

UPGMA

- Unweighted Pair Group Method using Arithmetic average
- Sequential Clustering Algorithm
- Join two nodes with minimum distance to create a new node
- Branch length =
 - height(top node) – height(bottom node)
- Distance $d_{i,j}$ between clusters C_i and C_j

$$d_{i,j} = \frac{1}{|C_i||C_j|} \sum_{p \in C_i, q \in C_j} d_{p,q}$$

The UPGMA Algorithm

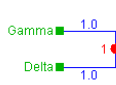
- Initialization**
 - Assign each sequence i to its own cluster C_i ,
 - Define one leaf of T for each sequence; place at height zero.
- Iteration** While more than two clusters, do
 - Determine the two clusters C_i, C_j for which d_{ij} is minimal.
 - Define a new cluster $C_k = C_i \cup C_j$; compute d_{kl} for all l .
 - Define a node k with children i and j ; place it at height $d_{ij}/2$.
 - Replace clusters C_i and C_j with C_k .
- Termination**
 - Join last two clusters, C_i and C_j ; place the root at height $d_{ij}/2$.

UPGMA Example (1)


	Alpha	Beta	Gamma	Delta
Alpha	0	4	3	4
Beta	4	0	6	5
Gamma	3	6	0	2
Delta	4	5	2	0

Select the pair of species with the smallest distance

UPGMA Example (2)



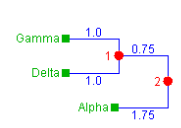
	Alpha	Beta	Gamma	Delta
Alpha	0	4	3	4
Beta	4	0	6	5
Gamma	3	6	0	2
Delta	4	5	2	0



	Alpha	Beta	I
Alpha	0	4	3.5
Beta	4	0	5.5
I	3.5	5.5	0

- $d_{Alpha, I} = (3 + 4) \div 2 = 3.5$
- $d_{Beta, I} = (6 + 5) \div 2 = 5.5$

UPGMA Example (3)



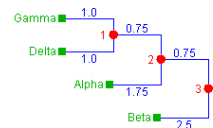
	Alpha	Beta	I
Alpha	0	4	3.5
Beta	4	0	5.5
I	3.5	5.5	0

Tree not drawn to scale

- $d_{Beta, 2} = (1 \times 4 + 2 \times 5.5) \div (1 \times 3) = 5$

	Beta	2
Beta	0	5
2	5	0

UPGMA Example (4)



- Obtain the final phylogenetic tree