

HW 8, Due on Thursday, April 7th (25 points)

For questions 1, 2, and 3, turn in a written version in class at 11:40AM.

Question 1

Align the two nucleotide sequences 'ACT' and 'AAC' using the global alignment method by drawing the 4×4 partial score matrix. Fill all cells and all corresponding “backwards-pointing straight-line segments” for each entry (if the same maximum score corresponds to multiple directions, draw all corresponding straight-line segments as shown in class). Identify the optimal path by thicker lines (or a different color), and write down the resulting optimal alignment and its score.

Use the following scoring system (slightly different than the one shown in class):

matches = +2;

indels = -3 (gap penalty);

for transitions (from A to G and vice versa, or from T to C and vice versa, so that the nucleotide type –purine or pyrimidine- is preserved) = -1;

and for transversions (substitutions in which the nucleotide type is modified) = -2.

In case there are multiple optimal alignments, show all of them.

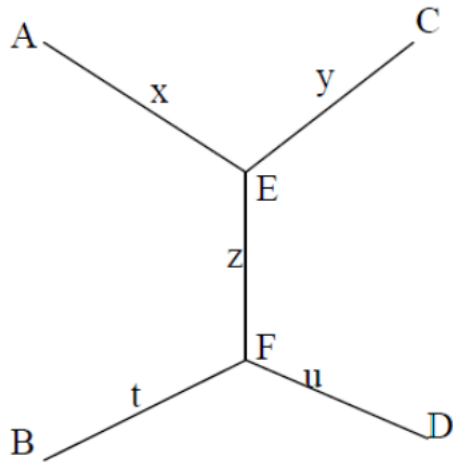
Question 2

Repeat Question 1, using the same scoring system, but this time use the local alignment method instead of global alignment method.

Question 3

It is given that the “Neighbor Joining” (NJ) algorithm was used to generate the following pairwise distance matrix for four DNA sequences A, B, C, D.

	A	B	C	D
A	0	0.3	0.5	0.6
B	0.3	0	0.6	0.5
C	0.5	0.6	0	0.9
D	0.6	0.5	0.9	0



This yields the following topology for an unscaled, unrooted tree from which the lengths of the branches are given by the variables x , y , z , t , u .

a) Find the lengths x , y , z , t , u . (Write down the systems of equations and their solutions).

b) Make the tree a rooted one by setting the root at the midpoint of the path connecting the two nodes of maximum distance from each other. Where will this root lie? Write down the Newick format notation of the resulting rooted tree.

c) Create a rooted scaled tree using the same distance matrix and the UPGMA method. Write down its Newick format notation of this tree. Is the topology identical to that of the rooted tree resulting from the NJ method? If not, explain why the two solutions are different.

HW8 Solutions

Question 1

		A	A	C
	0	-3	-6	-9
A	-3	2	-1	-4
C	-6	-1	0	1
T	-9	-4	-3	-1

Optimal global alignment (score: -1):

ACT

AAC

Question 2

		A	A	C
	0	0	0	0
A	0	2	2	0
C	0	0	0	4
T	0	0	0	1

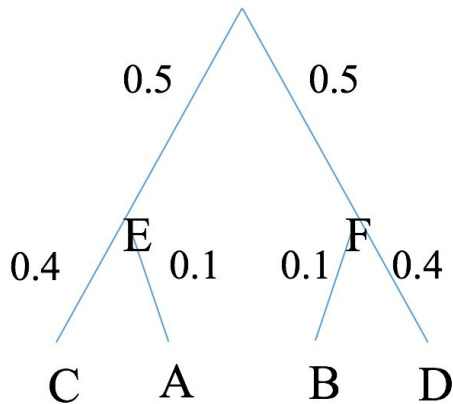
Optimal local alignment (score 4):

AC

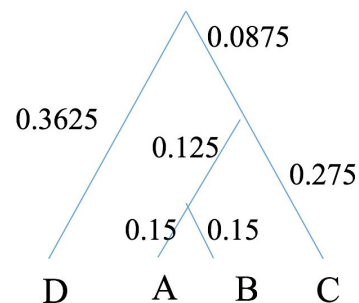
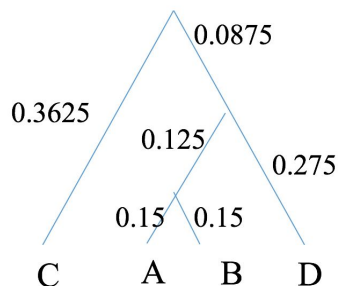
AC

SECOND:

- Making the unrooted tree rooted by setting the root at the midpoint of the path connecting the two nodes of maximum distance from each other.
- Nodes C and D are farthest from each other (0.9).
 - Therefore, the root will be $0.9 / 2 = 0.45$ (the midpoint connecting these two nodes)
 - **Newick format of this tree ((C, A), (B, D))**



THIRD: The topology of the rooted scaled tree using the same distance matrix and the UPGMA method *results in a different tree*. The Newick format is as follows (there are two versions technically). The reason that these two trees are different is because the algorithms for NJ and UPGMA are not identical. Though both use similar distance-based techniques, the UPGMA assignment of the closest pair of leaves to become neighbors does not necessarily apply to the problem. For example, UPGMA may fail to identify a pair of neighboring leaves, one of which contains an exceptionally long edge.



2 versions of the same tree via UPGMA (where C and D are switched). Newick format is (D, ((A, B), C)) and (C, ((A, B), D)).