Building Phylogenetic Trees

UPGMA & NJ

UPGMA

UPGMA

Unweighted Pair-Group Method with Arithmetic mean

- Unweighted = all pairwise distances contribute equally.
- Pair-Group = groups are combined in pairs.
- Arithmetic mean = pairwise distances to each group (clade) are mean distances to all members of that group.

Sokal R &Michener C (1958). A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38:1409-1438.

UPGMA: Principle

UPGMA

Principle

A B C D E

Start with unjoined ndoes and
a pair-wise distance matrix

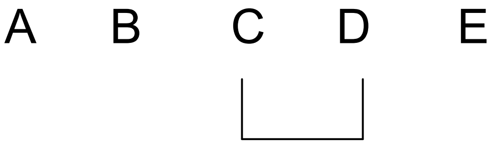
	A	B	C	D	E
A	-				
B	$d_{A,B}$	-			
C	$d_{A,C}$	$d_{B,C}$	-		
D	$d_{A,D}$	$d_{B,D}$	$d_{C,D}$	-	
E	$d_{A,E}$	$d_{B,E}$	$d_{C,E}$	$d_{D,E}$	-

Find the 2 nodes with the
shortest distance (here: C+D)

Join the 2 nodes

→

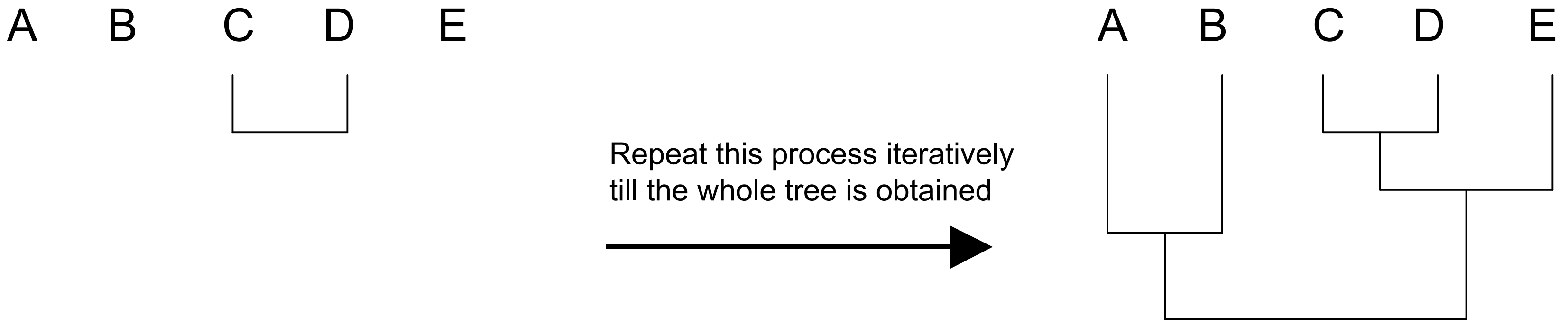
Compute the branch lengths
($d_{CD,A}$, $d_{CD,B}$, $d_{CD,E}$)



UPGMA: Principle

UPGMA

Principle



Repeat this process iteratively
till the whole tree is obtained

UPGMA: Example

	A	B	C	D	E	F	G
A	-						
B	19	-					
C	27	31	-				
D	8	18	26	-			
E	33	36	41	31	-		
F	18	1	32	17	35	-	
G	13	13	29	14	28	12	-

Distance matrix

(can be obtained from pair-wise sequence alignments)

The following example is from Dr Richard J. Edwards
http://www.southampton.ac.uk/~re1u06/teaching/upgma/

UPGMA: Example

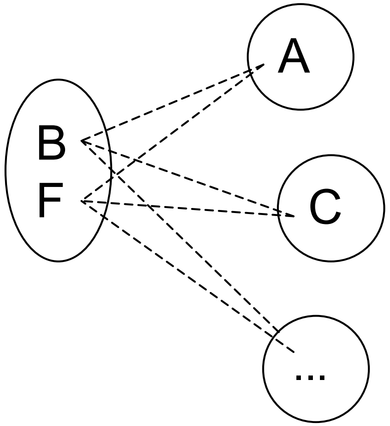
	A	B	C	D	E	F	G
A	-						
B	19	-					
C	27	31	-				
D	8	18	26	-			
E	33	36	41	31	-		
F	18	1	32	17	35	-	
G	13	13	29	14	28	12	-

- Find the shortest distance. Here the shortest distance is 1 (between B and F)
- Join the "nodes" (sequences) with the shortest distance: Here we join B and F to create node BF.
- Depth of the new branch = 1/2 of the shortest distance (so that the node-to-node path length is equal to the shortest distance). Here: $d_{BF}/2 = 0.5$.

UPGMA: Example

	A	BF	C	D	E	F	G
A	-						
BF	?	-					
C	27	?	-				
D	8	?	26	-			
E	33	?	41	31	-		
F	18	-	32	17	35	-	
G	13	?	29	14	28	12	-

- Calculate mean pairwise distances with the other nodes (sequences)



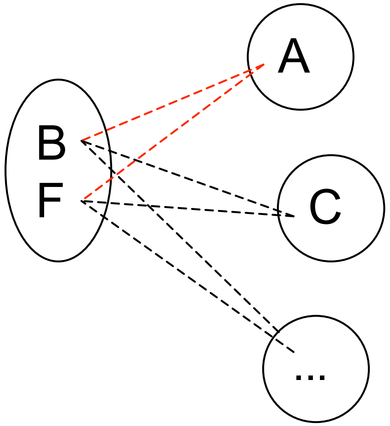
UPGMA: Example

	A	BF	C	D	E	F	G
A	-						
BF	**18.5**	-					
C	27	31.5	-				
D	8	17.5	26	-			
E	33	35.5	41	31	-		
F	18	-	32	17	35	-	
G	13	12.5	29	14	28	12	-

- Calculate mean pairwise distances with the other nodes (sequences)

Example

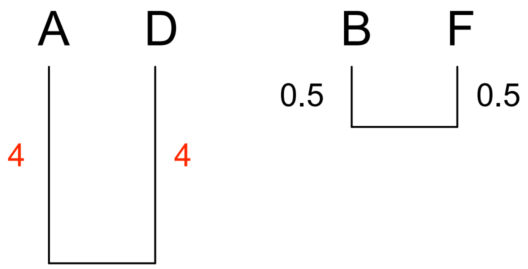
$d_{BF,A} = (d_{B,A} + d_{F,A}) / 2 = (19 + 18) / 2 = 18.5$



UPGMA: Example

	A	BF	C	D	E	G
A	-					
BF	18.5	-				
C	27	31.5	-			
D	8	17.5	26	-		
E	33	35.5	41	31	-	
G	13	12.5	29	14	28	-

- Repeat cycle with new shortest distances. Here, the next shortest distance is 8 (between A and D). We thus join A and D with branch length = 8 / 2 = 4.



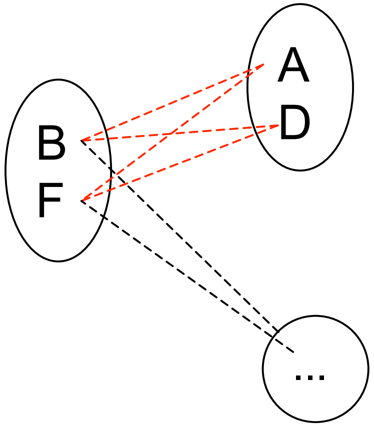
UPGMA: Example

	AD	BF	C	D	E	G
AD	-					
BF	18	-				
C	26.5	31.5	-			
D	8	17.5	26	-		
E	32	35.5	41	31	-	
G	13.5	12.5	29	14	28	-

- We join the closest nodes/groups and we recalculate the distances between nodes/groups.

Example

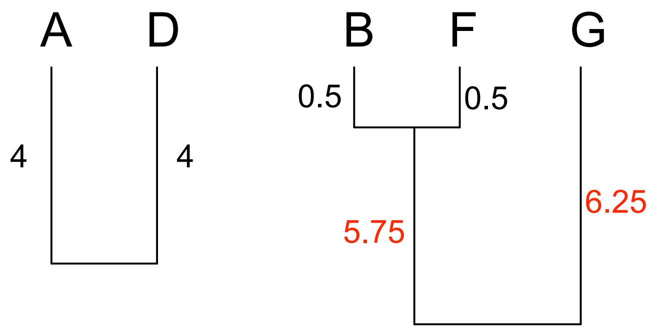
$$d_{BF,AD} = (d_{B,A} + d_{F,A} + d_{B,D} + d_{F,D}) / 4 =$$
$$= (19 + 18 + 18 + 17) / 4 = 18$$



UPGMA: Example

	AD	**BF**	**C**	**E**	**G**
AD	-				
BF	18	-			
C	26.5	31.5	-		
E	32	35.5	41	-	
G	13.5	12.5	29	28	-

- Repeat cycle with new shortest distances. Here, the next shortest distance is 12.5 (between BF and G). We thus join BF and G with branch length = 12.5 / 2 = 6.25.



UPGMA: Example

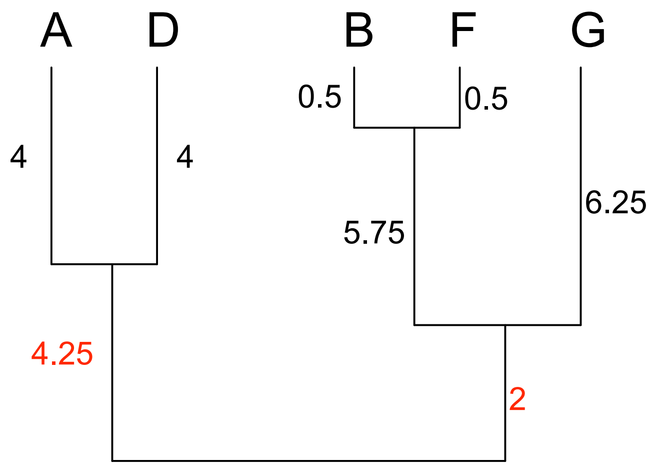
	AD	BFG	C	E	G
AD	-				
BFG	16.5	-			
C	26.5	30.67	-		
E	32	33.0	41	-	
G	13.5	12.5	29	28	-

- The distances between nodes/groups are recalculated.

UPGMA: Example

	AD	BFG	C	E
AD	-			
BFG	16.5	-		
C	26.5	30.67	-	
E	32	33.0	41	-

- The shortest disance is recalculated, the nodes/groups are joined and the branch length is calculated.

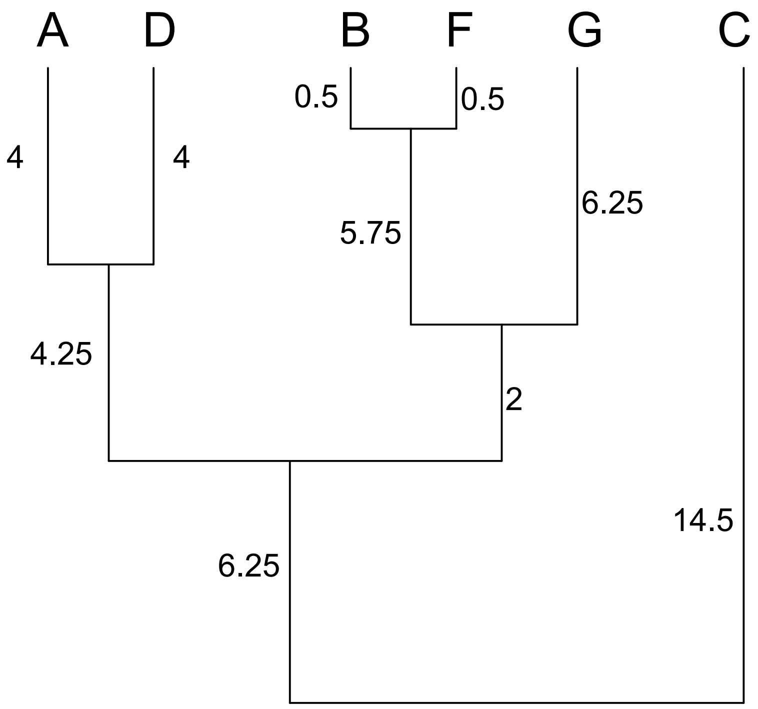


UPGMA: Example

	ADBFG	BFG	C	E
ADBFG	-			
BFG	16.5	-		
C	29	30.67	-	
E	32.6	33.0	41	-

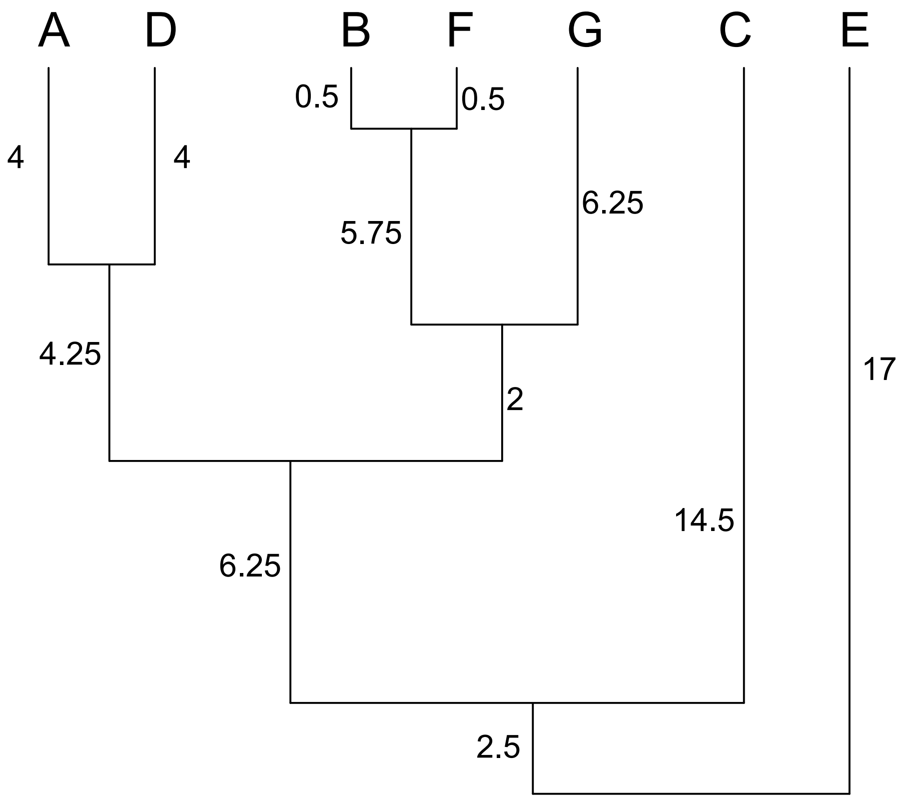
UPGMA: Example

	ADBFG	C	E
ADBFG	-		
C	29	-	
E	32.6	41	-



UPGMA: Example

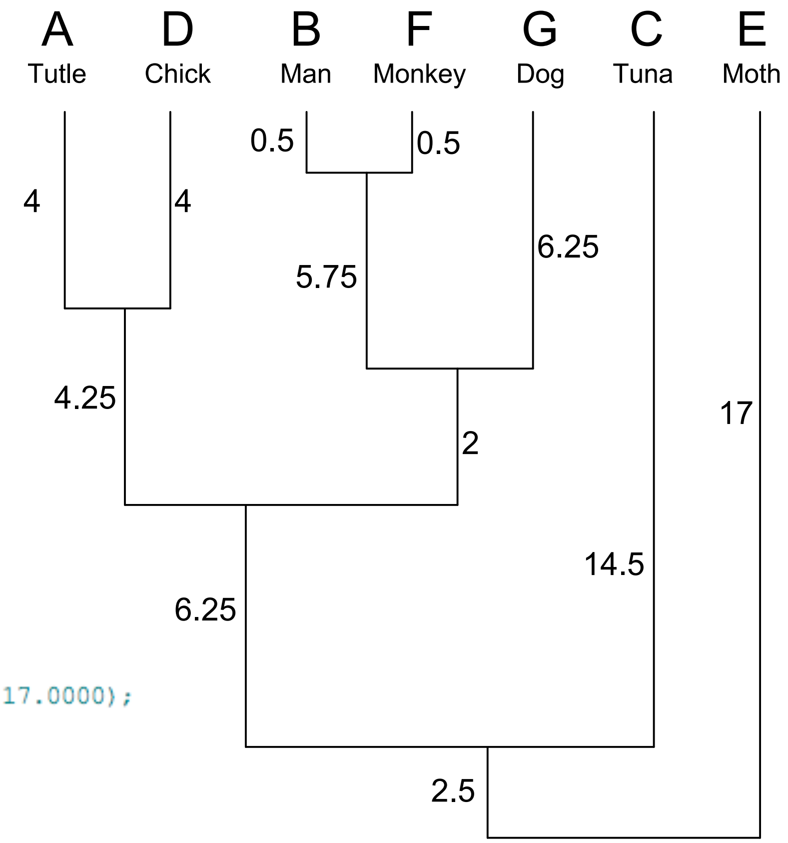
	ADBFGC	E
ADBFGC	-	
E	34	-



UPGMA: Example

Remark: The source data for this example is a selection of Cytochrome C distances from Table 3 of Fitch & Margoliash (1967) Construction of phylogenetic tree, *Science* 155:279-84

	A Turtle	B Human	C Tuna	D Chicken	E Moth	F Monkey	G Dog
A	-						
B	19	-					
C	27	31	-				
D	8	18	26	-			
E	33	36	41	31	-		
F	18	1	32	17	35	-	
G	13	13	29	14	28	12	-



Newick representation:

```
((((Monkey:0.5000,Man:0.5000):5.7500,Dog:6.2500):2.0000,
(Turtle:4.0000,Chicken:4.0000):4.2500):6.2500,Tuna:14.5000):2.5000,Moth:17.0000);
```

Source: Dr Richard J. Edwards
Slides: http://www.southampton.ac.uk/~re1u06/teaching/upgma/
Software: http://bioware.soton.ac.uk/upgma.html