

1)

D	A	B	C	D	
A	0	1	0.5	1.3	2.8
B	1	0	1.1	1.9	4
C	0.5	1.1	0	1.2	2.8
D	1.3	1.9	1.2	0	4.4

D*	A	B	C	D
A	0	-4.8	-4.6	-4.6
B	-4.8	0	-4.6	-4.6
C	-4.6	-4.6	0	-4.8
D	-4.6	-4.6	-4.8	0

D	m	C	D
m	0	0.3	1.1
C	0.3	0	1.2
D	1.1	1.2	0

D*	m	C	D
m	0	-2.6	-2.6
C	-2.6	0	-2.6
D	-2.6	-2.6	0

D	m	n
m	0	0.1
n	0.1	0

m 0.1 n

m 0.1 n 0.2 C
\1.0 D

A 0.2 m 0.1 n 0.2 C
B 0.8 / \1.0 D

The matrix is additive.

- Yes, because by definition, the limb length of a leaf is its distance to its parent. If the distance between 2 leaves is the sum of the limb lengths, then they must share a parent, and the matrix must be additive since the limb lengths add up to the distance between the 2 leaves.
- The additive phylogeny algorithm has runtime of $O(m^2)$ since it loops over every pair of leaves, while the neighbor-joining algorithm has a runtime of $O(m)$ since it loops over every leaf only twice.
- No, as the algorithm simply counts the number of characters and returns the character that occurs the most, which may not give the optimal answer all the time. Also, the recursive call of the right child is made to CountCharacter instead of FuzzySmallParsimony.
- Use the FarthestFirstTraversal to assign the k centers. Then, for each pair of points in n that are in different clusters, we find the distance between them. Then, we check to see if any distances are repeated, if a distance is repeated there cannot be a solution.
- Trivially, we can see that for a cluster of 2 points, the maximum distance between a point and a center will be the distance between those centers, and with an optimal center, the maximum distance will be half that distance. When we move to 3 points, if we assume that the 3 points are on a line, the maximum distance in the first case will be the distance between the 2 outermost points.

For the optimal center, if 1 point is equidistant from the other 2, the maximum distance is exactly half that of the distance between the 2 outermost points. If the center point moves closer to one point, so will the center, and the maximum distance from a point to the center will increase, thus still fulfilling the equation.

- 7) Start with k randomly assigned centers. Then, we find the distances of each point from each center, and assign each point to a cluster and find the center of that cluster. Then, we recur until we converge to an answer.
- 8) $0.6^8(1 - 0.6)^2 = 0.0026873856$
 $0.5^8(1 - 0.5)^2 = 0.0009765625$
- 9) Yes, as if the distance between clusters is the minimum distance between 2 points from different clusters, any pairs of points with a smaller distance than that must not be in different clusters as that distance would then be the minimum distance, thus they are in the same cluster.
- 10) Suffix array:

Sorted suffixes	Position
\$	11
i\$	10
ippi\$	7
issippi\$	4
ississippi\$	1
mississippi\$	0
pi\$	9
ppi\$	8
sippi\$	6
sissippi\$	3
ssippi\$	5
ssissippi\$	2

