Advance Algorithm in Bioinformatics – CSC 8540 – Fall 2022

September 18 2022

Pardis Sadatian Moghaddam

Panther ID: 002722641

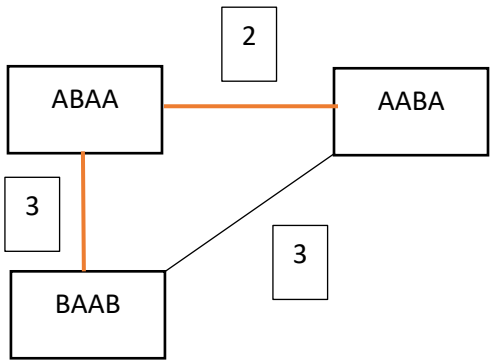
1. For the following matrix M, can you find the parsimony tree using the approximation algorithm? Is this the maximum parsimony tree? If not, what is the maximum parsimony tree?

M	C1	C2	C3	C4
S1	1	0	1	1
S2	1	1	0	1
S3	0	1	1	0

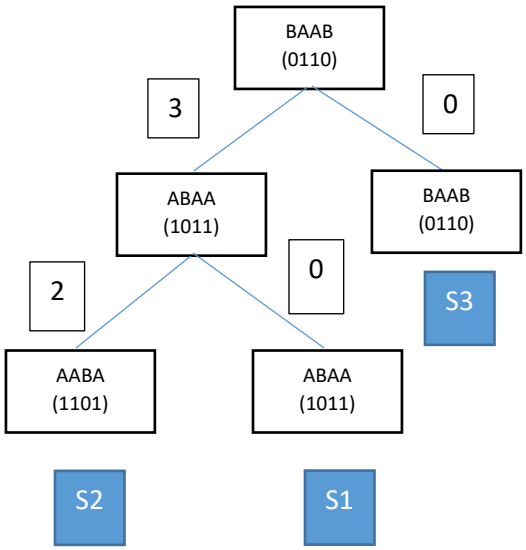
We have the matrix M of a set S of n taxa:
S = 3
We need to create a weighted complete graph G(S):
1) The vertex set is S
2) The weight of every edge (ai,aj) equauls the Hamming distance (That is, the number of changes) between the characters of ai and aj where ai,aj ∈ S.

We write down the matrix like this:

M	C1	C2	C3	C4
S1	A	B	A	A
S2	A	A	B	A
S3	B	A	A	B



We have the minimum spanning tree.



This is the maximum parsimony tree.

2. Can you solve the large compatibility problem, that is, find the largest set of characters which admit the perfect phylogeny? Can you check if the following matrix has a perfect phylogeny? If yes, can you report the corresponding tree?

M	C1	C2	C3	C4
S1	0	0	0	1
S2	1	1	0	0
S3	0	0	0	1
S4	0	1	1	0
S5	0	1	1	1

We have S of n taxa, S1 and S2 and S3 and S4 and S5.
Let Oi be the set of taxa with state 1:
O1 = {2}
O2 = {2,4,5}
O3 = {4,5}
O4 = {1,3,5}

For any pair of characters i and j, they are private compatible if Oi and Oj are disjoint or one of them contains the other.

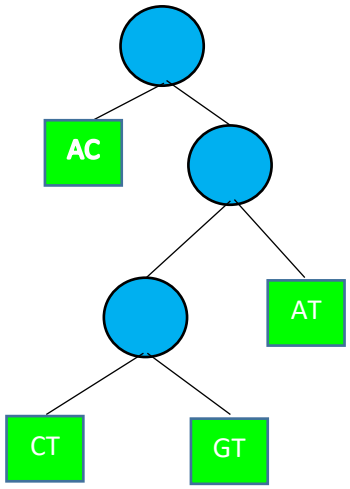
C1 and C2 are pairwise compatible since O1 contains O2.
C2 and C3 are pairwise compatible since O2 contains O3.

C1 and C3 are pairwise compatible since O1 and O3 are disjoint.
C1 and C4 are pairwise compatible since O1 and O4 are disjoint.

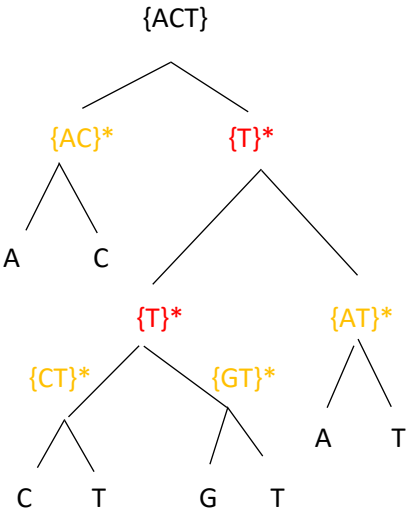
M admits a perfect phylogeny if and only if every pair of characters i and j are pairwise compatible.

C2 and C4 are not pairwise compatible. It is not the perfect phylogeny. We can't report the corresponding tree.

3. Consider the following tree topology for taxa {AC, CT, GT, AT }. Can you compute the parsimony length? Also, can you give the correspond□ing labeling for the internal nodes?



This is the small parsimony problem. We can do this by Fitch algorithm. We have this:



Parsimony Sore = Parsimony length = 4

4. For the above question, do we have another labeling for the internal nodes such that we have the same parsimony length?

In here we are labeling the internal node