

Bioinformatics Summative Assignment 2018-19

clvp22

March 18, 2019

1 Question 2

1. We begin with the following distance matrix,

	a	b	c	d	e	f
a	0	15	24	29	25	37
b	15	0	32	31	23	43
c	24	32	0	30	43	49
d	29	31	30	0	45	57
e	25	23	43	45	0	55
f	37	43	49	57	55	0

2. Clearly species a and b are the closest pair of species so we cluster these species together to get a new species ab and the following matrix,

	ab	c	d	e	f
ab	0	28	30	24	40
c	28	0	30	43	49
d	30	30	0	45	57
e	24	43	45	0	55
f	40	49	57	55	0

3. We can begin to construct our phylogenetic tree as follows,

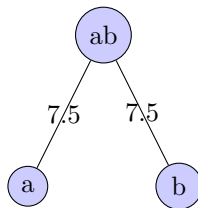


Figure 1: Phylogenetic Tree

4. Now species ab and e are the closest pair of species in our matrix so we reduce the matrix by creating a new cluster of species abe . Because we are following the UPGMA algorithm we must use the following formula to calculate the distances,

$$d(abe, x) = \frac{2 \cdot d(ab, x) + 1 \cdot d(e, x)}{2 + 1}$$

for some x .

Consequently we get the following matrix,

	abe	c	d	f
abe	0	33	35	45
c	33	0	30	49
d	35	30	0	57
f	45	49	57	0

5. We continue to build our phylogenetic tree,

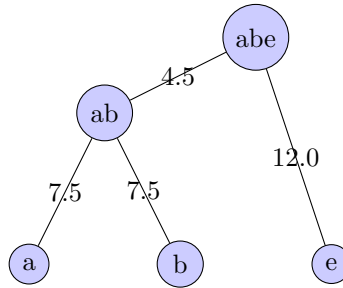


Figure 2: Phylogenetic Tree

6. We reduce our matrix again by clustering the new closest pair of species c and d into cd to get the following matrix,

	abe	cd	f
abe	0	34	45
cd	34	0	53
f	45	53	0

7. We continue to build our phylogenetic tree,

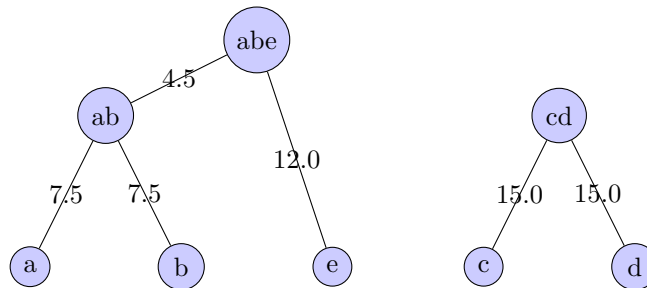


Figure 3: Phylogenetic Tree

8. We now reduce our matrix by clustering the clusters *abe* and *cd*, again we are following the UPGMA algorithm so,

$$d(abecd, f) = \frac{3 \cdot d(abe, f) + 2 \cdot d(cd, f)}{3 + 2}$$

. We get the following matrix,

	abecd	f
abecd	0	48.2
f	48.2	0

9. We continue to build our phylogenetic tree,

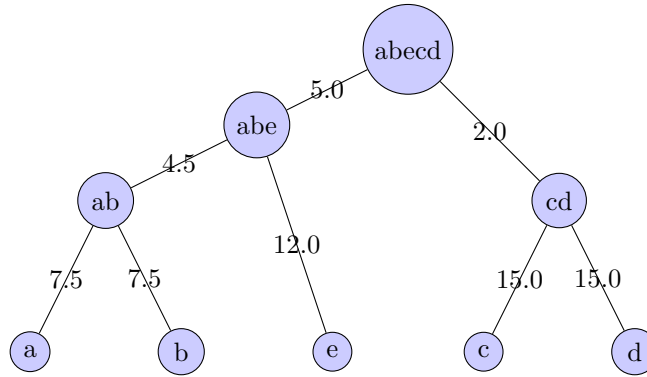


Figure 4: Phylogenetic Tree

10. Finally, we need not reduce our matrix any further, because we can simply add our remaining species *f* to the phylogenetic tree as follows to get our solution,

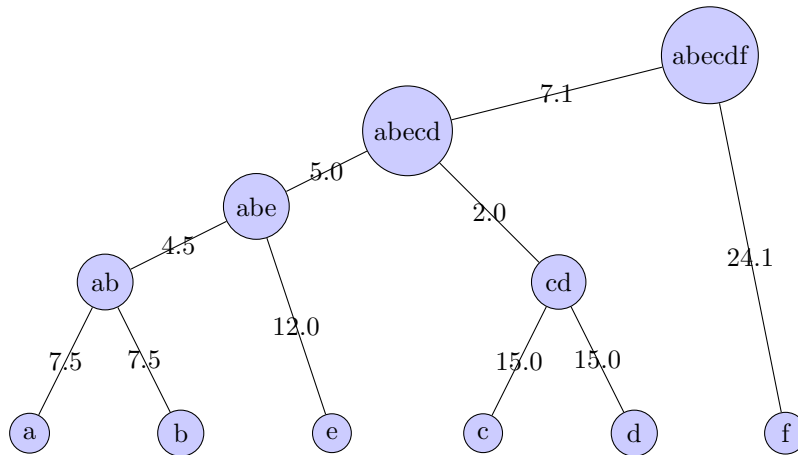


Figure 5: Final Phylogenetic Tree