

Parsimony and Compatibility in Phylogenetics

1. Problem 1

Can you find the parsimony tree using the approximation algorithm for the following matrix M ? Is this the maximum parsimony tree? If not, what is the maximum parsimony tree?

Given matrix M with $S = 3$ taxa:

$$M = \begin{pmatrix} \text{C1} & \text{C2} & \text{C3} & \text{C4} \\ \text{S1} & 1 & 0 & 1 & 1 \\ \text{S2} & 1 & 1 & 0 & 1 \\ \text{S3} & 0 & 1 & 1 & 0 \end{pmatrix}$$

Translated into character matrix:

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

Solution 1

We create a weighted complete graph $G(S)$:

1. Vertices = $\{S1, S2, S3\}$
2. Edge weights = Hamming distances between character strings

$$d(S1, S2) = 2$$

$$d(S1, S3) = 3$$

$$d(S2, S3) = 3$$

Using these weights, the Minimum Spanning Tree (MST) connects: S1–S2 (2), and S1–S3 (3).

Note for Problem 1

The MST-based approximation does not guarantee the optimal (maximum parsimony) tree. The optimal solution can be verified using character-based evaluation.

Below is the maximum parsimony tree with internal node labeling and edge lengths:

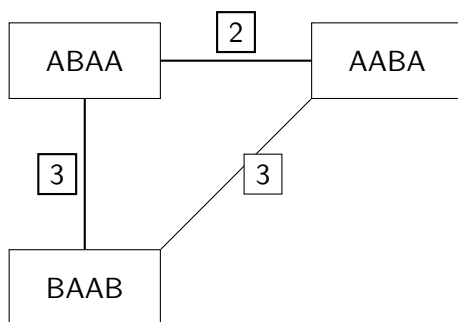


Figure 1: Minimum Spanning Tree

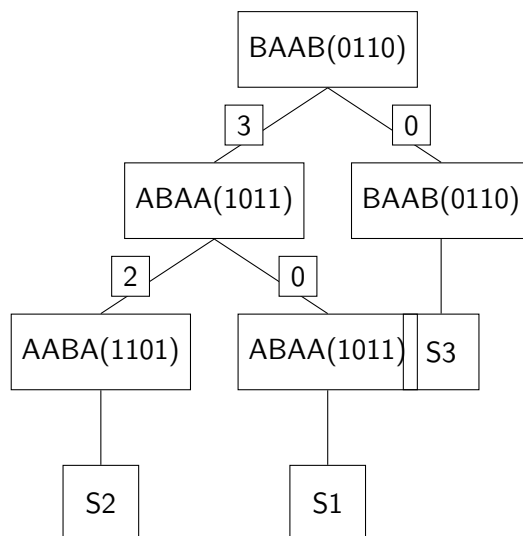


Figure 2: Maximum Parsimony Tree

2. Problem 2

Can you find the largest subset of characters that admit a perfect phylogeny? Does the matrix M admit a perfect phylogeny?

$$M = \begin{pmatrix} \text{C1} & \text{C2} & \text{C3} & \text{C4} \\ \text{S1} & 0 & 0 & 0 & 1 \\ \text{S2} & 1 & 1 & 0 & 0 \\ \text{S3} & 0 & 0 & 0 & 1 \\ \text{S4} & 0 & 1 & 1 & 0 \\ \text{S5} & 0 & 1 & 1 & 1 \end{pmatrix}$$

Solution 2

We define:

$$O_1 = \{2\}$$

$$O_2 = \{2, 4, 5\}$$

$$O_3 = \{4, 5\}$$

$$O_4 = \{1, 3, 5\}$$

Check pairwise compatibility:

1. C1 and C2: Compatible since $O_1 \subset O_2$
2. C2 and C3: Compatible since $O_3 \subset O_2$
3. C1 and C3: Compatible (disjoint)
4. C1 and C4: Compatible (disjoint)
5. C2 and C4: Not compatible (intersect but no containment)

Note for Problem 2

Matrix M does not admit a perfect phylogeny due to the incompatibility between C2 and C4. The largest compatible subset would exclude one of them—e.g., remove C4.

3. Problem 3

Can you compute the parsimony length for the following topology? Use Fitch's algorithm to assign internal node labels.

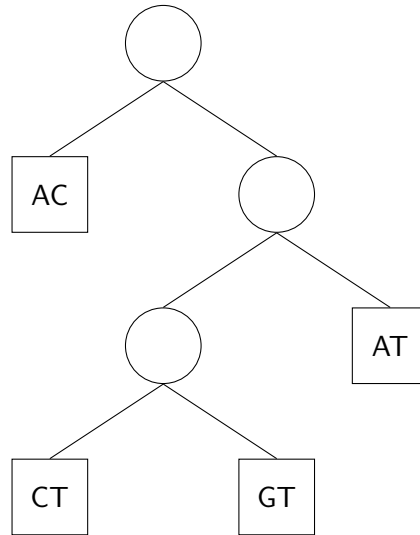


Figure 3: Tree Topology for Taxa {AC, CT, GT, AT}

Solution 3

Apply Fitch's algorithm bottom-up:

- $CT \cap GT = \{T\}$
- Internal label: T (no change)
- $(T) \cap AT = \{T\}$
- Internal label: T (no change)
- $(AC) \cap (T) = \emptyset \Rightarrow \{A, C, T\}$, increment score

Note for Problem 3

Total changes = 4 (parsimony score). See diagram below for full Fitch steps.

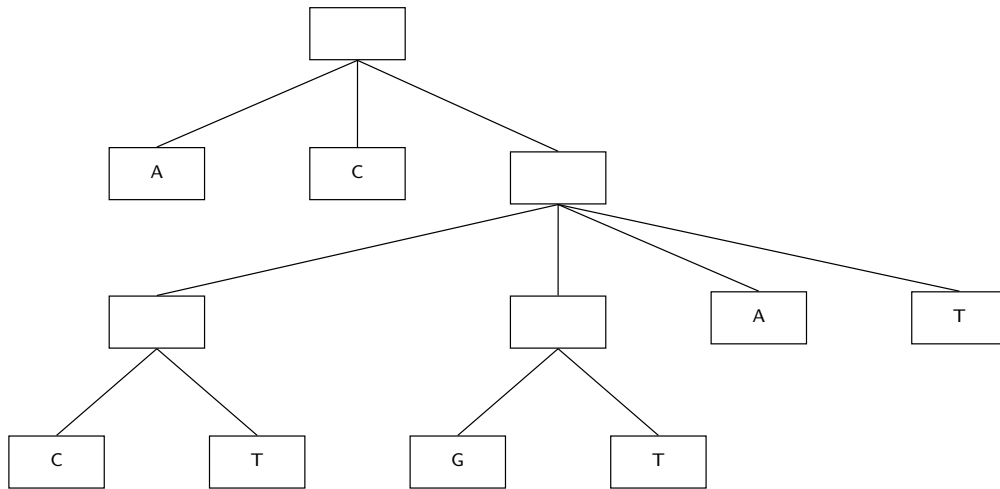


Figure 4: Fitch Steps for Parsimony Length = 4

4. Problem 4

Can we provide an alternative labeling of internal nodes that results in the same parsimony length?

Solution 4

Yes. Different internal labelings can still yield the same number of changes.

Note for Problem 4

The figure below shows an alternative labeling of internal nodes where the parsimony length remains 4.

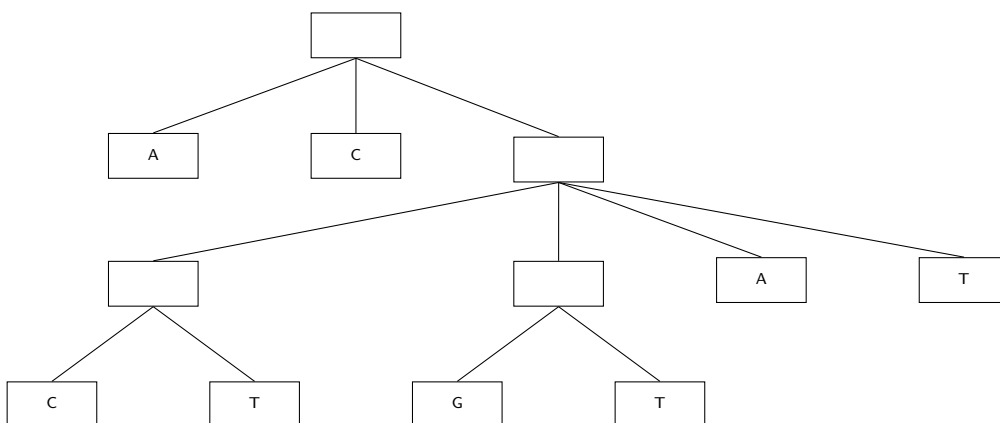


Figure 5: Alternative Internal Labeling (Parsimony Length = 4)