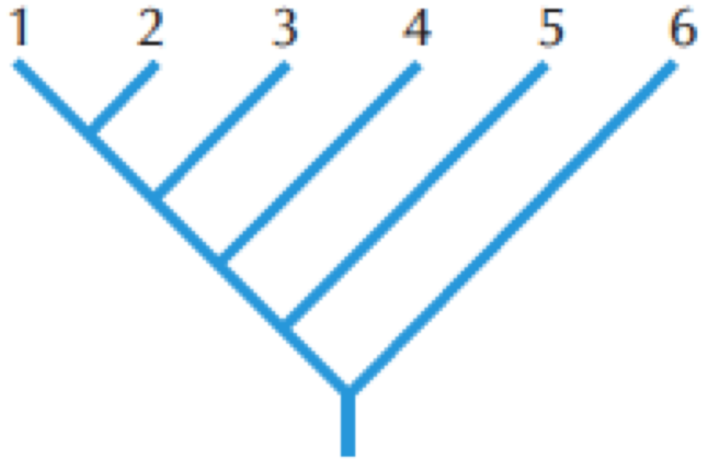
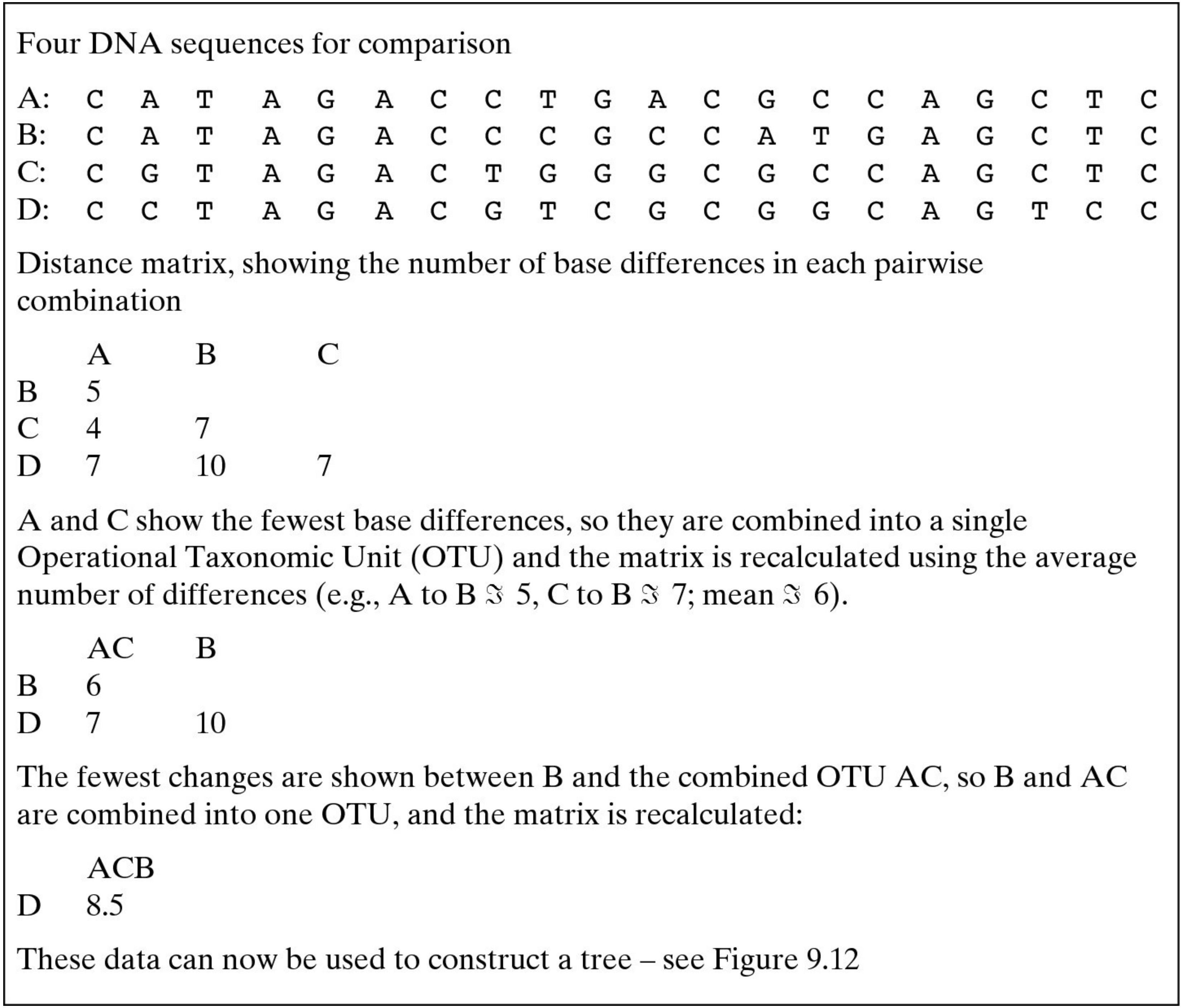
Steps in building a phylogenetic tree.

Method	Example
Choose gene(s) of interest ↓	5
Identify homologs ↓	2 1 3 4 6 5
Align sequences ↓	
Calculate gene tree	



Four DNA sequences for comparison

A: C A T A G A C C T G A C G C C A G C T C
B: C A T A G A C C C G C C A T G A G C T C
C: C G T A G A C T G G G C G C C A G C T C
D: C C T A G A C G T C G C G G C A G T C C

Distance matrix, showing the number of base differences in each pairwise combination

	A	B	C
B	5		
C	4	7	
D	7	10	7

A and C show the fewest base differences, so they are combined into a single Operational Taxonomic Unit (OTU) and the matrix is recalculated using the average number of differences (e.g., A to B ℑ 5, C to B ℑ 7; mean ℑ 6).

	AC	B
B	6	
D	7	10

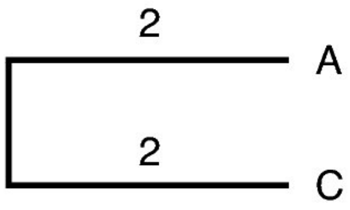
The fewest changes are shown between B and the combined OTU AC, so B and AC are combined into one OTU, and the matrix is recalculated:

	ACB
D	8.5

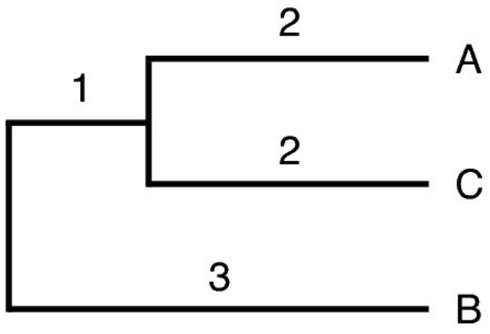
These data can now be used to construct a tree – see Figure 9.12

Figure 9.11 Phylogenetic trees based on DNA sequences; sample data and Unweighted Pair Group Method with Arithmetic means (UPGMA) calculation.

UPGMA :



Following recalculation of the matrix, AC and B are combined into one OTU. Again, the mean number of differences (6) is divided equally between the two branches. Add the numbers together as you trace the path from A or C to B



Finally, ACB is combined with D into a single OTU

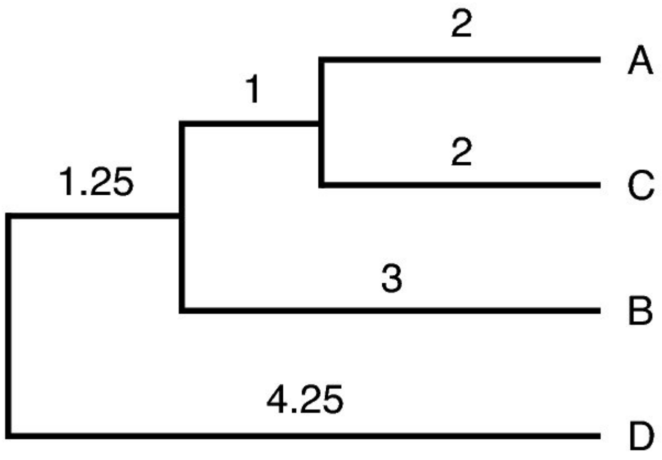


Figure 9.12 Derivation of a phylogenetic tree using Unweighted Pair Group Method with Arithmetic means (UPGMA).

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

UPGMA:

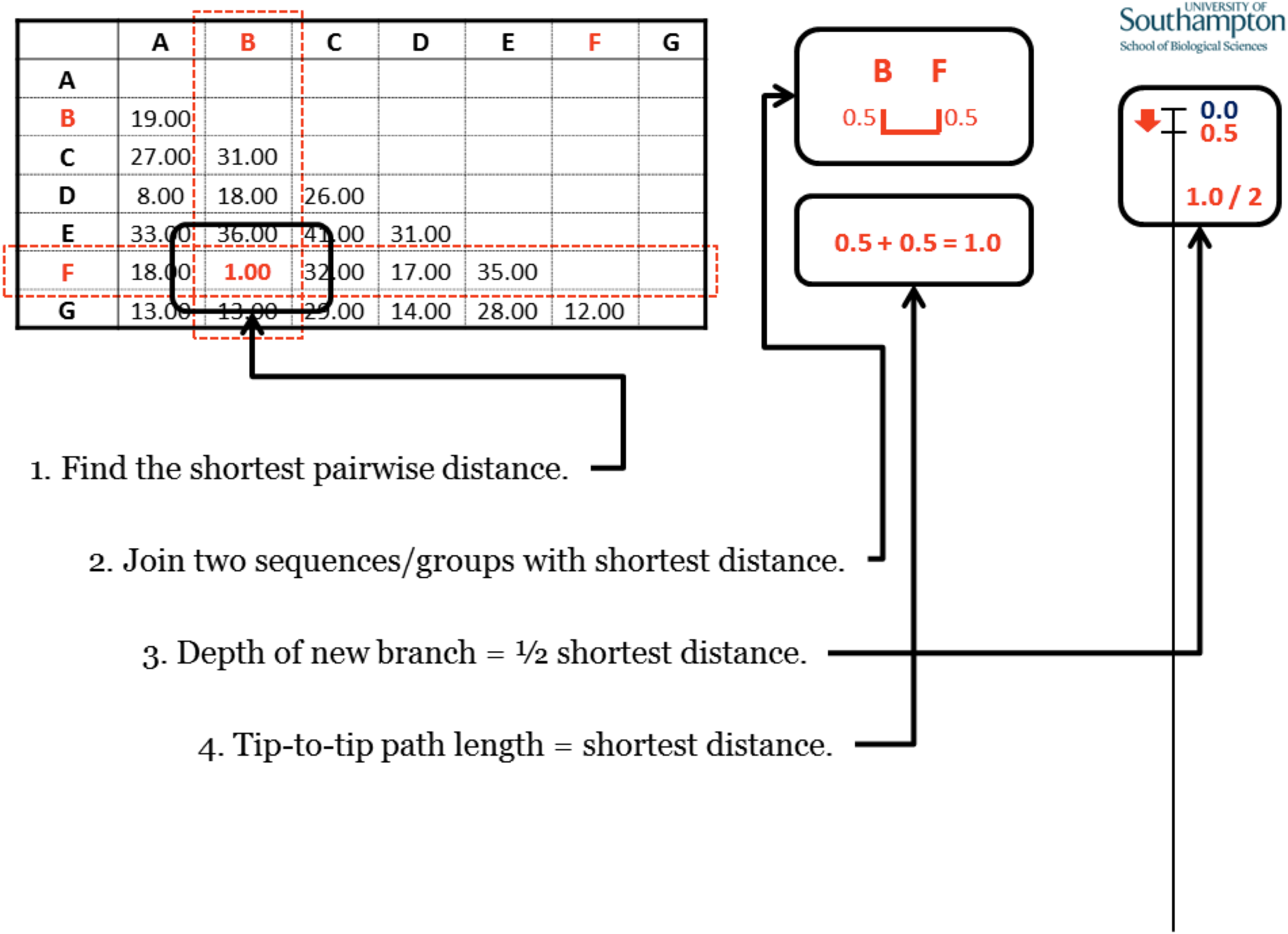
Unweighted Pair-Group Method with Arithmetic mean

Unweighted – all pairwise distances contribute equally.

Pair-Group – groups are combined in pairs (dichotomies only).

Arithmetic mean – pairwise distances to each group (clade) are mean distances to all members of that group.

(Ultrametric – assumes molecular clock)



	A	**B**	C	D	E	**F**	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	**1.00**	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

1. Find the shortest pairwise distance.
2. Join two sequences/groups with shortest distance.
3. Depth of new branch = ½ shortest distance.
4. Tip-to-tip path length = shortest distance.

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

5. Calculate mean pairwise distances with other sequences in new matrix.

(19 + 18) / 2 = 18.5

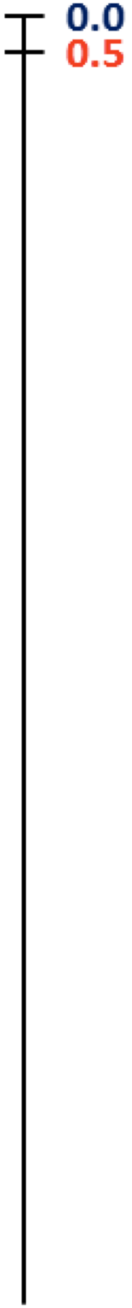
(31 + 32) / 2 = 31.5

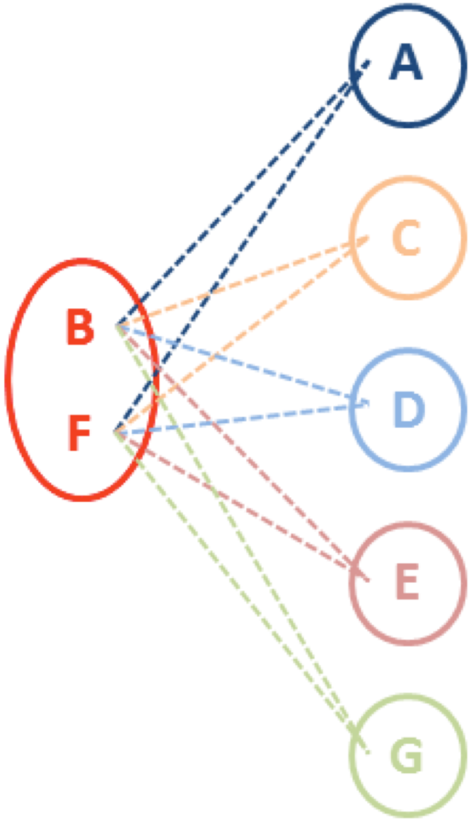
	A	BF	C	D	E	G
A						
BF	18.50					
C	27.00	31.50				
D	8.00	17.50	26.00			
E	33.00	35.50	41.00	31.00		
G	13.00	12.50	29.00	14.00	28.00	

(18 + 17) / 2 = 17.5

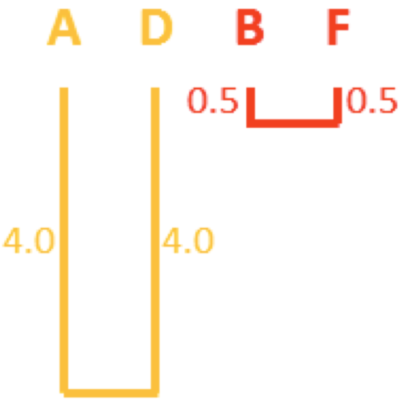
(36 + 35) / 2 = 35.5

(13 + 12) / 2 = 12.5

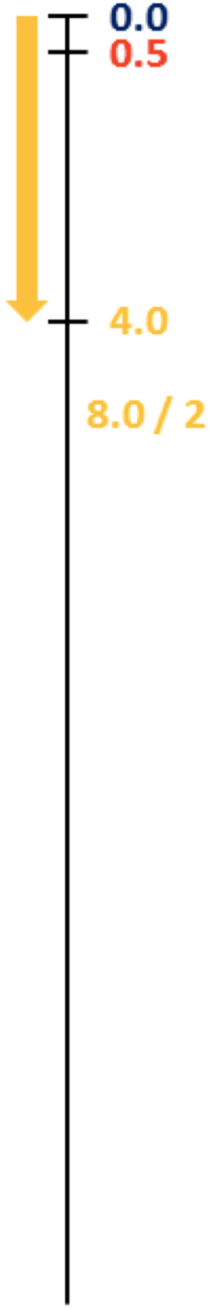




	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	



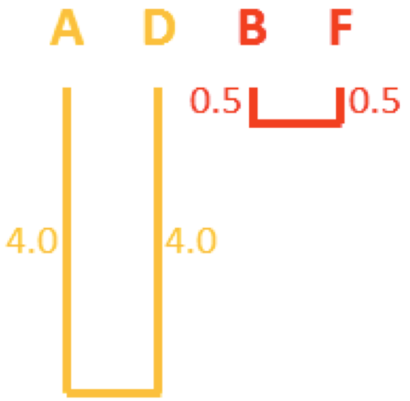
4.0 + 4.0 = 8.0

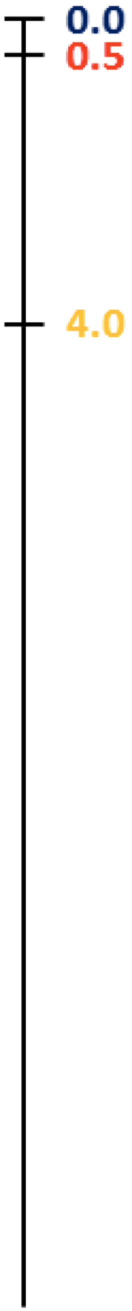


	A	BF	C	D	E	G
A						
BF	18.50					
C	27.00	31.50				
D	8.00	17.50	26.00			
E	33.00	35.50	41.00	31.00		
G	13.00	12.50	29.00	14.00	28.00	

6. Repeat cycle with new shortest distance.

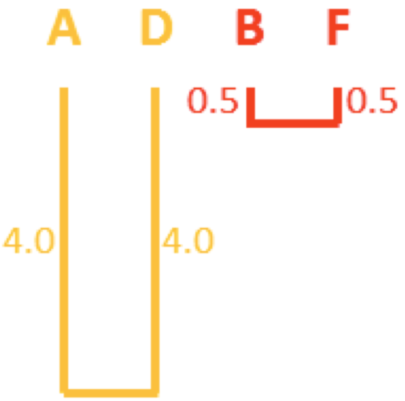
	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	





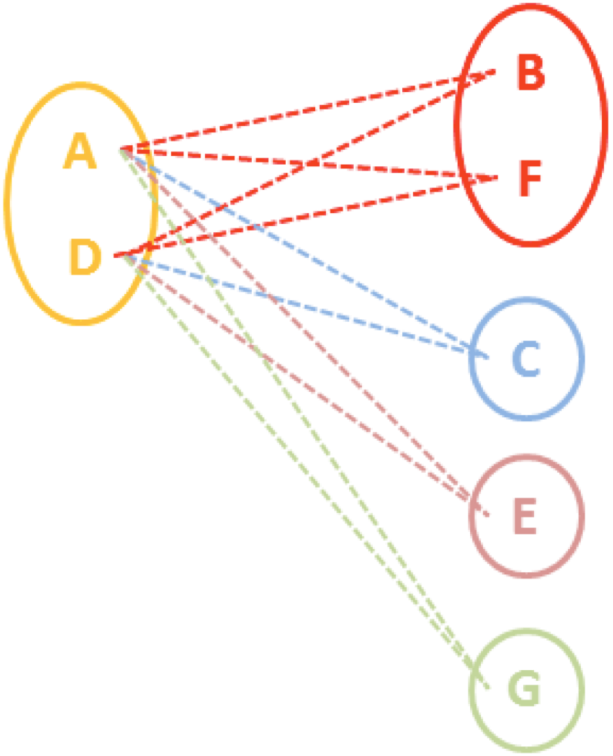
	A	BF	C	D	E	G
A						
BF	18.50					
C	27.00	31.50				
D	8.00	17.50	26.00			
E	33.00	35.50	41.00	31.00		
G	13.00	12.50	29.00	14.00	28.00	

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	



(19 + 18 + 18 + 17) / 4 = 18.0

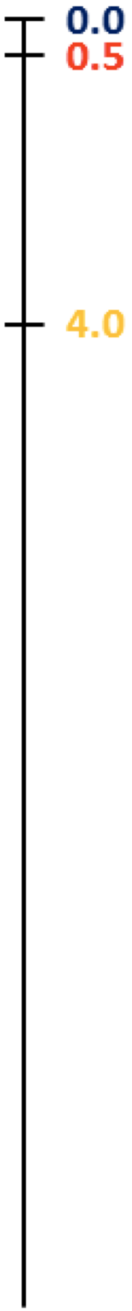
	AD	BF	C	E	G
AD					
BF	18.00				
C	26.50	31.50			
E	32.00	35.50	41.00		
G	13.50	12.50	29.00	28.00	



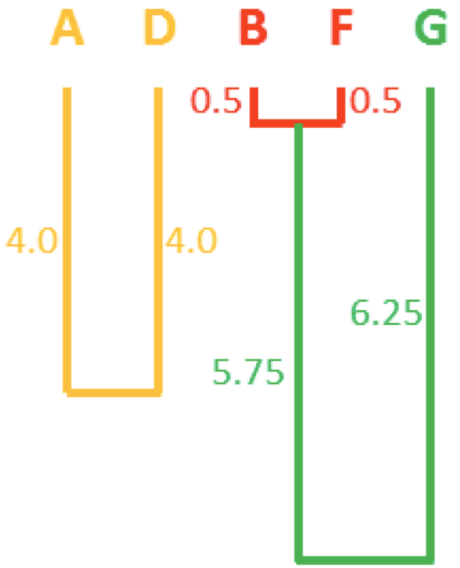
(27 + 26) / 2 = 26.5

(33 + 31) / 2 = 32.0

(13 + 14) / 2 = 13.5



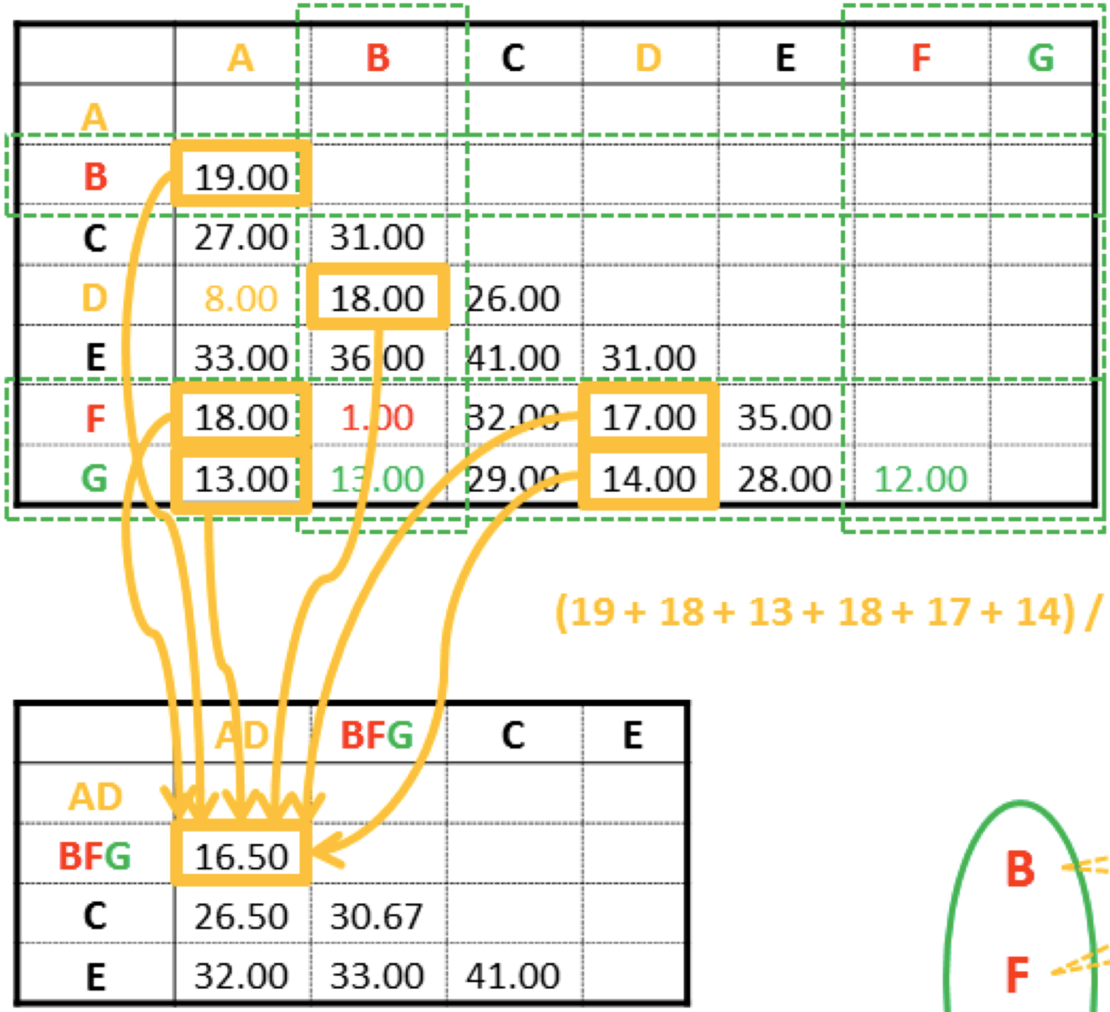
	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	



	AD	BF	C	E	G
AD					
BF	18.00				
C	26.50	31.50			
E	32.00	35.50	41.00		
G	13.50	**12.50**	29.00	28.00	

0.5 + 5.75 + 6.25 = 12.5

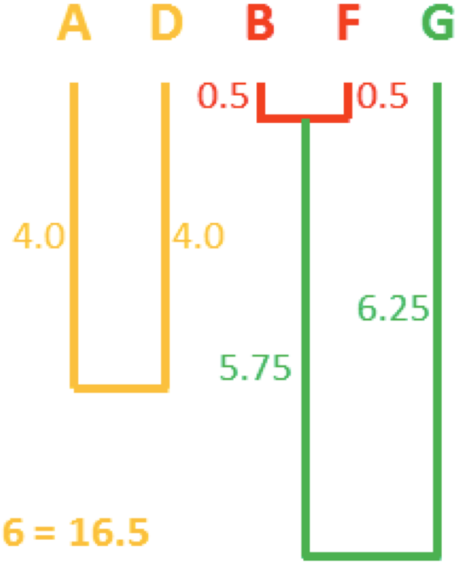
12.5 / 2

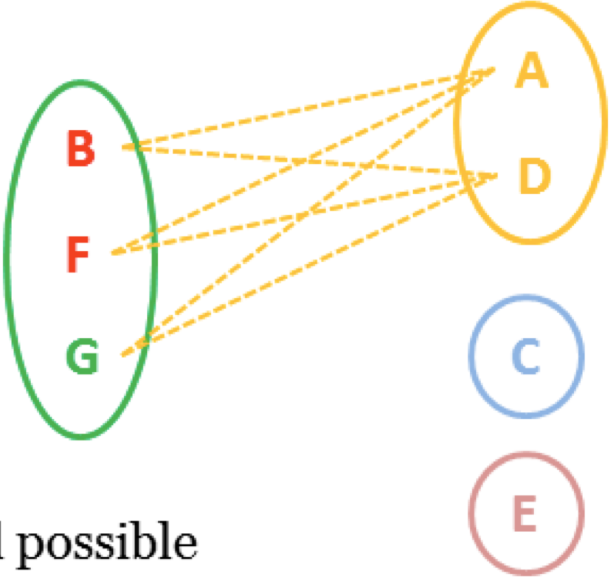


	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

(19 + 18 + 13 + 18 + 17 + 14) / 6 = 16.5

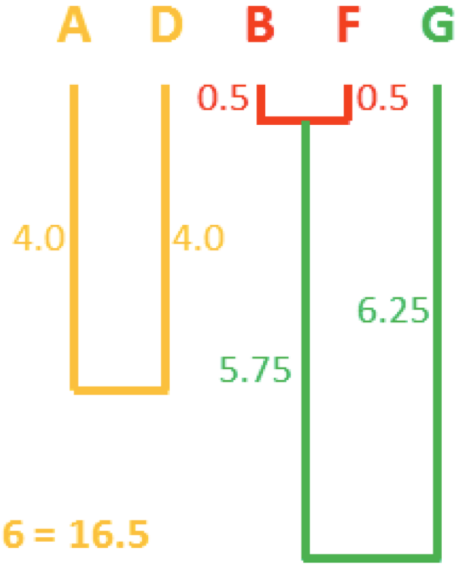
	AD	BFG	C	E
AD				
BFG	16.50			
C	26.50	30.67		
E	32.00	33.00	41.00	





New distances are mean values for all possible pairwise distances **between** groups.

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

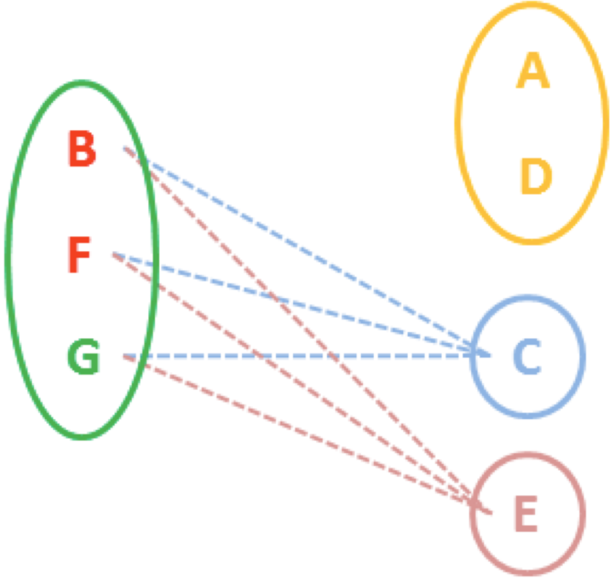


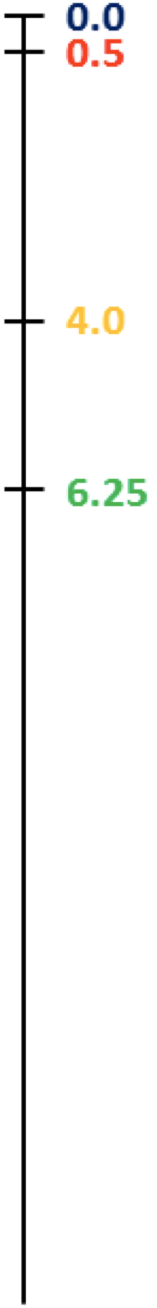
(19 + 18 + 13 + 18 + 17 + 14) / 6 = 16.5

	AD	BFG	C	E
AD				
BFG	16.50			
C	26.50	30.67		
E	32.00	33.00	41.00	

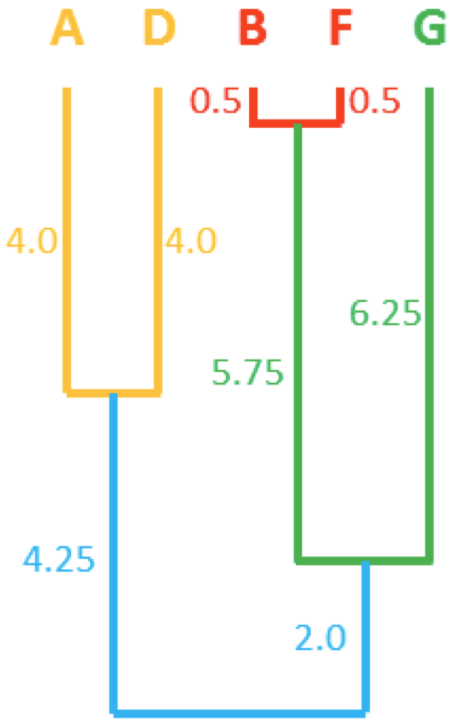
(31 + 32 + 29) / 3 = 30.67

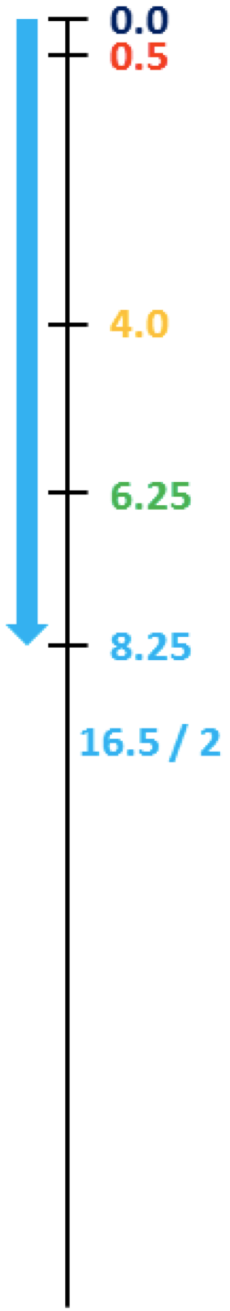
(36 + 35 + 28) / 3 = 33.0





	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	





	AD	BFG	C	E
AD				
BFG	16.50			
C	26.50	30.67		
E	32.00	33.00	41.00	

0.5 + 5.75 + 4.25 = 16.5

4.0 + 4.25 +

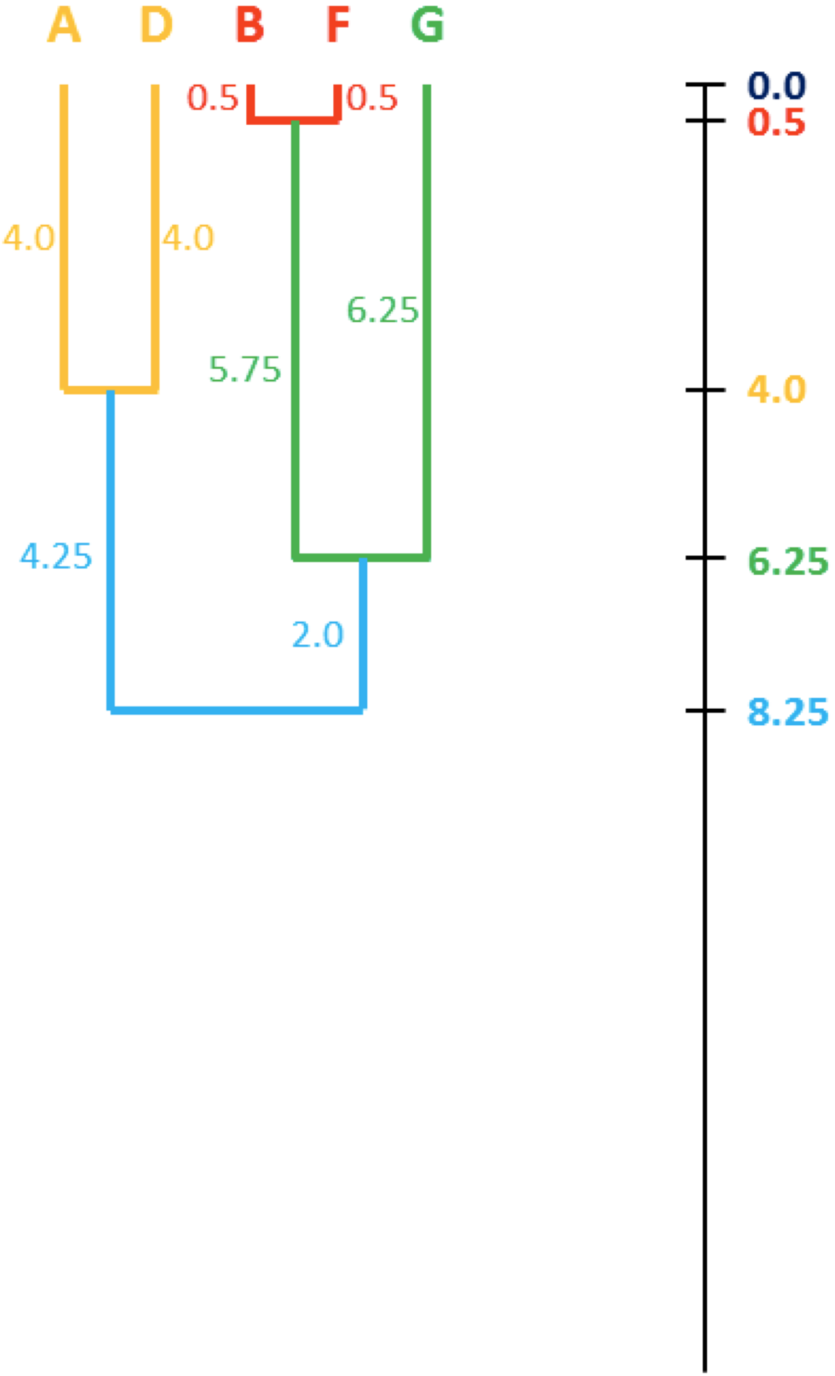
6.25 + 2.0 = 16.5

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

(27 + 31 + 26 + 32 + 29) / 5 = 29.00

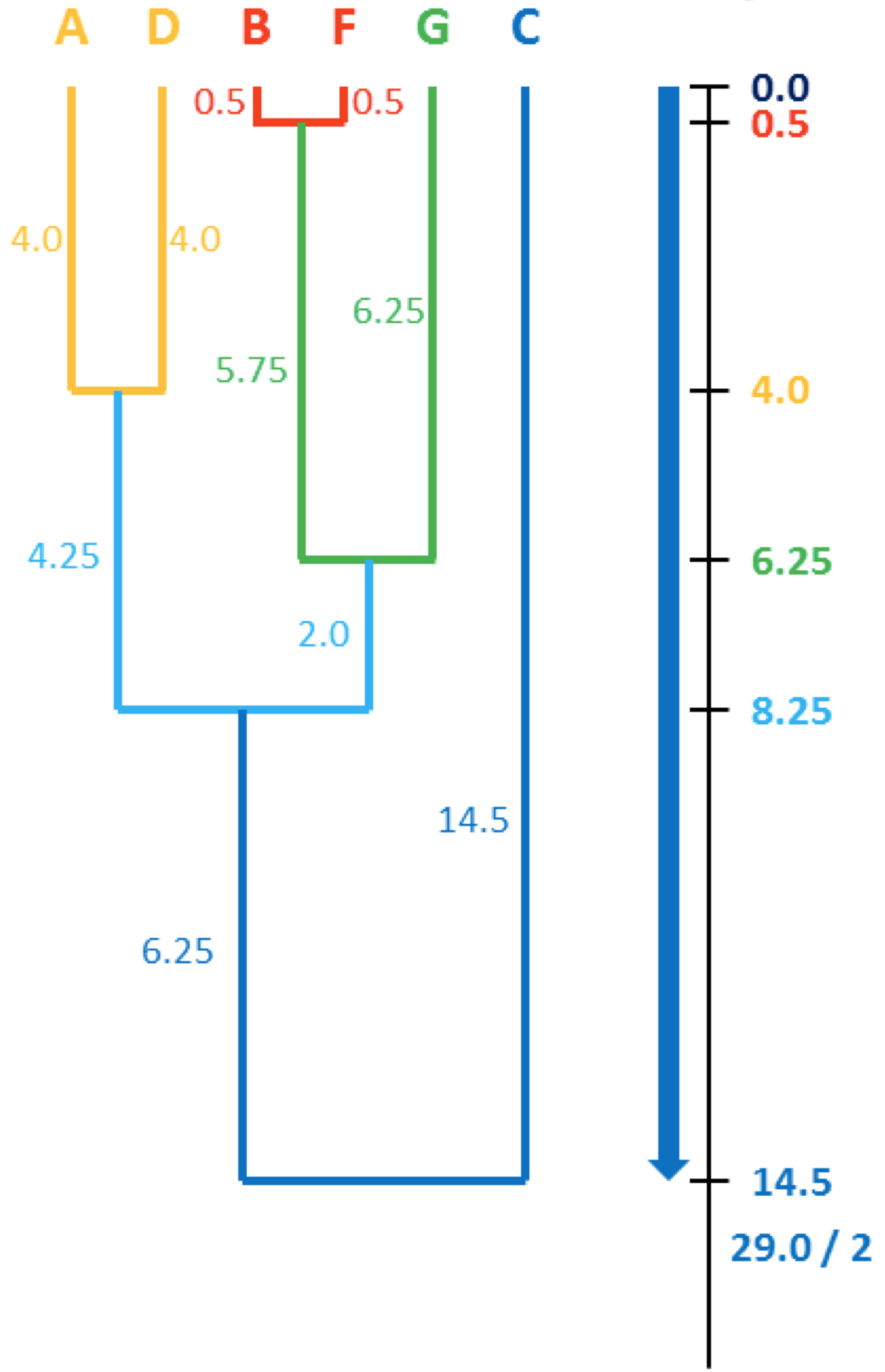
	ADBFG	C	E
ADBFG			
C	29.00		
E	32.60	41.00	

(33 + 36 + 31 + 35 + 28) / 5 = 32.60



	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	ADBFG	C	E
ADBFG			
C	**29.00**		
E	32.60	41.00	

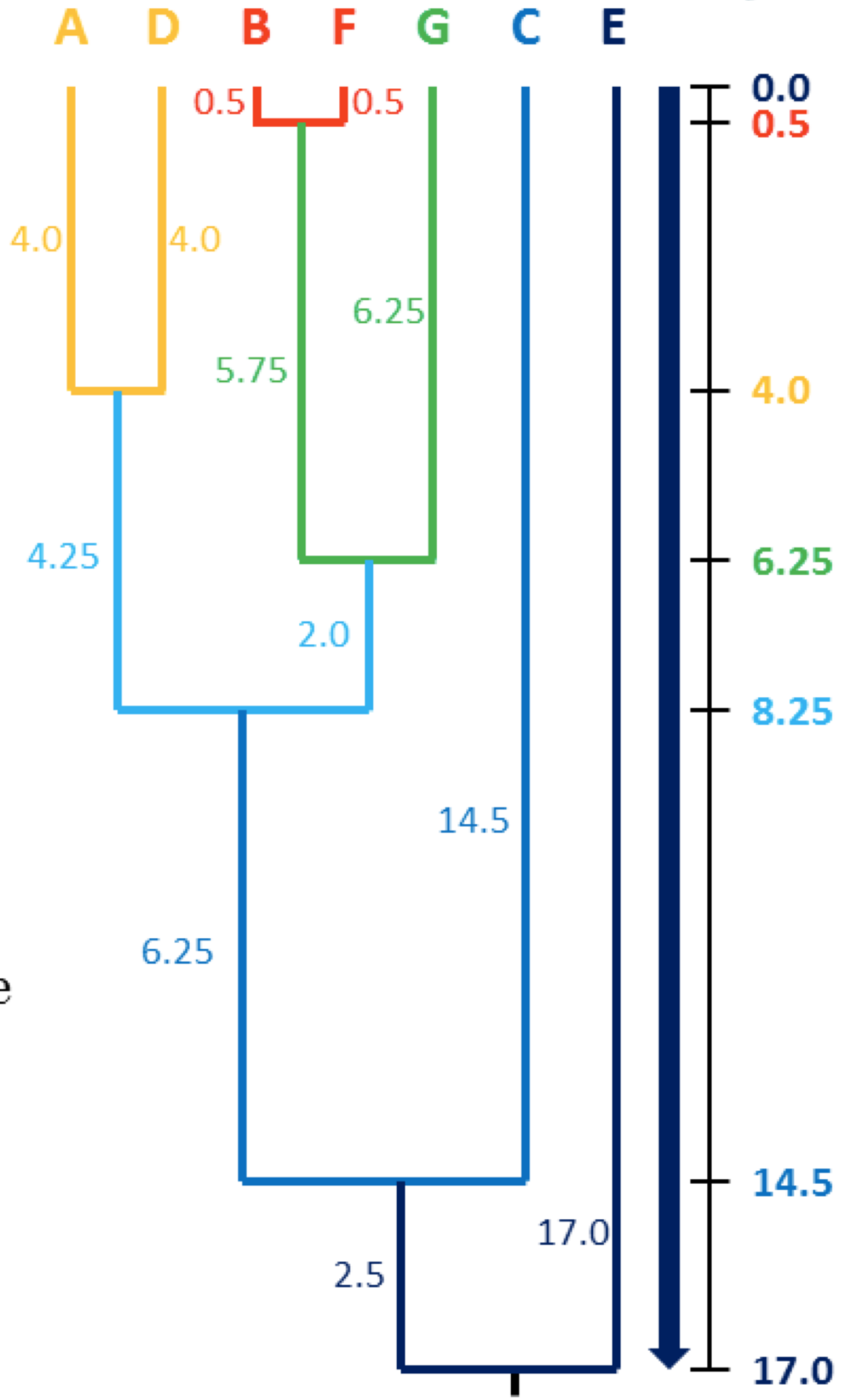


	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

(33 + 36 + 41 +31 + 35 + 28) / 6 = 34.00

	ADBFGC	E
ADBFGC		
E	34.00	

UPGMA assumes a molecular clock. The tree is rooted with the final joining of clades. All tip-to-tip distances via the root will have the same total distance, equal to the final mean distance.



	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	**1.00**	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

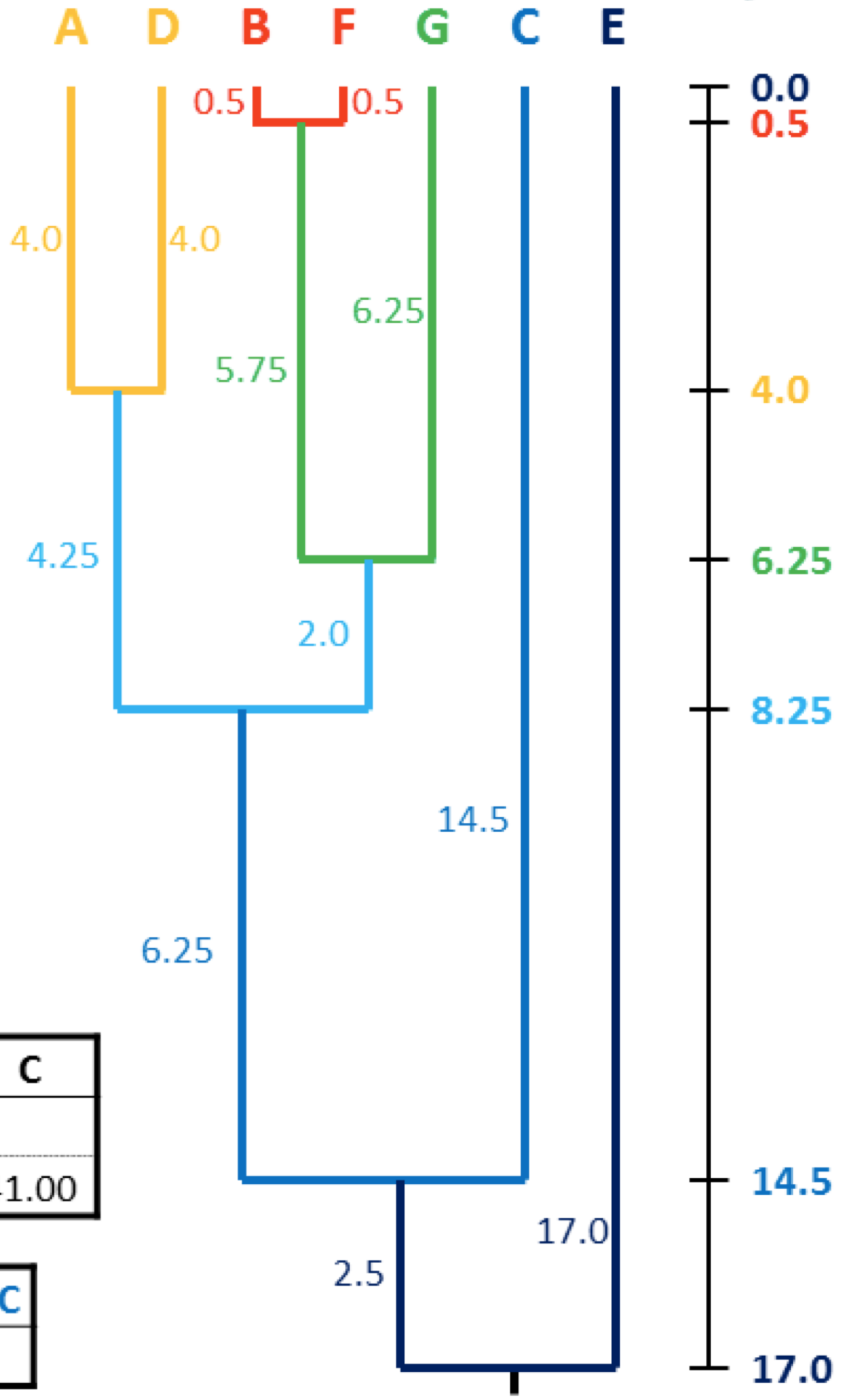
	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	**8.00**	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	**12.50**	29.00	28.00

	AD	BFG	C
BFG	**16.50**		
C	26.50	30.67	
E	32.00	33.00	41.00

	ADBFG	C
C	**29.00**	
E	32.60	41.00

	ADBFGC
E	**34.00**



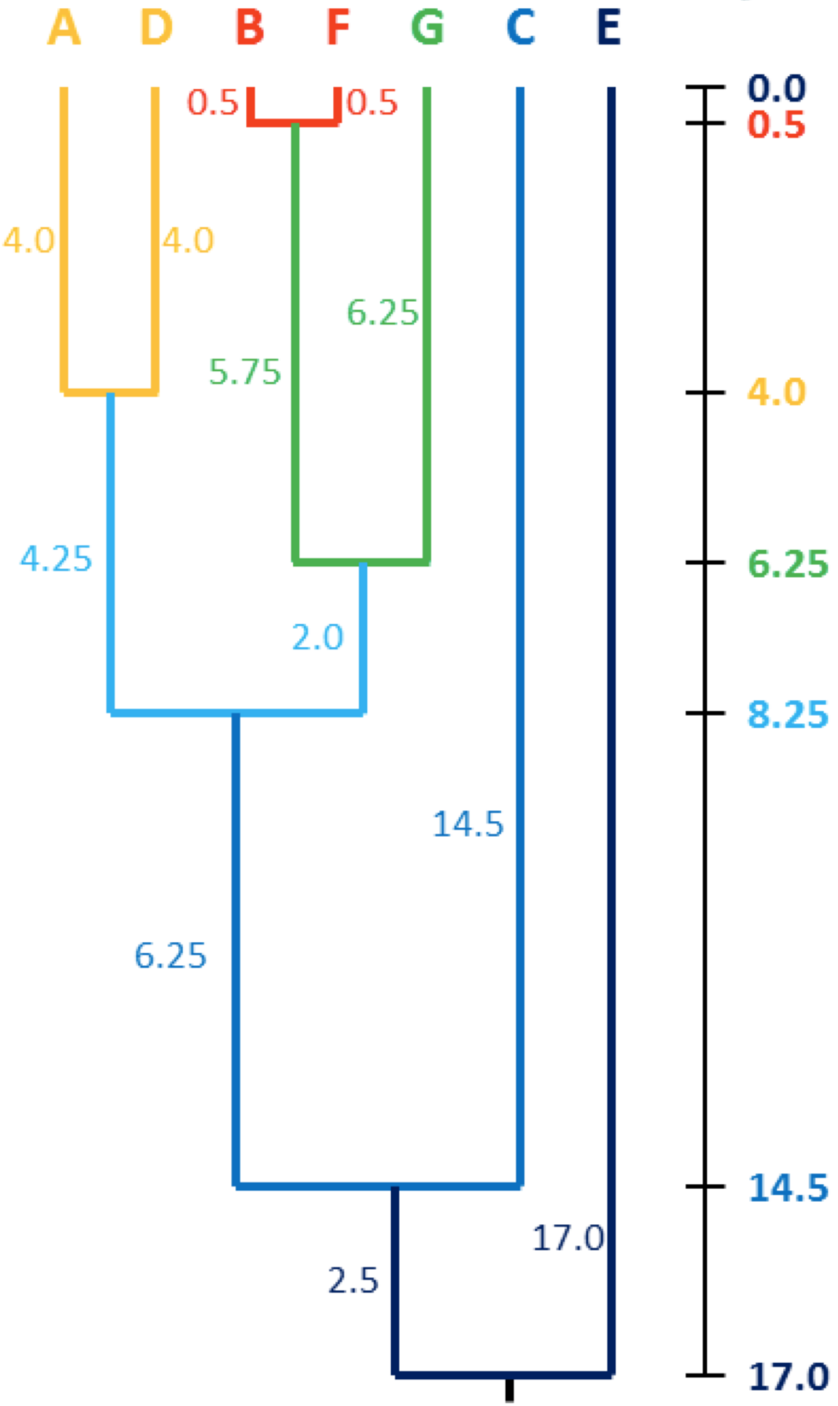
UNIVERSITY OF Southampton
School of Biological Sciences

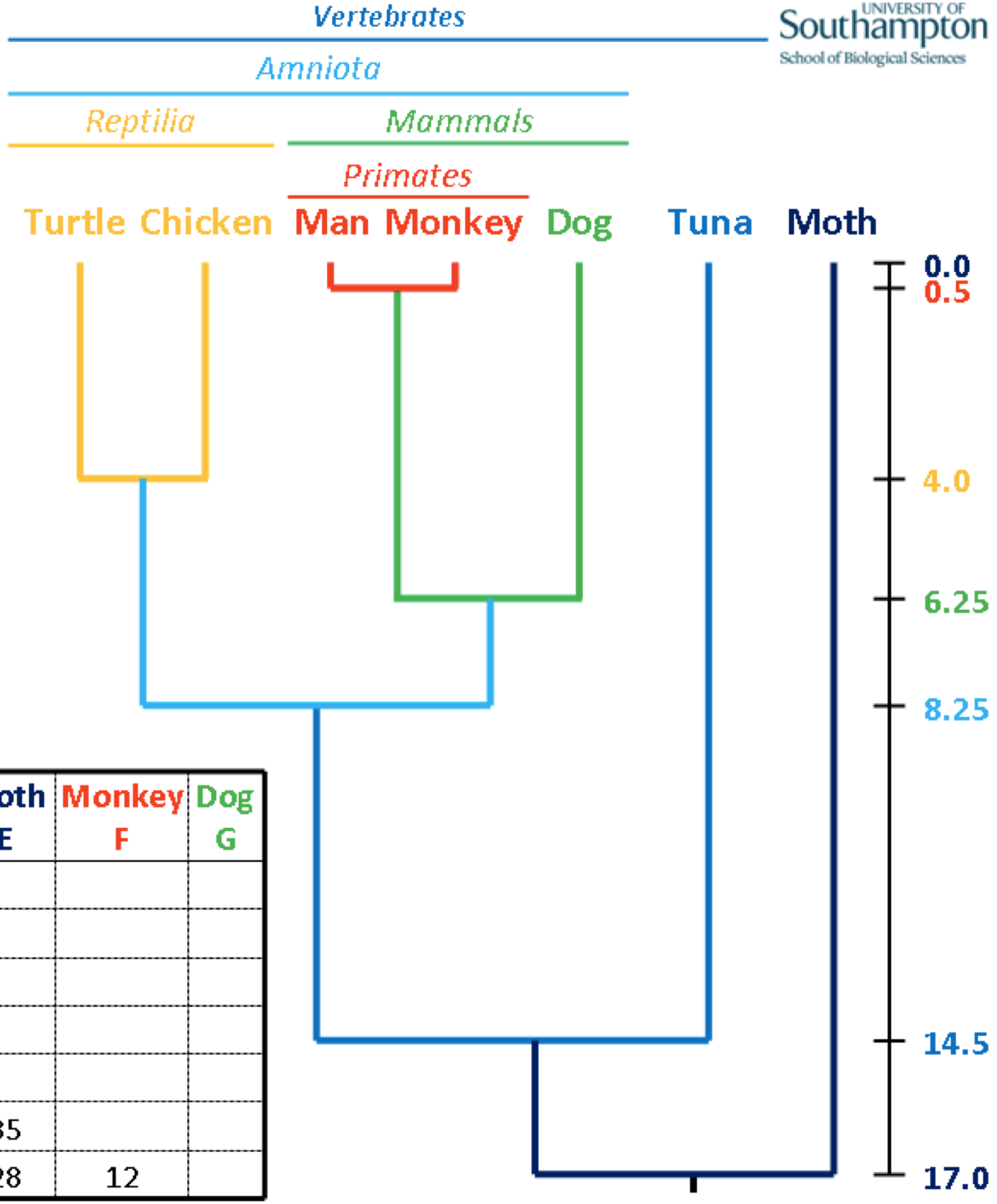
	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

The source data for this worked example is a selection of Cytochrome C distances from Table 3 of one of the seminal phylogenetic papers: Fitch WM & Margoliash E (1967). Construction of phylogenetic trees. *Science* **155**:279-84.

http://www.ncbi.nlm.nih.gov/pubmed/5334057

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





The UPGMA tree based on this Cytochrome C data supports the known evolutionary relationships of these organisms.

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