

# Solution to Homework 2: Multiple Sequence Alignment

BCH4300B, Winter 2014  
 Assigned: March 13, 2014  
 Due: March 20, 2014

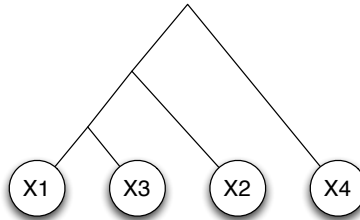
**1A)** At the beginning, sequences  $X_1$  and  $X_3$  are the most similar, with a score of 7. Thus, they would be grouped first. After this, the reduced distance matrix is:

	$X_1, X_3$	$X_2$	$X_4$
$X_1, X_3$	-	3.5	-1.5
$X_2$	-	-	1
$X_4$	-	-	-

In the next iteration, the highest similarity is between group  $X_1, X_3$  and sequence  $X_2$ . So, those are joined, and the reduced distance matrix becomes:

	$X_1, X_2, X_3$	$X_4$
$X_1, X_2, X_3$	-	$-2/3$
$X_4$	-	-

So, of course, in the final round,  $X_4$  is joined to the group  $X_1, X_2, X_3$ . The guide tree thus looks like:



**1B)** The dynamic programming table I offered to get you started on the HW2 was incorrect in the first row—for reasons I will explain in class. The correct one is below.

		C	T	G	G	-
		C	T	-	G	T
	0	1	2	0	1	-1
C	-2	9	10	8	9	7
A	-4	7	8	7	9	7
G	-6	5	6	9	16	14
G	-8	3	4	7	18	16

The optimal alignment is therefore:

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