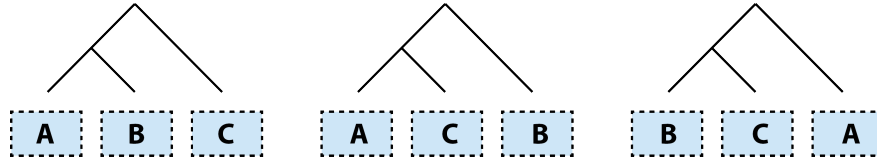# 9 Exercise solutions – Phylogenetic tree

1. **Tree topology**

   A rooted phylogenetic tree can have three topologically different trees when m is 3.

   (a) Fill the labels A, B, or C to satisfy three topologically distinct trees.

A	B	C		A	C	B		B	C	A

2. **UPGMA**

UPGMA is an unweighted version of PGMA (pair-group method using arithmetic mean) for reconstructing a phylogenetic tree. Pairwise sequence alignments are used to calculate the distances among four sequences A, B, C, and D.

	A	B	C	D
A	0	2	7	7
B		0	5	9
C			0	8
D				0

Below are two examples of the distance calculation that can be used for UPGMA.

$$d_{(\alpha\beta),\gamma} = \frac{d_{\alpha,\gamma} + d_{\beta,\gamma}}{2}, \quad d_{(\alpha\beta\gamma),\delta} = \frac{d_{\alpha,\delta} + d_{\beta,\gamma} + d_{\delta,\gamma}}{3}$$
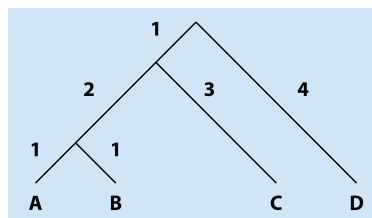
(a) Identify the first internal node and update the distance matrix.

	(AB)	C	D
(AB)	0	6	8
C		0	8
D			0

(b) Identify the second internal node and update the distance matrix accordingly.

	(ABC)	D
(ABC)	0	8
D		0

(c) Reconstrut a rooted tree from the calcualted distances.

(d) Reconstruct a rooted tree by using UPGMA.

	A	B	C	D
A	0	2	6	8
B		0	6	8
C			0	8
D				0

3. **Maximum parsimony**

The maximum parsimony uses column-wise operations of union and intersection. The number of union operations is counted to find the tree that minimizes the evolutionary changes.
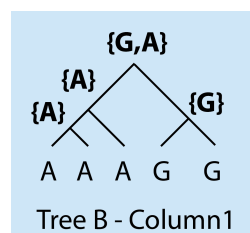

s1 ACT	
s2 GCT	
s3 ACT	
s4 GCT	
s5 ACT	
Column 1	Tree A

Tree A - Column1

s1 s5 s3 s2 s4

Tree B
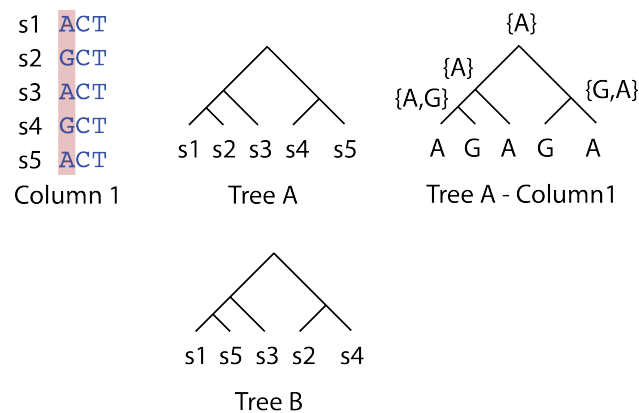
Use the first column of the MSA and also Tree A and B to answer the following questions.

(a) How many union operations are necessary to estimate the labels of the root and the internal nodes of Tree A?

Solution: 2

(b) Estimate the labels of the root and the internal nodes of Tree B.

Tree B - Column1

(c) How many union operations are necessary to estimate the root and the internal nodes of Tree B?

Solution: 1

(d) Which tree, Tree A or B, indicates less evolutionary changes when only the first column is considered?

Solution: Tree B

4. **Maximum likelihood**

The maximum likelihood is a method to find the most suitable phylogenetic tree for a given MSA. Assume that all necessary probabilities are pre-calculated.

Tree A   Tree B
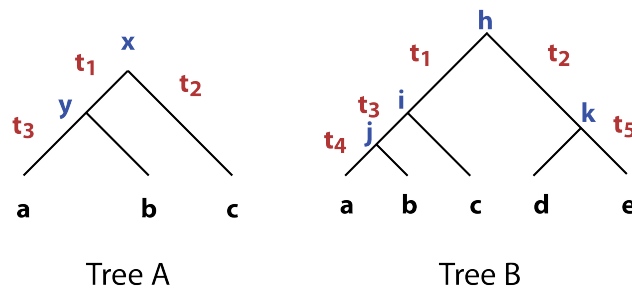
Use the first column of the MSA and also Tree A and B to answer the following questions.

Likelihood of Tree A: $L(H = TreeA|D = MSA) = p_{xy}(t1)p_{ya}(t3)p_{yb}(t3)p_{xc}(t2)$

(a) What is the likelihood of Tree B? Use the same format for Tree A above.

$L(H = TreeB|D = MSA)$

Solution: $p_{hi}(t1)p_{hk}(t2)p_{ij}(t3)p_{ja}(t4)p_{ke}(t5)p_{jb}(t4)p_{ic}(t3 + t4)p_{kd}(t5)$

(b) What is the log-likelihood of Tree A? Use base 2 and the following values for the calculation.

- $p_{xy}(t1) = 0.25$
- $p_{ya}(t3) = 0.0625$
- $p_{yb}(t3) = 0.0625$
- $p_{xc}(t2) = 0.125$

- $2^{-2} = 0.25$
- $2^{-3} = 0.125$
- $2^{-4} = 0.0625$

$L(H = TreeA|D = MSA)$

Solution:
$\log_2 p_{xy}(t1) + \log_2 p_{ya}(t3) + \log_2 p_{yb}(t3) + \log_2 p_{xc}(t2)$
$= \log_2 0.25 + \log_2 0.0625 + \log_2 0.0625 + \log_2 0.125$
$= -2 - 4 - 4 - 3$
$= -13$