

# Molecular Phylogenetics

Dr. Krishna K Ojha

Department of Bioinformatics

Central University of South Bihar

# Tree Building using UPGMA

- UPGMA is an acronym of unweighted pair-group method with arithmetic mean.
- This is a Distance based method of phylogenetic tree construction.
- It is very simple and handy and very fast.
- The great disadvantage of UPGMA is that it assumes the same evolutionary speed on all lineages, i.e. the rate of mutations is constant over time and for all lineages in the tree. This is called a 'molecular clock hypothesis'.
- This would mean that all leaves (terminal nodes) have the same distance from the root. In reality the individual branches are very unlikely to have the same mutation rate. Therefore, *UPGMA frequently generates wrong tree topologies.*
- Generates rooted trees (re-rooting is not allowed).
- Generates ultrametric trees.

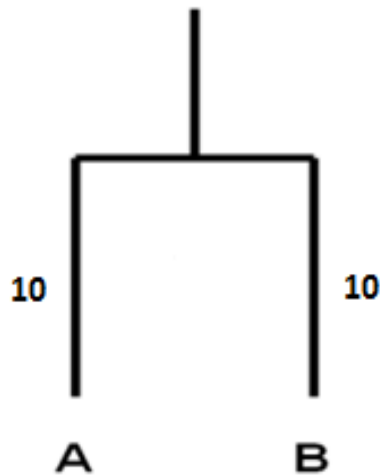
# UPGMA Method steps

- Prepare a distance matrix  $q$
- Repeat step 1 and step 2 until there are only two clusters  $q$  .
- **Step 1:** Cluster a pair of leaves (taxa) by shortest distance  $q$ .
- **Step 2:** Recalculate a new average distance with the new cluster and other taxa, and make a new distance matrix .

# Working Example of UPGMA

	A	B	C	D	E
A	0				
B	20	0			
C	60	50	0		
D	100	90	40	0	
E	90	80	50	30	0

- Suppose we have following distance matrix of 5 taxa.
- We have to find the two taxon pairs which have the smallest distance which is A and B in this case having distance 20.
- In next step we have to merge A and B and have to calculate the pair wise distance from all other taxon with (A,B)



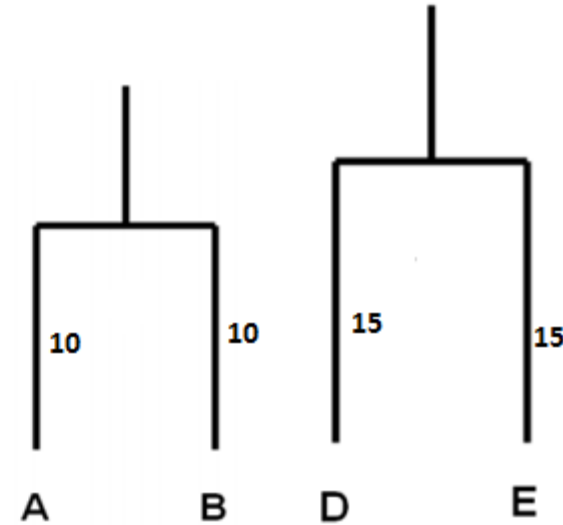
- We will combine A and B in a binary clad with a common parents.
- The distance from parents to A and B will be half to the distance between A and B. i.e  $20/2 = 10$
- Now we have to create next matrix from the parent one in which A & B will be fused together.

	AB	C	D	E
AB	0			
C	55	0		
D	95	40	0	
E	85	50	30	0

- Distance from AB to all other will be calculated using following method.
- New average distance between AB and C is:
- Distance  $(AC + BC) / 2$
- $= (60 + 50) / 2 = 55$
- Distance between D to AB is:  
D to AB =  $(100 + 90) / 2 = 95$
- Distance between E to AB is:  
E to AB =  $(90 + 80) / 2 = 85$

Again we will see the minimum distance between two group which is between D & E (30).

	AB	C	D	E
AB	0			
C	55	0		
D	95	40	0	
E	85	50	30	0

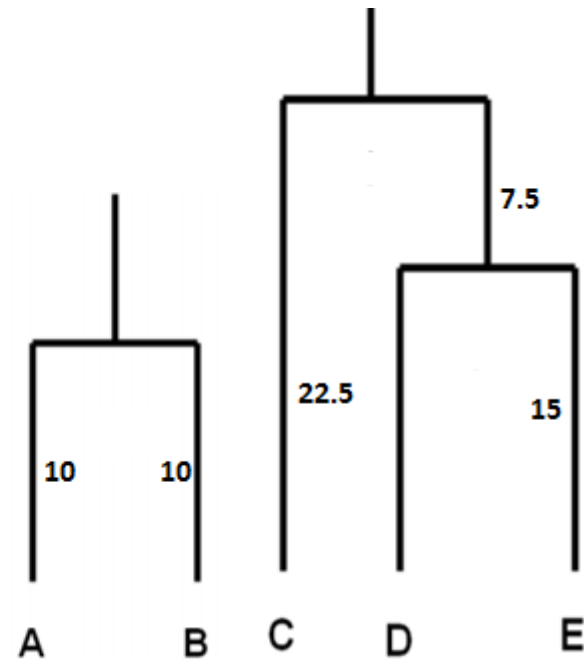


We will merge D&E together in one group, so next matrix will be { (AB), C (DE)} of size 3 X 3.

The distance of D and E are joined from a common parents so the distance form parents to D and E will be  $30/2 = 15$

Again we will see the minimum distance between two group which is between C & DE (45).

	AB	C	DE
AB	0		
C	55	0	
DE	90	45	0



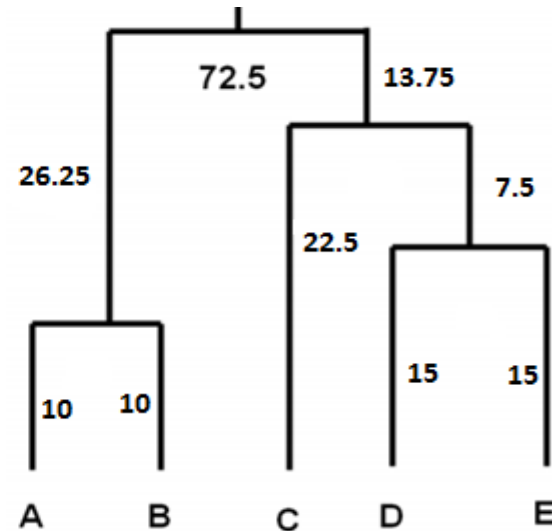
We will join C and DE together form a common parent, the distance from parent to C and D will be equal that is  $45/2=22.5$

Because DE already have 15 unit distance so the remaining distance will be  $22.5-15=7.5$



Finally we have a 2 X 2 matrix and we will end at this stage we have to merge  $\{(AB)\}$  with  $\{(CDE)\}$

	AB	CDE
AB	0	
CDE	72.5	0



□ There are only two clusters. so this completes the calculation!

The distance between  $\{(AB)\}$  with  $\{(CDE)\}$  will be divided in equal two half and adjusted as in previous step so that from both side it becomes  $72.5/2 = 36.25$

# Conclusion

- UPGMA Produce rooted , Ultrametric, binary tree.
- It follows Molecular clock Hypothesis.
- All branch of the tree terminate at same height it means rate of evolution is constant along all lineages.