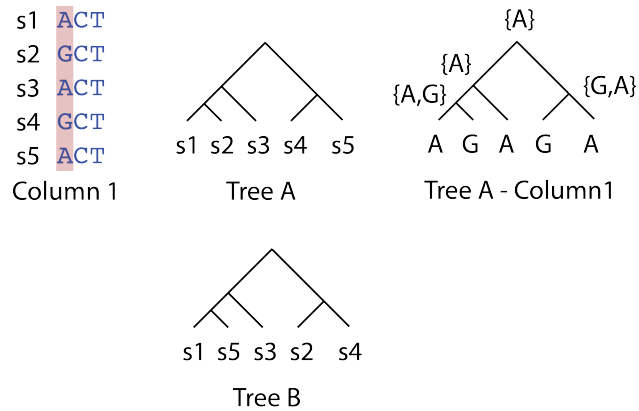# INF281 Exercise 07 solutions

1. **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.
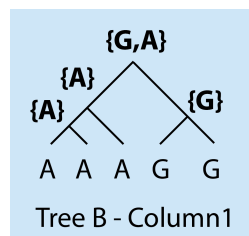

   Use the first column of the MSA and also Tree A and B to answer the following questions.

   (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.

   (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

2. **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
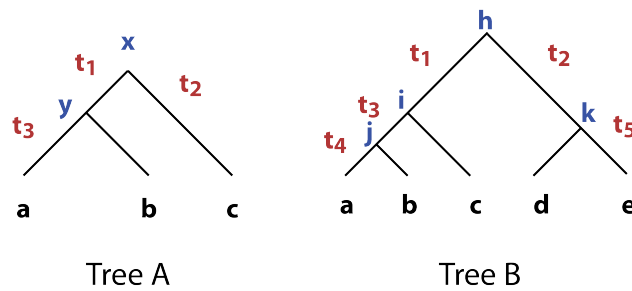
Use the first column of the MSA and also Tree A and B to answer the following questions.

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

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

3. **Linkage clustering method for progressive alignment**

Select two alignments from the three alignments, $\mathcal{A}^1 = s^1$, $\mathcal{A}^2 = s^2, s^3$, and $\mathcal{A}^3 = s^4, s^5$, for clustering.

Pairwise scores

	s^1	s^2	s^3	s^4	s^5
s^1	-	2	3	1	6
s^2		-	-	4	5
s^3			-	3	4
s^4				-	-
s^5					-

(a) Use the average linkage.

Solution: $\mathcal{A}^2$ and $\mathcal{A}^3$: $S(\mathcal{A}^1, \mathcal{A}^2) = 2.5$, $S(\mathcal{A}^1, \mathcal{A}^3) = 3.5$, $S(\mathcal{A}^2, \mathcal{A}^3) = 4$
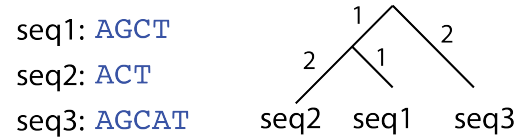
(b) Use the maximum linkage.

Solution: $\mathcal{A}^1$ and $\mathcal{A}^3$: $S(\mathcal{A}^1, \mathcal{A}^2) = 3$, $S(\mathcal{A}^1, \mathcal{A}^3) = 6$, $S(\mathcal{A}^2, \mathcal{A}^3) = 5$

(c) Use the minimum linkage.

Solution: $\mathcal{A}^2$ and $\mathcal{A}^3$: $S(\mathcal{A}^1, \mathcal{A}^2) = 2$, $S(\mathcal{A}^1, \mathcal{A}^3) = 1$, $S(\mathcal{A}^2, \mathcal{A}^3) = 3$

4. **Linear progressive alignment**

Construct an MSA from seq1, seq2, seq3 and a phylogenetic tree by using the progressive alignment method specified below.

<div align="center">
seq1: AGCT

seq2: ACT

seq3: AGCAT
</div>

- Clustering: Linear clustering
- Aligning method: Pair-guided alignment
- Aligning order: Use the specified tree
- Pairwise DP: Global alignment with linear gap penalty
- DP scoring scheme: match (10), mismatch (-5), gap penalty (10)

(a) What is the aligning order that can be defined by the given tree?

> **Solution:**
>
> ```
> 1: Seq2 & Seq1
> 2: Seq1 & Seq3
> ```

(b) Solve the first pairwise alignment.

> **Solution:**
>
> ```
> Seq2: A-CT
> Seq1: AGCT
> ```

(c) Solve the second pairwise alignment.

> **Solution:**
>
> ```
> Seq1: AGC-T
> Seq3: AGCAT
> ```

(d) Find the optimal MSA by combining the first and the second alignments.

> **Solution:**
>
> ```
> Seq1: AGC-T
> Seq2: A-C-T
> Seq3: AGCAT
> ```

(e) What is the SP score of the optimal MSA?

> **Solution:**
>
> ```
> Seq1 & Seq2: 20
> Seq1 & Seq3: 30
> Seq2 & Seq3: 10
>
> SP: 60
> ```