1) GCGTTA-
- 2) GCGTT-A
- 3) GCGT-TA
- 4) GCG-TTA
- 5) GC-GTTA
- 6) G-CGTTA
- 7) -GCGTTA

Alignments with two:

GGAGTATA

- 1) GCGTTA--
- 2) GCGTT-A-
- 3) GCGT-TA-
- 4) GCG-TTA-
- 5) GC-GTTA-

- 6) G-CGTTA-
- 7) -GCGTTA-
- 8) GCGT-T-A
- 9) GCG-TT-A
- 10) GC-GTT-A
- 11) G-CGTT-A
- 12) -GCGTT-A
- 13) GCG-T-TA
- 14) GC-GT-TA
- 15) G-CGT-TA
- 16) -GCGT-TA
- 17) GC-G-TTA
- 18) G-CG-TTA
- 19) -GCG-TTA
- 20) G-C-GTTA
- 21) -GC-GTTA
- 22) -G-CGTTA
- 23) --GCGTTA

TA Signature: _____

- 5) Now fill in the following table with Example 1 from above and

Example 3: GGAGTATAA and GCGTTA

Note that Example 3 has one more character in the longer string than in Example 1.

	Example 1	Example 3 (Extra Credit)
Number of possible alignments with gaps in shorter sequence	30	
Number of comparisons per alignment	8	
Total number of comparisons	240	

TA Signature: _____