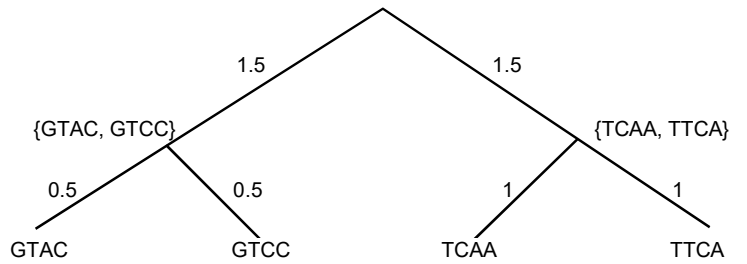


Examples

E1. In the following phylogenetic tree;

- a. If the distance between GTAC and GTCC is 1, what will be the distance between cluster "{GTAC, GTCC}" and cluster "{TCAA, TTCA}"? Explain how to find.
- b. Calculate weights of each gene.



E2. In **Expectation-Maximization** algorithm there are two steps: Expectation step and Maximization step.

In the expectation step, background residue frequencies are calculated *based on those residues that are not in the initially aligned sites (calculated as the ratio of the number of a specific residue in the background and total number of background residues)*.

Frequency of column specific residues is calculated for each position in the initial motif alignment (calculated as the ratio of number of residues at each position in the motifs and number of sequences).

An initial, random alignment for 4 sequences are given here. The motif we are searching for is 4 bases wide (motifs at each sequence are highlighted). In this case;

- a. Fill in the observed residue counts table by calculating background and observed residue counts of each residue for initial alignment.
- b. Using the count information table obtained in (a), fill in the residue frequency table.

C	A	G	T	T	A	T	A	A	A
A	G	C	C	G	C	C	C	T	C
T	G	A	G	G	G	A	G	C	A
G	G	T	A	C	T	G	C	T	G

a. Observed residue counts of initial alignment

Nucleotide	Motif position (0 = Background)				
	0	1	2	3	4
A					
C	6				
G				2	
T					

b. Residue frequencies of initial alignment

Nucleotide	Motif position (0 = Background)				
	0	1	2	3	4
A					
C	0.25				
G				0.5	
T					

=====

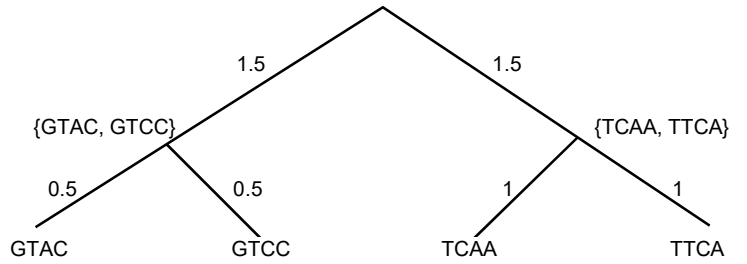
E3. The following table shows a pairwise distance matrix for sequences A, B, C, D, and E.
Plot the phylogenetic tree by using a simple clustering procedure to derive phylogenetic tree.

Seq.	A	B	C	D	E
A	0	9	8	12	15
B	9	0	11	15	18
C	8	11	0	10	13
D	12	15	10	0	5
E	15	18	13	5	0

Example Solutions

E1. In the following phylogenetic tree;

- c. If the distance between GTAC and GTCC is 1, what will be the distance between cluster "{GTAC, GTCC}" and cluster "{TCAA, TTCA}"? Explain how to find. (07)
- d. Calculate weights of each gene. (08)



- a. If the distance between GTAC and GTCC is 3, which is $(0.5 + 0.5)$, distance between clusters "{GTAC,GTCC}" and "{TCAA,TTCA}" will be 3, which is $(1.5 + 1.5)$.
- b.
 - $W_{GTAC} = 0.5 + (1.5/2) = 1.25$
 - $W_{GTCC} = 0.5 + (1.5/2) = 1.25$
 - $W_{TCAA} = 1.0 + (1.5/2) = 1.75$
 - $W_{TTCA} = 1.0 + (1.5/2) = 1.75$

E2. In **Expectation-Maximization** algorithm there are two steps: Expectation step and Maximization step.

In the expectation step, background residue frequencies are calculated *based on those residues that are not in the initially aligned sites* (calculated as the ratio of the number of a specific residue in the background and total number of background residues).

Frequency of column specific residues is calculated for each position in the initial motif alignment (calculated as the ratio of number of residues at each position in the motifs and number of sequences).

An initial, random alignment for 4 sequences are given here. The motif we are searching for is 4 bases wide (motifs at each sequence are highlighted). In this case;

C	A	G	T	T	A	T	A	A	A
A	G	C	C	G	C	C	C	T	C
T	G	A	G	G	G	A	G	C	A
G	G	T	A	C	T	G	C	T	G

- c. Fill in the observed residue counts table by calculating background and observed residue counts of each residue for initial alignment. (07)
- d. Using the count information table obtained in (a), fill in the residue frequency table. (08)

b. Observed residue counts of initial alignment

Nucleotide	Motif position (0 = Background)				
	0	1	2	3	4
A	7	1	1	0	1
C	6	1	1	1	1
G	7	1	1	2	1
T	4	1	1	1	1

b. Residue frequencies of initial alignment

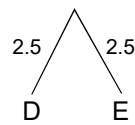
Nucleotide	Motif position (0 = Background)				
	0	1	2	3	4
A	0.29	0.25	0.25	0	0.25
C	0.25	0.25	0.25	0.25	0.25
G	0.29	0.25	0.25	0.5	0.25
T	0.17	0.25	0.25	0.25	0.25

E3. The following table shows a pairwise distance matrix for sequences A, B, C, D, and E.
 Plot the phylogenetic tree by using a simple clustering procedure to derive phylogenetic tree. (20)

Seq.	A	B	C	D	E
A	0	9	8	12	15
B	9	0	11	15	18
C	8	11	0	10	13
D	12	15	10	0	5
E	15	18	13	5	0

Seq.	A	B	C	D	E
A	0	9	8	12	15
B	9	0	11	15	18
C	8	11	0	10	13
D	12	15	10	0	5
E	15	18	13	5	0

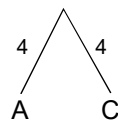
The smallest distance is 5 between D and E. The first cluster is (DE). New distance matrix and cluster are given below



Seq.	A	B	C	DE
A	0	9	8	13.5
B	9	0	11	16.5
C	8	11	0	11.5
DE	$\frac{12+15}{2} = 13.5$	$\frac{15+18}{2} = 16.5$	$\frac{10+13}{2} = 11.5$	0

Seq.	A	B	C	DE
A	0	9	8	13.5
B	9	0	11	16.5
C	8	11	0	11.5
DE	13.5	16.5	11.5	0

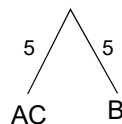
The smallest distance is 8 between A and C. The second cluster is (AC). New distance matrix and cluster are given below



Seq.	B	AC	DE
B	0	10	16.5
AC	$\frac{9+11}{2} = 10$	0	$\frac{13.5+11.5}{2} = 12.5$
DE	16.5	12.5	0

Seq.	B	AC	DE
B	0	10	16.5
AC	10	0	12.5
DE	16.5	12.5	0

The smallest distance is 10 between (AC) and B. The third cluster is (AC)B. New distance matrix is given below



Seq.	(AC)B	DE
(AC)B	0	14.5
DE	$\frac{16.5+12.5}{2} = 14.5$	0

Final phylogenetic tree is

