

Here is a list of six “organisms” identified by their DNA sequences:

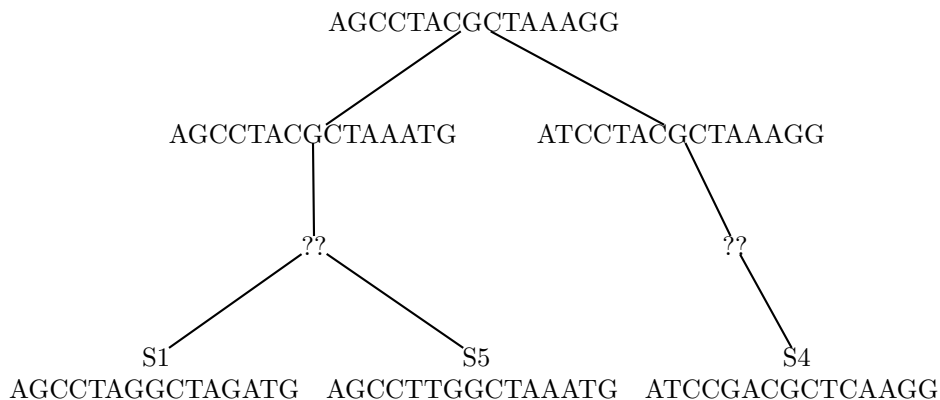
S1: AGCCTAGGCTAGATG
 S2: AGTTTACGCCAAGGA
 S3: CGTCTACGCCAAGGA
 S4: ATCCGACGCTCAAGG
 S5: AGCCTTGGCTAAATG
 S6: ATCCTACTTTAAAGG

The following is a partially-filled-out matrix of phylogenetic distances between these organisms:

	S1	S2	S3	S4	S5	S6
S1	0	?	8	6	2	6
S2	?	0	?	8	8	8
S3	8	?	0	8	8	8
S4	6	8	8	0	6	?
S5	2	8	8	6	0	6
S6	6	8	8	?	6	0

- (a) Determine the missing entries in the above matrix. Check this by showing that the distances between the three organisms S2, S3, and S4 satisfy the three properties of phylogenetic distances.

Below is a partial tree obtained by performing the first few steps of the phylogenetic tree algorithm.



- (b) Put in the DNA sequences into the tree above for the missing ancestors so as to satisfy the properties of a phylogenetic tree.
- (c) Continue the phylogenetic tree algorithm to obtain the complete phylogenetic tree. *Do not* fill in the DNA sequences for the rest of this tree.

Solution: (a) Determine the missing entries in the above matrix. Check this by showing that the distances between the three organisms S2, S3, and S4 satisfy the three properties of phylogenetic distances.

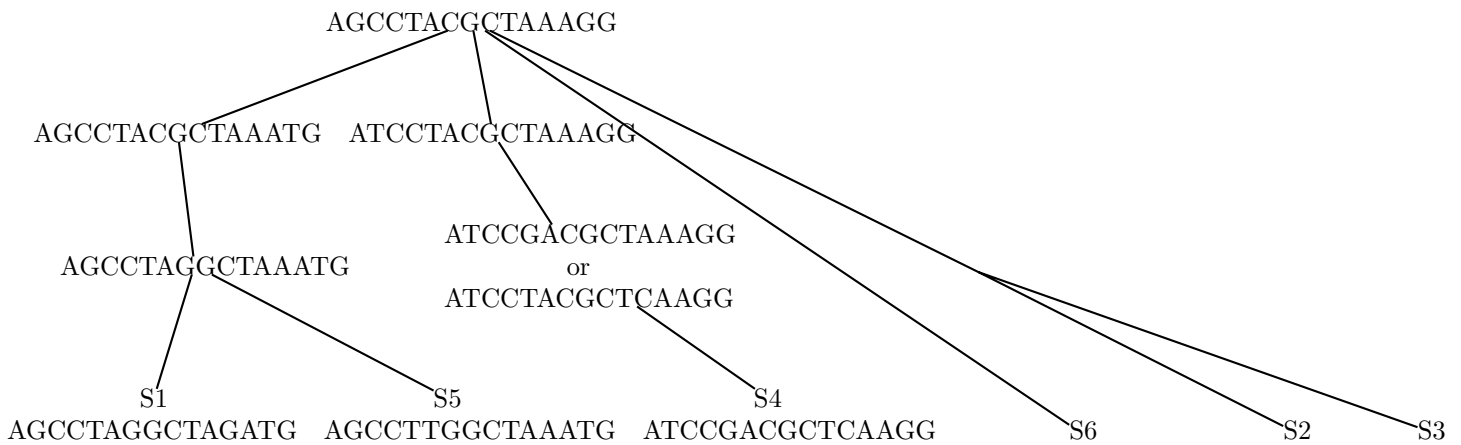
	S1	S2	S3	S4	S5	S6
S1	0	8	8	6	2	6
S2	8	0	2	8	8	8
S3	8	2	0	8	8	8
S4	6	8	8	0	6	4
S5	2	8	8	6	0	6
S6	6	8	8	4	6	0

Distance from S2 to S3: 2

Distance from S2 to S4: 8 two largest values are equal

Distance from S3 to S4: 8

Below is a partial tree obtained by performing the first few steps of the phylogenetic tree algorithm.



(b) Put in the DNA sequences into the tree above for the missing ancestors so as to satisfy the properties of a phylogenetic tree.

(c) Continue the phylogenetic tree algorithm to obtain the complete phylogenetic tree. *Do not* fill in the DNA sequences for the rest of this tree.