

Biol 47800/59500 2012 Homework 3

1. Work individually on this assignment. Feel free to write a program to perform the alignments if you like (it took me about 1.5 hr to do so).

2.

3. ACGCTAGCTG

4. TCCCCTGATG

5. Using the sequences above, show the global dynamic programming matrix (scores and pointers) and alignment using an identity scoring system

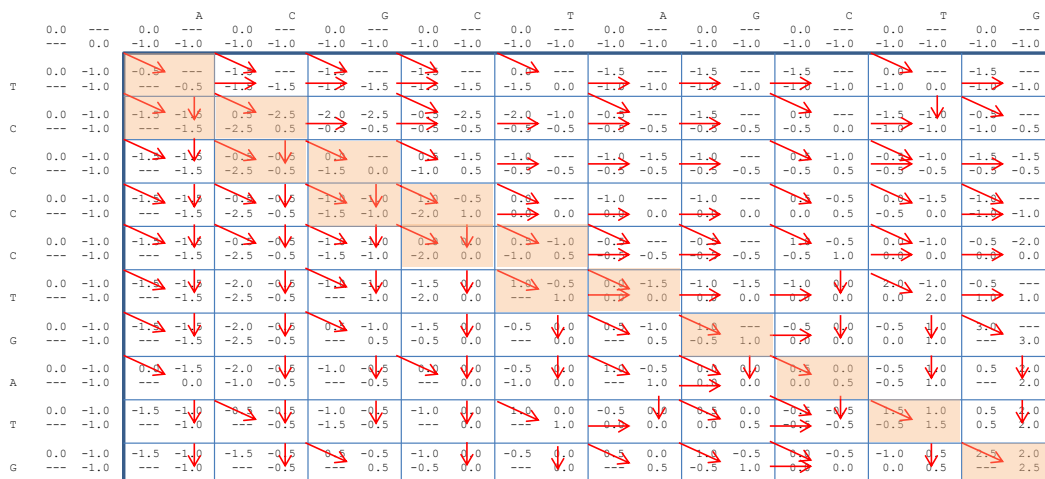
match = +1
 mismatch = -0.5
 gap open = -1 *nb*, length independent penalty
 gap extend = 0

Hint: you can avoid fractions by multiplying by 10.

6. Using the first two sequences, show the global dynamic programming matrix (scores and pointers) and alignment using an identity scoring system

match = +1
 mismatch = -0.5
 gap open = -1 *nb*, length independent penalty
 gap extend = 0

Hint: you can avoid fractions by multiplying by 10.



A C G C T A G C T G	5 matches = 5	A C G C . T A G C T G	
	5 mis == -2.5	A C G . C T A G C T G	5 matches = 6
T C C C C T G A T G	0 gaps = 0	A C . G C T A G C T G	3 mis = -1.5
	score = 2.5	A . C G C T A G C T G	2 gaps = -2
		. A C G C T A G C T G	score = 2.5
		T C C C C T . G A T G	

7. Show the dynamic programming matrix for the local alignment of sequence 1, "ACG", and sequence 2, "AACCGCG", using an identity scoring system

match=+1
 mismatch=-0.8
 gap open=-1
 gap extend = 0).

Scores shown below have been multiplied by 10 to avoid using fractions.

		A	A	C	C	G	C	G	
		0	0	0	0	0	0	0	
A		10	-10	-10	-10	-8	-10	-8	-10
	C	-10	10	0	10	0	0	0	0
C		-8	0	2	0	20	-10	10	-10
	G	-10	0	-10	2	-8	20	10	10
G		-8	0	-8	-8	-6	10	12	0
	C	0	-10	0	-10	0	10	0	12
	G					20	0	2	0
	C					2	20	10	10
	G							20	0
	C							10	20

In the upper left there are two ways to reach cell 3,1 (dotted arrow)

- a) Give the positions of all the maximum score(s) (e.g., position 1,1 would be the first letter in sequence 1 aligned with the first letter of sequence 2).

(3,5) (3,7) (2,3)

- b) For each maximum, show all the local alignments that can be traced using the traceback strategy discussed in class.

You can double check that all have score = 2 (20).

(2, 3) AC
 AC

(3, 5) CG AC.G A.CG
 CG ACCG ACCG

(3, 7) CG AC...G A...CG
 CG ACCGCG ACCGCG