

Βιοπληροφορική

Ενότητα 10: Κατασκευή φυλογενετικών δέντρων

Αν. καθηγητής Αγγελίδης Παντελής

e-mail: paggelidis@uowm.gr

ΕΕΔΙΠ Μπέλλου Σοφία

e-mail: sbellou@uowm.gr

Τμήμα Μηχανικών Πληροφορικής και Τηλεπικοινωνιών



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ψηφιακά Μαθήματα στο Πανεπιστήμιο Δυτικής Μακεδονίας**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



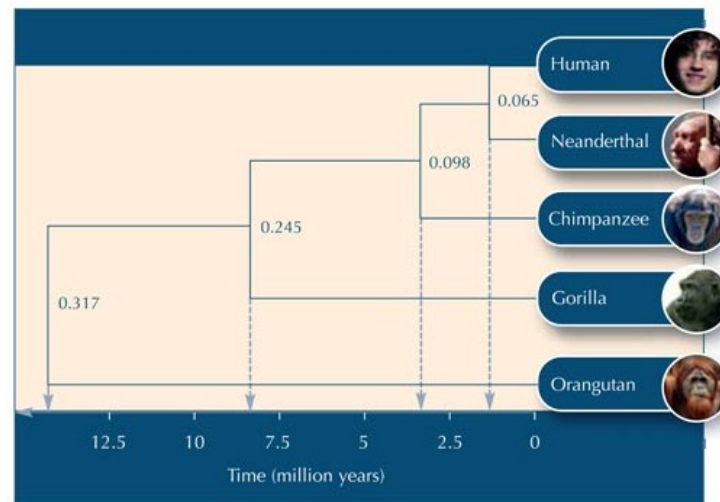
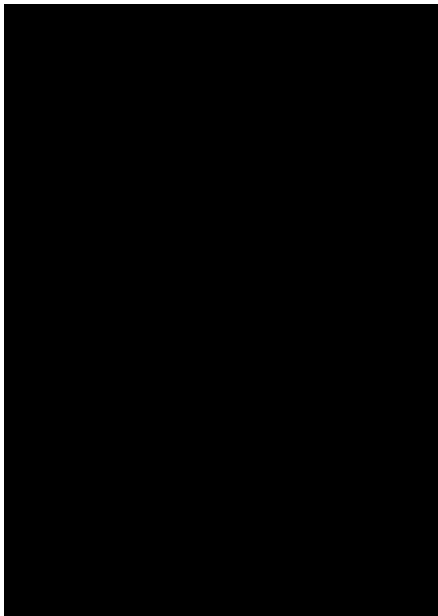
Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Κατασκευή φυλογενετικών δέντρων



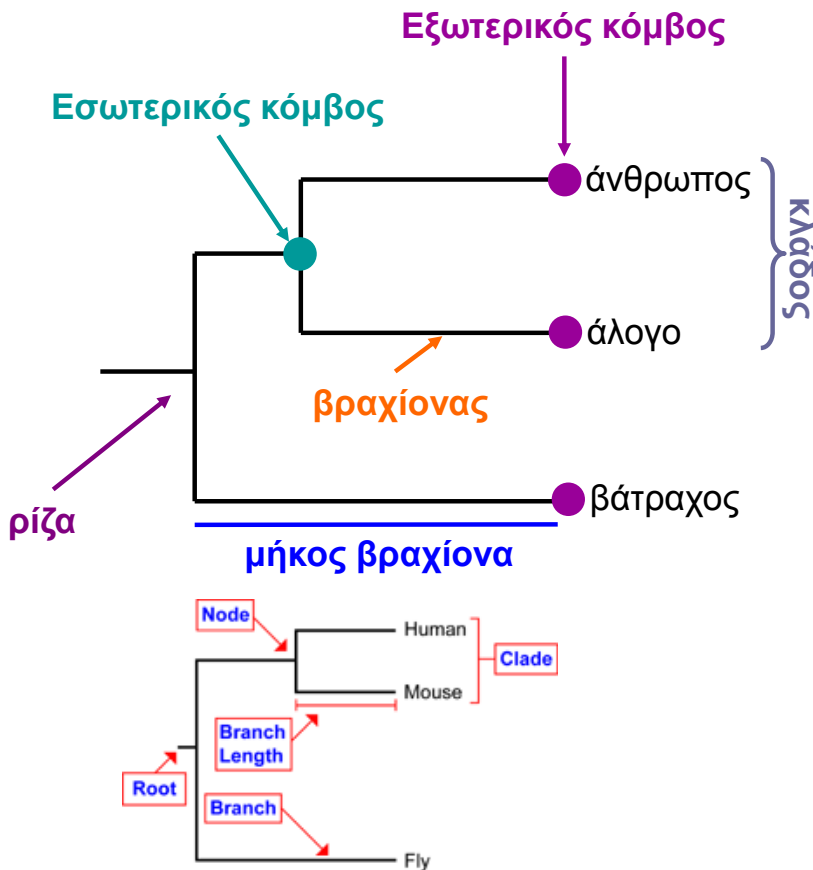
Φυλογενετική ανάλυση

– Αρχή και σκοπός

- Εξαρτάται από μία πολύ καλή πολλαπλή στοίχιση αλληλουχιών.
- Οι αλληλουχίες με παρόμοιες αντικαταστάσεις ομαδοποιούνται με σκοπό να κατασκευαστεί ένα φυλογενετικό δέντρο.
- **ΣΚΟΠΟΣ:** Η ανακάλυψη των σχέσεων μεταξύ των βραχιόνων καθώς και το μήκος των βραχιόνων.
- Αλληλουχίες που μοιράζονται μεγάλη ομοιότητα τοποθετούνται σε γειτονικούς βραχίονες.
- Κατανόηση της εξελεγκτικής ιστορίας:
 - Origin of Europeans.
- Πρόβλεψη λειτουργίας γονιδίου.
- Μελέτη των αλλαγών που συμβαίνουν σε ραγδαία μεταβαλλόμενους οργανισμούς, π.χ. ιούς (επιδημιολογία).



Φυλογενετικά δέντρα



- **Φύλλα:** Ταξινομικές μονάδες, όπως ομάδες οργανισμών, ειδών, γονιδίων (αλληλουχίες).
- **Κόμβος:** Τον πιο πρόσφατο κοινό πρόγονο των ταξινομικών μονάδων. Αντιπροσωπεύει τη δημιουργία 2 διαφορετικών ειδών (ειδογένεση).
- **Βραχίονες:** Το μήκος τους μπορεί να δείχνει το πλήθος των αλλαγών (εξελεγκτική απόσταση).
- **Κλάδος:** Περιλαμβάνει έναν κόμβο και όλες τις ταξινομικές μονάδες που προκύπτουν από αυτόν.



Branches

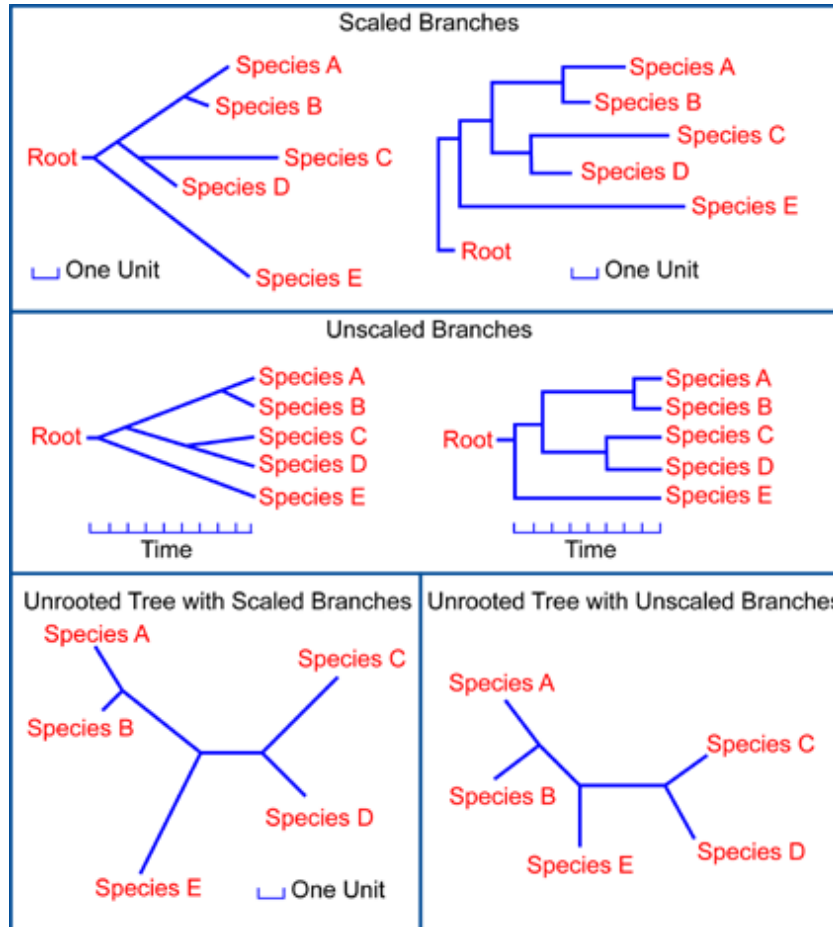


Image: <http://www.ncbi.nlm.nih.gov/About/primer/phylo.html>



Μέθοδοι κατασκευής φυλογενετικών δέντρων (1/2)

- **Μέθοδοι βασισμένες σε χαρακτήρες – εξελεγκτικά μοντέλα:**
 - Μέθοδος μέγιστης συντήρησης (Maximum parsimony method).
- **Μέθοδοι βασισμένες στην απόσταση:**
 - Fitch and Margoliash Algorithm.
 - UPGMA (Unweighted pair-group method).



Μέθοδος ομαδοποίησης αστάθμητων ζευγών με αριθμητικούς μέσους όρους

- **UPGMA:** Unweighted Pair Group Method with Arithmetic mean.
- Προτάθηκε αρχικά (Sneth & Sokal 1973) για ταξινομήσεις με βάση μορφολογικά δεδομένα, αλλά αργότερα χρησιμοποιήθηκε σε μοριακά δεδομένα χωρίς να μειωθεί η αξία της.
- Βασική αρχή για την κατασκευή δέντρου αποτελεί η προϋπόθεση **σταθερού ρυθμού εξέλιξης** μεταξύ όλων των κλάδων του δέντρου.
- Ο αλγόριθμος της μεθόδου καταλήγει με απλούς μαθηματικούς υπολογισμούς πάντα σε ένα **φυλογενετικό δέντρο με ρίζα**.



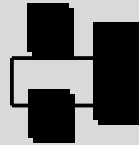
Μέθοδος ομαδοποίησης αστάθμητων ζευγών με αριθμητικούς μέσους όρους – Μέθοδος

1. Υπολογίζονται οι γενετικές αποστάσεις για όλα τα ζεύγη των αλληλουχιών και οι τιμές τους τοποθετούνται σε μία μήτρα αποστάσεων.
2. Από το σύνολο των αποστάσεων της μήτρας αρχικά επιλέγεται το ζεύγος των αλληλουχιών με τη μικρότερη τιμή. Οι δύο αλληλουχίες συγκροτούν μία καινούργια μονάδα, η οποία αναφέρεται ως σύνθετη, και αποτελεί την πρώτη ομάδα του δέντρου.
3. Η σύνθετη ομάδα σε συνδυασμό με τις υπόλοιπες αλληλουχίες συνθέτουν μία καινούργια μήτρα αποστάσεων από την οποία επιλέγεται και πάλι το ζεύγος με τη μικρότερη απόσταση.
4. Η διαδικασία επαναλαμβάνεται έως το τέλος.



Μέθοδος ομαδοποίησης αστάθμητων ζευγών με αριθμητικούς μέσους όρους – Παράδειγμα (1/3)

- Επιλέγεται το ζευγάρι με τη μικρότερη απόσταση, δηλ. το **AB**.
- Επειδή ο ρυθμός εξέλιξης είναι ο ίδιος, η απόσταση των A και B από τον κοινό τους πρόγονο είναι $2/2=1$.



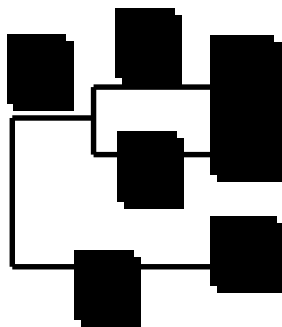
- Οι A και B θεωρούνται μία ενιαία σύνθετη μονάδα και υπολογίζεται μία καινούργια μήτρα αποστάσεων:
 $\text{dist}(A,B),C=(\text{dist}AC+\text{dist}BC)/2=4,$
 $\text{dist}(A,B),D=\text{dist}(AD+BD)/2=6$
 $\text{dist}(A,B),E=(\text{dist}AE+\text{dist}BE)/2=6,$
 $\text{dist}(A,B),F=\text{dist}(AF+BF)/2=8$
- **Κατασκευάζεται μία καινούργια μήτρα αποστάσεων.**

	A	B	C	D	E	F
A						
B	2					
C	4	4				
D	6	6	6			
E	6	6	6	4		
F	8	8	8	8	8	



Μέθοδος ομαδοποίησης αστάθμητων ζευγών με αριθμητικούς μέσους όρους – Παράδειγμα (2/3)

- Το επόμενο ζευγάρι είναι C και AB.
 - Με την ίδια διαδικασία έχουμε:



	A,B,C	D,E	F
A,B,C			
D,E	6		
F	8	8	

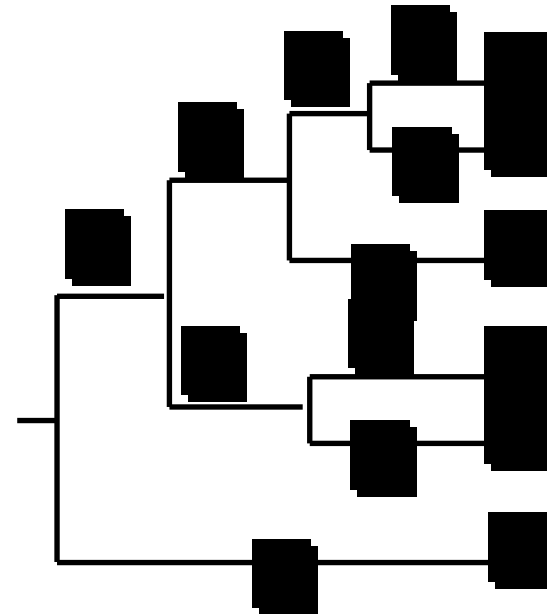
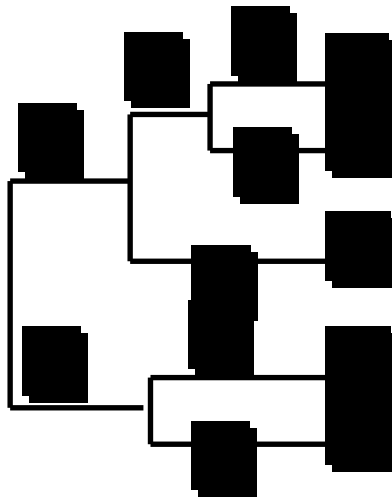
	A	B	C	D	E	F
A						
B	2					
C	4	4				
D	6	6	6			
E	6	6	6	4		
F	8	8	8	8	8	

Μέθοδος ομαδοποίησης αστάθμητων ζευγών με αριθμητικούς μέσους όρους – Παράδειγμα (3/3)

- Και τέλος, έχουμε:

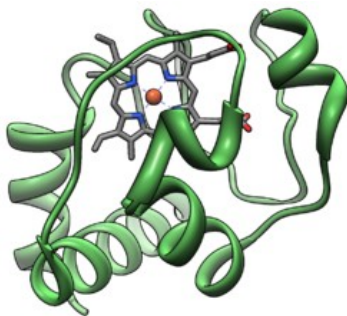
	A,B,C,D,E	F
A,B,C,D,E		
F	8	

	A,B,C	D,E	F
A,B,C			
D,E	6		
F	8	8	



URPGMA – 2nd Example (1/15)

- Η πρωτεΐνη cytochrome C εκφράζεται σε 7 διαφορετικούς οργανισμούς (A, B, C, D, E, F, & G).
- Οι αλληλουχίες της πρωτεΐνης από τους 7 οργανισμούς συγκρίθηκαν μεταξύ τους (όλοι οι συνδυασμοί).
- Ο πίνακας δείχνει την **απόσταση** μεταξύ των αλληλουχιών.
- Να κατασκευαστεί το φυλογενετικό δέντρο των οργανισμών που προκύπτει από την παρακάτω μήτρα αποστάσεων.

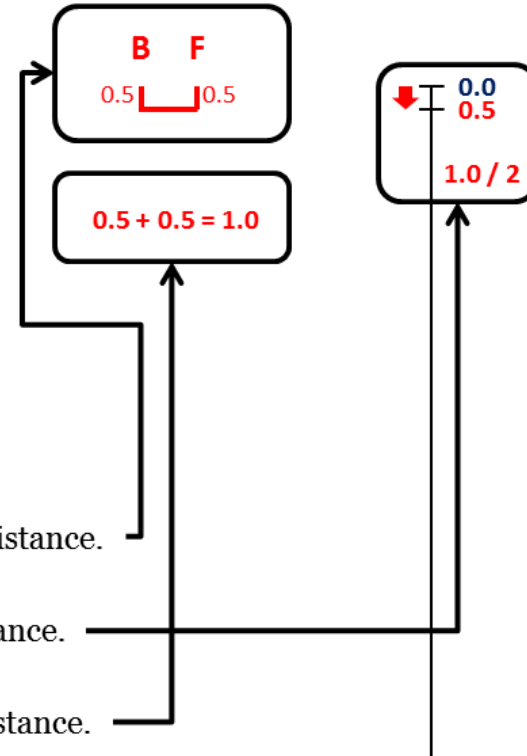


	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

UPGMA – 2nd Example (2/15)

	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 = $\frac{1}{2}$ shortest distance.
4. Tip-to-tip path length = shortest distance.



UPGMA – 2nd Example (3/15)

	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	



0.0
0.5

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

$$(19 + 18) / 2 = 18.5$$

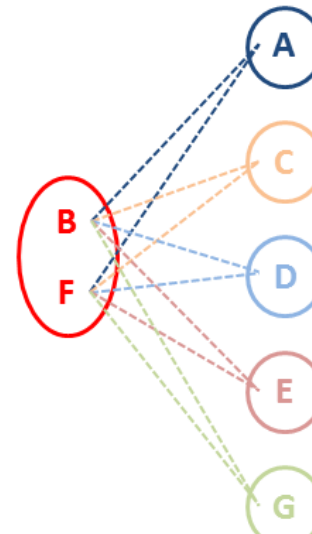
$$(31 + 32) / 2 = 31.5$$

	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	

$$(18 + 17) / 2 = 17.5$$

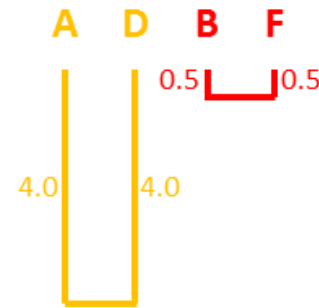
$$(36 + 35) / 2 = 35.5$$

$$(13 + 12) / 2 = 12.5$$



UPGMA – 2nd Example (4/15)

	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	



$$4.0 + 4.0 = 8.0$$

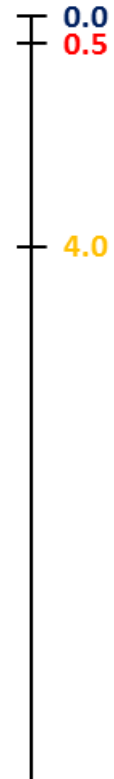
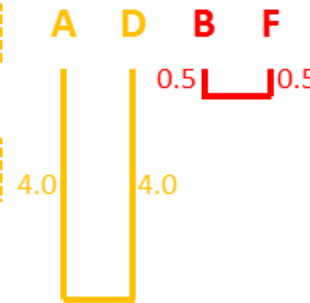


	A	BF	C	D	E	G
A						
BF	18.50					
C	27.00	31.50				
D	8.00	17.50	26.00			
E	33.00	35.50	41.00	31.00		
G	13.00	12.50	29.00	14.00	28.00	



UPGMA – 2nd Example (5/15)

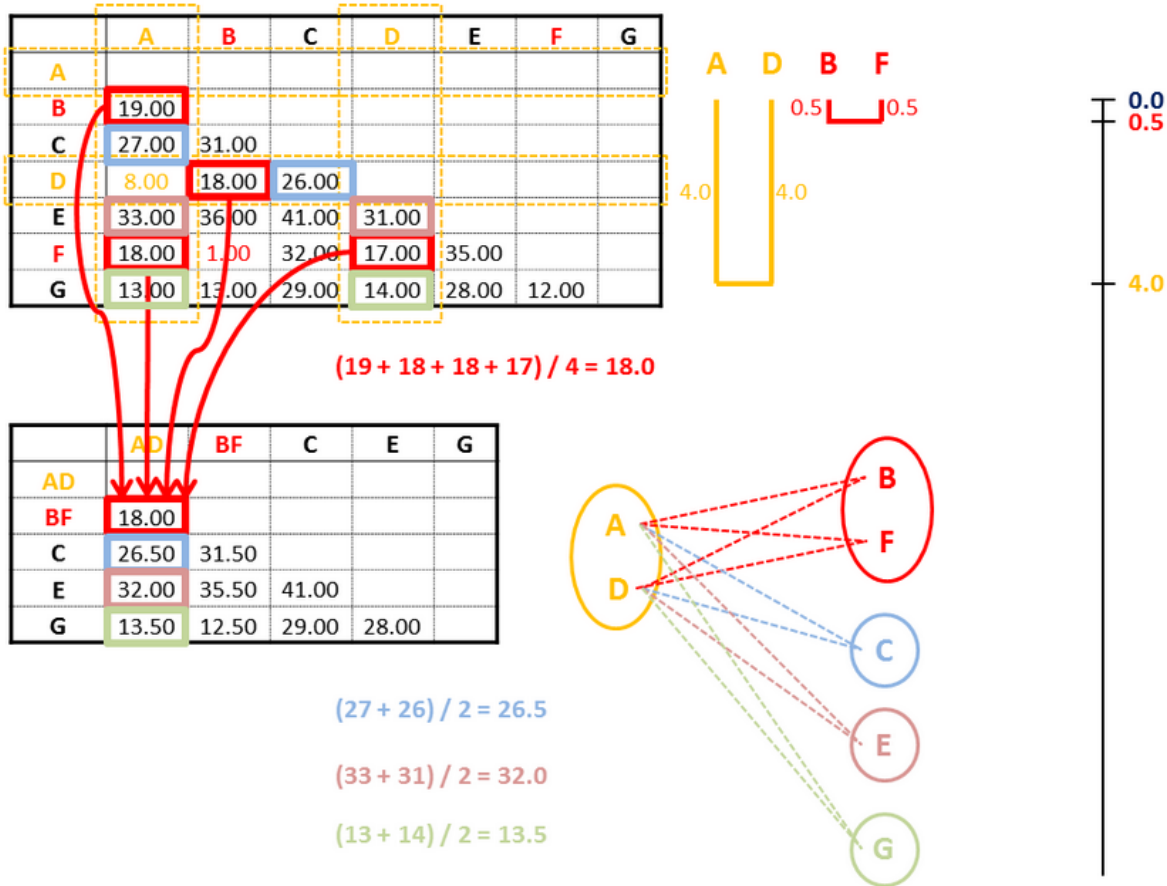
	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	



	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	

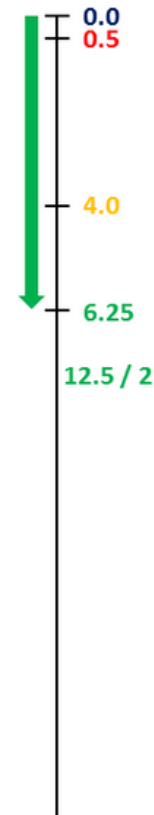
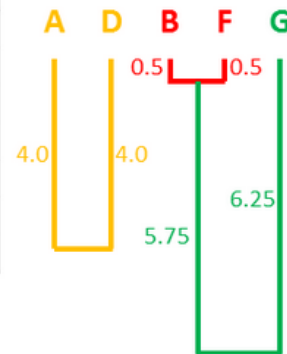


UPGMA – 2nd Example (6/15)



UPGMA – 2nd Example (7/15)

	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	

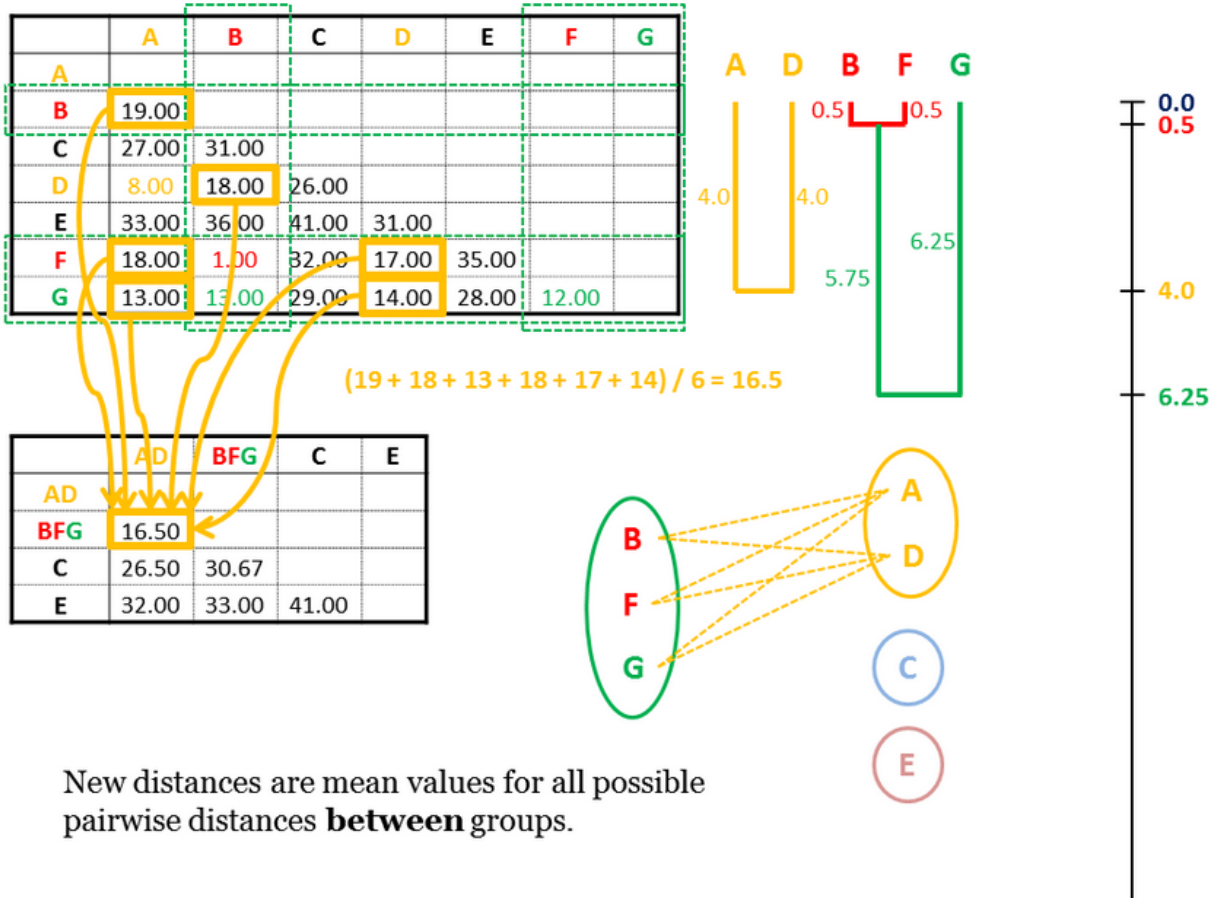


$$0.5 + 5.75 + 6.25 = 12.5$$

	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	



UPGMA – 2nd Example (8/15)



UPGMA – 2nd Example (9/15)

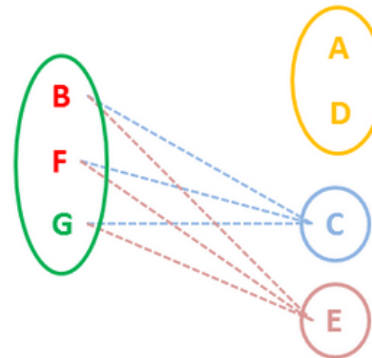
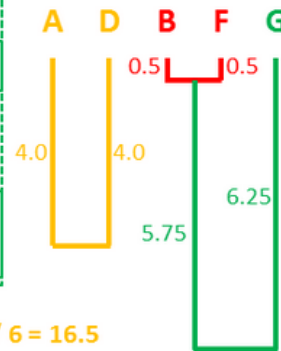
	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

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

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

$$(31 + 32 + 29) / 3 = 30.67$$

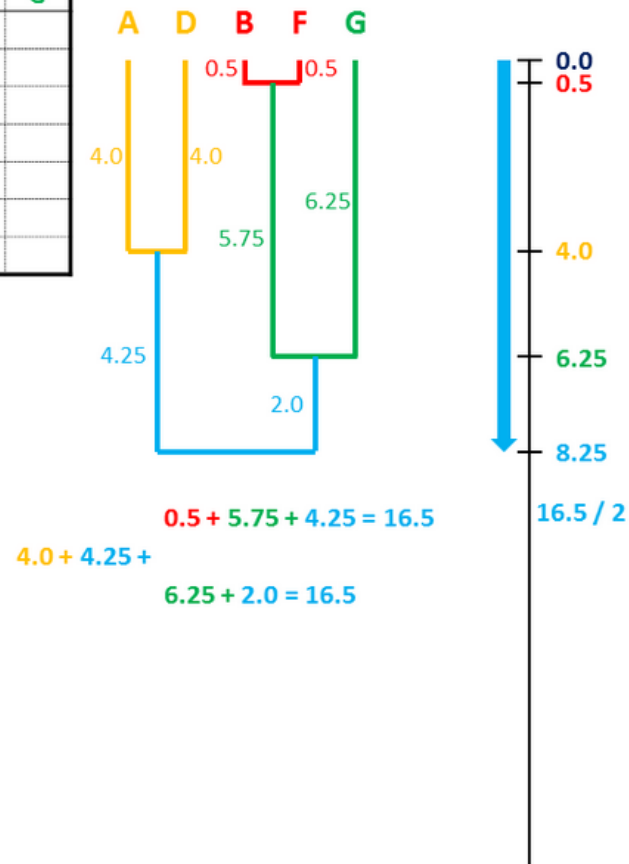
$$(36 + 35 + 28) / 3 = 33.0$$



UPGMA – 2nd Example (10/15)

	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	



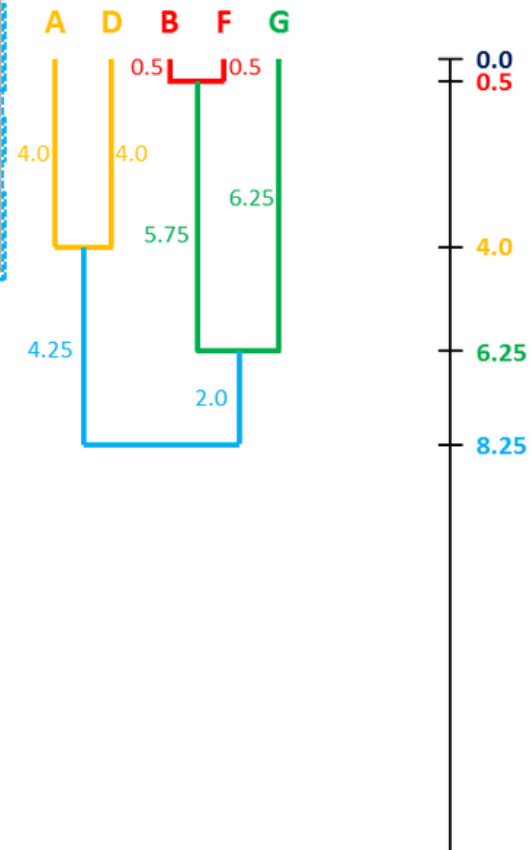
UPGMA – 2nd Example (11/15)

	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	

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

	AD BFG	C	E
AD BFG			
C	29.00		
E	32.60	41.00	

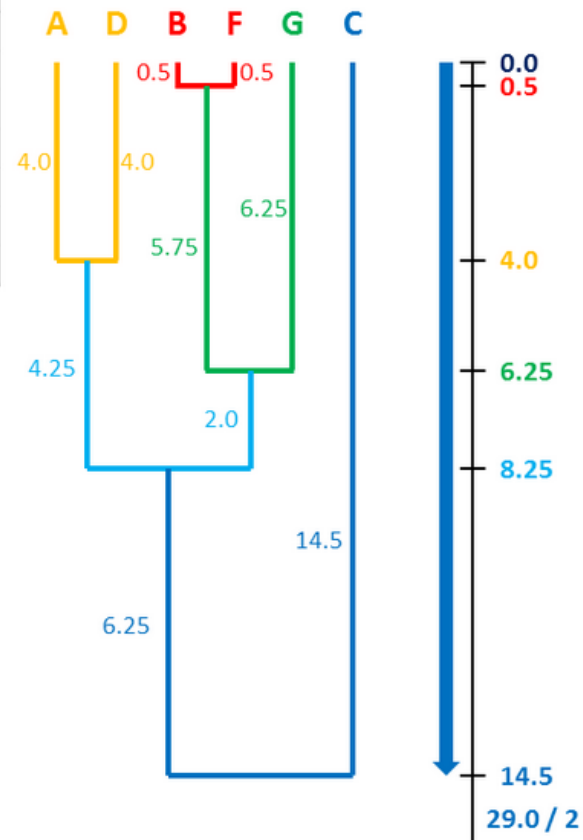
$$(33 + 36 + 31 + 35 + 28) / 5 = 32.60$$



UPGMA – 2nd Example (12/15)

	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 D B F G	C	E
A D B F G			
C	29.00		
E	32.60	41.00	



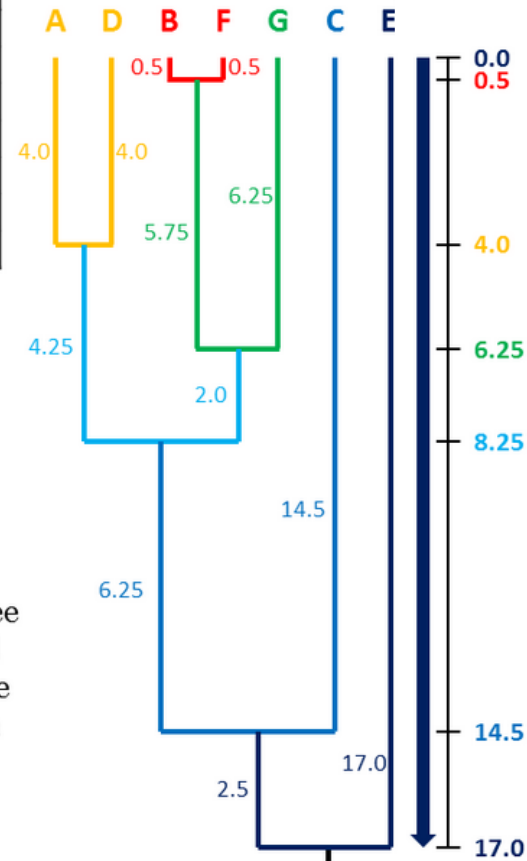
UPGMA – 2nd Example (13/15)

	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.



UPGMA – 2nd Example (14/15)

	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	

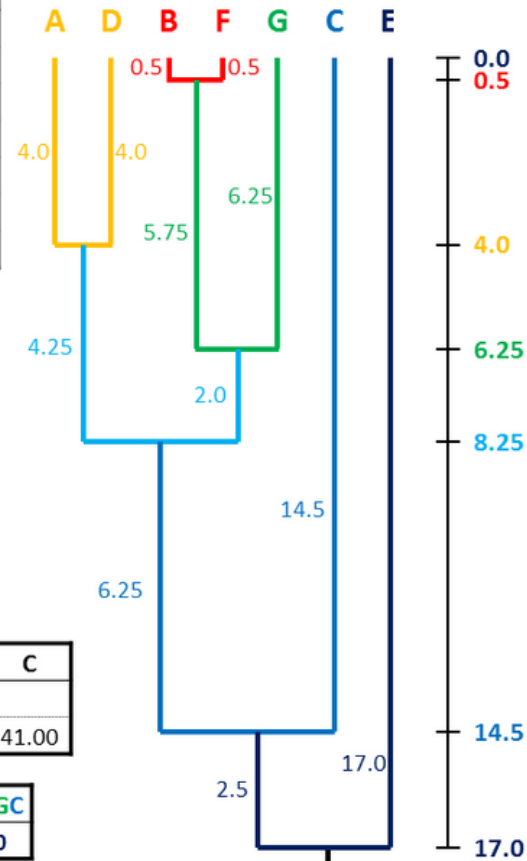
	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	8.00	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	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

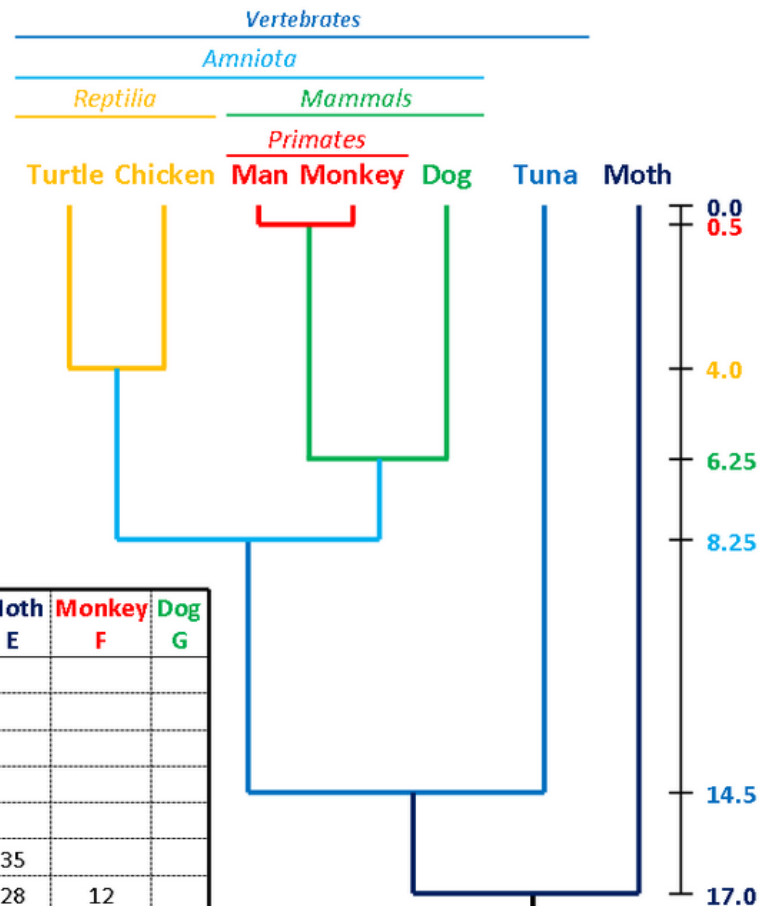
	ADBFGC
E	34.00



UPGMA – 2nd Example (15/15)

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	



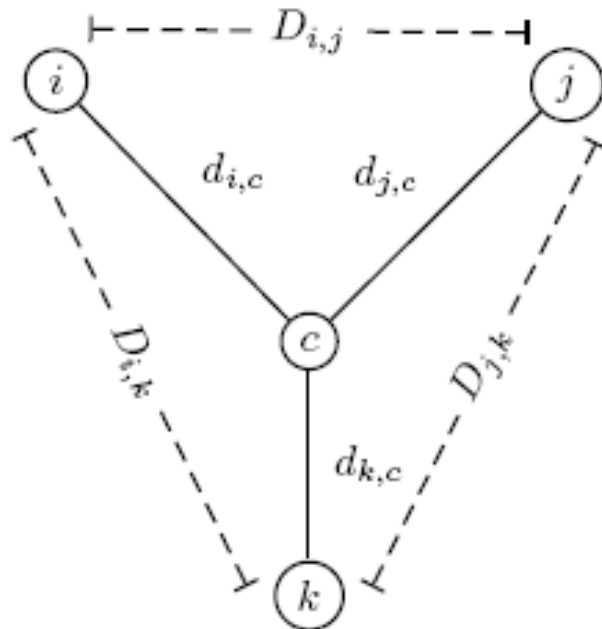
Μέθοδοι κατασκευής φυλογενετικών δέντρων (2/2)

- **Μέθοδοι βασισμένες σε χαρακτήρες – εξελεγκτικά μοντέλα:**
 - Μέθοδος μέγιστης συντήρησης (Maximum parsimony method).
- **Μέθοδοι βασισμένες στην απόσταση:**
 - Fitch and Margoliash Algorithm.
 - UPGMA (Unweighted pair-group method).



Υπολογισμός αποστάσεων με δεδομένη μήτρα διαστάσεων D , 3×3 - Fitch and Margoliash algorithm (1/2)

Διαδικό δέντρο **χωρίς ρίζα** με 3 φύλλα και ένα κέντρο



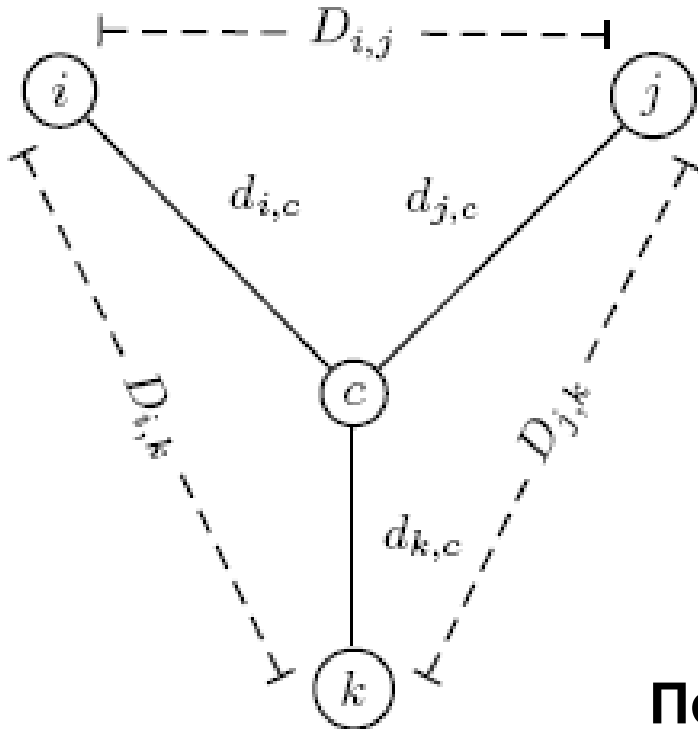
Παρατηρούμε ότι:

$$d_{ic} + d_{jc} = D_{ij}$$

$$d_{ic} + d_{kc} = D_{ik}$$

$$d_{jc} + d_{kc} = D_{jk}$$

Υπολογισμός αποστάσεων με δεδομένη
μήτρα διαστάσεων D , 3×3 - Fitch and Margoliash algorithm (2/2)



$$d_{ic} + d_{jc} = D_{ij}$$

$$+ \underline{d_{ic}} + \underline{d_{kc}} = \underline{D_{ik}}$$

$$2d_{ic} + \underbrace{d_{jc} + d_{kc}} = D_{ij} + D_{ik}$$

$$2d_{ic} + D_{jk} = D_{ij} + D_{ik}$$

$$d_{ic} = (D_{ij} + D_{ik} - D_{jk})/2$$

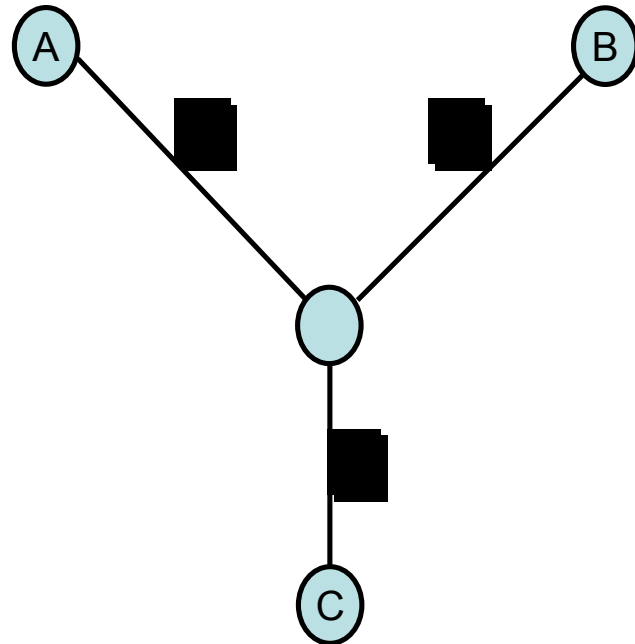
Παρομοίως,

$$d_{jc} = (D_{ij} + D_{jk} - D_{ik})/2$$

$$d_{kc} = (D_{ki} + D_{kj} - D_{ij})/2$$

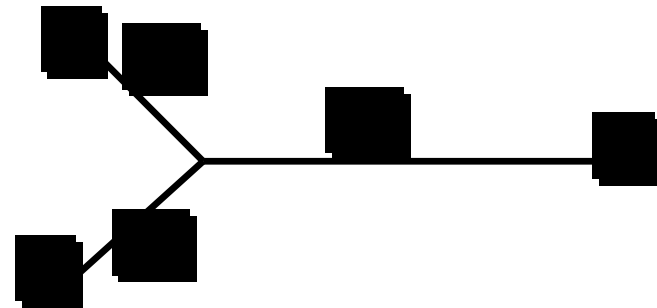


Υπολογισμός αποστάσεων με δεδομένη μήτρα διαστάσεων D , 3×3 - Παράδειγμα



	A	B	C
A	-	22	39
B	-	-	41
C	-	-	-

- $D_{AB} = a + b = 22$ (1)
 - $D_{AC} = a + c = 39$ (2)
 - $D_{BC} = b + c = 41$ (3)
- (2) – (3): $a - b = -2$ (4)
- (1) + (4): $2a = 20$, $a = 10$
- Από (1) και (2): $b = 12$, $c = 29$

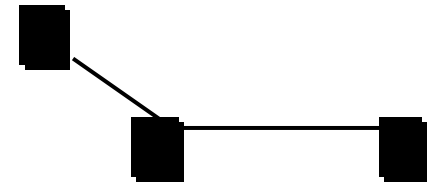
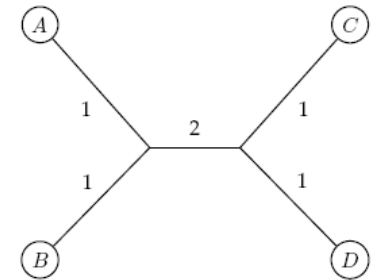


Προσθετικές και μη προσθετικές μήτρες (1/2)

- Για $n=4$: Επίλυση έξι εξισώσεων με 5 μεταβλητές.
- Αυτό δεν είναι πάντα εφικτό για $n > 3$.
- Η μήτρα D είναι προσθετική εάν υπάρχει δέντρο T με $d_{ij}(T) = D_{ij}$.
- ή διαφορετικά:
 - $d(a,b) + d(b,c) = d(a,c)$.
 - για το μονοπάτι εξέλιξης μίας αλληλουχίας $a \rightarrow b \rightarrow c$.



	A	B	C	D
A	0	2	4	4
B	2	0	4	4
C	4	4	0	2
D	4	4	2	0



Προσθετικές και μη προσθετικές μήτρες (2/2)

Μη προσθετική
μήτρα



	A	B	C	D
A	0	2	2	2
B	2	0	3	2
C	2	3	0	2
D	2	2	2	0

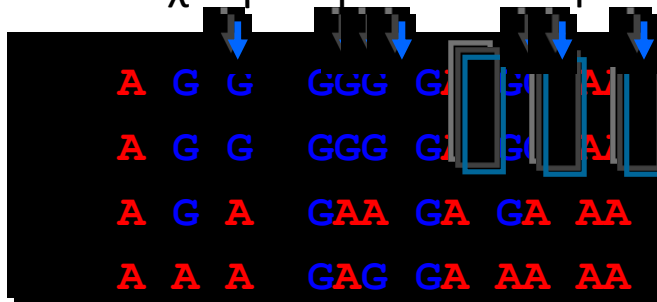
?

Each tree does correspond to one distance matrix,
but the opposite is not always true



Example of Distance Analysis (1/3)

A. Έστω ότι έχουμε την πολλαπλή στοίχιση:



B. Οι αποστάσεις μεταξύ των αλληλουχιών είναι:

$$n_{AB} = 3$$

$$n_{AC} = 7$$

$$n_{AD} = 8$$

$$n_{BC} = 6$$

$$n_{BD} = 7$$

$$n_{CD} = 3$$



	A	B	C	D
A	-	3	7	8
B	-	-	6	7
C	-	-	-	3
D	-	-	-	-



Example of Distance Analysis (2/3)

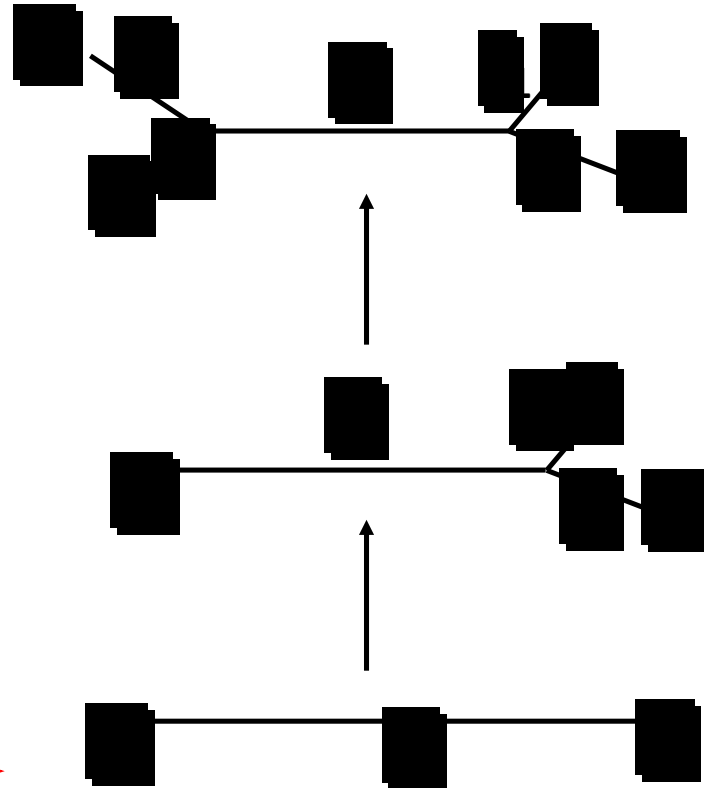
	A	B	C	D
A	-	3	7	8
B	-	-	6	7
C	-	-	-	3
D	-	-	-	-



	A	C	D
A	-	7	8
C	-	-	3
D	-	-	-

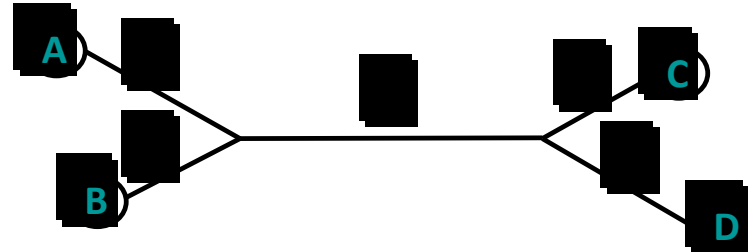


	A	C
A	-	7
C	-	-



Example of Distance Analysis (3/3)

	A	B	C	D
A	-	3	7	8
B	-	-	6	7
C	-	-	-	3
D	-	-	-	-



- Το προτεινόμενο φυλογενετικό δέντρο για τις αλληλουχίες A-D. Το άθροισμα των μηκών των ακμών μεταξύ δύο αλληλουχιών στο δέντρο έχει την ίδια τιμή με την απόσταση μεταξύ των αλληλουχιών.

Fitch and Margoliash

Algorithm – Μέθοδος (1/2)

1. Βρίσκουμε τις αλληλουχίες με τη μεγαλύτερη ομοιότητα, για παράδειγμα την A και τη B.
2. Θεωρούμε τις υπόλοιπες αλληλουχίες σαν μία αλληλουχία.
3. Υπολογίζουμε τη μέση απόσταση της A από όλες τις άλλες αλληλουχίες και τη μέση απόσταση της B από όλες τις άλλες αλληλουχίες.
4. Χρησιμοποιώντας αυτές τις σχέσεις υπολογίζουμε τα a και b.
5. Στη συνέχεια συμπεριφερόμαστε στις A και B σαν μία αλληλουχία AB, υπολογίζουμε τη μέση απόσταση της AB από τις υπόλοιπες αλληλουχίες και κατασκευάζουμε καινούργια μήτρα αποστάσεων.



Fitch and Margoliash Algorithm – Μέθοδος (2/2)

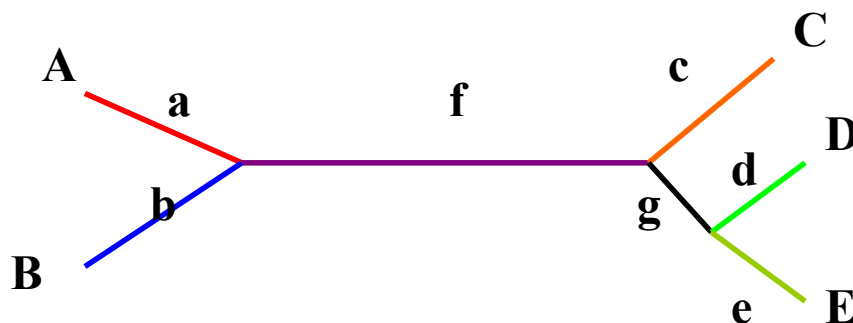
6. Βρίσκουμε το επόμενο ζευγάρι αλληλουχιών με τη μεγαλύτερη ομοιότητα και επαναλαμβάνουμε την παραπάνω διαδικασία.
7. Επαναλαμβάνω τη συνολική διαδικασία ξεκινώντας με όλα τα πιθανά ζεύγη των αλληλουχιών A και B, A και C, A και D, κτλ.
8. Υπολογίζουμε την προβλεπόμενη απόσταση μεταξύ κάθε ζεύγους αλληλουχιών για κάθε δέντρο και προτείνουμε το βέλτιστο δέντρο.



Fitch and Margoliash

Algorithm (5 sequences) (1/7)

- Ο αλγόριθμος μπορεί να εφαρμοστεί και σε περισσότερες αλληλουχίες. Έστω ότι έχουμε τις παρακάτω αποστάσεις.
- Εντοπίζουμε τις πιο κοντινές αλληλουχίες:

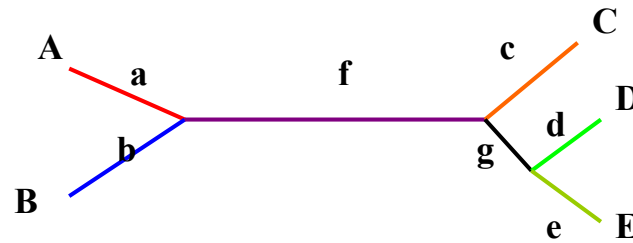


	A	B	C	D	E
A	--	22	39	39	41
B	--	--	41	41	43
C	--	--	--	18	20
D	--	--	--	--	10
E	--	--	--	--	--



Fitch and Margoliash Algorithm (5 sequences) (2/7)

	A	B	C	D	E
A	--	22	39	39	41
B	--	--	41	41	43
C	--	--	--	18	20
D	--	--	--	--	10
E	--	--	--	--	--



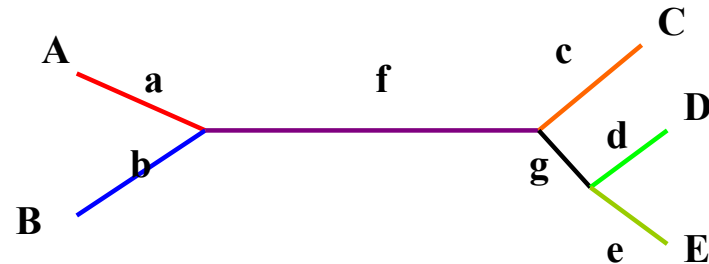
	D	E	AVG ABC
D	--	10	32.7
E	--	--	34.7
AVG ABC	--	--	--

- Δημιουργώ καινούργιο πίνακα αποστάσεων συνδυάζοντας τις υπόλοιπες αλληλουχίες (A, B, C):
 - **Απόσταση από D στο ABC:** Μέση απόσταση του καθένα από το D, $((39 + 41 + 18) / 3 = 32.7)$.
 - **Απόσταση από E στο ABC:** Μέση απόσταση του καθένα από το E, $((41+43+20)/3 = 34.7)$.



Fitch and Margoliash Algorithm (5 sequences) (3/7)

	D	E	AVG ABC
D	--	10	32.7
E	--	--	34.7
AVG ABC	--	--	--



$$d + e = 10 \quad (1)$$

$$\frac{DA + DB + DC}{3} = \frac{d + g + f + a}{3} + \frac{d + g + f + b}{3} + \frac{d + g + c}{3} = d + g + \frac{a + b + f + c}{3} = 32.7 \quad (2)$$

$$\frac{EA + EB + EC}{3} = \frac{e + g + f + a}{3} + \frac{e + g + f + b}{3} + \frac{e + g + c}{3} = e + g + \frac{a + b + f + c}{3} = 34.7 \quad (3)$$

$$(2) - (3) \quad d - e = -2.7 \quad (4)$$

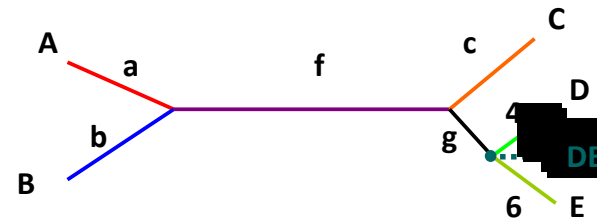
$$(1) + (4) \quad 2d = 7.3 \Rightarrow d = 3.65 \quad e = 6.35$$



Fitch and Margoliash Algorithm (5 sequences) (4/7)

- Θεωρώ την DE μία αλληλουχία.
- Δημιουργώ καινούργια μήτρα αποστάσεων.
- Η απόσταση από την A στην DE είναι η μέση τιμή της A προς την D και της A προς την E.

	A	B	C	D	E
A	--	22	39	39	41
B	--	--	41	41	43
C	--	--	--	18	20
D	--	--	--	--	10
E	--	--	--	--	--

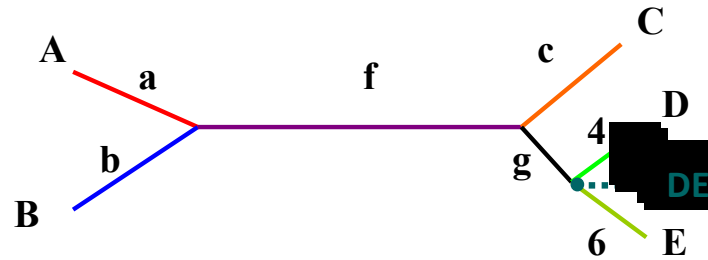


	A	B	C	(DE)
A	--	22	39	40
B	--	--	41	42
C	--	--	--	19
(DE)	--	--	--	--



Ή πιο αναλυτικά...

	A	B	C	D	E
A	--	22	39	39	41
B	--	--	41	41	43
C	--	--	--	18	20
D	--	--	--	--	10
E	--	--	--	--	--



$$\frac{AD}{2} + \frac{AE}{2} = \frac{a+f+g+l}{2} + \frac{a+f+g+i}{2} = a+f+g+i$$

$$a+f+g = AD - l = 39 - l = 35$$

$$a+f+g = AE - i = 41 - i = 35$$

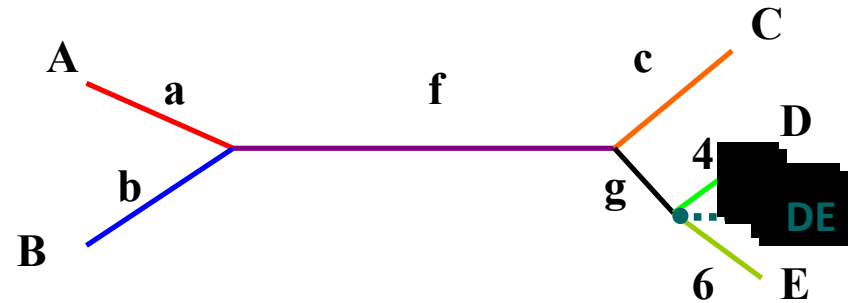
A to DE = 40

	A	B	C	(DE)
A	--	22	39	40
B	--	--	41	42
C	--	--	--	19
(DE)	--	--	--	--



Fitch and Margoliash Algorithm (5 sequences) (5/7)

	A	B	C	(DE)
A	--	22	39	40
B	--	--	41	42
C	--	--	--	19
(DE)	--	--	--	--

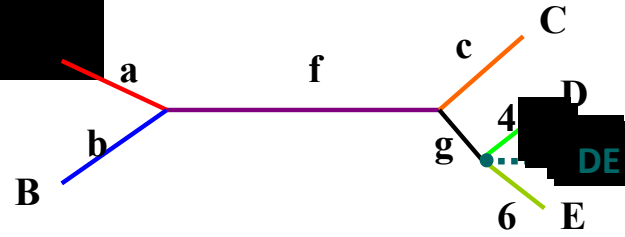


- Επομένως χρησιμοποιούμε την C,DE σαν μία αλληλουχία.
- Δημιουργώ καινούργιο πίνακα συνδυάζοντας τις υπόλοιπες αλληλουχίες, δηλ, τις A,B:
 - Μέση απόσταση της DE από την AB: $(40+42)/2 = 41$.
 - Μέση απόσταση της C από την AB: $(39+41)/2 = 40$.



Fitch and Margoliash Algorithm (5 sequences) (6/7)

[Redacted]



[Redacted]

$$\frac{a+f+g+i}{2} + \frac{b+f+g+i}{2} = f+g+i + \frac{a+b}{2} = 1 \quad (1)$$

[Redacted]

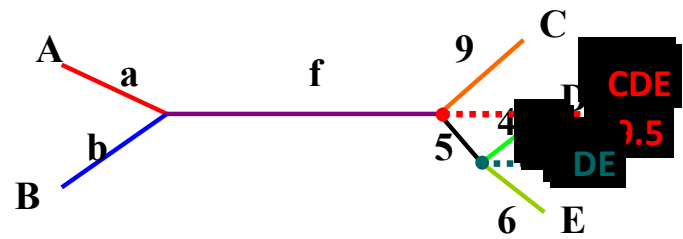
$$\frac{c+f+a}{2} + \frac{c+f+b}{2} = c+f + \frac{a+b}{2} = 1 \quad (2)$$

$$c-g = . \quad (3)$$

$$c+g = 9 \quad (4)$$

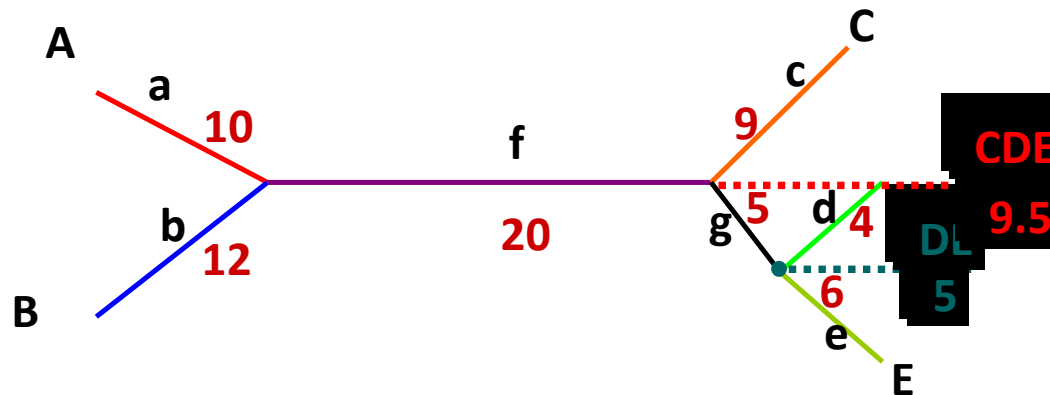
[Redacted]

	A	B	C	(DE)
A	--	22	39	40
B	--	--	41	42
C	--	--	--	19
(DE)	--	--	--	--



Fitch and Margoliash Algorithm (5 sequences) (7/7)

- Συνεχίζοντας με τον ίδιο τρόπο υπολογίζουμε όλα τα μήκη:



Fitch and Margoliash Algorithm – 2nd Example (1/4)

	A	B	C	D
A		3	7	8
B			6	7
C				3
D				

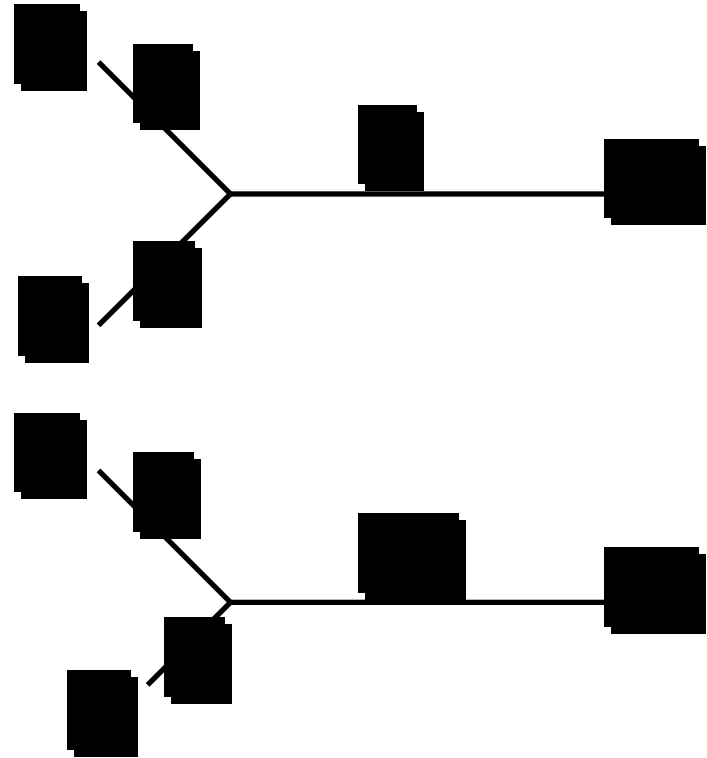
$$a+b=3 \quad (1)$$

$$\text{dist}(A,CD) = a+x = (AC+AD)/2 = 7.5 \quad (2)$$

$$\text{dist}(B,CD) = b+x = (BC+BD)/2 = 6.5 \quad (3)$$

$$(2)-(3) \rightarrow a-b = 1, (1) \rightarrow a+b=3 \rightarrow$$

$$a = 2, b = 1 \ \& \ x=5.5$$



Fitch and Margoliash

Algorithm – 2nd Example (2/4)

	A	B	C	D
A		3	7	8
B			6	7
C				3
D				

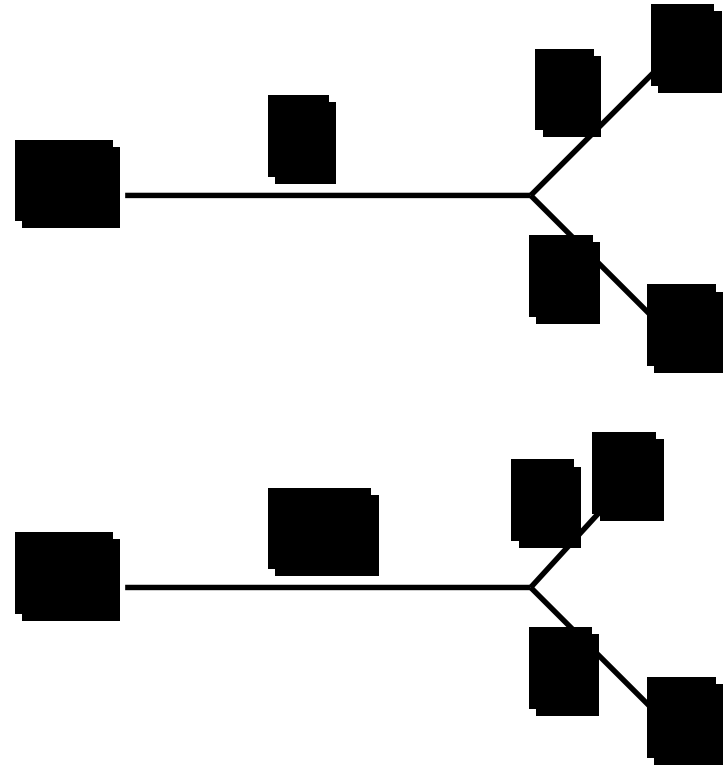
$$c+d=3 \quad (1)$$

$$\text{dist}(C,AB) = c+y = (CA+CB)/2 = 6.5 \quad (2)$$

$$\text{dist}(D,AB) = d+y = (DA+DB)/2 = 7.5 \quad (3)$$

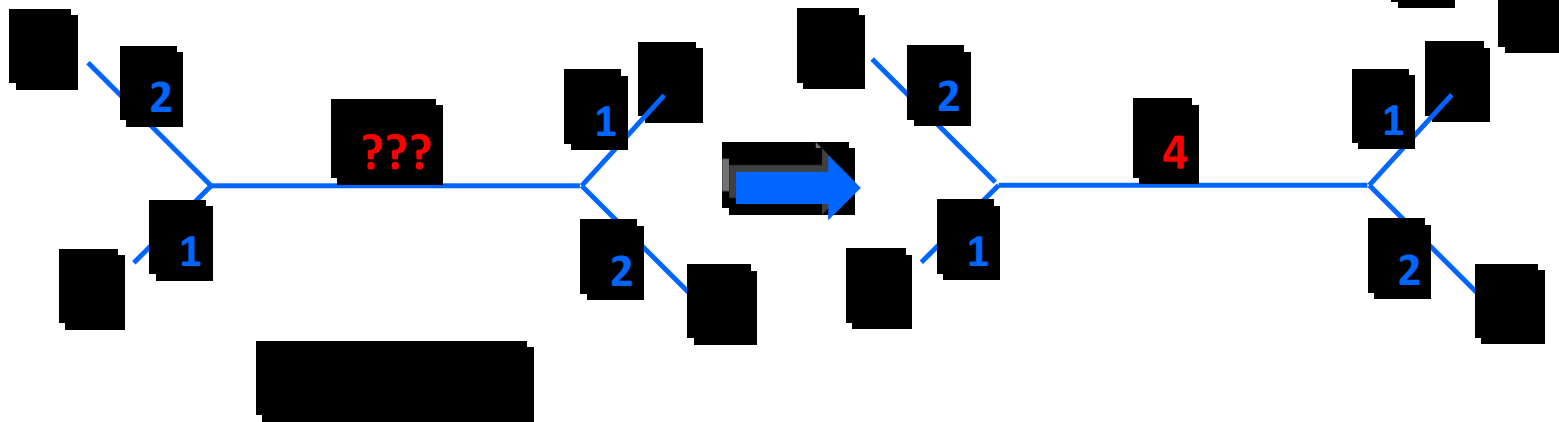
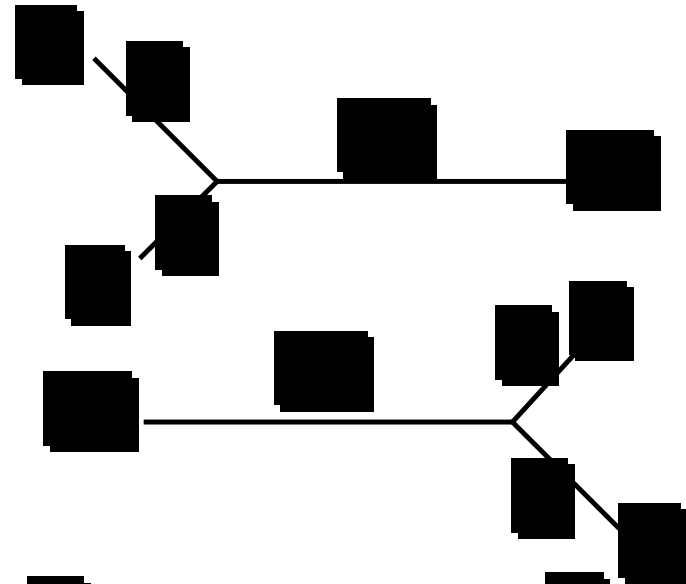
$$(3)-(2) \rightarrow d-c = 1, (1) \rightarrow c+d=3 \rightarrow$$

$$\mathbf{d = 2, c = 1 \ \& \ y=5.5}$$



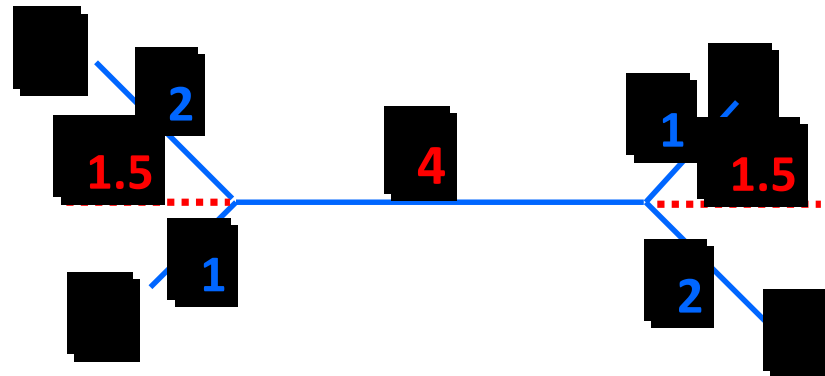
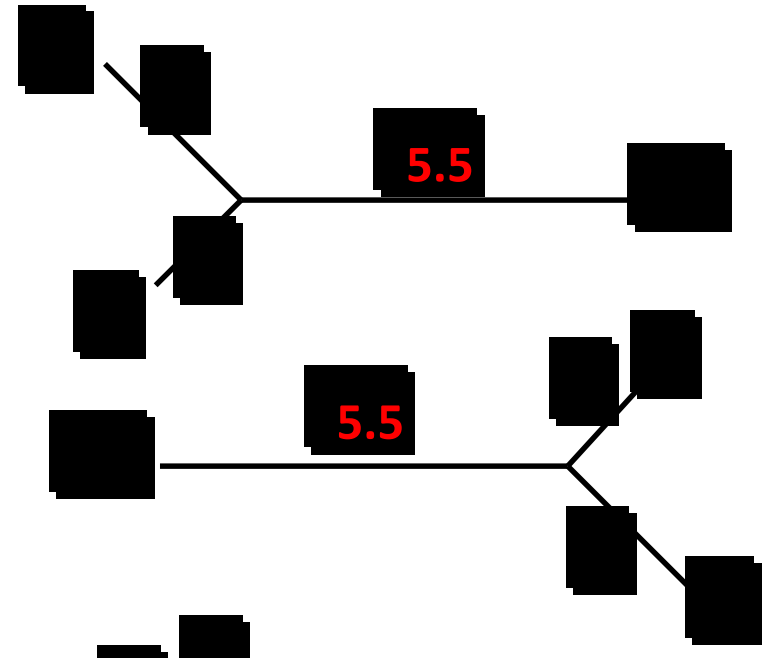
Fitch and Margoliash Algorithm – 2nd Example (3/4)

	A	B	C	D
A		3	7	8
B			6	7
C				3
D				

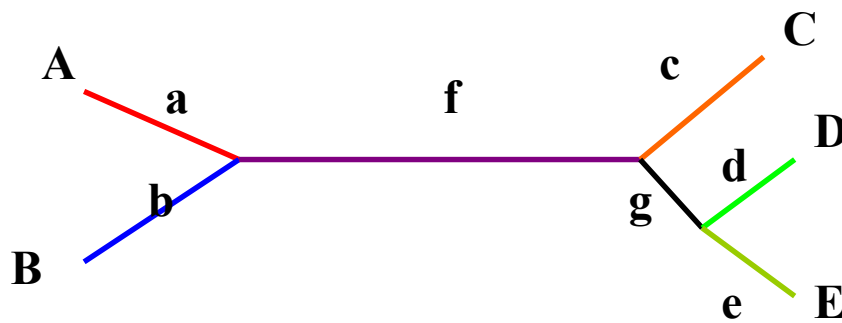


Fitch and Margoliash Algorithm – 2nd Example (4/4)

	A	B	C	D
A		3	7	8
B			6	7
C				3
D				



Fitch and Margoliash Algorithm (1/5)

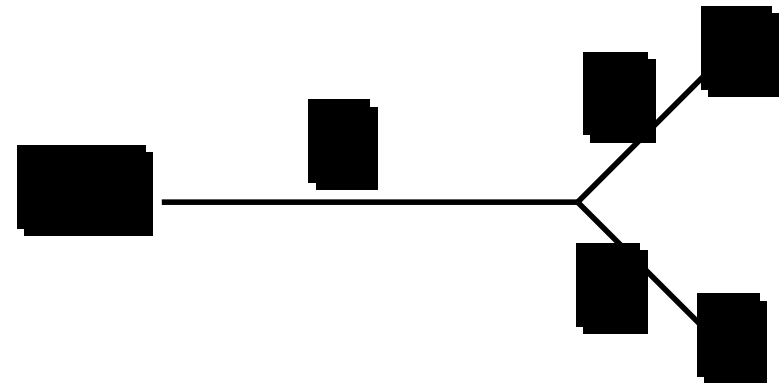


	A	B	C	D	E
A	--	22	39	39	41
B	--	--	41	41	43
C	--	--	--	18	20
D	--	--	--	--	10
E	--	--	--	--	--



Fitch and Margoliash Algorithm (2/5)

	A	B	C	D	E
A	--	22	39	39	41
B	--	--	41	41	43
C	--	--	--	18	20
D	--	--	--	--	10
E	--	--	--	--	--



$$d+e=10 \quad (1)$$

$$\text{dist}(D,ABC) = d+x = (DA+DB+DC)/3 = 32.67 \quad (2)$$

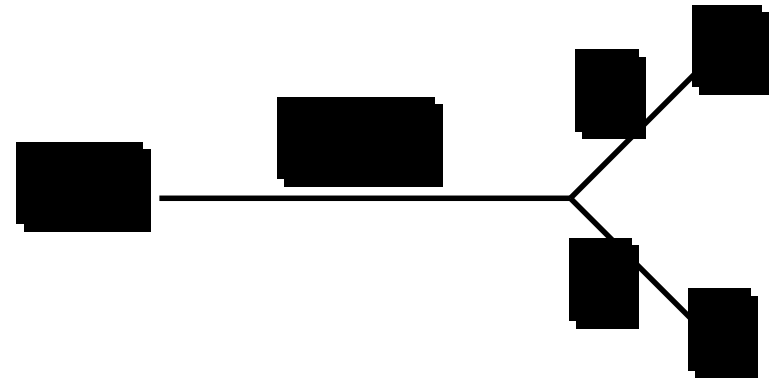
$$\text{dist}(E, ABC) = e+x = (EA+EB+EC)/3 = 34.67 \quad (3)$$

$$(1) \quad (2) \quad (3) \rightarrow e = 6, d = 4, x = 28,67$$



Fitch and Margoliash Algorithm (3/5)

	A	B	C	D	E
A	--	22	39	39	41
B	--	--	41	41	43
C	--	--	--	18	20
D	--	--	--	--	10
E	--	--	--	--	--



(1) (2) (3) $\rightarrow e = 6, d = 4, x = 28,67$

$$A \rightarrow DE: (AD+AE)/2=40$$

$$B \rightarrow DE: (BD+BE)/2=42$$

$$C \rightarrow DE: (CD+CE)/2=19$$

	A	B	C	(DE)
A	--	22	39	40
B	--	--	41	42
C	--	--	--	19
(DE)	--	--	--	--



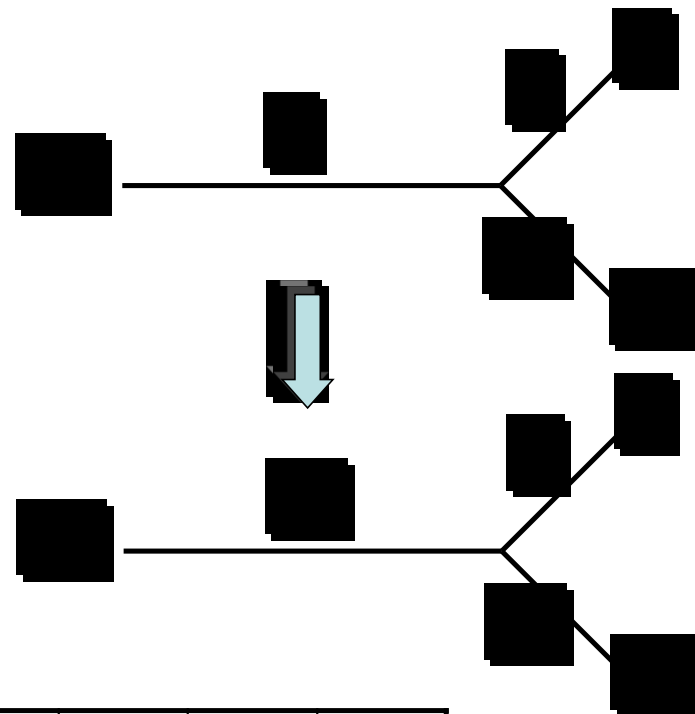
Fitch and Margoliash Algorithm (4/5)

	A	B	C	(DE)
A	--	22	39	40
B	--	--	41	42
C	--	--	--	19
(DE)	--	--	--	--

→ c = 9, de = 10, γ = 31

A → CDE: $(AC+AD+AE)/3=39.67$

B → CDE: $(BC+BD+BE)/3=41.67$



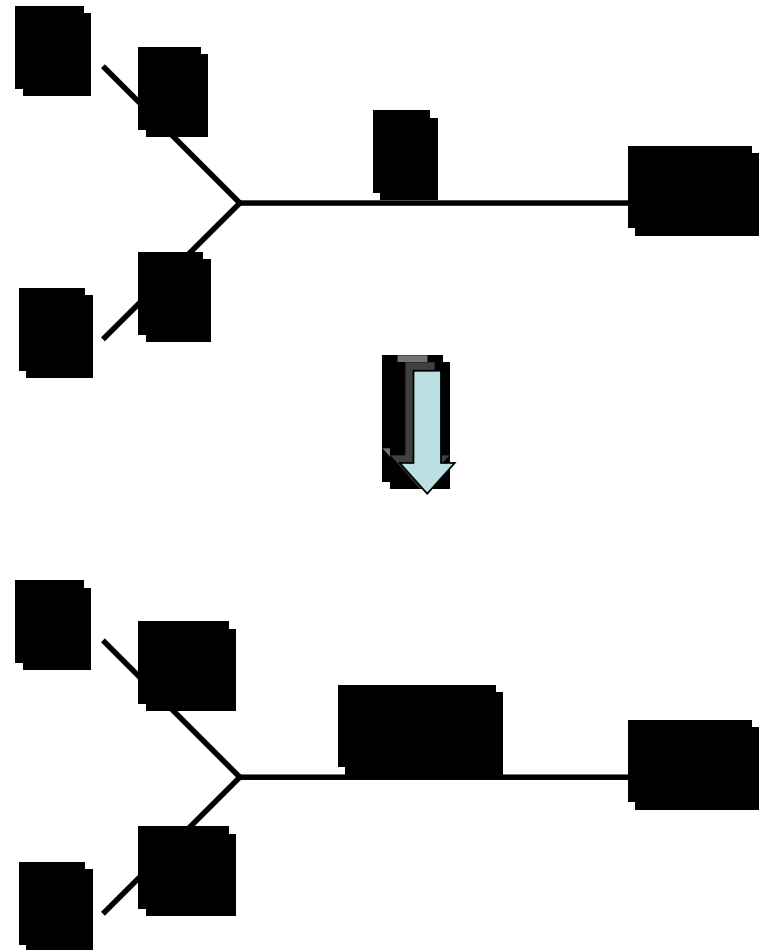
	A	B	CDE
A		22	39.67
B			41.67
CDE			



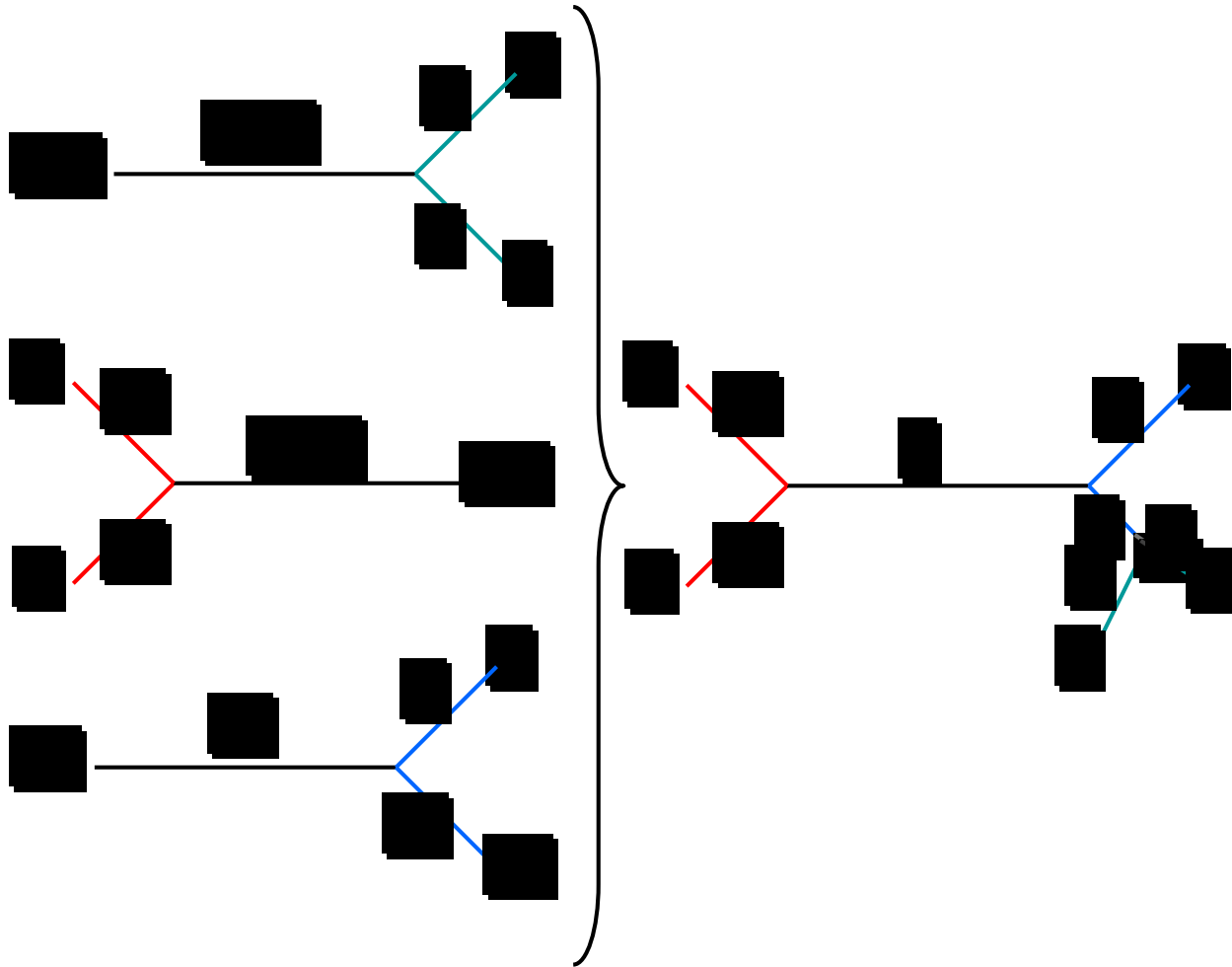
Fitch and Margoliash Algorithm (5/5)

	A	B	CDE
A		22	39.67
B			41.67
CDE			

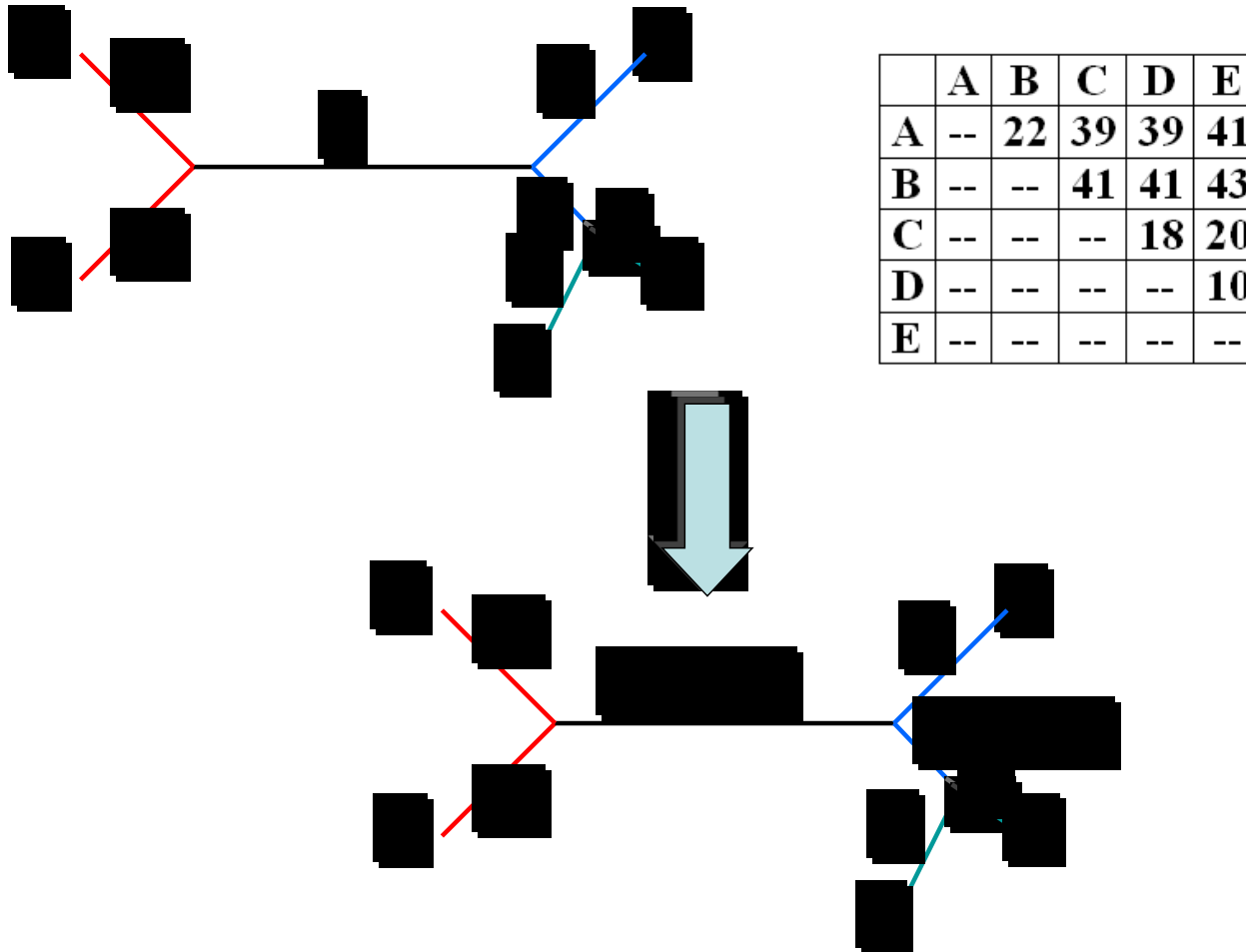
→ $a = 10$, $b = 12$, $z = 29.67$



Τελικά... (1/2)



Τελικά... (2/2)



Τέλος Ενότητας



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Σημείωμα Αναφοράς

- Copyright Πανεπιστήμιο Δυτικής Μακεδονίας, Τμήμα Μηχανικών Πληροφορικής και Τηλεπικοινωνιών, Αγγελίδης Παντελής. «**Βιοπληροφορική**». Έκδοση: 1.0. Κοζάνη 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <https://eclass.uowm.gr/courses/ICTE102/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Όχι Παράγωγα Έργα Μη Εμπορική Χρήση 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Ως Μη Εμπορική ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους
υπερσυνδέσμους.

