

Βιοπληροφορική

Ενότητα 9: Φυλογενετική ανάλυση

Αν. καθηγήτης 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.



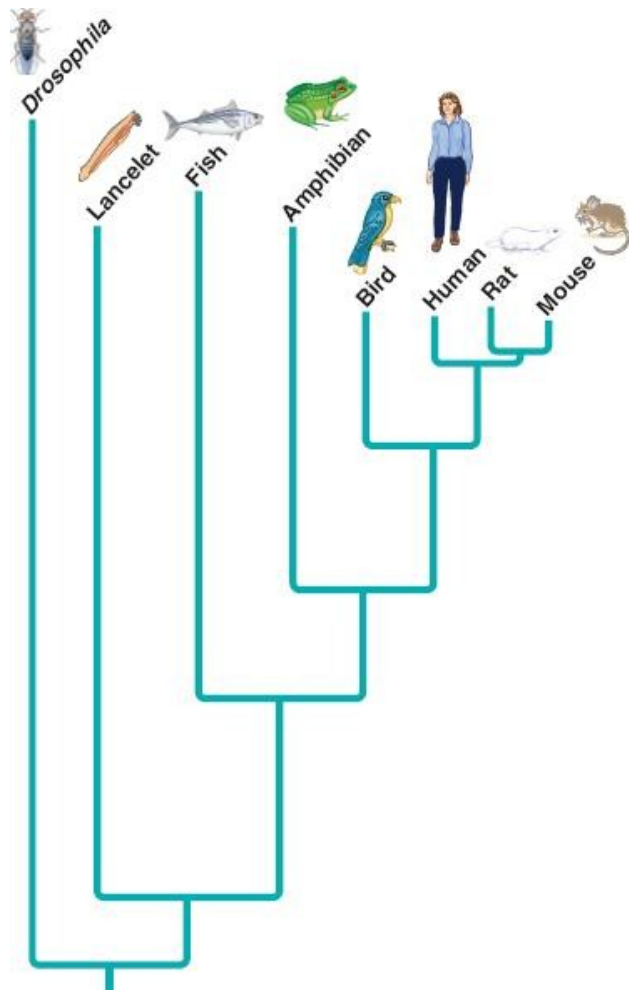
Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Φυλογενετική ανάλυση – Ορισμός



- Η φυλογενετική ανάλυση μίας οικογένειας DNA/RNA ή πρωτεϊνών ορίζει τον τρόπο με τον οποίο προήλθε η οικογένεια κατά την εξέλιξη.
- Οι εξελεγκτικές σχέσεις μεταξύ των αλληλουχιών αναπαρίστανται τοποθετώντας τις αλληλουχίες σε βραχίονες δέντρων.
- Δύο παρόμοιες αλληλουχίες τοποθετούνται σε γειτονικούς βραχίονες, οι οποίοι ενώνονται σε έναν κοινό βραχίονα.
- Η φυλογενετική ανάλυση στηρίζεται στη στοίχιση πολλαπλών ακολουθιών.



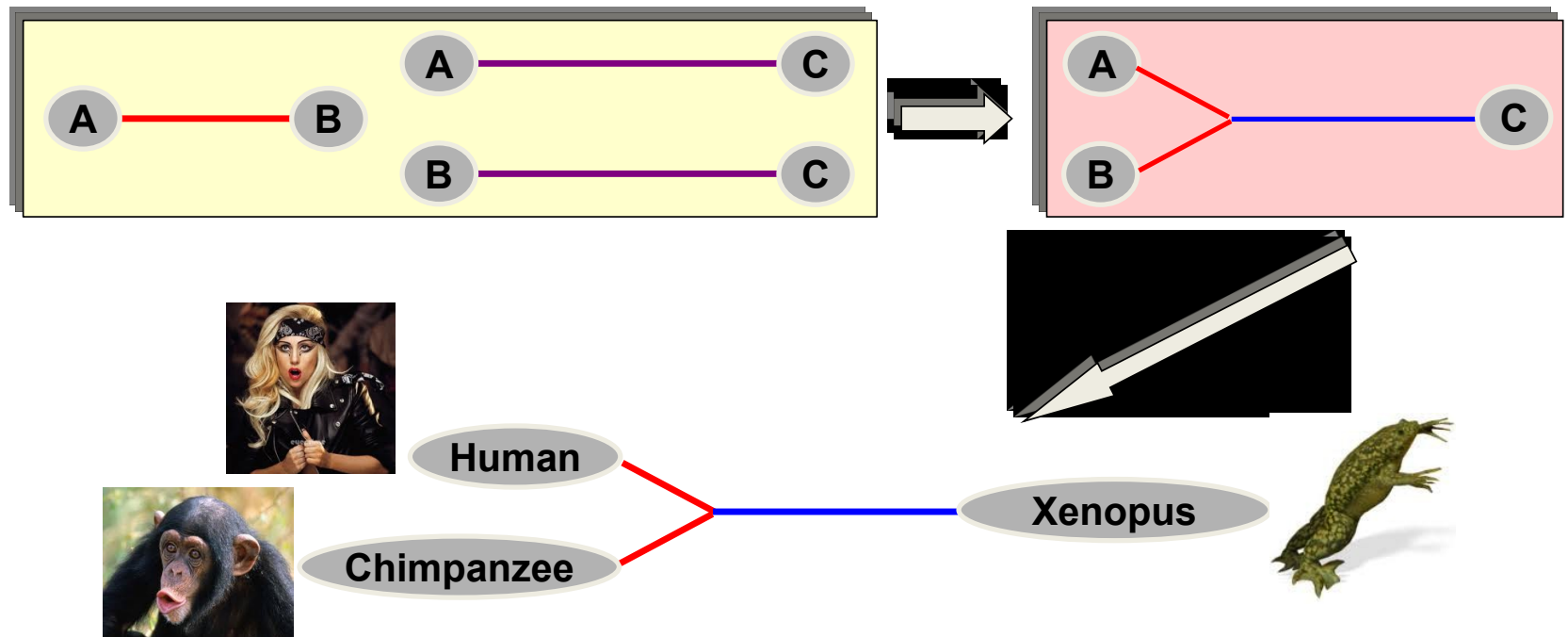
Σχέση φυλογενετικής ανάλυσης και σύγκρισης αλληλουχιών (1/2)

- Έστω δύο αλληλουχίες DNA ή πρωτεϊνών, protein A & protein B:
 - ανήκουν σε διαφορετικά είδη, π.χ. άνθρωπος και χιμπατζής,
 - μοιράζονται αυξημένη ομοιότητα.
- Πιθανά να μοιράζονται τον ίδιο πρόγονο.
- Έστω ότι μία 3^η πρωτεΐνη από ένα τρίτο είδος, π.χ. βάτραχος, συγκριθεί με τις υπόλοιπες και διαφέρει.
- Πιθανά, ο 3^{ος} οργανισμός στον οποίο ανήκει η 3^η πρωτεΐνη να «απέχει εξελεγκτικά» πολύ από τους άλλους 2 οργανισμούς.



Σχέση φυλογενετικής ανάλυσης και σύγκρισης αλληλουχιών (2/2)

- Protein A – Human
- Protein B – Chimpanzee
- Protein C – Xenopus



Παρελθόν: Φυλογενετική ανάλυση βάσει μορφολογικών χαρακτηριστικών

- Η κατάταξη των οργανισμών γινόταν με βάση τη **συνολική τους ομοιότητα**.
- Η κατάταξη των οργανισμών με βάση τη μορφολογική τους ομοιότητα μπορεί να αντανακλά τη φυλογένεση των εξελεγκτικών μονάδων. Εκείνα που μοιάζουν περισσότερο έχουν μεγάλη πιθανότητα να προέρχονται από κοινό πρόγονο.
- Ωστόσο αυτό δεν είναι πάντα σωστό, γι' αυτό...
- **ΣΗΜΕΡΑ** οι βιολόγοι βασίζονται στις αλληλουχίες DNA για την ανακατασκευή των εξελεγκτικών δέντρων.



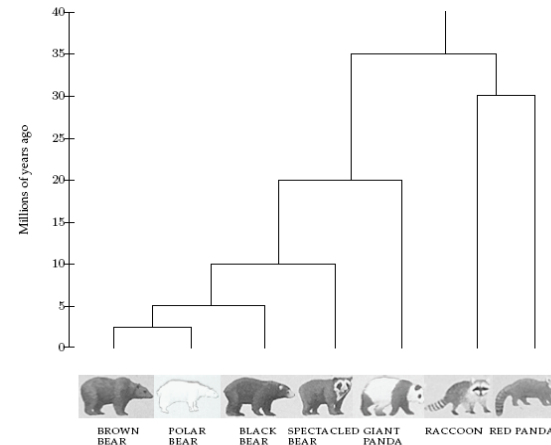
Εξέλιξη και ανάλυση DNA: Το αίνιγμα του γιγαντιαίου πάντα (1/2)

- Για περισσότερα από 100 χρόνια οι επιστήμονες δεν μπορούσαν να συμφωνήσουν αν το γιγαντιαίο πάντα έπρεπε να ταξινομηθεί στις αρκούδες ή στα ρακούν.
- Το 1870 εξετάστηκαν οστά από πάντα, τα οποία έμοιαζαν περισσότερο με τα οστά ενός κόκκινου πάντα παρά με εκείνα των αρκούδων.
- Τα κόκκινα πάντα ανήκουν στην οικογένεια των ρακούν.
- Το γιγαντιαίο πάντα μοιάζει στην αρκούδα αλλά έχει ορισμένα ασυνήθιστα χαρακτηριστικά για αρκούδες και συνηθισμένα για ρακούν, π.χ. δεν πέφτει σε χειμερία νάρκη και βγάζει κραυγές σαν το ρακούν.



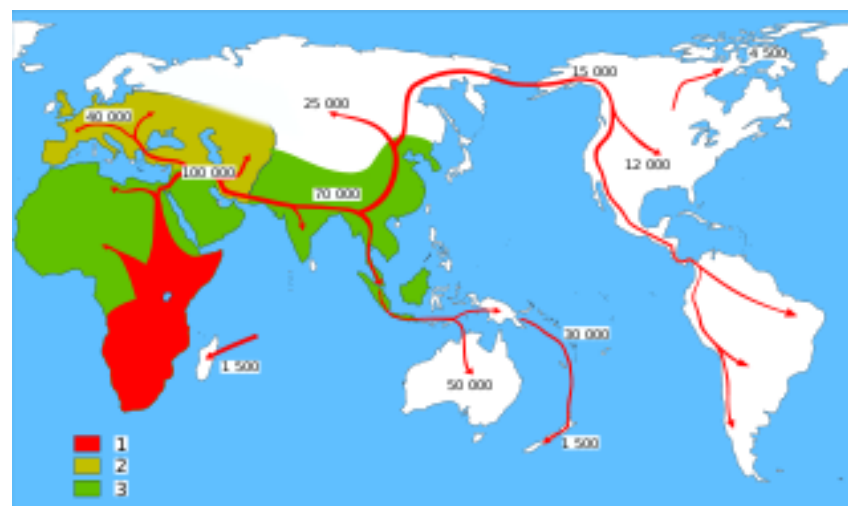
Εξέλιξη και ανάλυση DNA: Το αίνιγμα του γιγαντιαίου πάντα (2/2)

- Το 1985, ο Steven O'Brien και οι συνεργάτες του έλυσαν την ταξινόμηση του γιγαντιαίου πάντα χρησιμοποιώντας αλληλουχίες DNA και κατάλληλους αλγόριθμους στοίχισης.
- Χρησιμοποιήθηκαν περίπου 500,000 νουκλεοτίδια.

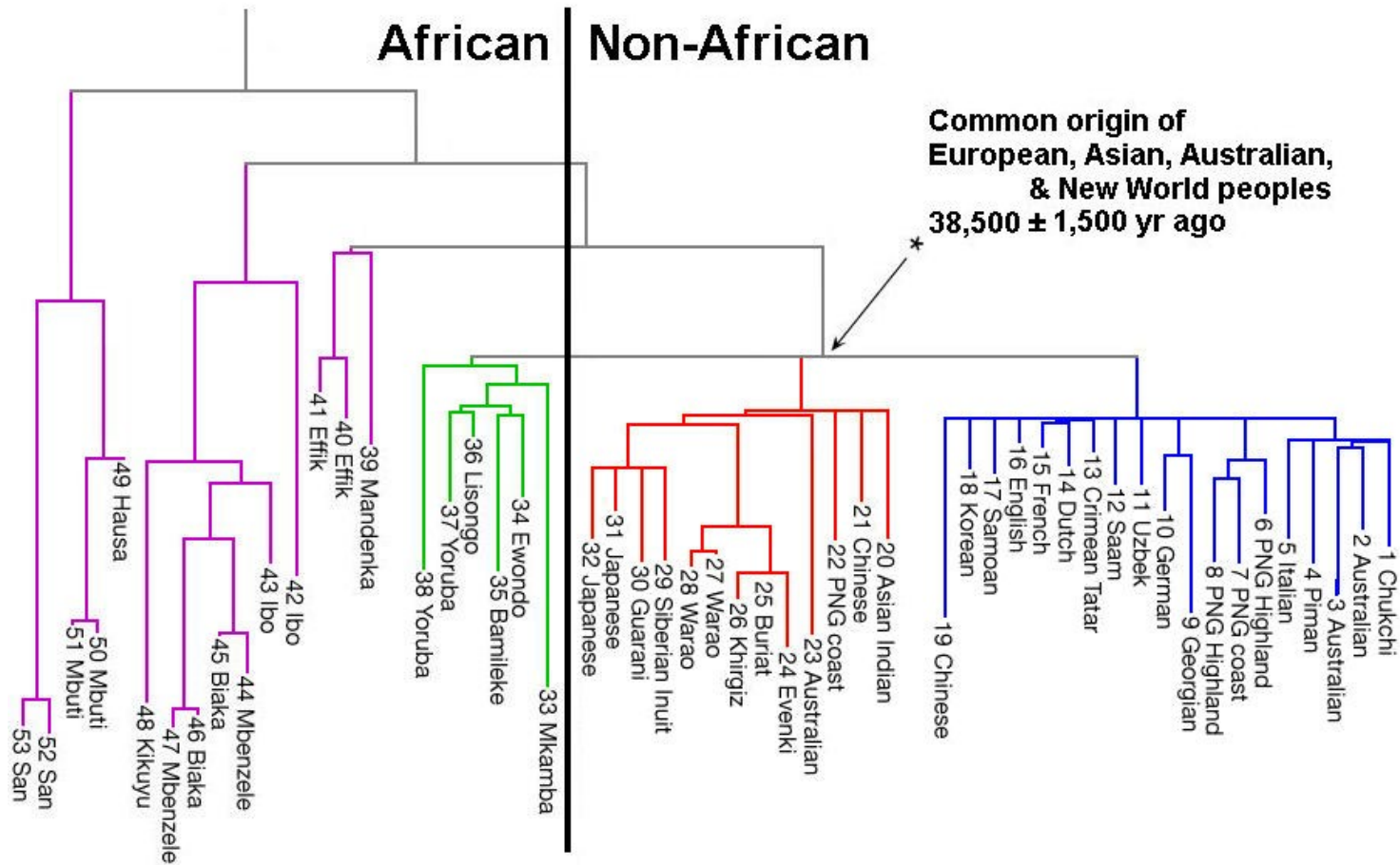


“Out of Africa” theory

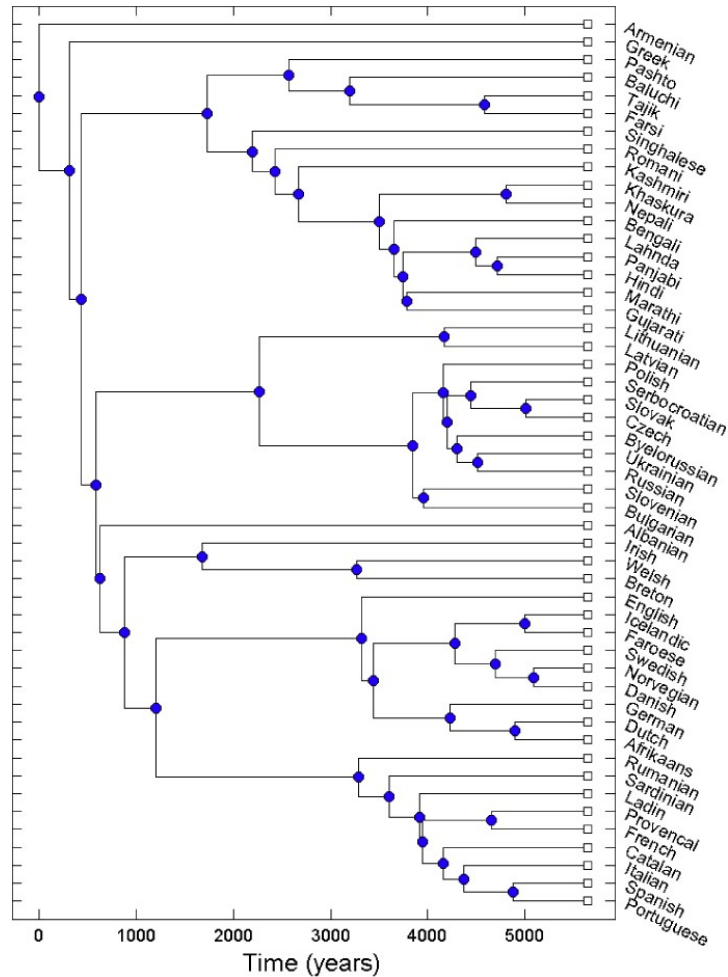
- Την ίδια περίπου εποχή η Rebecca Cann, ο Mark Stoneking και ο Allan Wilson διατύπωσαν τη Θεωρία «Out of Africa», σύμφωνα με την οποία ο πιο αρχαίος πρόγονος του ανθρώπου έζησε στην Αφρική πριν από 200,000.
- Εξετάστηκαν 650,000 γενετικοί δείκτες από 1000 άτομα που ανήκαν σε 51 διαφορετικούς πληθυσμούς.
- **Αποτέλεσμα: Κατασκευάστηκε ένα εξελεγκτικό δέντρο.**



Εξελεγκτικό δέντρο (1/2)



Εξελεγκτικό δέντρο (2/2)



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).



Identity (ομοιότητα) & homology (ομολογία)

- **Ομοιότητα** παρουσιάζουν οι αλληλουχίες που αποτελούνται από ίδια – ταυτόσημα κατάλοιπα (νουκλεοτίδια, αμινοξέα).
- **Ομολογία** παρουσιάζουν οι αλληλουχίες που **προέρχονται από τον ίδιο πρόγονο**. Για παράδειγμα, πρωτεΐνες με στατιστικά σημαντικό ποσοστό ομοιότητας και, πιθανά, παρόμοια δομή και λειτουργία.
- Όταν δύο αλληλουχίες παρουσιάζουν ομοιότητα:
 - >35%, είναι πιθανά και ομόλογες.
 - 35-20%, δεν είναι ασφαλή τα συμπεράσματα.
 - <20%, τυχαία ομοιότητα.

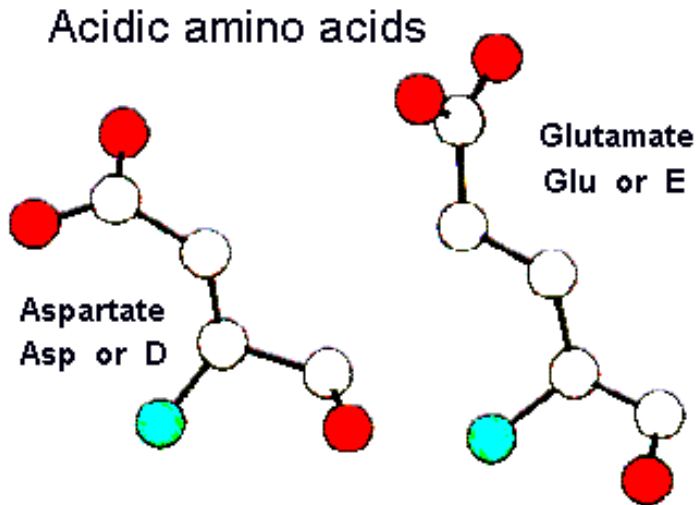


Identity and Similarity

Identity

Sequence 1	M	A	L	V	D	D	M	F	R
Match	M	A		V	D	+	M	F	R
Sequence 2	M	A	C	V	D	E	M	F	R

Similarity



Blue = Nitrogen
White = Carbon
Red = Oxygen

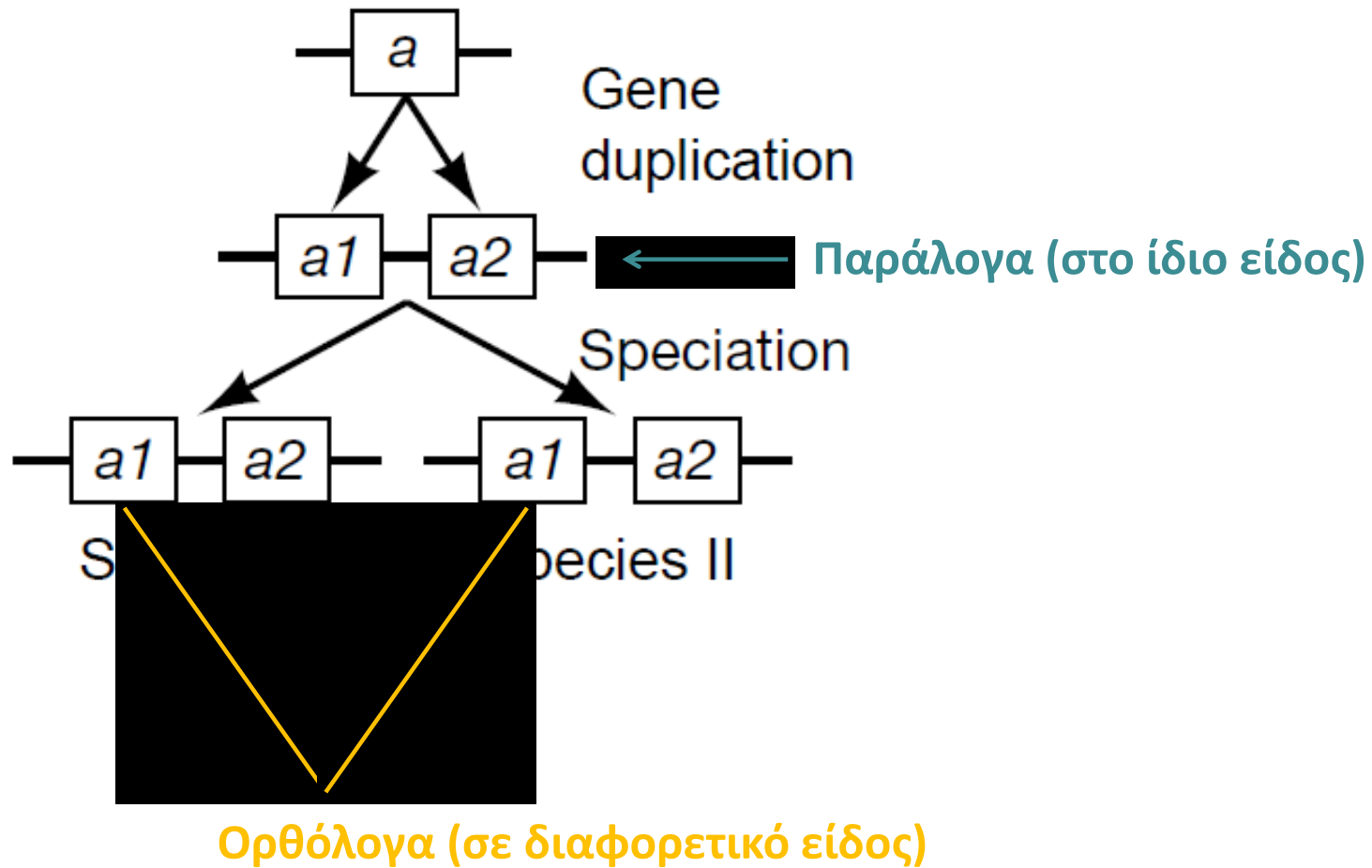


Είδη ομόλογων ακολουθιών – Ορθόλογα & παράλογα

- **Ομόλογες:** Όταν δυο αλληλουχίες DNA ή πρωτεϊνών είναι παρόμοιες, τότε, προέρχονται από τον ίδιο πρόγονο (εξελεγκτική συγγένεια).
- **Ορθόλογα:** Γονίδια που παρουσιάζουν ομοιότητα και **ανήκουν σε διαφορετικό είδος**. Αυτά τα γονίδια είναι πολύ πιθανό να κωδικοποιούν πρωτεΐνες με **παρόμοια δομή και λειτουργία**.
- **Παράλογα:** Γονίδια που προέρχονται από τον διπλασιασμό γονιδίων σε ένα γονιδίωμα. **Εμφανίζονται σε ένα είδος** και έχουν διαφορετικές ιδιότητες από τα γονίδια από τα οποία προήλθαν.



Ορθόλογα / Παράλογα



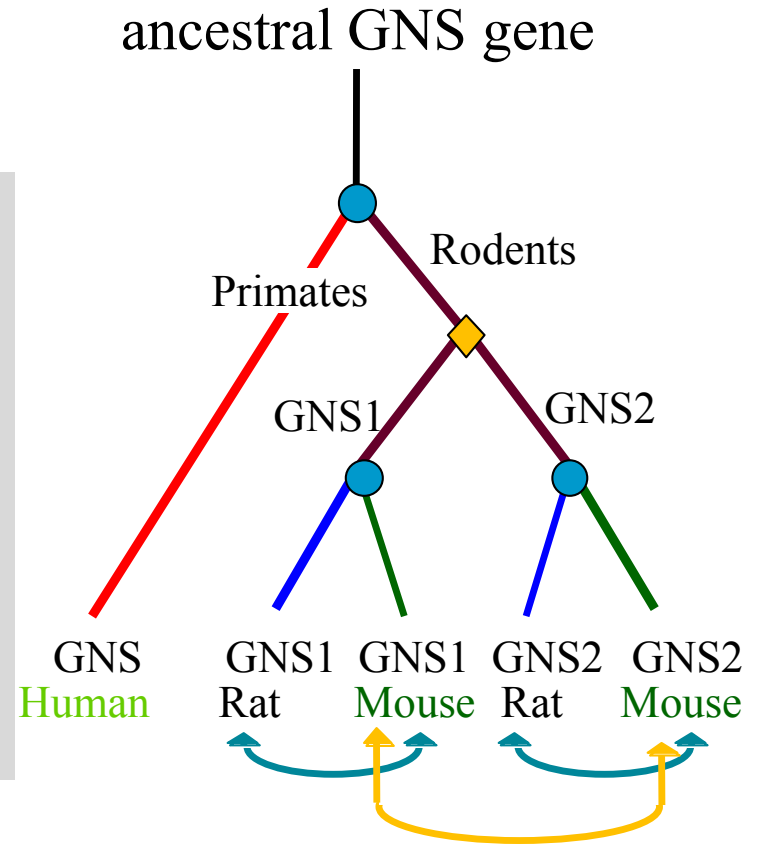
Ορθόλογα / Παράλογα - Παράδειγμα

- speciation
- ◆ duplication

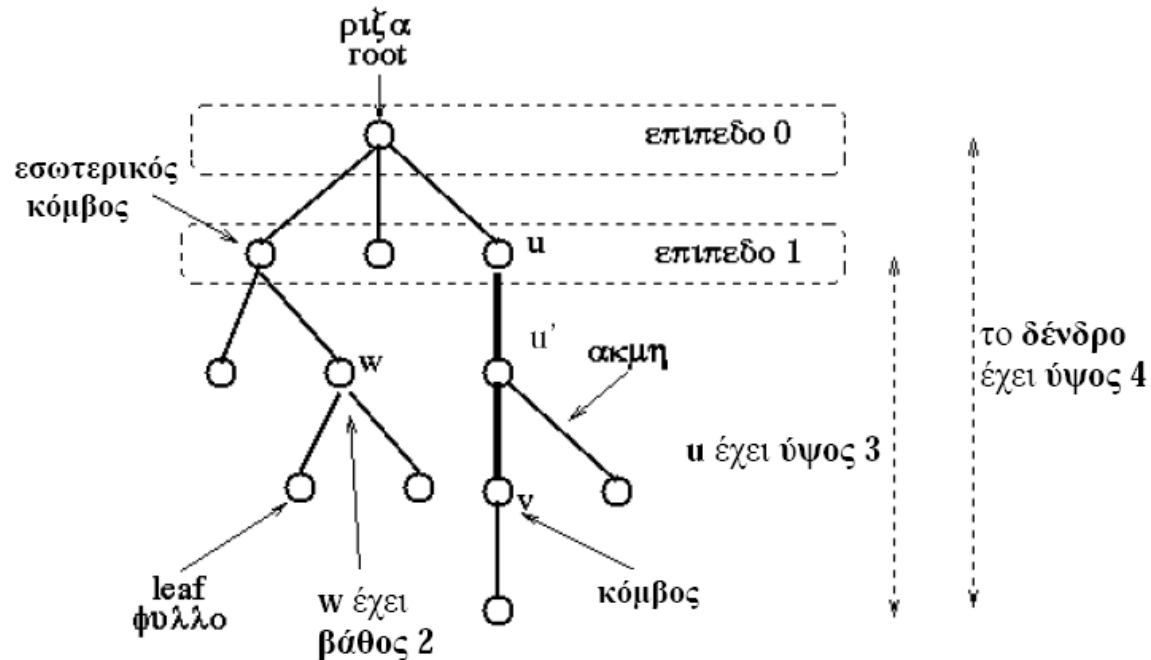
Ομόλογα: Δύο γονίδια που έχουν κοινό πρόγονο.

Ορθόλογα: Δύο γονίδια προερχόμενα από γεγονότα ειδογένεσης (speciation). **Ανήκουν σε διαφορετικό είδος.** Συνήθως έχουν κοινή λειτουργία.

Παράλογα: Γονίδια προερχόμενα από ένα κοινό προγονικό γονίδιο το οποίο διπλασιάστηκε σε ένα είδος και ύστερα απέκλινε. Ανήκουν στο ίδιο είδος. Συνήθως έχουν διαφορετική λειτουργία.



Δέντρα δομών (1/2)

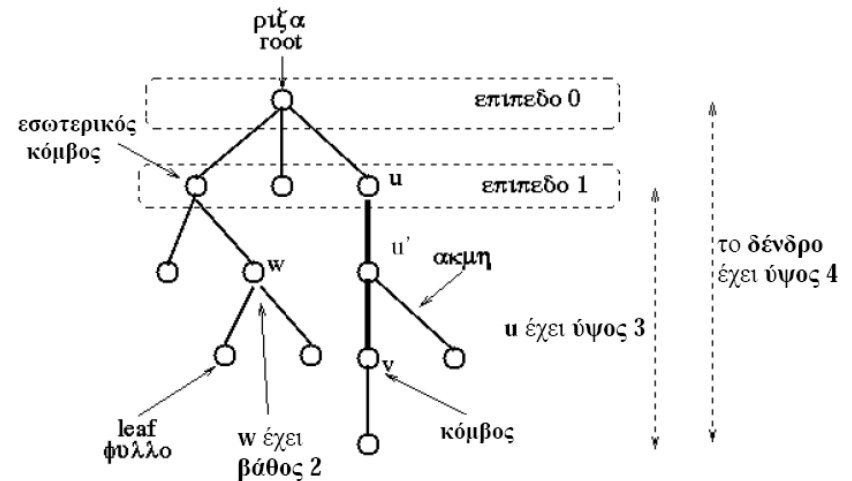


- Ρίζα (root)
- Κόμβος (node)
- Ακμή (edge)
- Γονέας (parent)
- Παιδί (child)
- Διαδρομή (path)
- Μήκος διαδρομής (πλήθος ακμών)
- Πρόγονος (ancestor)
- Απόγονος (descendant)
- Φύλλο (leaf)



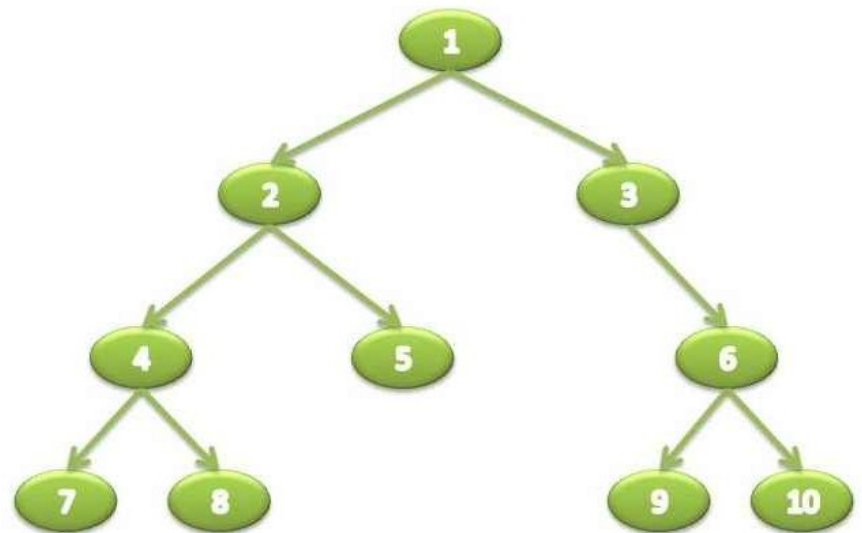
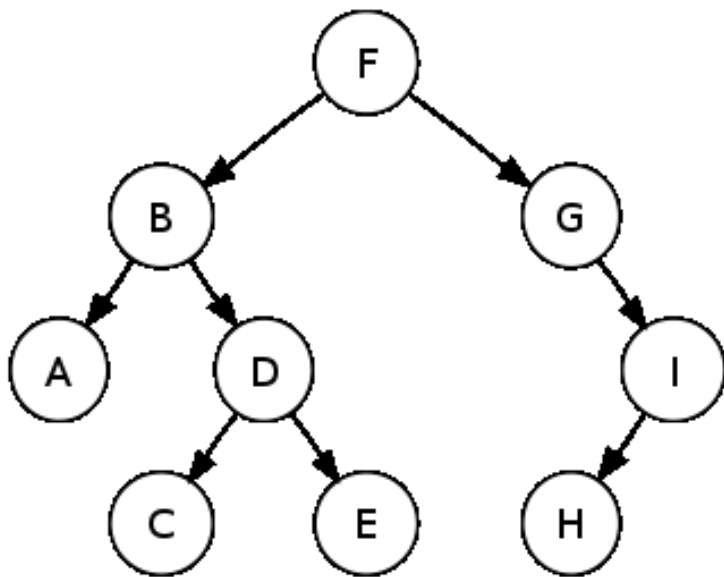
Δέντρα δομών (2/2)

- **Βαθμός κόμβου:** Το πλήθος των παιδιών του κόμβου.
- **Επίπεδο:** Σύνολο από κόμβους. Η ρίζα βρίσκεται στο επίπεδο 0.
- **Ύψος κόμβου:** Μήκος μακρύτερου μονοπατιού από τον κόμβο σε κάποιο φύλλο.
- **Ύψος δέντρου:** Το ύψος της ρίζας του δέντρου.
- **Βάθος κόμβου:** Μήκος μονοπατιού από τη ρίζα στον κόμβο.



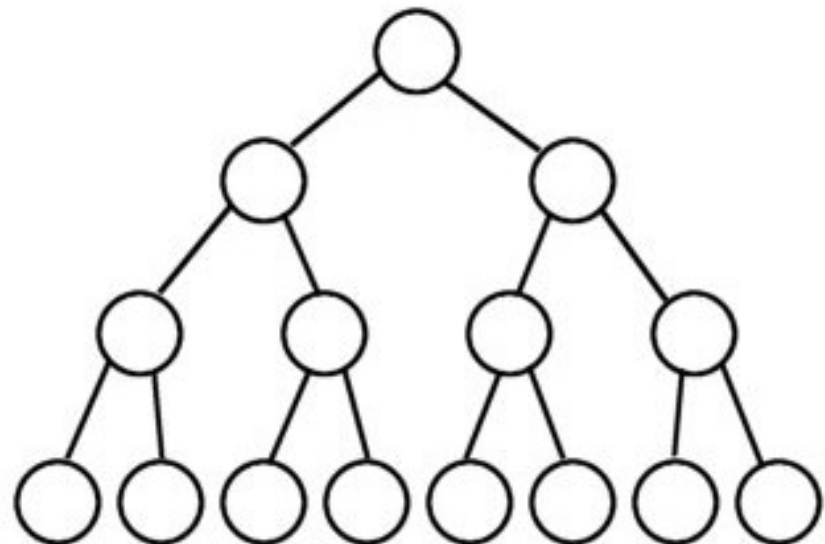
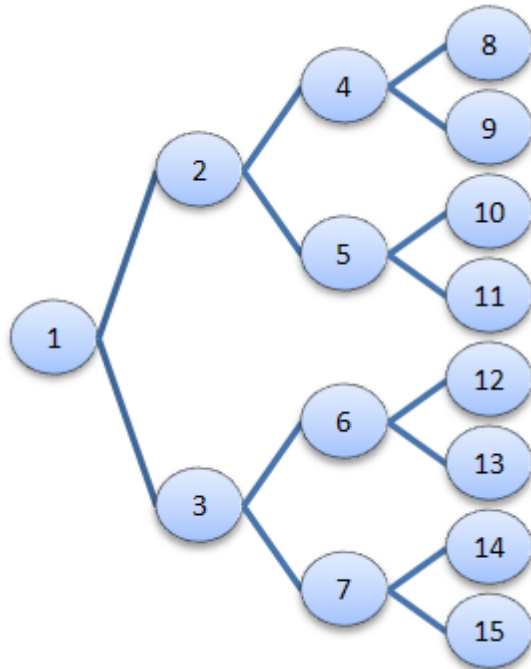
Δυαδικά δέντρα - Binary trees

- Διατεταγμένο δέντρο στο οποίο κάθε κόμβος έχει μέγιστο βαθμό 2 ή έχει το πολύ 2 παιδιά.



Full binary tree

Γεμάτο διατεταγμένο δέντρο: κάθε κόμβος έχει βαθμό 2 ή 0 (ή έχει 2 ή 0 παιδιά) → Κάθε εσωτερικός κόμβος έχει 2 παιδιά.



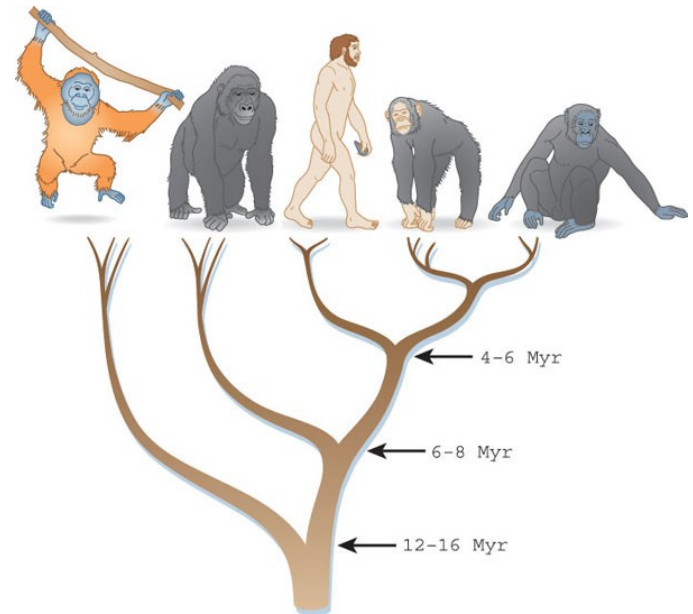
Φυλογενετικά δέντρα (1/2)

- **Τι είναι τα φυλογενετικά δέντρα;**
 - Η ανασύσταση της εξελεγκτικής ιστορίας των οργανισμών, δηλ. μας δείχνουν τις εξελεγκτικές σχέσεις μεταξύ των αντικειμένων μιας ομάδας
 - Παρέχουν πληροφορίες για την αλληλουχία των εξελεγκτικών γεγονότων
 - Τα στοιχεία που χρησιμοποιούνται για την ανασύσταση των φυλογενετικών δέντρων μπορεί να βασίζονται στη μορφολογία, τη φυσιολογία, τις διατροφικές ή άλλες συνήθειες, τη βιοχημεία ή τη γενετική.

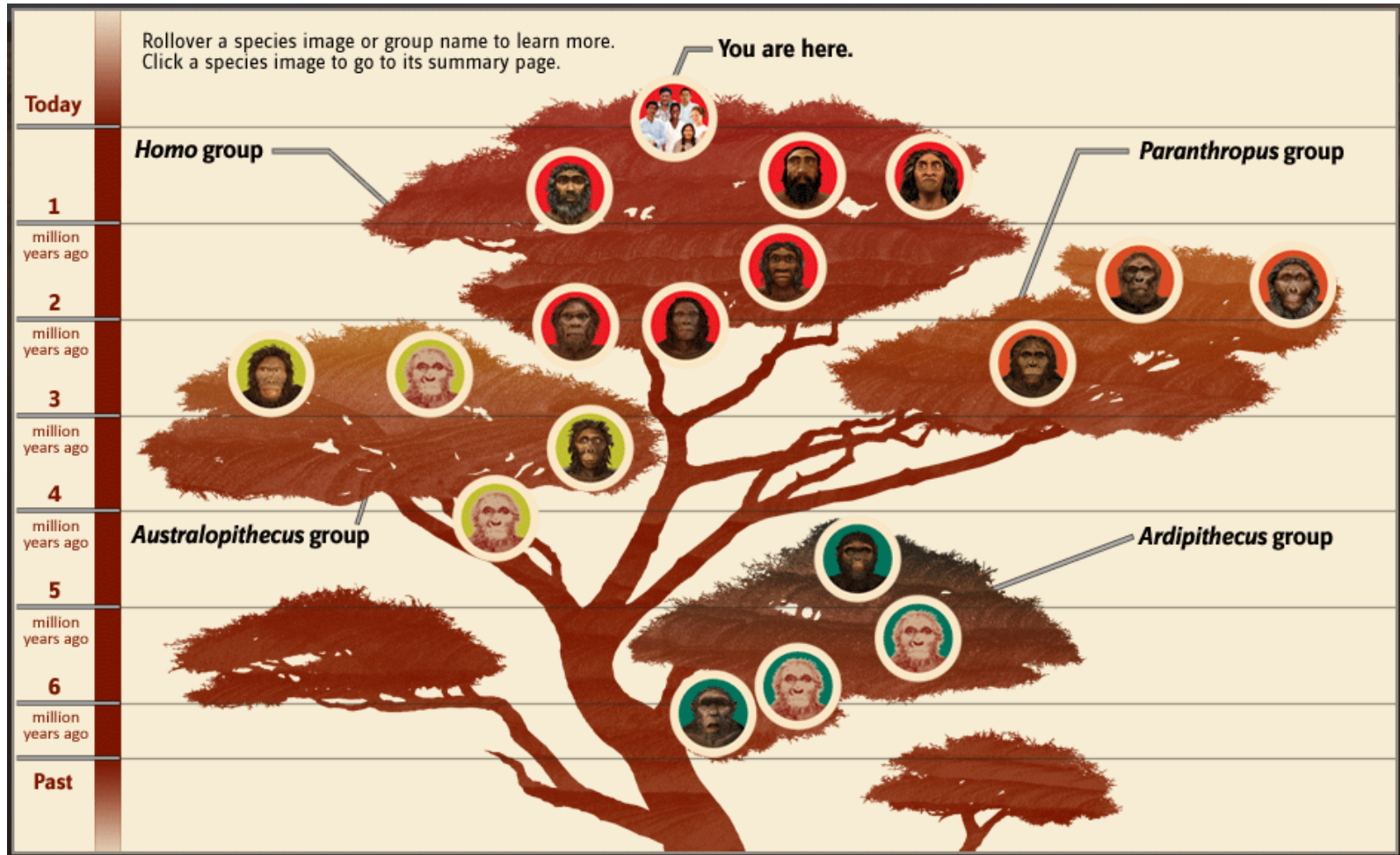


Φυλογενετικά δέντρα από μοριακά δεδομένα (1/2)

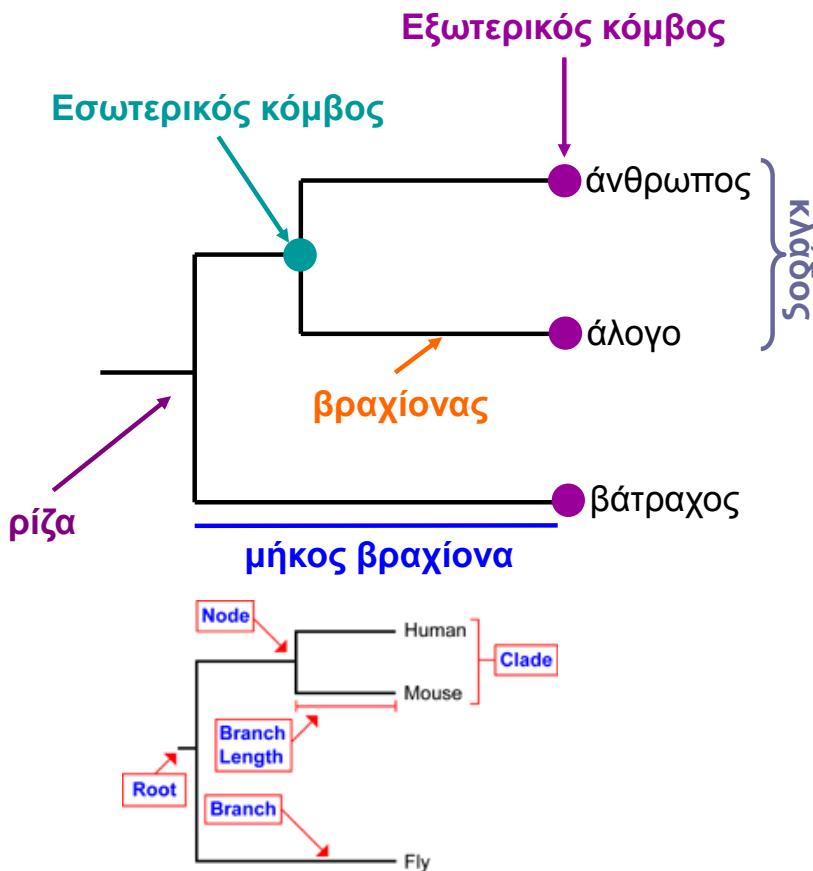
- Ο προσδιορισμός ενός φυλογενετικού δέντρου αποτελεί ένα **στατιστικό πρόβλημα**, η λύση του οποίου είναι μία εκτίμηση της πραγματικότητας.
- Όταν ένα φυλογενετικό δέντρο αντικατοπτρίζει τα εξελεγκτικά μονοπάτια μιας ομάδας ειδών, τότε το δέντρο αυτό χαρακτηρίζεται ως **δέντρο ειδών (species tree)**.



Φυλογενετικά δέντρα από μοριακά δεδομένα (2/2)



Φυλογενετικά δέντρα (2/2)



- **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.



Διαφορετικά είδη δέντρων

- **Βαθμός της εξέλιξης (Order of evolution):**
 - Με ρίζα: Υποδεικνύει την κατεύθυνση της εξέλιξης (indicates direction of evolution).
 - Χωρίς ρίζα: Δείχνει μόνο την εξελεγκτική απόσταση.
- **Ρυθμός της εξέλιξης (Rate of evolution):**
 - Μήκος των ακμών (Edge lengths): Απόσταση (scaled trees):
 - Molecular clock: constant rate of evolution.

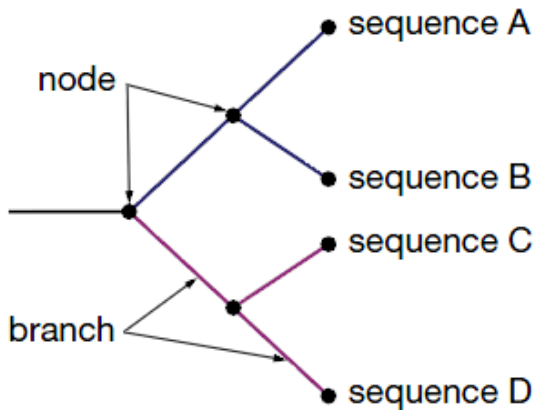


Δέντρο με ρίζα

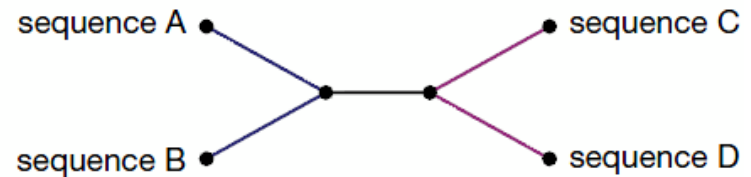
έναντι δένδρου χωρίς ρίζα (1/2)

- A. Υποδεικνύεται ο πιο κοινός πρόγονος από τον οποίο οι υπό εξέταση ταξινομικές μονάδες προήλθαν. Μπορούν να θεωρηθούν κατευθυνόμενα γραφήματα αν κατευθύνουμε τις ακμές του δέντρου από τη ρίζα προς τα φύλλα.
- B. Δεν κάνουμε καμία υπόθεση για τη θέση ενός εξελεγκτικού προγόνου (ρίζας) στο δέντρο.

A. Rooted tree

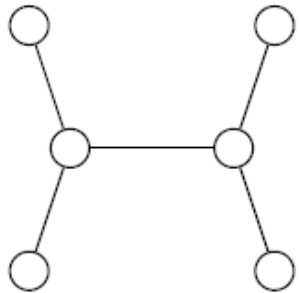


B. Unrooted tree

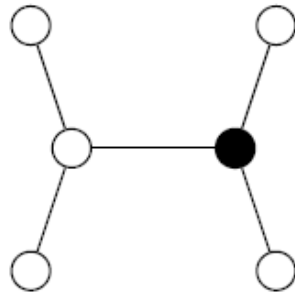


Δέντρο με ρίζα

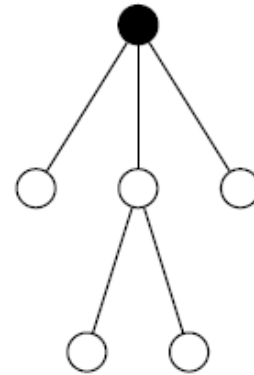
έναντι δένδρου χωρίς ρίζα (2/2)



(a) Unrooted tree



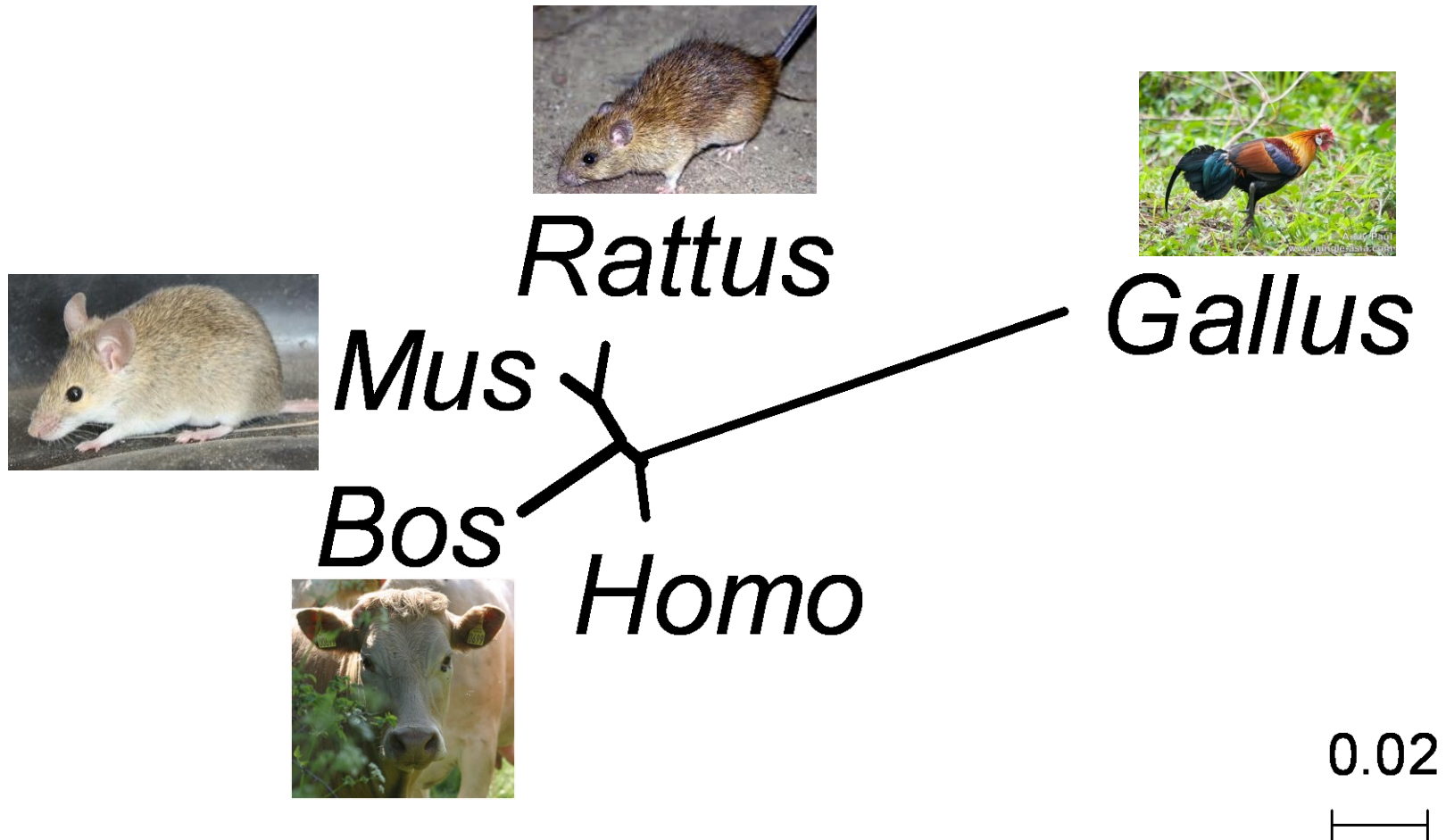
(b) Rooted tree



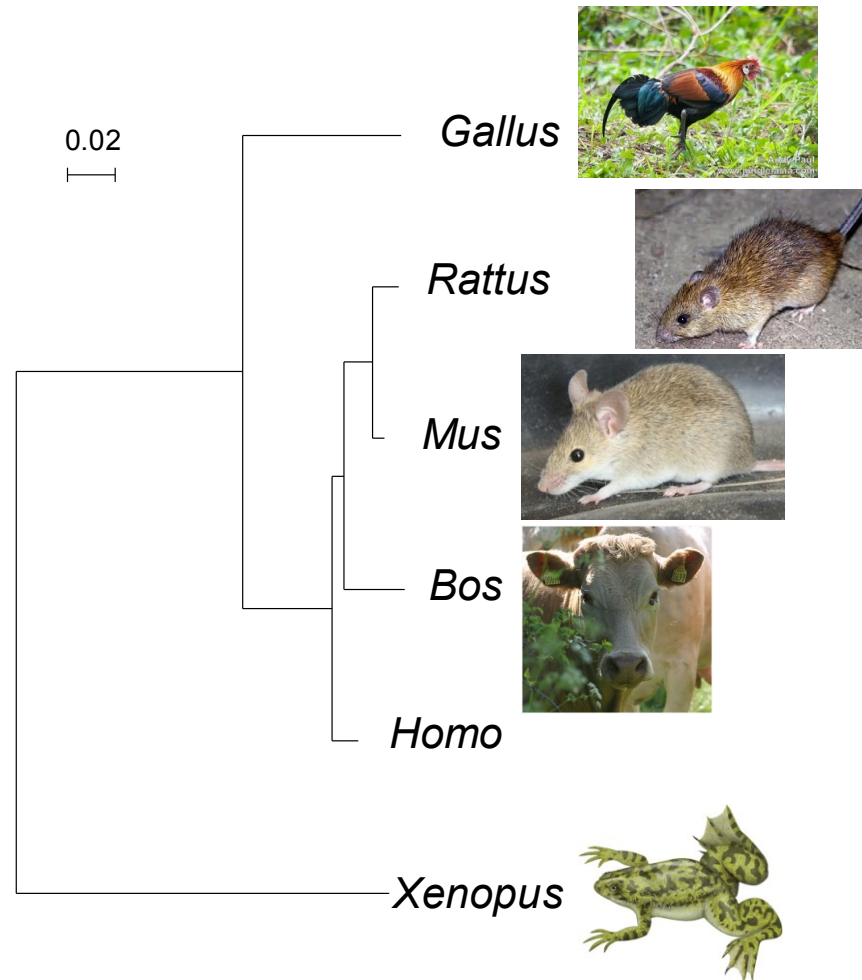
(c) The same rooted tree

- Οι πιο πολλές φυλογενετικές μέθοδοι κατασκευάζουν δέντρα χωρίς ρίζα. Αυτό συμβαίνει γιατί βρίσκουν τις διαφορές μεταξύ αλληλουχιών, αλλά δεν είναι δυνατός ο ορισμός κατεύθυνσης της αλλαγής σχετικά με τον χρόνο.

Unrooted Tree - Example

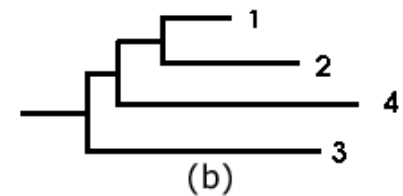
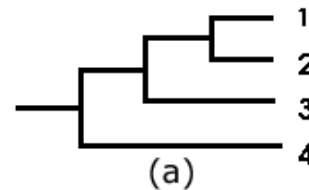


Rooted Tree



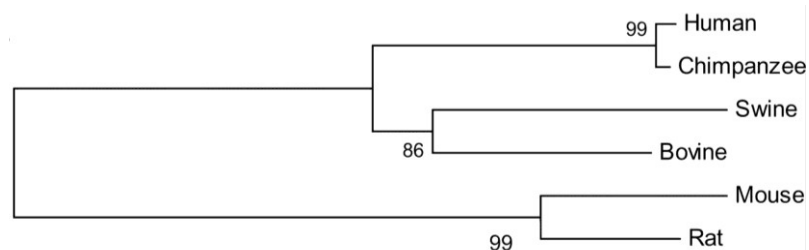
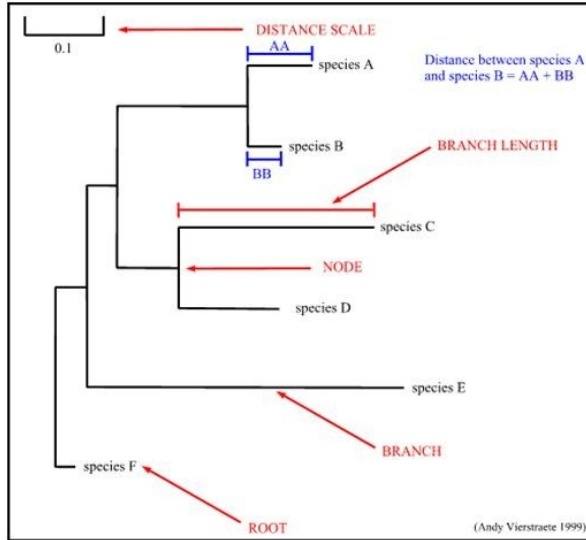
Trees with scaled and unscaled branches (1/2)

- Οι ακμές των φυλογενετικών δέντρων μπορούν να αναπαρασταθούν με 2 διαφορετικούς τρόπους:
 - **Μη κλιμακωμένες ακμές** – Όλες οι ακμές του δέντρου έχουν το ίδιο μήκος.
 - **Κλιμακωμένες ακμές** – Οι ακμές έχουν διαφορετικό μήκος βασισμένο στον αριθμό των μεταλλάξεων κατά την εξέλιξη ή στην απόσταση.



Trees with scaled and unscaled branches (2/2)

- Συντελεστής στάθμισης μιας ακμής (v,w) (μήκος της ακμής (v,w)) μπορεί να αντικατοπτρίζει:
 - α. Τον αριθμό των μεταλλάξεων στην εξελεγκτική διαδρομή από την κορυφή v προς την κορυφή w, ή
 - β. Μία χρονική εκτίμηση για την εξέλιξη του είδους v στο είδος w.



Branches

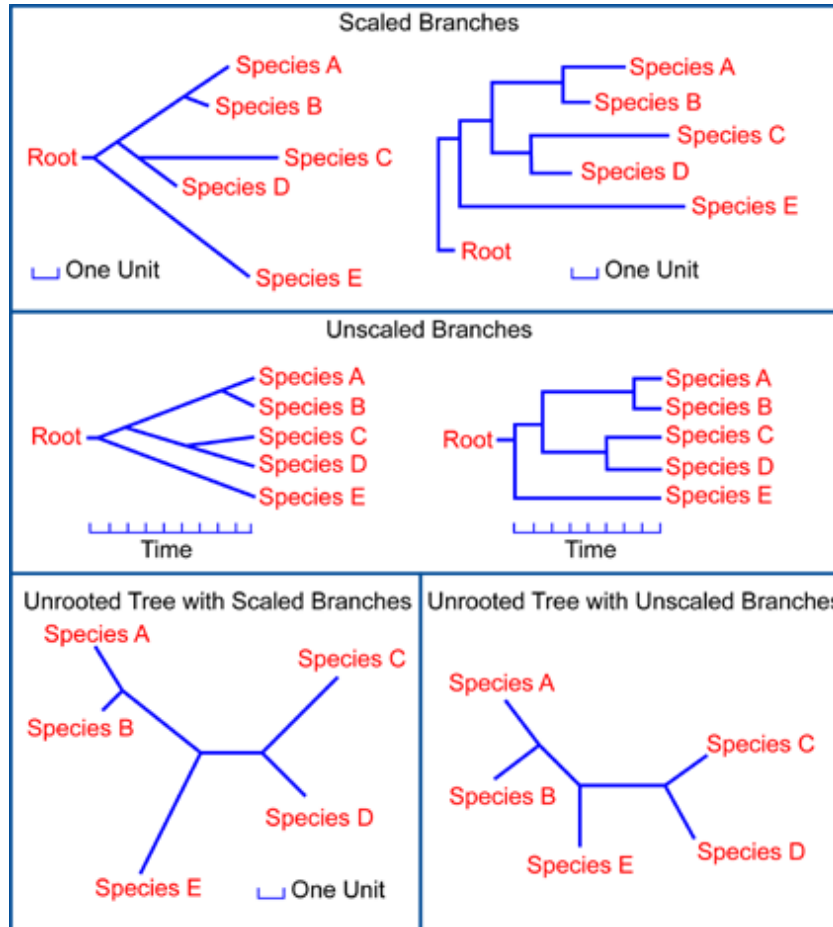


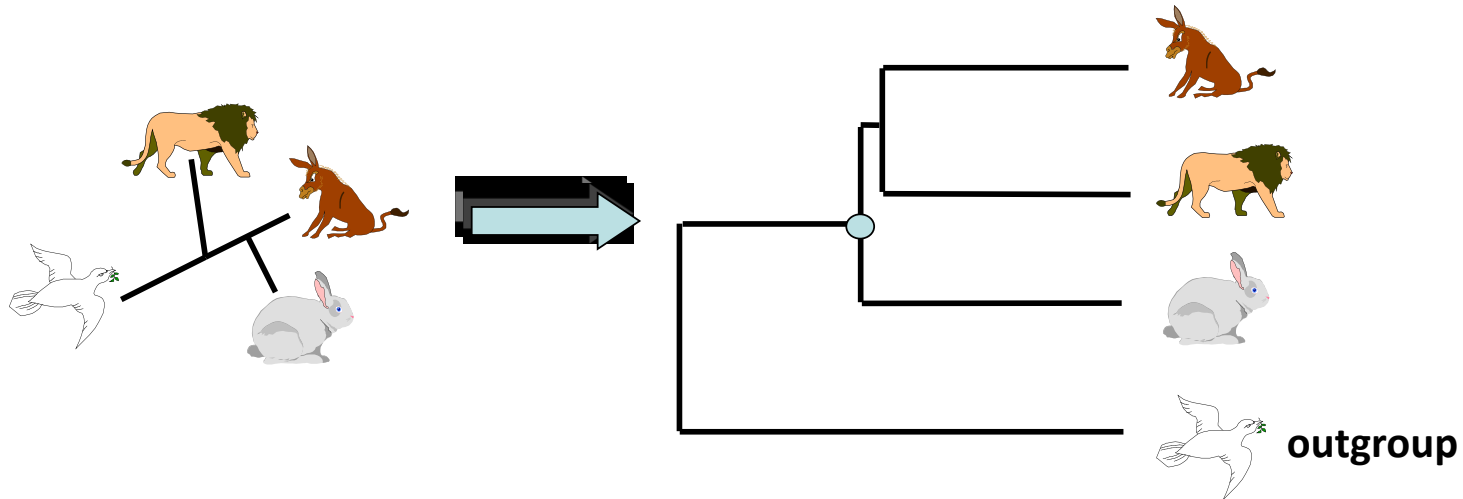
Image: <http://www.ncbi.nlm.nih.gov/About/primer/phylo.html>



Ορισμός ρίζας σε δέντρο χωρίς ρίζα – μέθοδος 1

Δύο μέθοδοι για τον ορισμό ρίζας σε δέντρο χωρίς ρίζα:

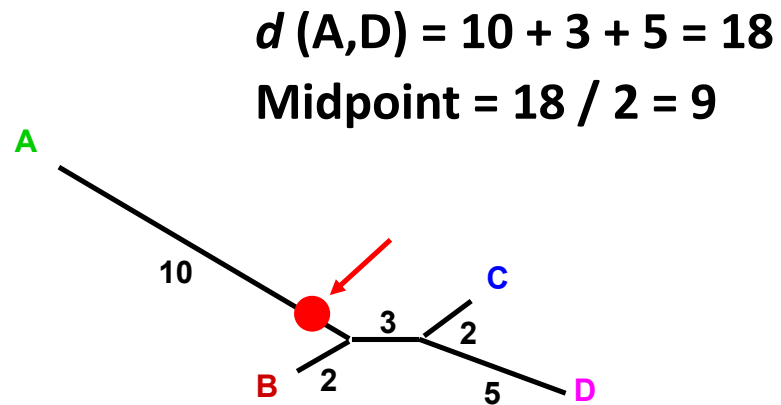
- 1. The outgroup method:** Στην ανάλυση περιλαμβάνεται και μία ομάδα αλληλουχιών που είναι εξ ορισμού εξωτερική της υπό μελέτης ομάδα. Η ρίζα τοποθετείται αναγκαστικά στον κόμβο που συνδέει τη μη σχετική ομάδα με τις υπόλοιπες αλληλουχίες (σχετικές μεταξύ τους).



Ορισμός ρίζας σε δέντρο χωρίς ρίζα – μέθοδος 2


- **By the molecular clock hypothesis:** Όλα τα είδη προήλθαν με την ίδια ταχύτητα από τη στιγμή της απόκλισης από τον κοινό τους πρόγονο. Η ρίζα τοποθετείται στο μέσο της απόστασης που χωρίζει τις δύο πιο απομακρυσμένες ταξινομικές μονάδες του δέντρου, βάσει το μήκος των βραχιόνων.

By midpoint or distance:



Αριθμός πιθανών δέντρων σε συνάρτηση του αριθμού των αλληλουχιών

Εξελικτικές μονάδες	Πλήθος Δέντρων χωρίς ρίζα	Πλήθος Δέντρων με ρίζα
4	3	15


$$N_r = \frac{(2m-1)!}{2^{m-1} (m-1)!} \quad N_u = \frac{(2m-2)!}{2^{m-2} (m-2)!}$$



Μέθοδοι κατασκευής φυλογενετικών δέντρων (1/3)

- **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).



Κατασκευή φυλογενετικού δέντρου που βασίζονται σε χαρακτήρες

- Αναζήτηση δέντρων που βελτιστοποιούν την προσαρμογή των δεδομένων των χαρακτήρων σε ένα φυλογενετικό δέντρο.
- Χρησιμοποιούν τα δεδομένα των χαρακτήρων με τέτοιο τρόπο ώστε η θέση κάθε χαρακτήρα να εκτιμάται σε συνάρτηση με τις θέσεις των υπολοίπων χαρακτήρων.



Μέθοδος μέγιστης συντήρησης (Maximum parsimony method)

- Προβλέπει το φυλογενετικό δέντρο που ελαχιστοποιεί τα βήματα για να φτάσουμε στις παρατηρούμενες διαφορές μεταξύ των ακολουθιών (minimum evolution method).
- Είναι απαραίτητη η πολλαπλή στοίχιση αλληλουχιών για να αναγνωριστούν οι θέσεις των αλληλουχιών που συμφωνούν.
- Εφαρμόζεται στην περίπτωση μικρού αριθμού αλληλουχιών με μεγάλη ομοιότητα.



Μέθοδος μέγιστης συντήρησης - Βήματα

Βήματα:

1. Πραγματοποιείται **πολλαπλή στοίχιση αλληλουχιών**.
2. Για κάθε στοίχιση αναγνωρίζονται τα φυλογενετικά δέντρα σύμφωνα με τα οποία απαιτείται ο **μικρότερος αριθμός εξελεγκτικών αλλαγών** για να προκύψουν οι παρατηρούμενες αλλαγές στις αλληλουχίες.
3. Κατασκευάζονται φυλογενετικά δέντρα που δημιουργούν τον **ελάχιστο αριθμό αλλαγών** για όλες τις θέσεις των αλληλουχιών.
4. Προτείνεται το **βέλτιστο δέντρο** – με τις ελάχιστες μεταλλάξεις.



Μέθοδος μέγιστης συντήρησης – Παράδειγμα (1/4)

1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G

- 4 αλληλουχίες

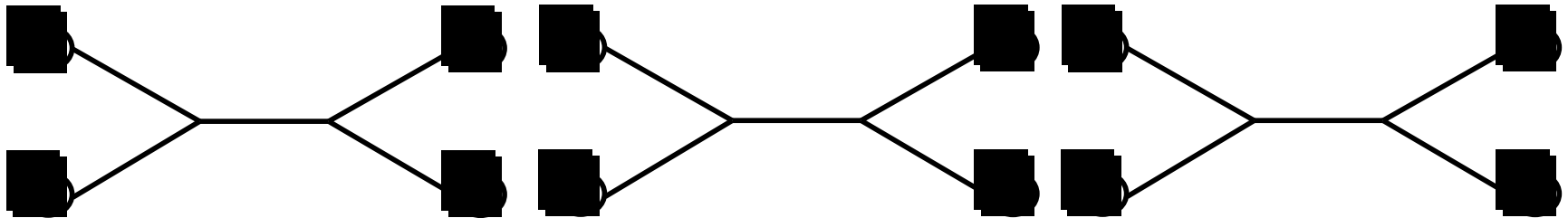
$$N_u = \frac{(2m -)!}{2^{m-} (m -)!}$$

3 πιθανά δέντρα χωρίς ρίζα
((1,2),(3,4)), ((1,3),(2,4)), (1,4),(2,3))



Μέθοδος μέγιστης συντήρησης – Παράδειγμα (2/4)

3 πιθανά δέντρα...



3 πιθανά δέντρα χωρίς ρίζα
 $((1,2),(3,4)), ((1,3),(2,4)), (1,4),(2,3))$

Informative site

- **Ακριβής ορισμός:** Η στοίχιση ενός καταλοίπου είναι φυλογενετικά χρήσιμη εάν «προτιμά» ένα υποσύνολο πιθανών δέντρων σε σχέση με τα υπόλοιπα.
- **Λειτουργικός ορισμός:** Η στοίχιση ενός καταλοίπου είναι φυλογενετικά χρήσιμη μόνο όταν υπάρχουν σε αυτή τουλάχιστον 2 διαφορετικά είδη νουκλεοτιδίων, και το καθένα από αυτά τα νουκλεοτίδια παρουσιάζονται σε τουλάχιστον 2 από τις αλληλουχίες υπό εξέταση.

1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G



Μέθοδος μέγιστης συντήρησης – Παράδειγμα (3/4)

1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G



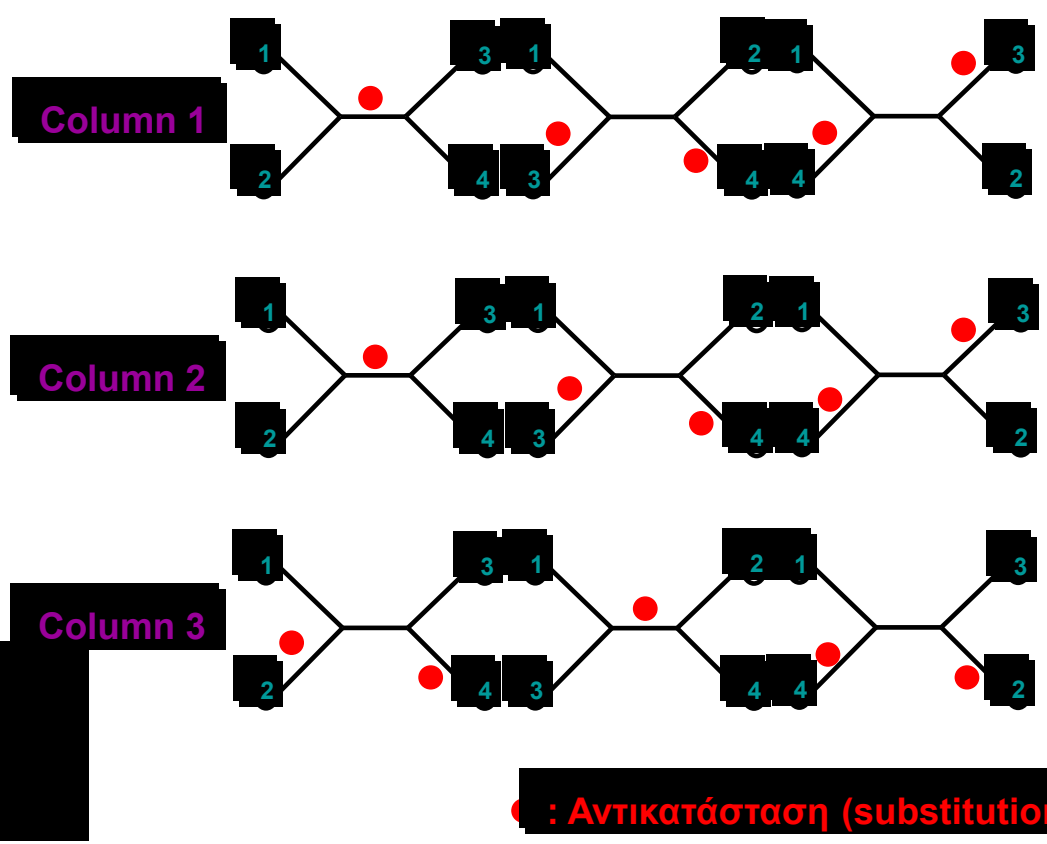
1	G	G	A
2	G	G	G
3	A	C	A
4	A	C	G



Μέθοδος μέγιστης συντήρησης – Παράδειγμα (4/4)

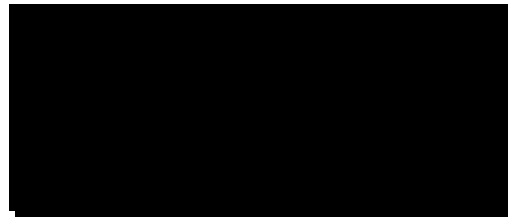
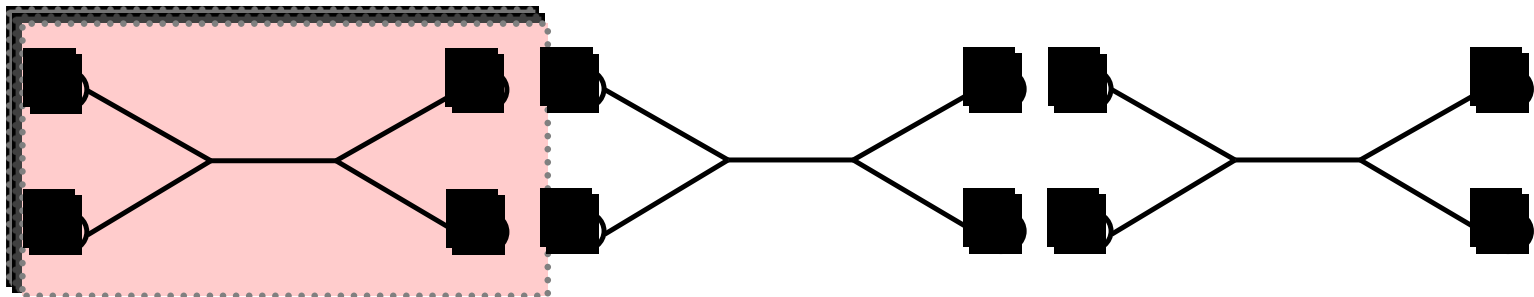
1	G	G	A
2	G	G	G
3	A	C	A
4	A	C	G

Tree 1 Tree 2 Tree 3



Μέθοδος μέγιστης συντήρησης – Προτείνεται:

1 **A** **A** **G** **A** **G** **T** **G** **C** **A**
2 **A** **G** **C** **C** **G** **T** **G** **C** **G**
3 **A** **G** **A** **T** **A** **T** **C** **C** **A**
4 **A** **G** **A** **G** **A** **T** **C** **C** **G**



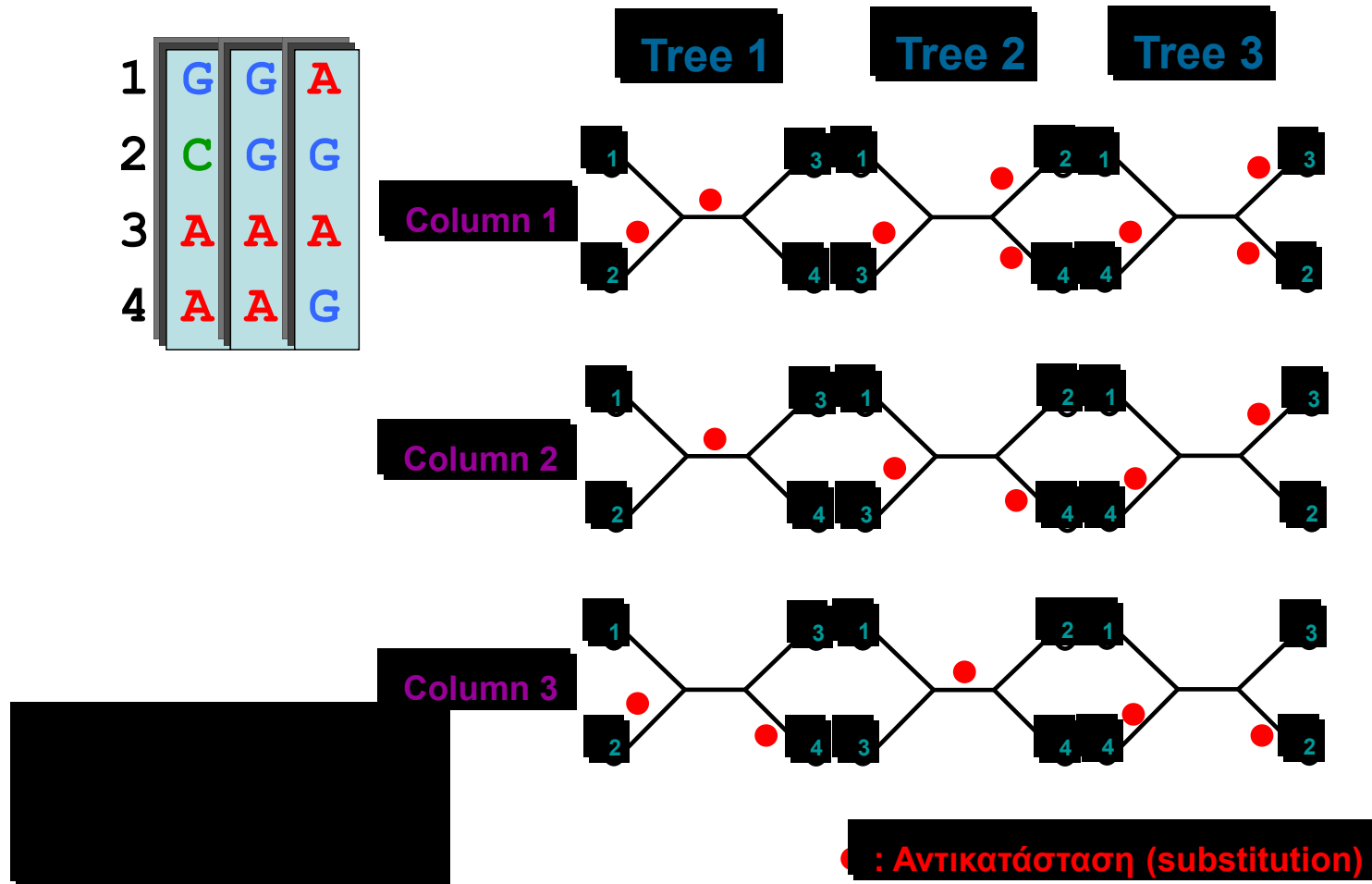
Αν είχαμε επιλέξει διαφορετικές στήλες; (1/2)

1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G



1	G	G	A
2	C	G	G
3	A	A	A
4	A	A	G

Αν είχαμε επιλέξει διαφορετικές στήλες; (2/2)



Αν είχαμε επιλέξει διαφορετικές στήλες πάλι; (1/2)

1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G



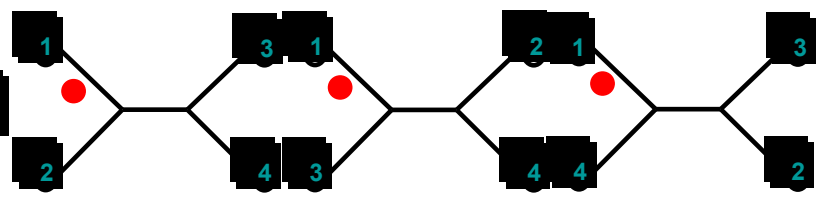
1	A	G	G
2	G	C	G
3	G	A	C
4	G	A	C

Αν είχαμε επιλέξει διαφορετικές στήλες πάλι; (2/2)

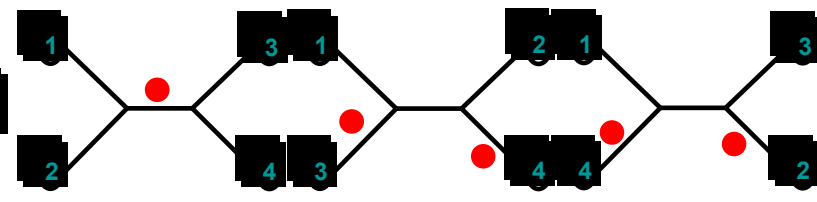
1	A	G	G
2	G	C	G
3	G	A	C
4	G	A	C

Tree 1 Tree 2 Tree 3

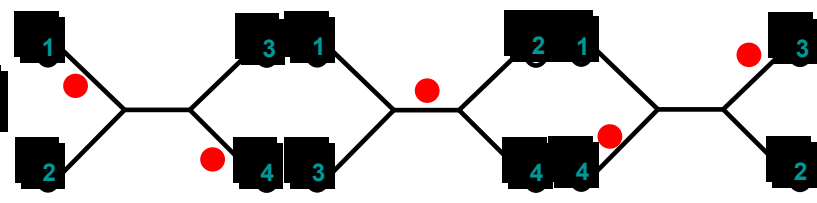
Column 1



Column 2



Column 3



????????????????



Μέθοδος μέγιστης συντήρησης - Συμπεράσματα

Maximum Parsimony (positive points):

- Οι στήλες που παρουσιάζουν μέτριες μεταβολές επικρατούν στην ανάλυση.
- Βασίζεται σε όμοια και «μοιρασμένα» χαρακτηριστικά.
- Δεν μειώνει την πληροφορία των αλληλουχιών.
- Παρέχει πληροφορίες για τις αρχέγονες αλληλουχίες.
- Αξιολογεί διαφορετικά δέντρα.

Maximum Parsimony (negative points):

- Δεν χρησιμοποιεί όλη την πληροφορία της πολλαπλής στοίχισης, μόνο ορισμένων στηλών.
- Αποδίδει καλά μόνο στην περίπτωση που οι αλληλουχίες μοιράζονται μεγάλο ποσοστό ομοιότητας.
- Δεν διορθώνει για πολλαπλές μεταλλάξεις (δεν προτείνεται ένα μοντέλο εξέλιξης).
- Δεν δίνει πληροφορίες για το μήκος των ακμών του δέντρου.



Μέθοδοι κατασκευής φυλογενετικών δέντρων (2/3)

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



Βασικοί ορισμοί (1/2)

- **Ζυγισμένη απόσταση μετασχηματισμού (Weighted Edit Distance) μεταξύ δύο συμβολοσειρών:** Το ελάχιστο πλήθος των πράξεων μετασχηματισμού που απαιτούνται για να μετασχηματίσουμε την πρώτη συμβολοσειρά στη δεύτερη.
- Κάθε πράξη μετασχηματισμού έχει συγκεκριμένο κόστος-βάρος
- Έστω ότι οι βασικές πράξεις μετασχηματισμού έχουν τα ακόλουθα βάρη:
 - Ένθεση ή διαγραφή: d ,
 - Αντικατάσταση: r
- **Παράδειγμα:** $\text{Weighted-edit-distance}(S_1 \rightarrow S_2) = 1r + 4d$

R	I	M	D	M	D	M	M	I
V		I	N	T	N	E	R	
W	R	I		T		E	R	S



Βασικοί ορισμοί (2/2)

- Εφαρμόζεται κυρίως στα προβλήματα στοίχισης ακολουθιών DNA και πρωτεϊνών, όπου χρησιμοποιούνται συγκεκριμένοι πίνακες αντικατάστασης, οι οποίοι ορίζουν το κόστος μετασχηματισμού του κάθε χαρακτήρα.
- Στην περίπτωση που η αντικατάσταση με την ένθεση/διαγραφή έχει το ίδιο κόστος-βάρος, τότε:
 - Απόσταση μεταξύ VINTERS & WRITERS: 5.

R	I	M	D	M	D	M	M	I
V		I	N	T	N	E	R	
W	R	I		T		E	R	S



Κατασκευή φυλογενετικού δέντρου με βάση την απόσταση (1/2)

1. Υπολογίζεται η απόσταση μεταξύ κάθε ζεύγους αλληλουχιών του αρχικού συνόλου. Τα ζεύγη των αλληλουχιών που έχουν τις λιγότερες αλλαγές στις αλληλουχίες (μικρότερη απόσταση) ονομάζονται **γειτονικά**. Στο φυλογενετικό δέντρο μοιράζονται ένα κόμβο ή έναν κοινό πρόγονο.
2. Οι ακολουθίες ομαδοποιούνται σύμφωνα με καθορισμένες αποστάσεις και κατασκευάζεται ένα δέντρο.

Συχνά, η εύρεση των πιο κοντινών «γειτόνων» μεταξύ αλληλουχιών βάσει της απόστασης αποτελεί το πρώτο βήμα της πολλαπλής στοίχισης αλληλουχιών (CLUSTALW).



Κατασκευή φυλογενετικού δέντρου με βάση την απόσταση (2/2)

- Προσοχή! Στην πολλαπλή στοίχιση αλληλουχιών μας ενδιαφέρει το ποσοστό ομοιότητας, όπως προκύπτει υπολογίζοντας τις απόλυτες ομοιότητες και τις συντηρητικές αντικαταστάσεις αμινοξέων, ενώ τα κενά αγνοούνται.
- Στην φυλογενετική ανάλυση μας ενδιαφέρει η απόσταση μεταξύ 2 αλληλουχιών:
 - Ο αριθμός των λάθος ταιριασμάτων (mismatches), ή
 - Ο αριθμός των αλλαγών που πρέπει να γίνουν σε μία αλληλουχία για να προκύψει η δεύτερη.
 - Τα κενά μπορούν να αγνοηθούν ή να υπολογιστούν σαν αντικαταστάσεις.

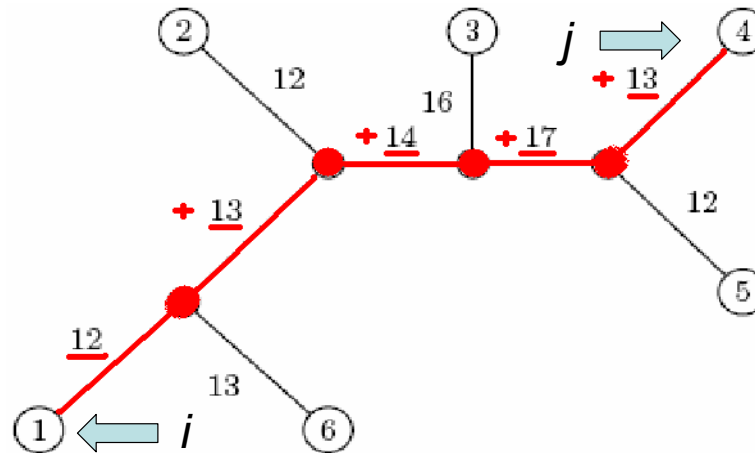


Αποστάσεις σε δέντρα

- Οι ακμές έχουν μήκος που δείχνει:
 - Το αριθμό των μεταλλάξεων στο εξελεγκτικό μονοπάτι από το ένα είδος στο άλλο.
 - Ο χρόνος που απαιτήθηκε για την εξέλιξη του ενός είδους από το άλλο.
- Σε ένα δέντρο απόστασης σημειώνουμε:
 - $d_{ij}(T)$ – το μήκος του μονοπατιού μεταξύ των i και j .
 - $d_{ij}(T)$ – **tree distance between i and j .**



Αποστάσεις σε δέντρα - Παράδειγμα



$$d_{1,4} = 12 + 13 + 14 + 17 + 12 = 68$$

- Με δεδομένο το δέντρο των αποστάσεων μπορούμε να υπολογίσουμε την εξελεγκτική απόσταση μεταξύ δύο αλληλουχιών που ανήκουν σε δύο διαφορετικά είδη.



Μήτρα αποστάσεων - Distance Matrix

- Συχνά οι βιολόγοι αντιμετωπίζουν το αντίθετο πρόβλημα.
- Γνωρίζουν ή υπολογίζουν τη μήτρα αποστάσεων D_{ij} , $n \times n$, και πρέπει να υπολογίσουν ένα δέντρο T που έχει n φύλλα και προσαρμόζεται στα δεδομένα.
- Δηλαδή να ισχύει: $d_{i,j} = D_{i,j}$ για κάθε ζεύγος φύλλων i και j .
- **Παράδειγμα μήτρας αποστάσεων:** Προσδιορισμός της αλληλουχίας ενός γονιδίου σε n είδη και ορισμός του $D_{i,j}$ ως η απόσταση μετασχηματισμού ανάμεσα σε αυτό το γονίδιο στο είδος i και στο ίδιο γονίδιο στο είδος j .
- D_{ij} – **edit distance between i and j .**



Fitting Distance Matrix

Lengths of path in an (*unknown*) tree T

- Fitting means $D_{ij} = d_{ij}(T)$

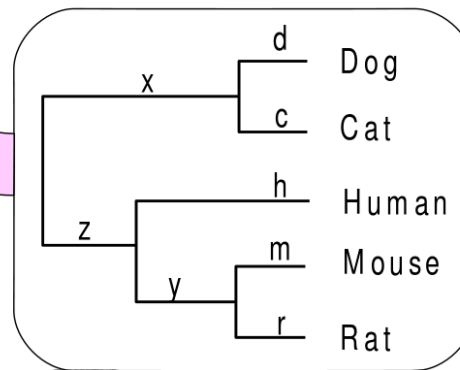
Edit distance between species (*known*)



Distance to trees

	Hum	Mou	Rat	Dog	Cat
Human	0	4	5	7	6
Mouse	h.y.m	0	3	8	5
Rat	h.y.r	m.r	0	9	7
Dog	h.z.x.d	m.y.z.x.d	r.y.z.x.d	0	2
Cat	h.z.x.c	m.y.z.x.c	r.y.z.x.c	d.c	0

Tree implies
a distance matrix
 M_{ij}



Map distances D_{ij}
to a tree

Algorithms:
UPGMA and
Neighbour Joining

Mapping from a tree to a distance matrix and vice versa

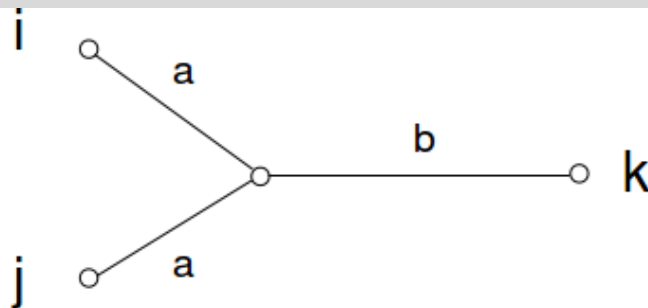


Μοντέλα με συγκεκριμένους περιορισμούς στον πίνακα αποστάσεων (1/2)

ULTRAMERIC:

- Για όλες τις τριπλέτες φύλλων (a,b,c) , δύο ζεύγη μεταξύ τους έχουν ίσες αποστάσεις και η τρίτη απόσταση είναι μικρότερη.
- Επομένως η τριπλέτα μπορεί να σημειωθεί i,j,k ώστε:

$$d_{ij} \leq d_{ik} = d_{jk}$$



where $a \leq b$



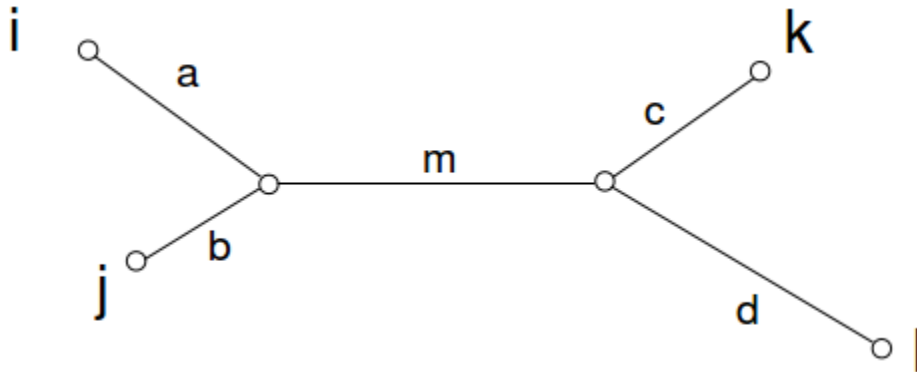
Μοντέλα με συγκεκριμένους περιορισμούς στον πίνακα αποστάσεων (2/2)

ADDITIVE:

- Όλα τα φύλλα του δέντρου μπορούν να σημανθούν i, j, k, l ώστε:

$$d_{ij} + d_{kl} \leq d_{ik} + d_{jl} = d_{il} + d_{jk}$$

$$(a+b) + (c+d) \leq (a+m+c) + (b+m+d) = (a+m+d) + (b+m+c)$$



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**.



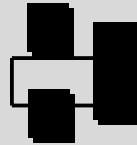
Μέθοδος ομαδοποίησης αστάθμητων ζευγών με αριθμητικούς μέσους όρους – Μέθοδος

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 (1/4)

- 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 (2/4)

$$\text{dist}(A,B),C=(\text{dist}AC+\text{dist}BC)/2=4,$$

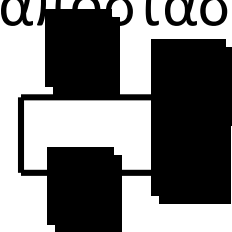
$$\text{dist}(A,B),D=(\text{dist}AD+\text{dist}BD)/2=6$$

$$\text{dist}(A,B),E=(\text{dist}AE+\text{dist}BE)/2=6,$$

$$\text{dist}(A,B),F=(\text{dist}AF+\text{dist}BF)/2=8$$

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

Το επόμενο πιο κοντινό ζευγάρι είναι το D και E, με απόσταση από τον κοινό τους πρόγονο ίση με $4/2=2$.

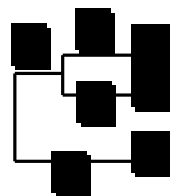


Μέθοδος ομαδοποίησης αστάθμητων ζευγών με αριθμητικούς μέσους όρους – Παράδειγμα (3/4)

- Επαναλαμβάνεται η προηγούμενη διαδικασία για το ζεύγος DE:
 - $\text{dist}(D,E),AB=(\text{dist}D(AB)+\text{dist}E(AB))/2=6$
 - $\text{dist}(D,E),C=\text{dist}(DC+EC)/2=6$
 - $\text{dist}(D,E),F=(\text{dist}DF+\text{dist}EF)/2=8$
- Οπότε έχουμε μία καινούργια μήτρα αποστάσεων:
 - Το επόμενο ζευγάρι είναι C και AB.
 - Με την ίδια διαδικασία έχουμε:

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

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



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

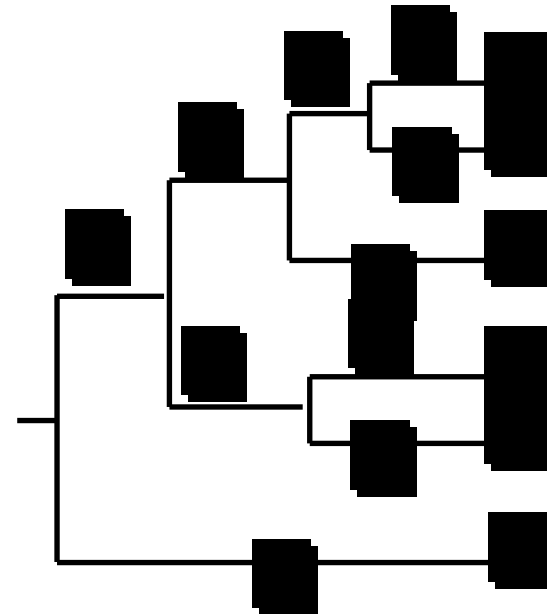
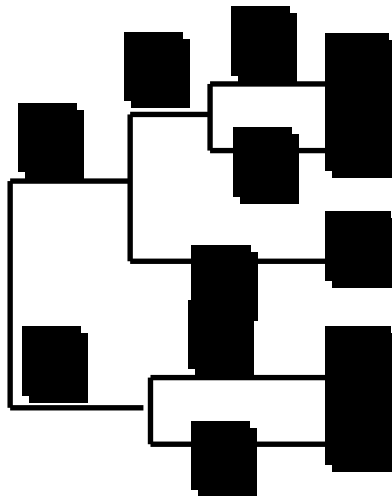


Μέθοδος ομαδοποίησης αστάθμητων ζευγών με αριθμητικούς μέσους όρους – Παράδειγμα (4/4)

- 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	



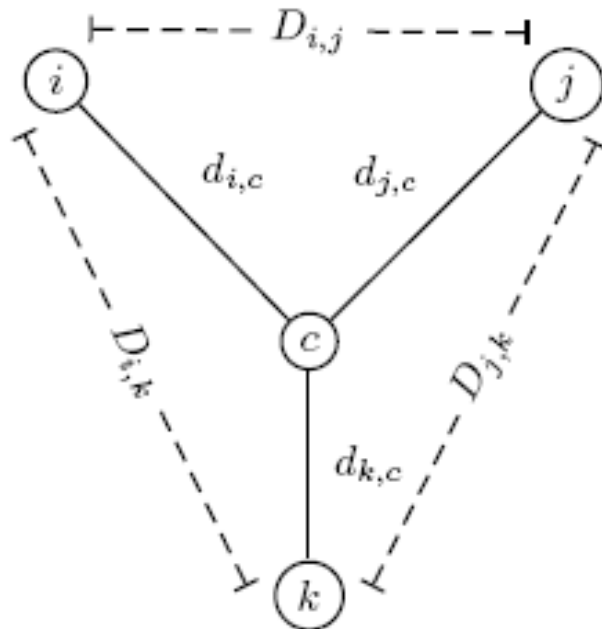
Μέθοδοι κατασκευής φυλογενετικών δέντρων (3/3)

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



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



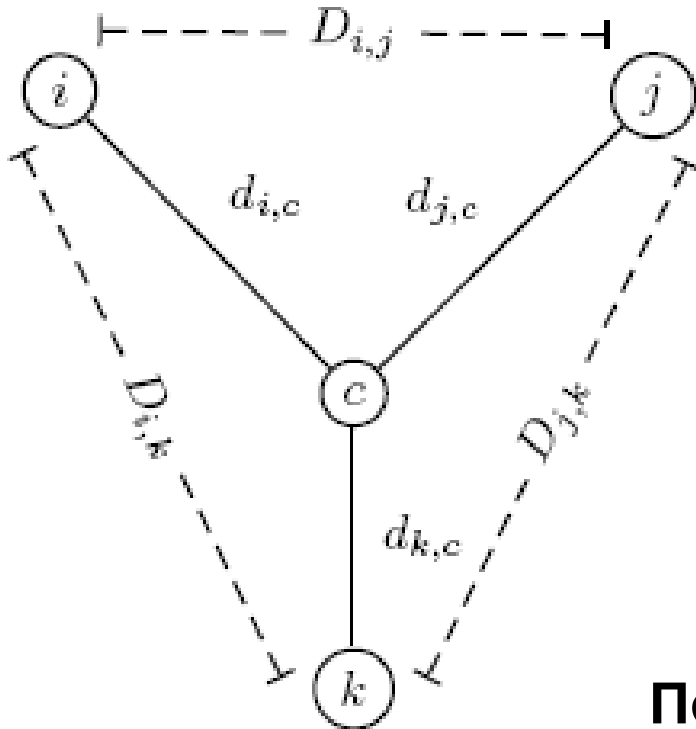
Παρατηρούμε ότι:

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

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

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

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



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

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

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

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

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

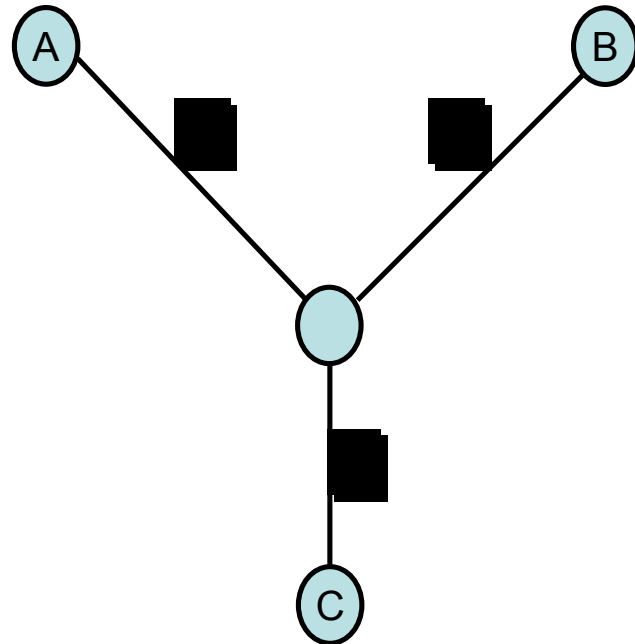
Παρομοίως,

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

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

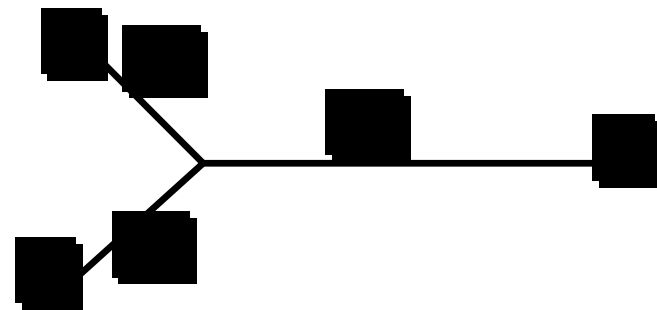


Υπολογισμός αποστάσεων με δεδομένη μήτρα διαστάσεων D, 3x3 - Παράδειγμα



	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): $b = 12$, $c = 29$

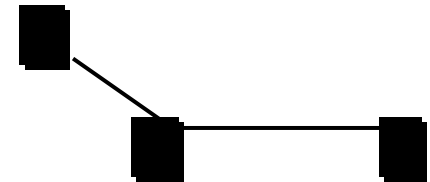
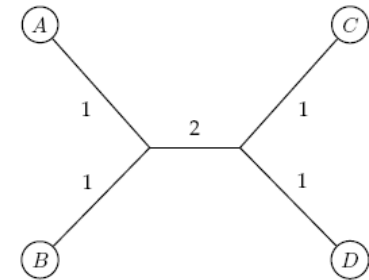


Προσθετικές και μη προσθετικές μήτρες (1/2)

- 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

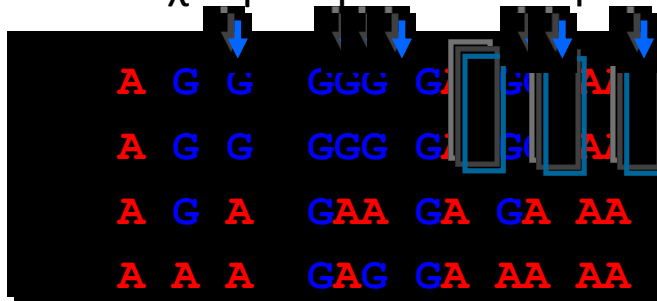
?

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)

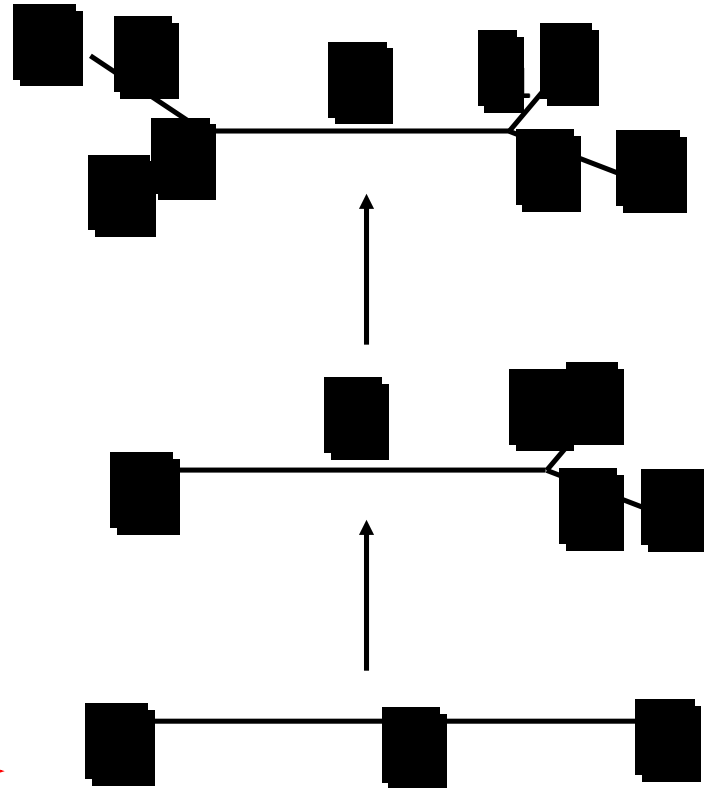
	A	B	C	D
A	-	3	7	8
B	-	-	6	7
C	-	-	-	3
D	-	-	-	-



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

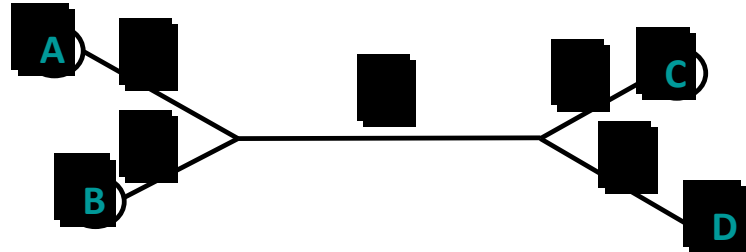


	A	C
A	-	7
C	-	-



Example of Distance Analysis (3/3)

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

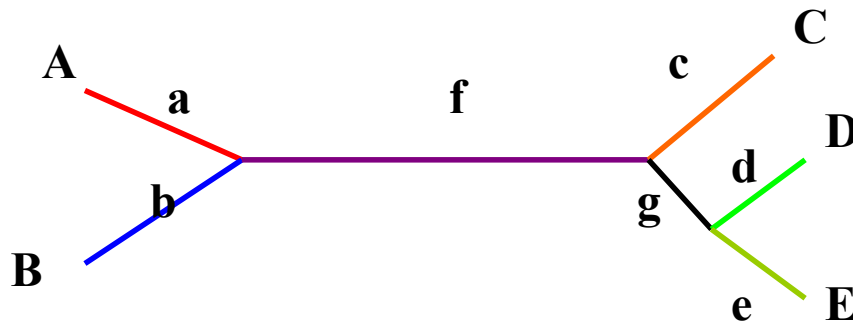


- Το προτεινόμενο φυλογενετικό δέντρο για τις αλληλουχίες A-D. Το άθροισμα των μηκών των ακμών μεταξύ δύο αλληλουχιών στο δέντρο έχει την ίδια τιμή με την απόσταση μεταξύ των αλληλουχιών.

Fitch and Margoliash

Algorithm (5 sequences) (1/7)

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

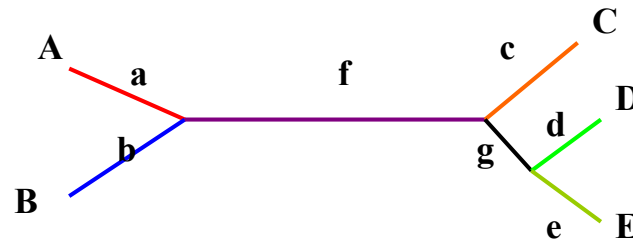


	A	B	C	D	E
A	--	22	39	39	41
B	--	--	41	41	43
C	--	--	--	18	20
D	--	--	--	--	10
E	--	--	--	--	--



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	--	--	--	--	--



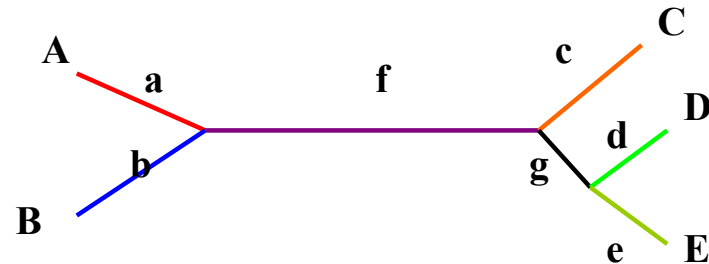
	D	E	AVG ABC
D	--	10	32.7
E	--	--	34.7
AVG ABC	--	--	--

- 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	--	--	--



[REDACTED]

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

[REDACTED]

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

[REDACTED]

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

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

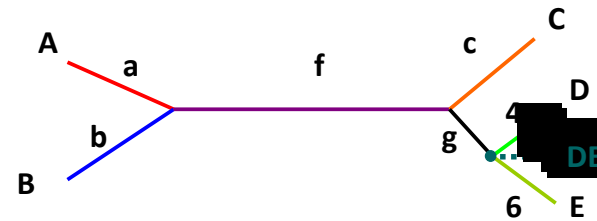
$$(1) + (4) \quad 2d = 8 \Rightarrow d = 4 \quad e = 6$$



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	--	--	--	--	--

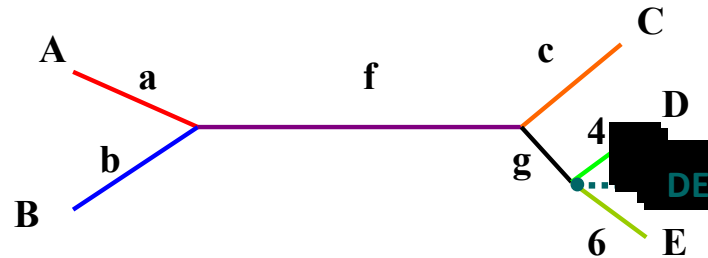


	A	B	C	(DE)
A	--	22	39	40
B	--	--	41	42
C	--	--	--	19
(DE)	--	--	--	--



Ή πιο αναλυτικά...

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



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

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

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

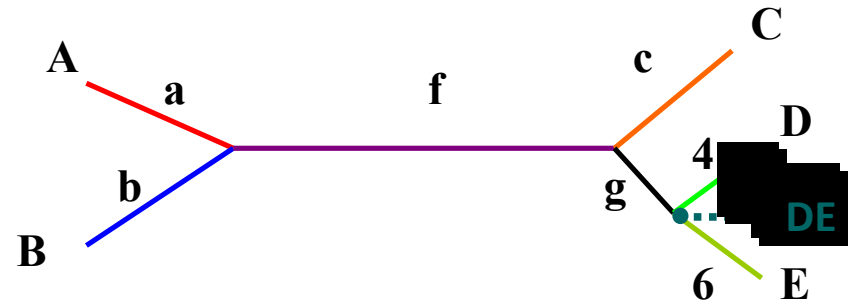
A to DE = 40

	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)	--	--	--	--

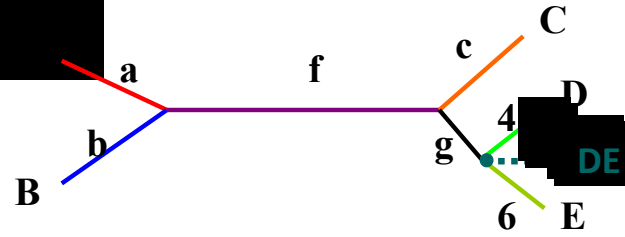


- Επομένως χρησιμοποιούμε την C,DE σαν μία αλληλουχία.
- Δημιουργώ καινούργιο πίνακα συνδυάζοντας τις υπόλοιπες αλληλουχίες, δηλ, τις A,B:
 - Μέση απόσταση της DE από την AB: $(40+42)/2 = 41$.
 - Μέση απόσταση της C από την AB: $(39+41)/2 = 40$.



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]

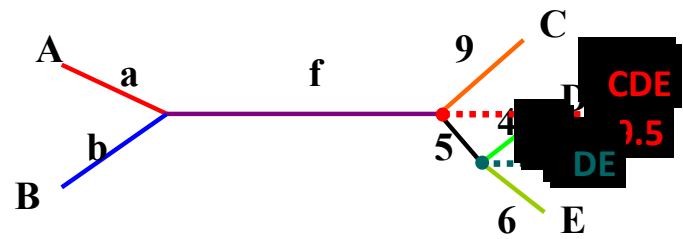
$$\frac{c+f+a}{2} + \frac{c+f+b}{2} = c+f + \frac{a+b}{2} = 1 \quad (2)$$

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

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

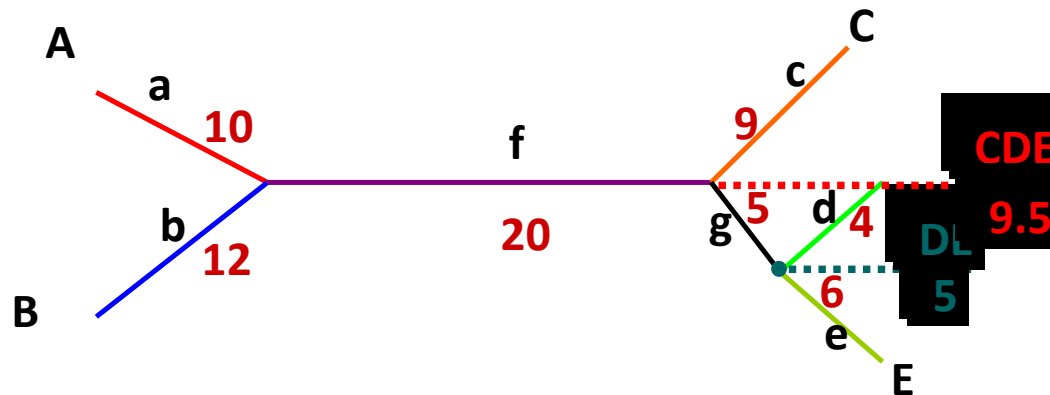
[Redacted]

	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 – 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)

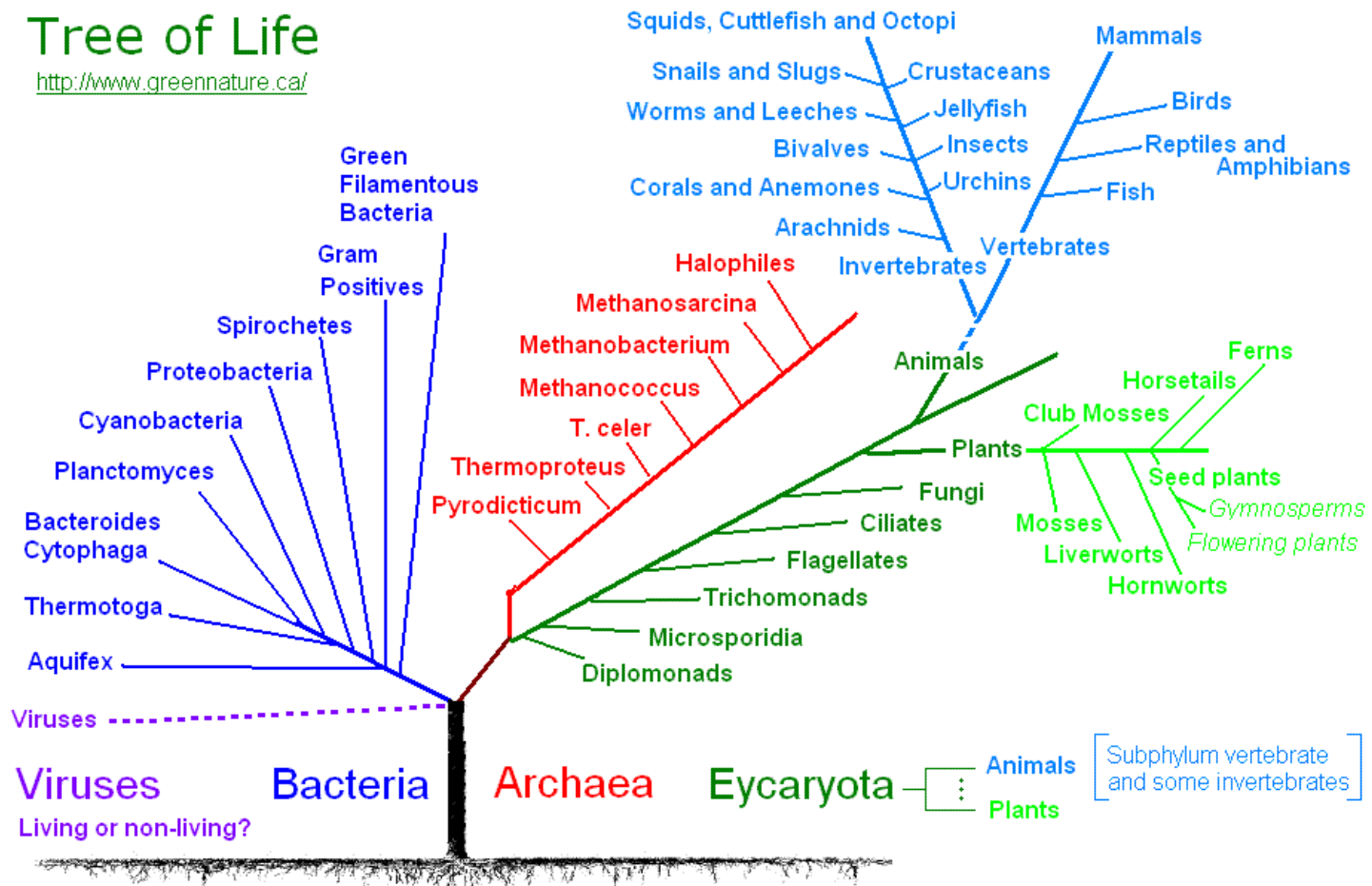
6. Βρίσκουμε το επόμενο ζευγάρι αλληλουχιών με τη μεγαλύτερη ομοιότητα και επαναλαμβάνουμε την παραπάνω διαδικασία.
7. Επαναλαμβάνω τη συνολική διαδικασία ξεκινώντας με όλα τα πιθανά ζεύγη των αλληλουχιών A και B, A και C, A και D, κτλ.
8. Υπολογίζουμε την προβλεπόμενη απόσταση μεταξύ κάθε ζεύγους αλληλουχιών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



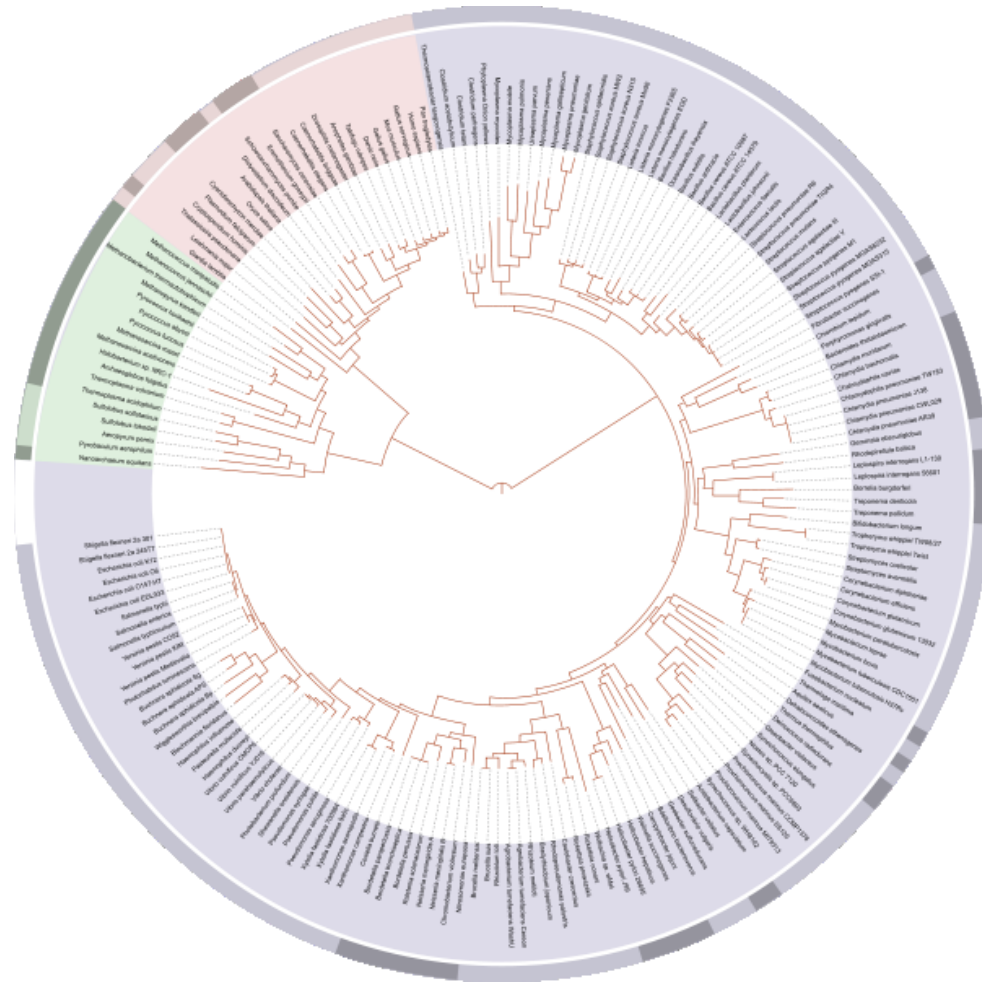
Tree of life

Tree of Life

<http://www.greennature.ca/>



Μοριακή Συστηματική



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.

