

Homework Assignment #3 (100 points, weight 15%)
 Due: Friday October 25

Multiple alignment and phylogeny reconstruction

1. (40 marks) (multiple sequence alignment)
 Consider the following set of sequences.

$$\begin{aligned} S_1 &= \text{CAGAGATCGA} \\ S_2 &= \text{CAGGATCGA} \\ S_3 &= \text{CAGAGAGCGA} \\ S_4 &= \text{CAGAGAGCCGA} \end{aligned}$$

Use a matching score of +1 and mismatch/indel score of -1.

- (a) (10 marks) Compute the optimal **pairwise** global alignment for each pair of strings from the above list. You don't have to run the algorithm step by step, and may write the answer directly when obvious.
 - (b) (10 marks) Compute the multiple sequence alignment for $\{S_1, S_2, S_3, S_4\}$ using the **center star method**. Please show your steps.
 - (c) (20 marks) Compute the multiple sequence alignment for $\{S_1, S_2, S_3, S_4\}$ using **ClustalW**. Please show your steps.
2. (20 marks) Consider the tree topology of exercise 3 page 193, for the given taxa. Compute the parsimony length and give the labeling for internal nodes.
3. (20 marks) Construct an additive tree for the following distance matrix.

<i>M</i>	<i>S</i> ₁	<i>S</i> ₂	<i>S</i> ₃	<i>S</i> ₄	<i>S</i> ₅
<i>S</i> ₁	0	10	9	16	8
<i>S</i> ₂		0	15	22	8
<i>S</i> ₃			0	13	13
<i>S</i> ₄				0	20
<i>S</i> ₅					0

4. (20 matrix) For the following character-state matrix, construct a phylogenetic tree using the approximation algorithm in Section 7.2.1.2 of the textbook:

<i>M</i>	<i>C</i> ₁	<i>C</i> ₂	<i>C</i> ₃
<i>S</i> ₁	0	0	1
<i>S</i> ₂	1	1	0
<i>S</i> ₃	1	0	1
<i>S</i> ₄	0	1	0
<i>S</i> ₅	0	1	1