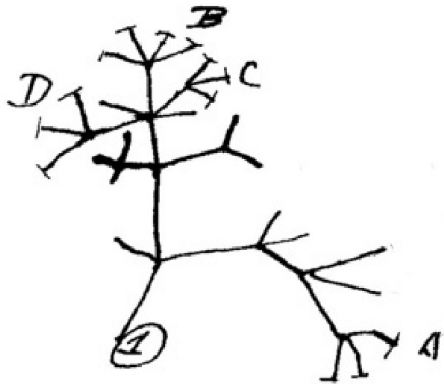


*I think*



ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΜΕΤΑΦΡΑΣΤΙΚΗ ΕΡΕΥΝΑ  
ΣΤΗΝ ΒΙΟΙΑΤΡΙΚΗ

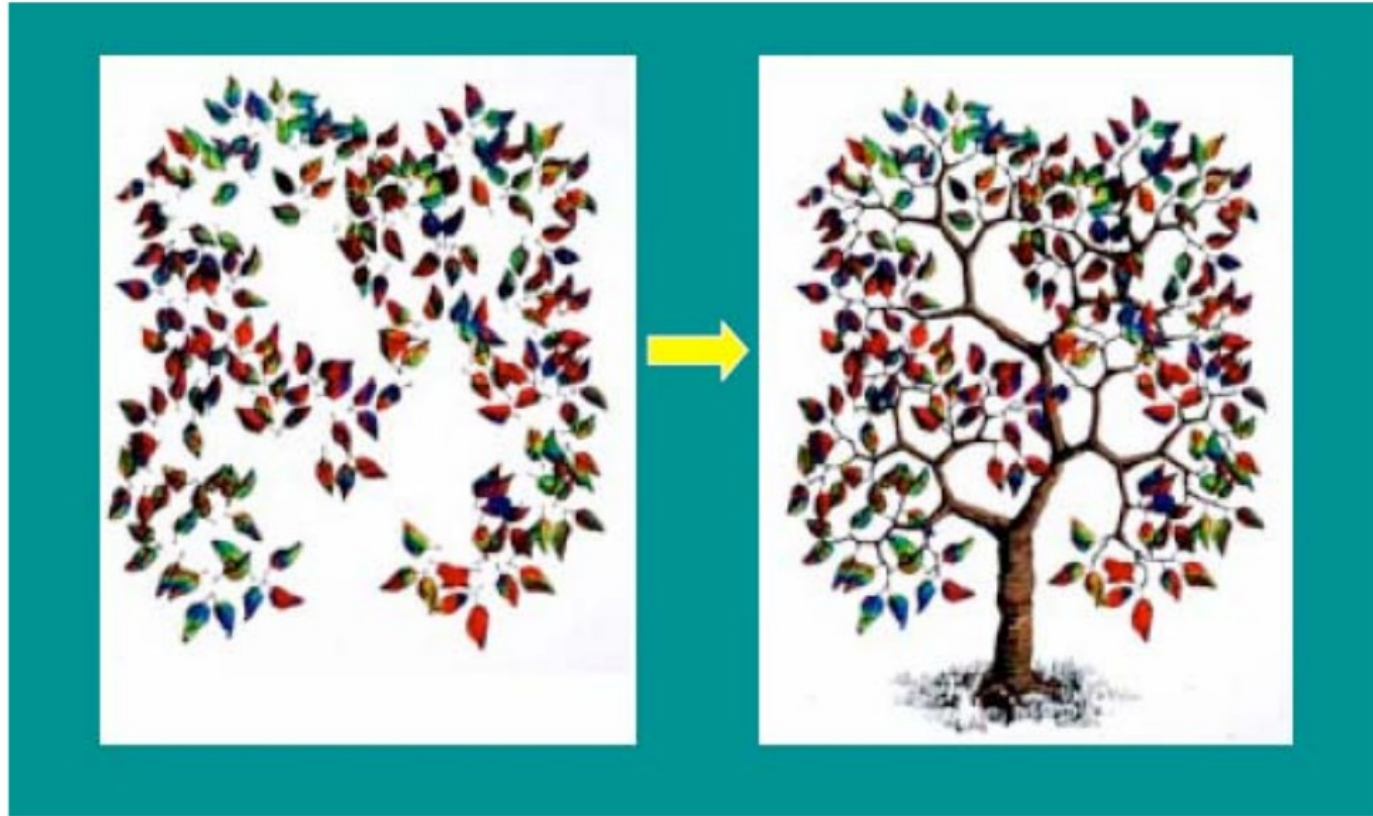
Μοριακή Διαγνωστική, Βιοδείκτες και  
Στοχευμένες Θεραπείες

MMBG103 Βιοπληροφορική

# Phylogenetic analysis and clustering *a very short introduction*

Αριστοτέλης Παπαγεωργίου, Τμ. ΜΒΓ ΔΠΘ, [aparage@mbg.duth.gr](mailto:aparage@mbg.duth.gr)

# Φυλογένεση

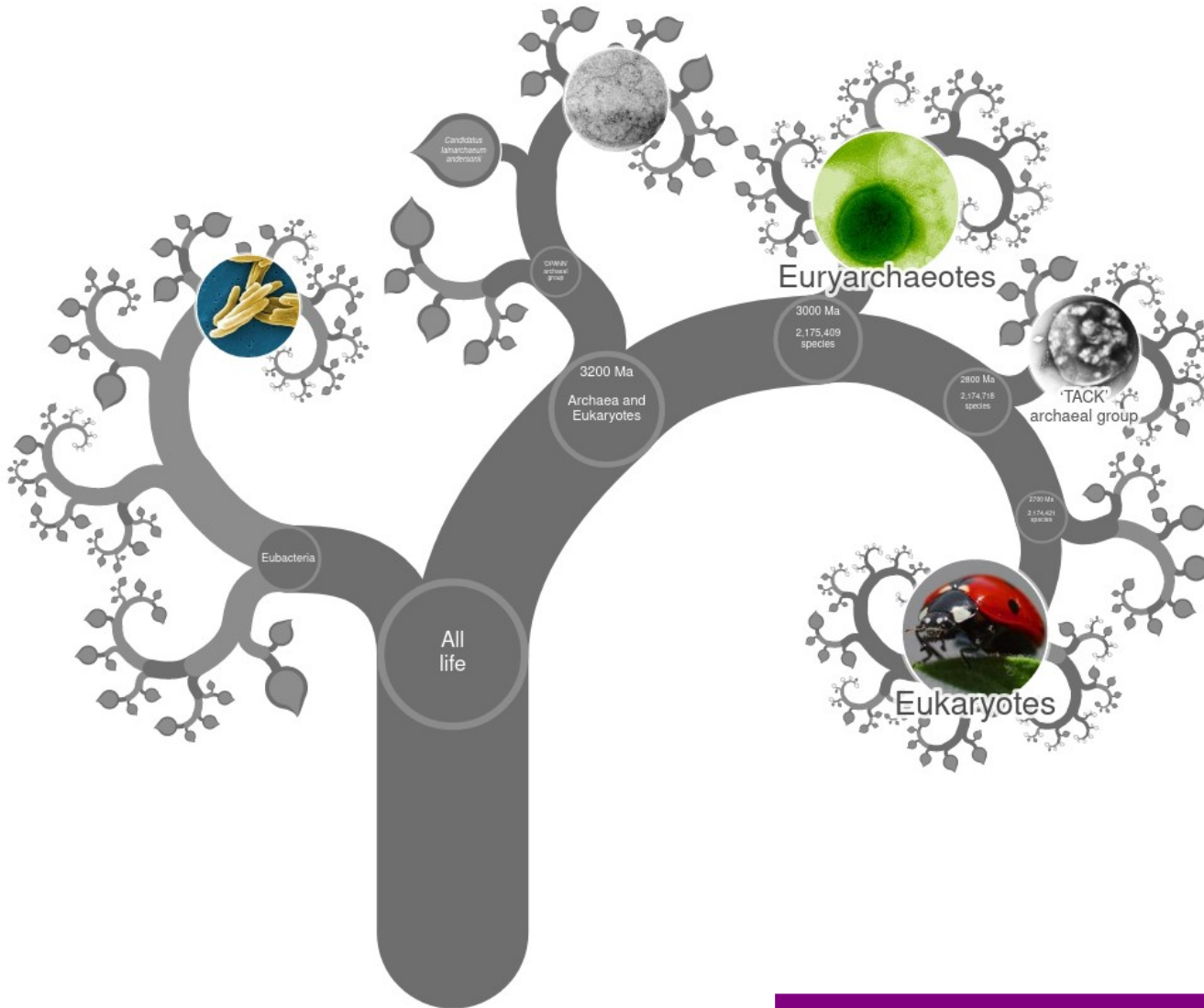


# Φυλογενετική

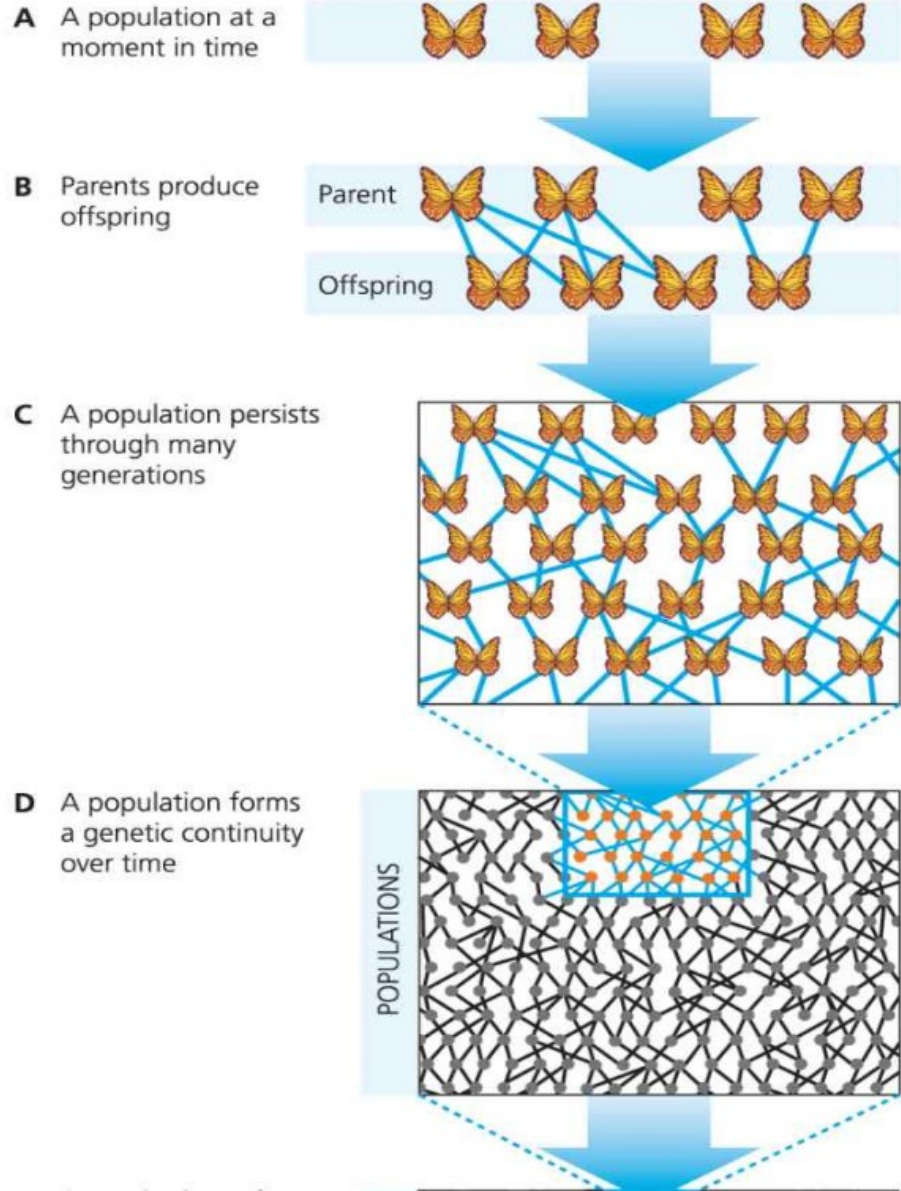
- Η αναζήτηση ομοιοτήτων ανάμεσα σε αλληλουχίες DNA και η στοίχιση πολλαπλών ακολουθιών οδηγούν στην ερώτηση:
  - Πώς σχετίζονται εξελικτικά αυτές οι ακολουθίες;
  - Πώς σχετίζονται εξελικτικά οι οργανισμοί από τους οποίους προέρχονται αυτές οι ακολουθίες;
- **Φυλογενετική:** η μελέτη των σχέσεων ανάμεσα σε προγόνους και απογόνους
  - Η κατάταξη των οργανισμών σύμφωνα με την εξελικτική τους ιστορία
- **Φυλογενετική υπόθεση:** ένα εξελικτικό σενάριο σχετικά με τις σχέσεις προγόνων και απογόνων
- **Φυλογενετικό δέντρο:** μια γραφική απεικόνιση μιας φυλογενετικής υπόθεσης

# Φυλογενετική υπόθεση

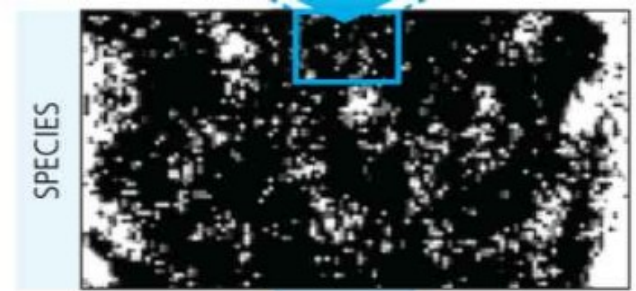
- Όλοι οι ζωντανοί οργανισμοί προέρχονται από έναν **κοινό πρόγονο** και συνδέονται μεταξύ τους με σχέσεις καταγωγής
- Οι φυλογενετικές υποθέσεις εξηγούν τη σημερινή ποικιλομορφία των μορφών ζωής
- Σήμερα οι φυλογενετικές υποθέσεις προκύπτουν από τη σύγκριση αλληλουχιών DNA, RNA και πρωτεϊνών



## Διαδραστικό φυλογενετικό δέντρο



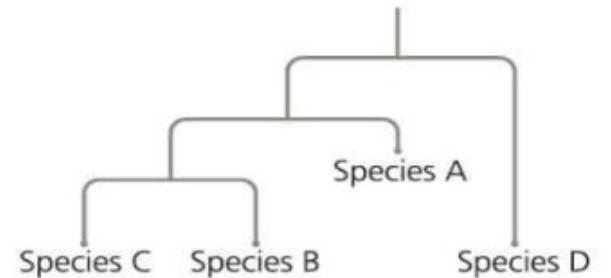
**E** A species is made up of many populations that persist through time



**F** A species diverges into new species



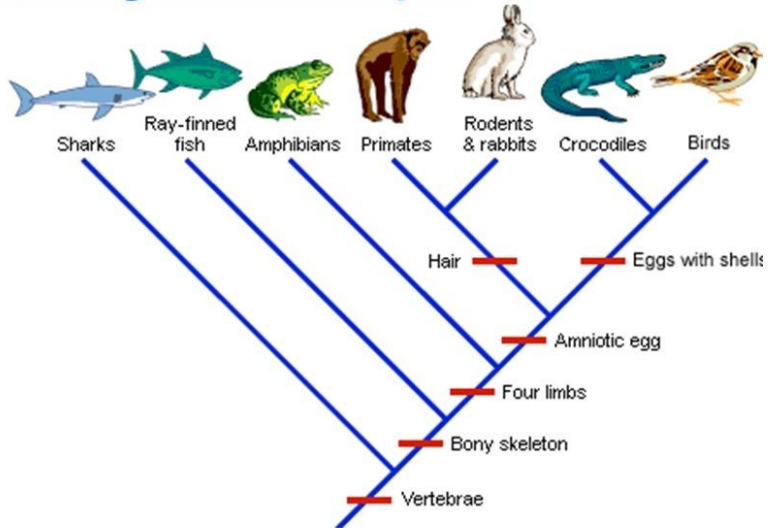
**G** The phylogeny of the four descendant species



Zimmer/Emlen, *Evolution: Making Sense of Life*, 3e, © 2020  
W. H. Freeman and Company

# Φυλογενετικοί χαρακτήρες

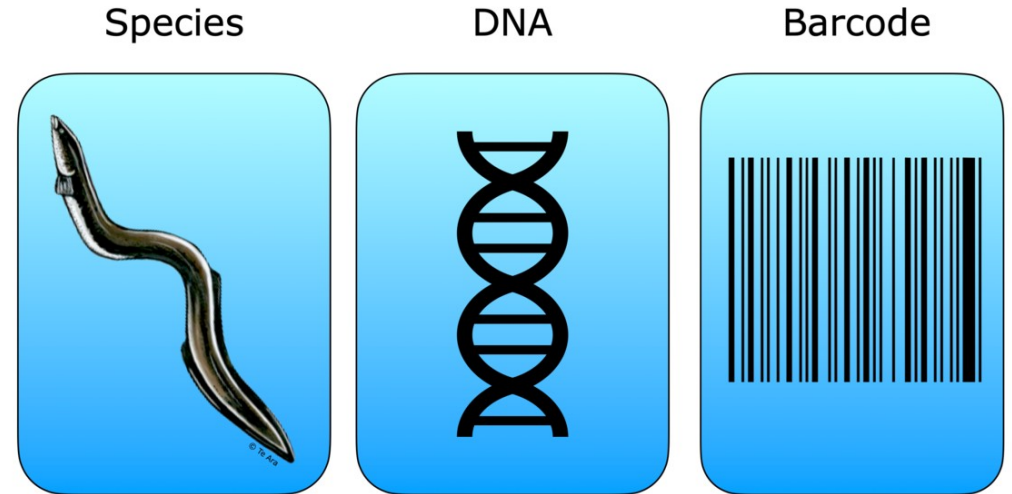
## Cladogram Example:



- Χαρακτηριστικά που χρησιμοποιούνται για την κατασκευή φυλογενετικών δέντρων
  - Ιδανικά παρέχουν πληροφορίες σχετικά με την εξελικτική ιστορία των taxa
- Μικρή ποικιλότητα μέσα στις ενότητες που εξετάζουμε
- Μεγάλη ποικιλότητα ανάμεσα στις ενότητες που εξετάζουμε

# Barcoding genes

- DNA barcode technology
  - molecular recognition technology
  - short and standard DNA fragments for species identification
  - Barcoding genes
  - Μικρή ποικιλομορφία μέσα στο είδος
  - Μεγάλη διαφοροποίηση ανάμεσα στα είδη



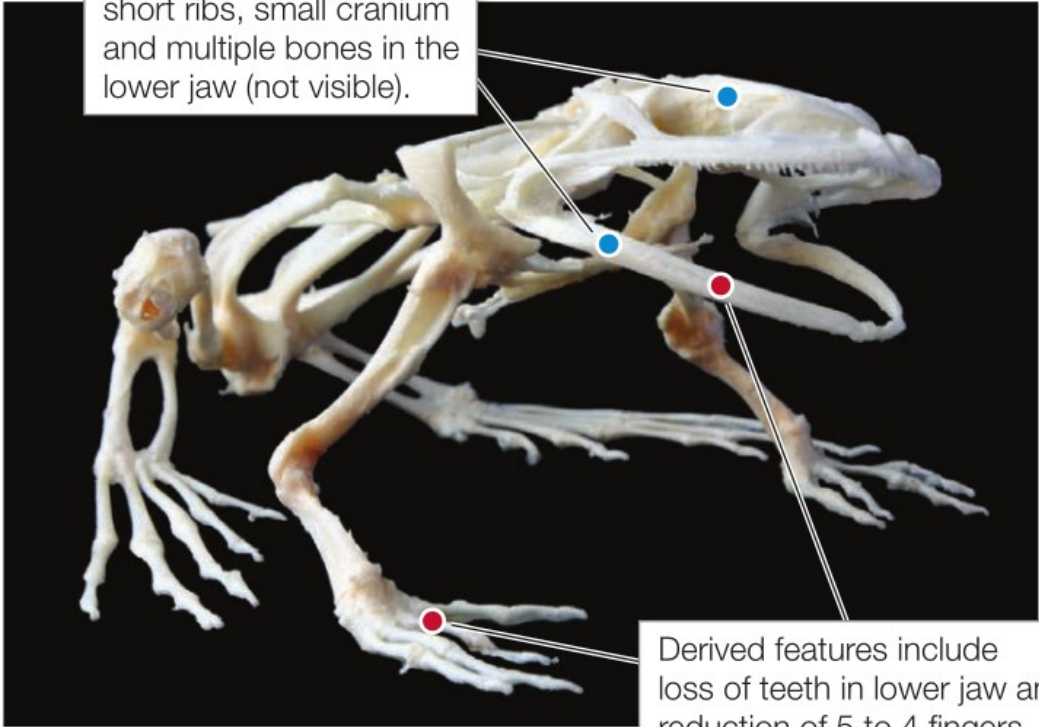


# Παράγωγος και προγονικός χαρακτήρας

- **Παράγωγος** (derived): χαρακτήρας που δεν υπήρχε στον κοινό πρόγονο (απομορφία)
- **Προγονικός** (ancestral): χαρακτήρας που κληρονομήθηκε από κοινό πρόγονο (πλησιομορφία)
- Μια παράγωγη κατάσταση χαρακτήρα που είναι κοινή σε μια ομάδα taxa λέγεται **συναπομορφία** (synapomorphy)
- Απόδειξη ότι τα taxa εξελίχτηκαν από τον ίδιο κοινό πρόγονο
- Το σύνολο των taxa που έχουν προέλθει από τον ίδιο κοινό πρόγονο αποτελούν μια **μονοφυλετική ομάδα** ή έναν **κλάδο** (monophyletic group, clade)

(A)

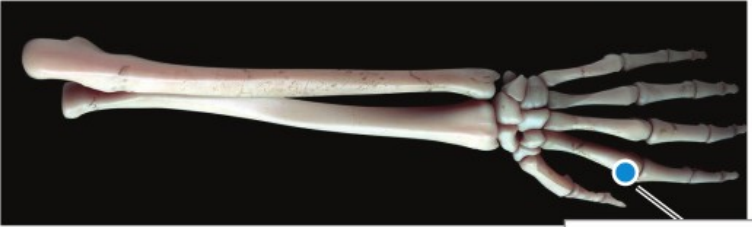
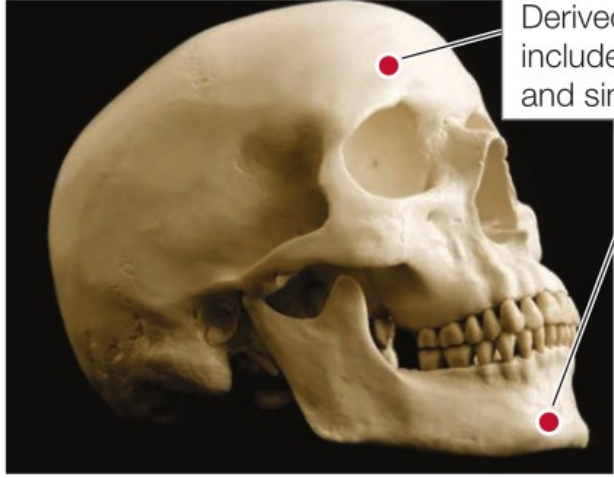
Ancestral features include short ribs, small cranium and multiple bones in the lower jaw (not visible).



Derived features include loss of teeth in lower jaw and reduction of 5 to 4 fingers

(B)

Derived features include large cranial size and single jaw bone

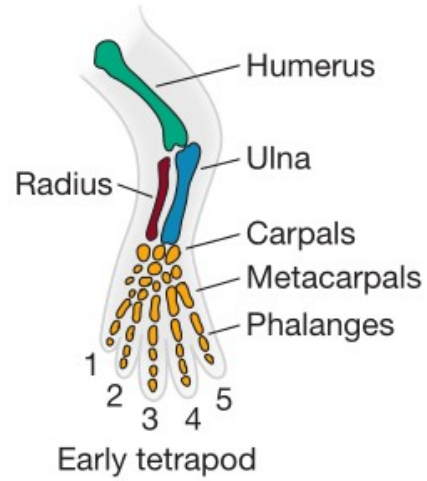


Ancestral features include 5 fingers on the hand.

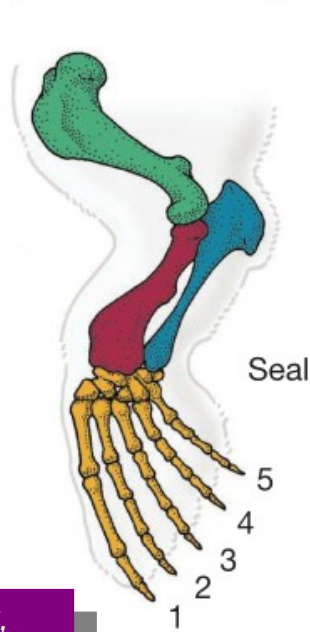
# Ομόλογοι χαρακτήρες

- Οι συναπόμορφοι χαρακτήρες λέγονται και **ομόλογοι**
  - Είναι κοινοί σε δύο (ή περισσότερα) taxa λόγω κληρονόμησης από κοινό πρόγονο
  - Είναι ιδιαίτερα χρήσιμοι για την κατασκευή φυλογενετικών δέντρων γιατί ορίζουν μονοφυλετικές ομάδες
- Ομόλογα γονίδια
  - Γονίδια που υπάρχουν σε δύο ή περισσότερους ορφανισμούς λόγω κληρονόμησης από κοινό πρόγονο
  - Ειδική περίπτωση ομόλογων χαρακτήρων

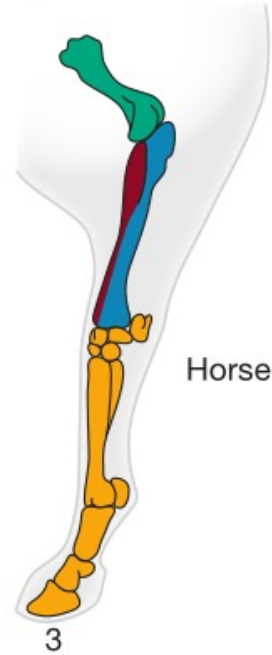
# Ομόλογοι χαρακτήρες



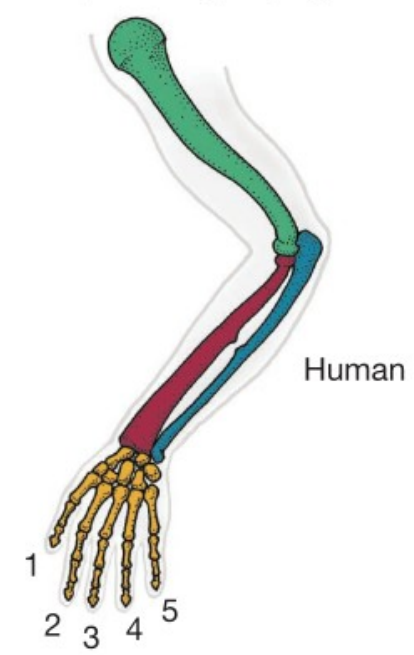
Adapted for swimming



Adapted for running



Adapted for grasping

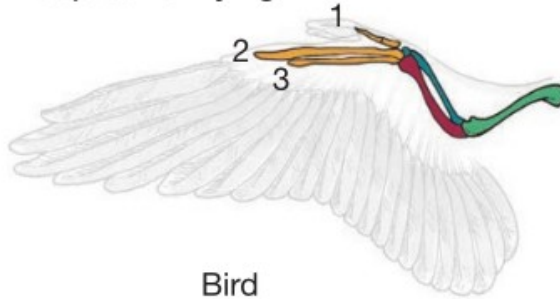


*Evolution 4th Edition, Futuyma & Kirkpatrick, Sinauer Associates 2017*

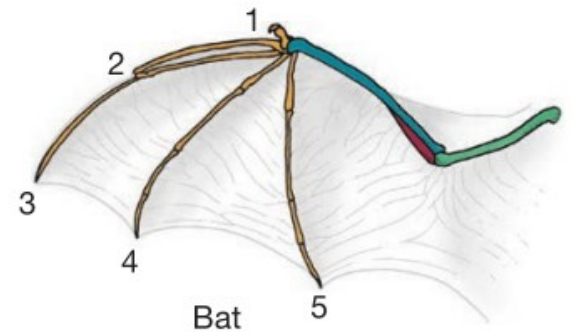


Pterosaur (extinct)

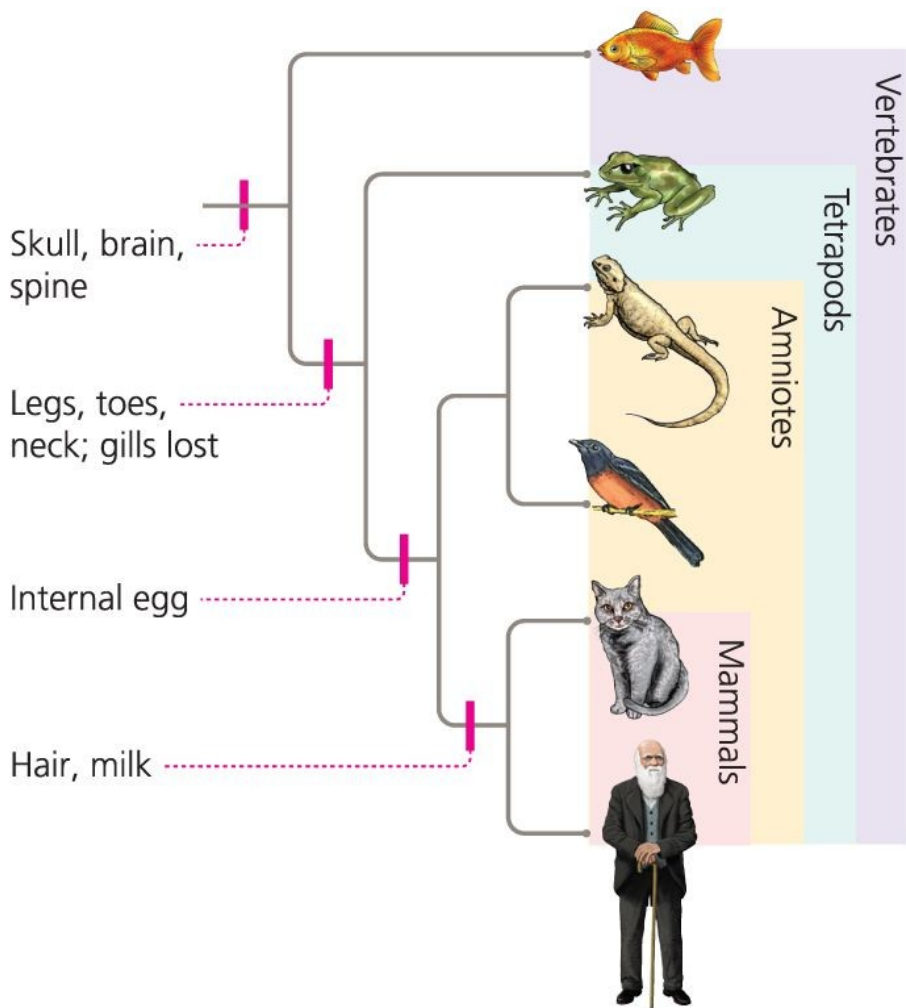
Adapted for flying



Bird

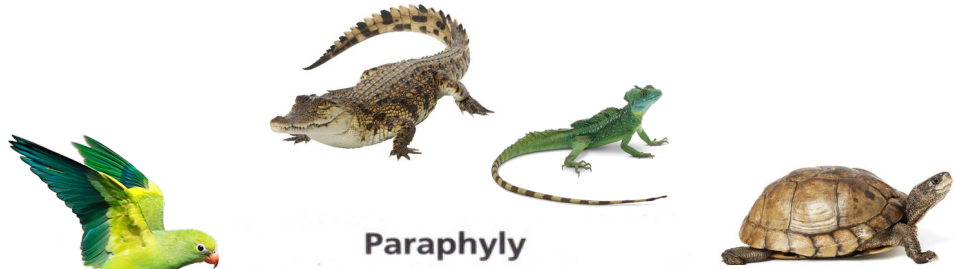


Bat



Οι ομόλογοι χαρακτήρες μας βοηθούν να ορίσουμε μονοφυλετικές ομάδες.

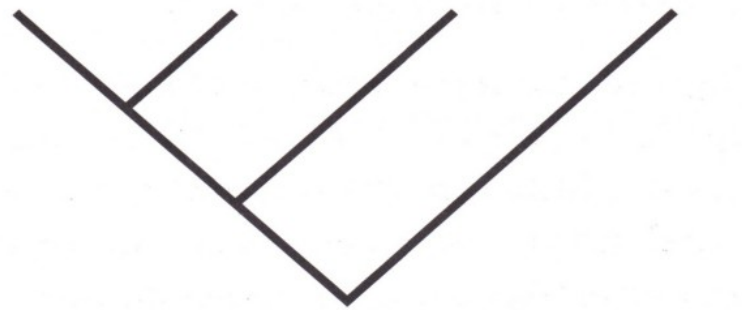
Η ταξινόμηση των taxa αποκτά εξελικτική τεκμηρίωση και σημασία



**Paraphyly**

'Reptiles'

Birds    Crocodiles    Lizards    Turtles



*Ciconia ciconia*

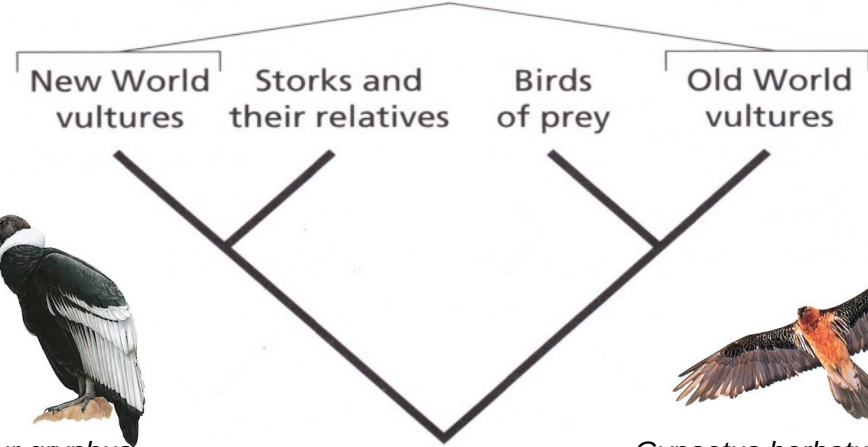


*Aquila chrysaetos*

**Polyphyly**

'Vultures'

New World vultures    Storks and their relatives    Birds of prey    Old World vultures



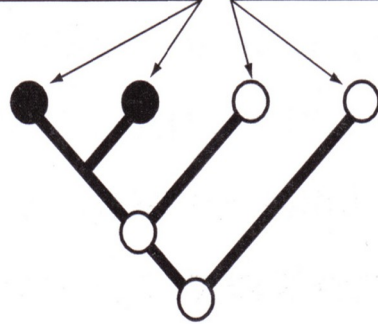
*Vultur gryphus*



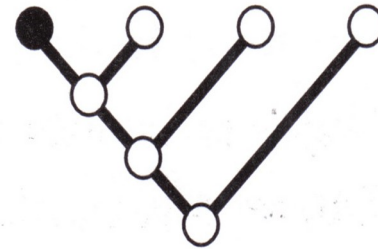
*Gypaetus barbatus*

# Καταστάσεις χαρακτήρων

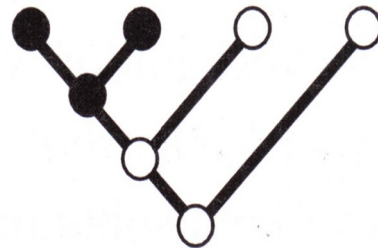
Apomorphy      Plesiomorphy



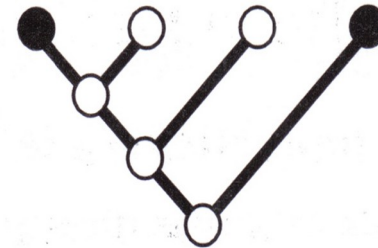
**Autapomorphy**



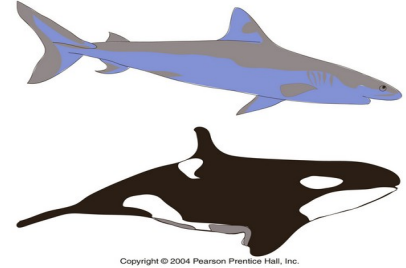
**Synapomorphy**



**Homoplasy**



# Ομολογία vs. Ομοπλασία



- Οι συναπόμορφοι χαρακτήρες (ή οι αλληλουχίες του DNA) λέγονται και **ομόλογοι**
  - Είναι κοινοί σε δύο taxa λόγω κληρονόμησης από κοινό πρόγονο
  - Είναι ιδιαίτερα χρήσιμοι για την κατασκευή φυλογενετικών δέντρων γιατί ορίζουν μονοφυλετικές ομάδες (κλάδους)
- Οι ομοπλασίες δυσκολεύουν την κατασκευή φυλογενετικών δέντρων γιατί οδηγούν σε λάθος ομαδοποιήσεις
  - Είναι όμοιοι χαρακτήρες που όμως εξελίχτηκαν χωριστά

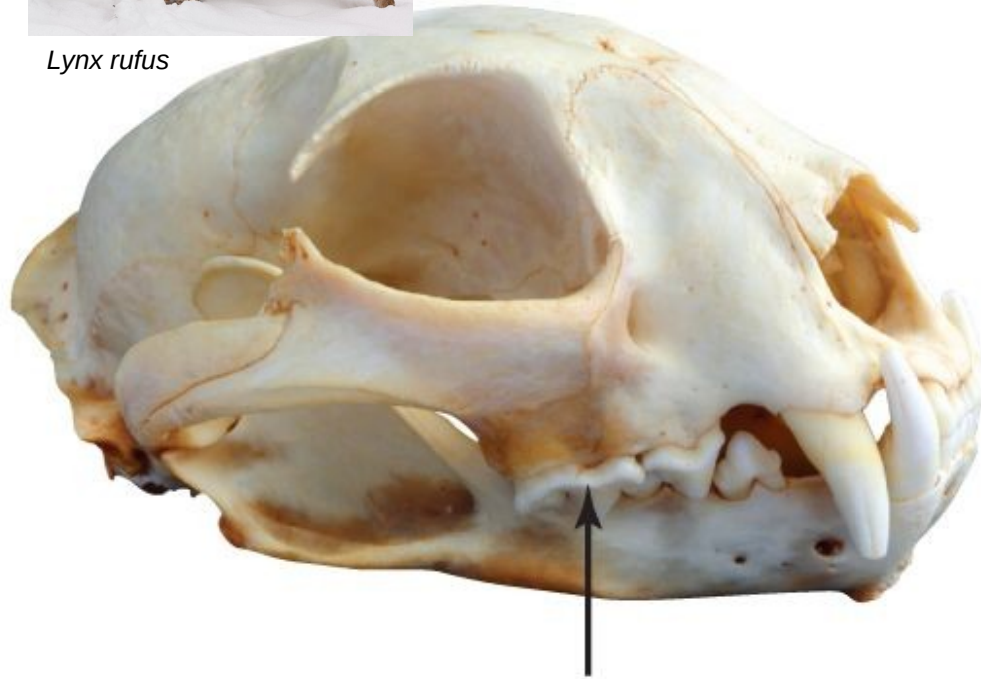


Ψάχνοντας για συναπτόμορφους χαρακτήρες: Οι **κοπτικοί γομφίοι** υπάρχουν τόσο στο *Lynx rufus* όσο και στον λύκο του Μεξικού (*Canis lupus baileyi*) και ορίζουν τον κλάδο των σαρκοφάγων θηλαστικών

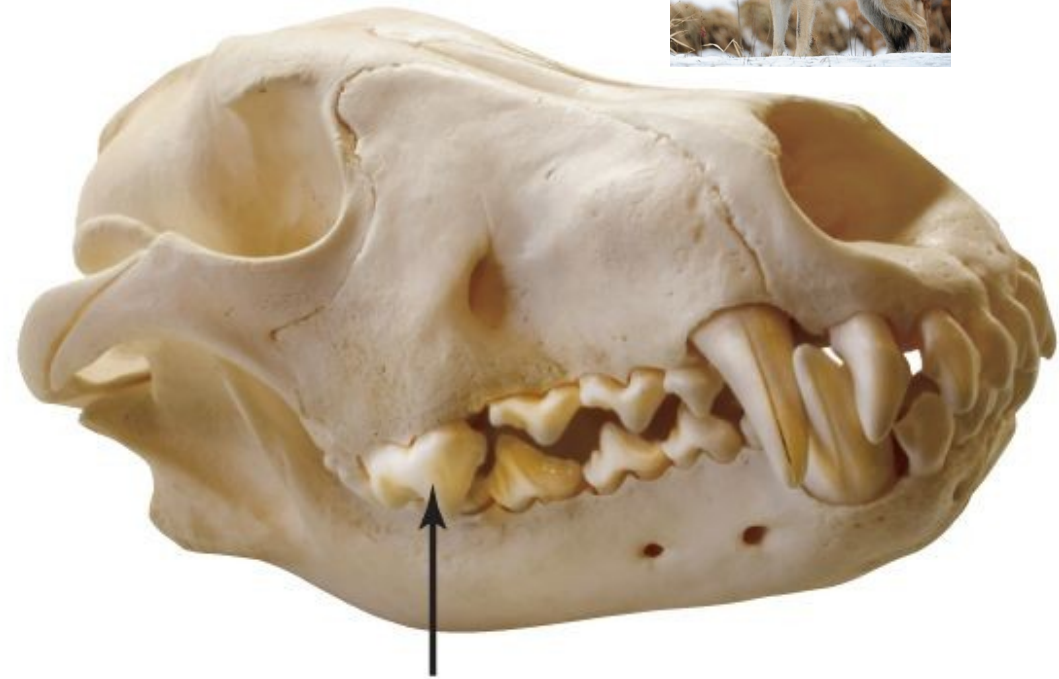


*Lynx rufus*

*Canis lupus baileyi*

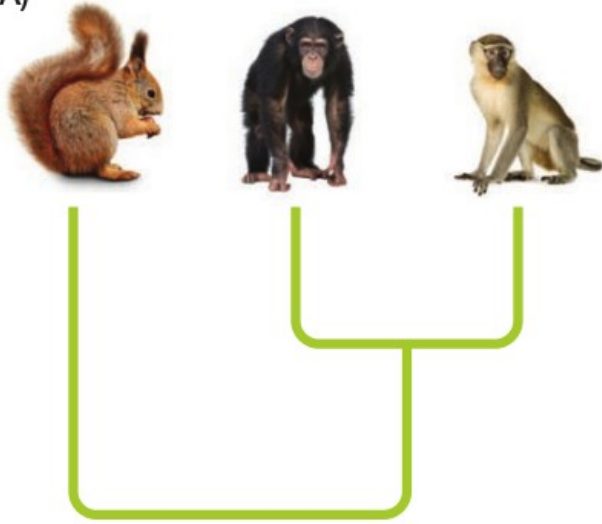


Carnassial

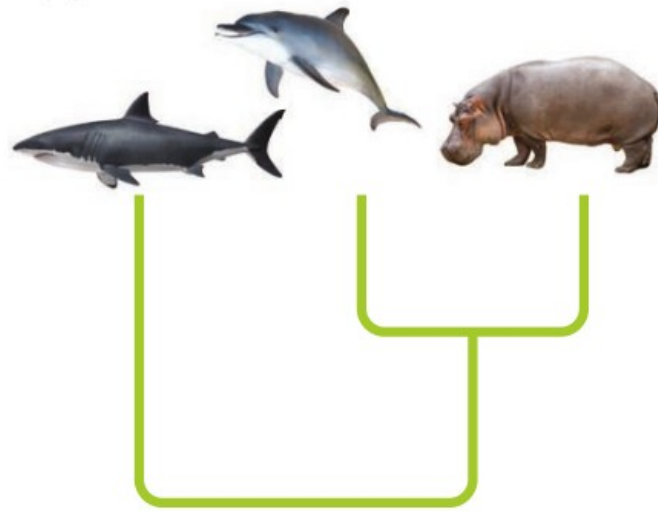


Carnassial

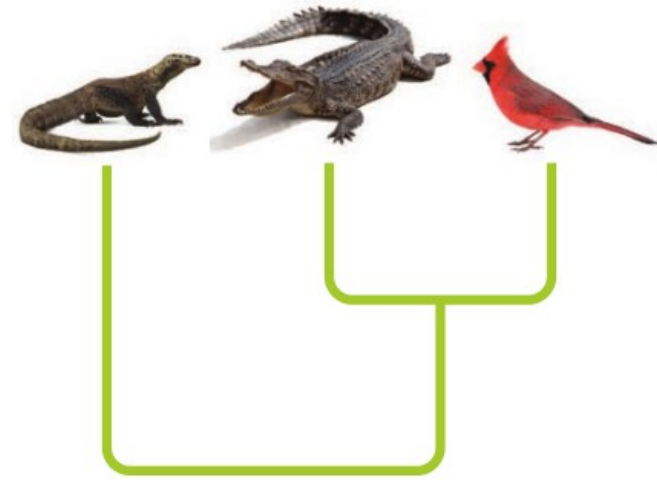
(A)



(B)



(C)



**FIGURE 2.7** Similarity versus relationship.

(A) Chimpanzees and monkeys are more closely related to each other than to rodents, and they are also more similar. (B) Dolphins are closely related to hipopotamuses and other mammals, even though they superficially look more like sharks. (C) Crocodiles and birds share a more recent common ancestor than either does with lizards, but birds look very different because they have undergone more evolutionary changes than crocodiles.

Η ομοιότητα δεν δηλώνει πάντα κοινή καταγωγή

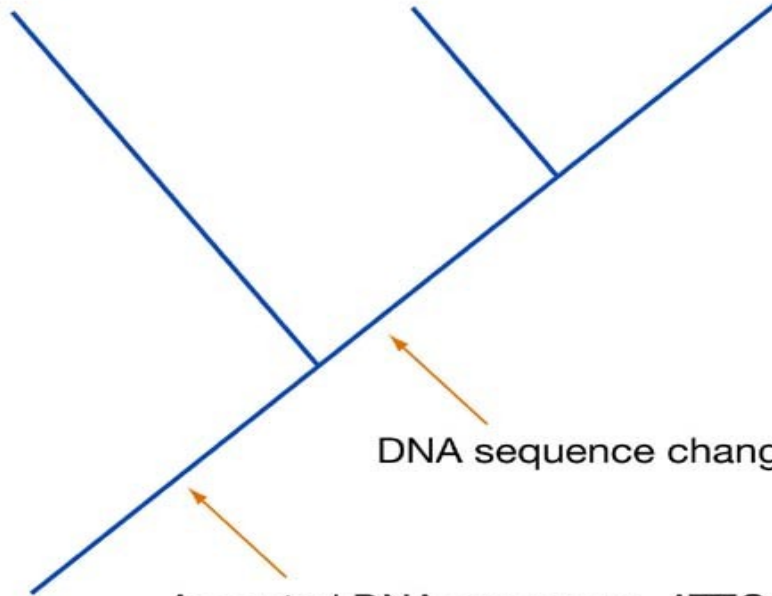
(a) Mutations can create synapomorphies

DNA sequences in descendants:

ATTGCT**A**TTC

ATTGCT**T**TTTC

ATTGCT**T**TTTC

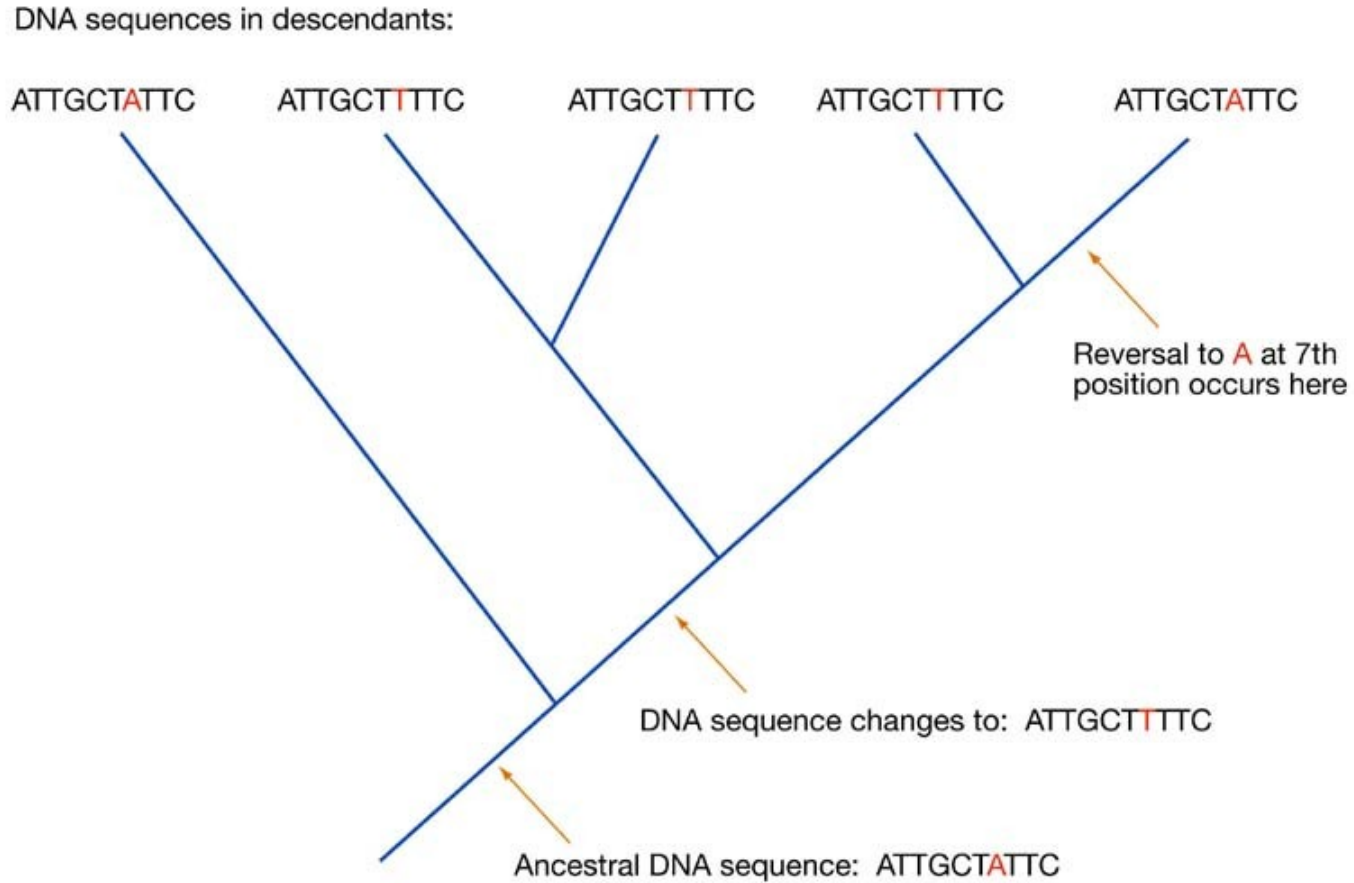


DNA sequence changes to: ATTGCT**T**TTTC

Ancestral DNA sequence: ATTGCT**A**TTC

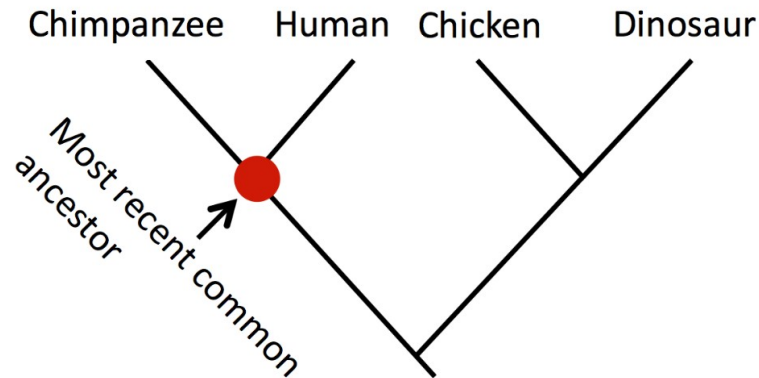
Copyright © 2004 Pearson Prentice Hall, Inc.

(b) Reversals ("back-mutations") can remove synapomorphies

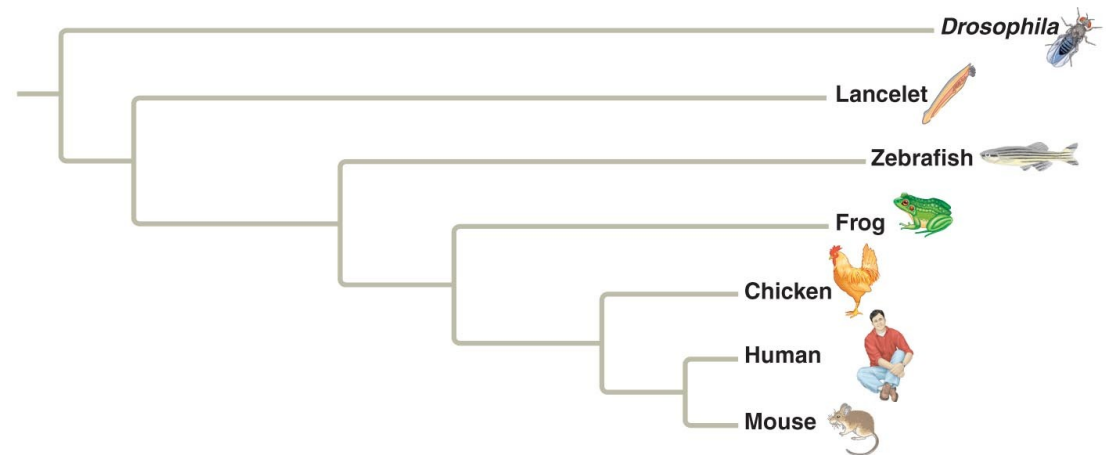


# Κλαδόγραμμα vs. φυλογενετικό δέντρο

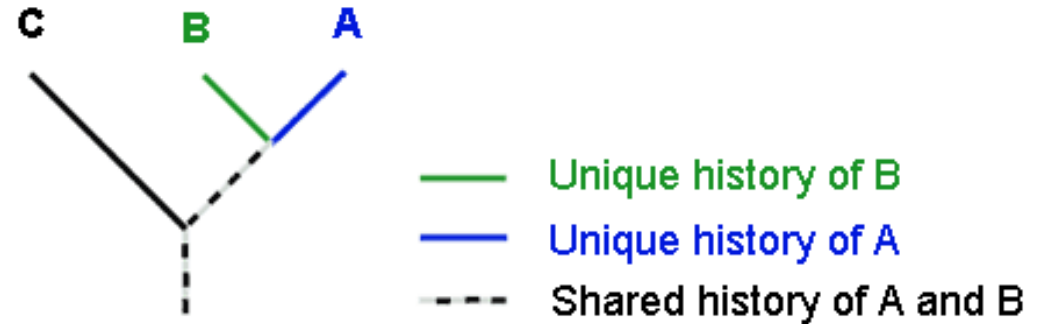
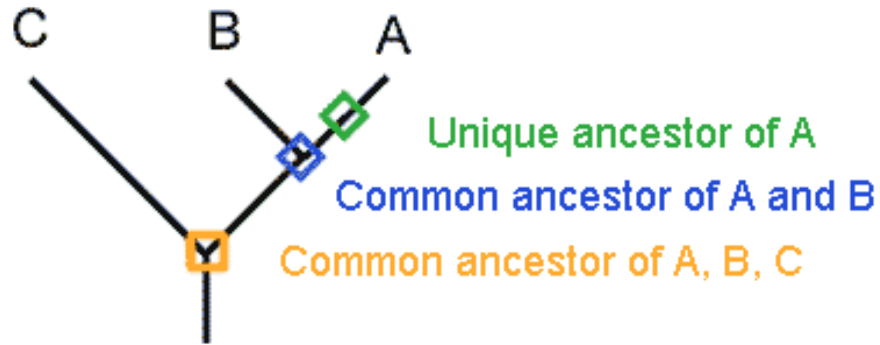
- **Κλαδόγραμμα:** δείχνει τις εξελικτικές σχέσεις ανάμεσα σε taxa



- **Φυλογενετικό δέντρο:** κάνει το ίδιο με τους κλάδους του να είναι ανάλογοι με τον χρόνο



# Διαβάζοντας ένα δέντρο



# Κατασκευή φυλογενετικού δέντρου

- Μέσω αποστάσεων (γενετικών)
  - Γενετική απόσταση: ο αριθμός των μεταλλάξεων που χωρίζουν δύο αλυσίδες DNA
  - Πιο γενικά, ο αριθμός των εξελικτικών συμβάντων ανάμεσα σε δύο taxa
- Μέσω χαρακτήρων
  - Αναζήτηση του κατάλληλου αλγόριθμου για τον υπολογισμό της πιθανότητας ένα X δέντρο να δώσει ένα συγκεκριμένο αποτέλεσμα αλληλουχιών (που εξετάζουμε) και στη συνέχεια επιλέγουμε από ένα σύνολο πιθανών δέντρων αυτό με τη μέγιστη πιθανότητα

# Δέντρο μέσω αποστάσεων

- UPGMA

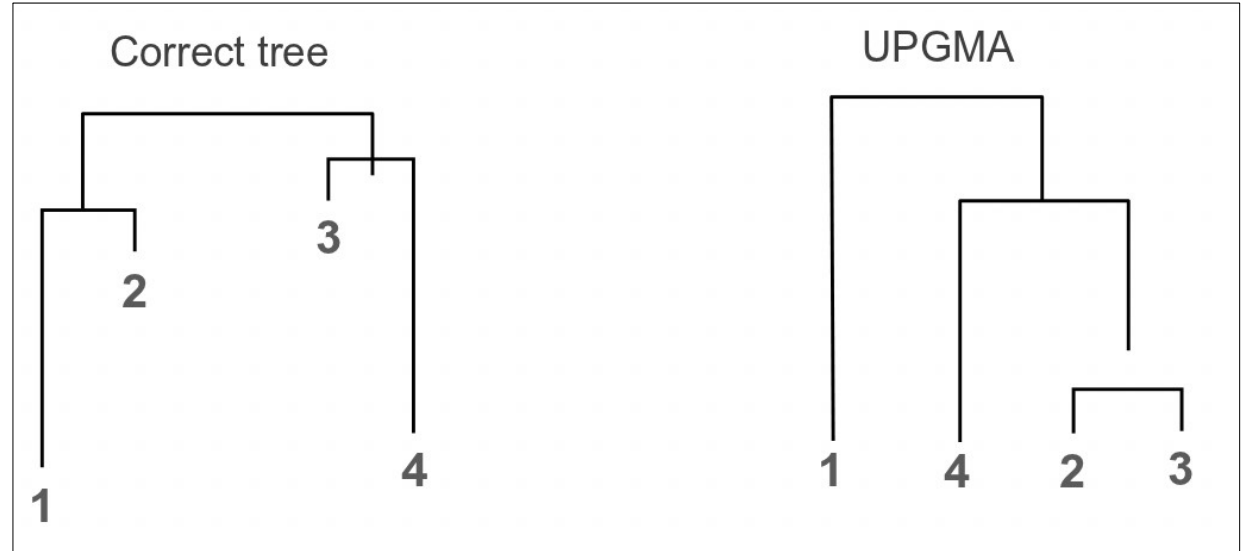
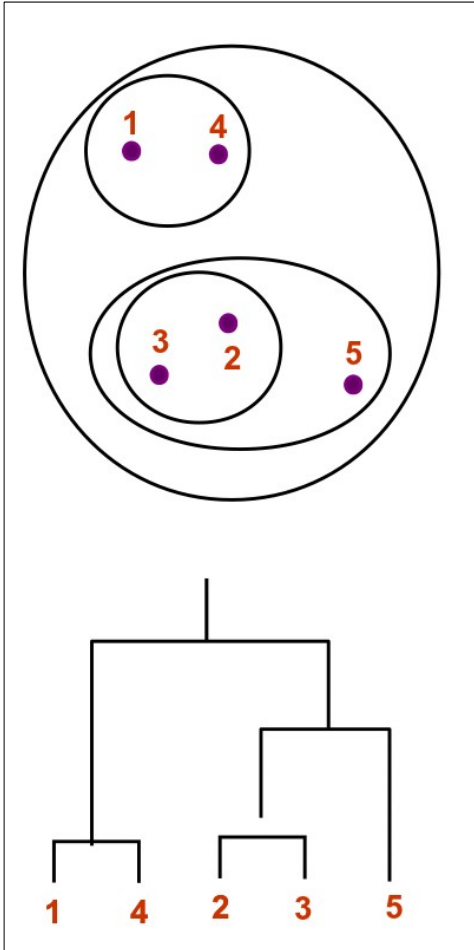
- Ομαδοποίηση taxa με την πιο μικρή απόσταση και υπολογισμός αποστάσεων της κάθε ομάδας με τα υπόλοιπα taxa ως ο μέσος όρος των δύο taxa
- Επανάληψη της διαδικασίας μειώνοντας διαστάσεις του πίνακα αποστάσεων
- Η πιο διαδεδομένη μέθοδος, χωρίς όμως υπολογισμό πραγματικών εξελικτικών πληροφοριών

- Neighbor Joining Method (NJ)

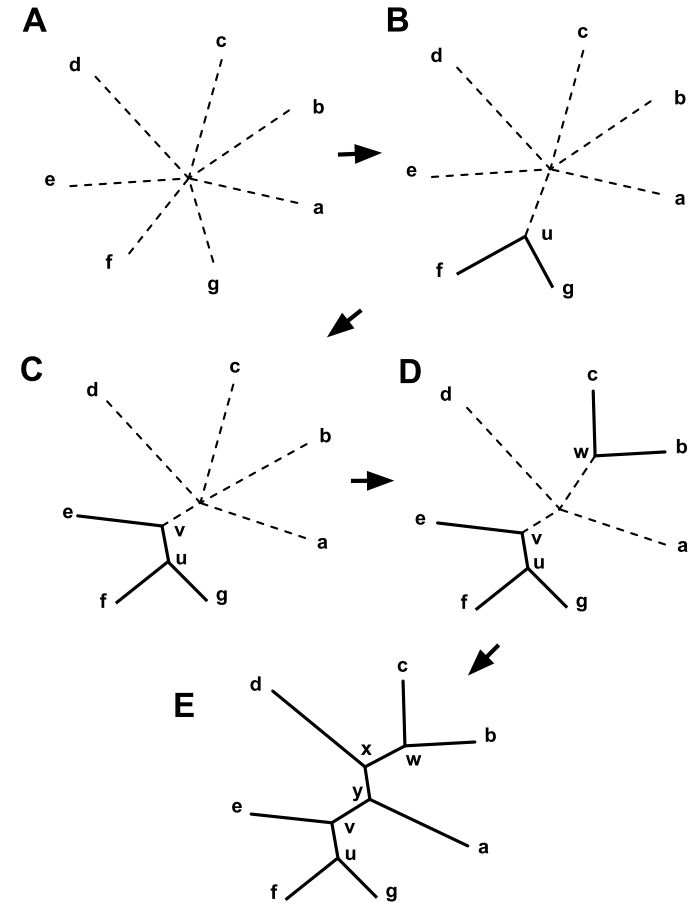
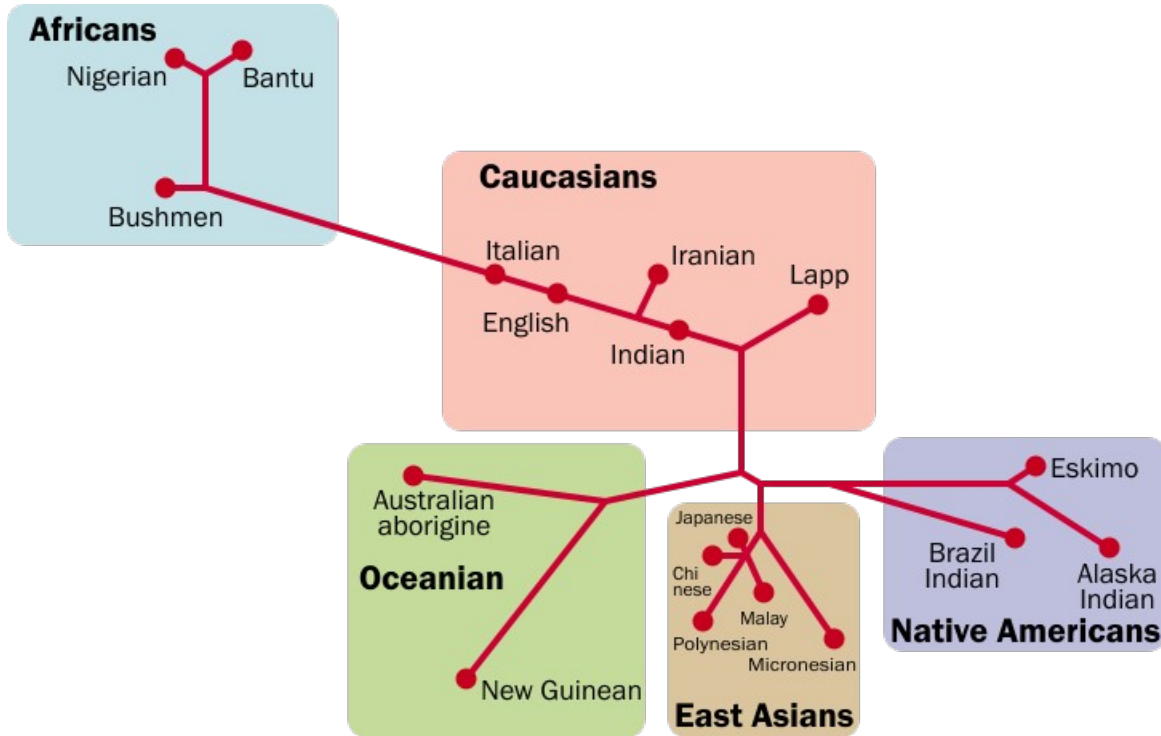
- Ξεκινά με ένα αδιαμόρφωτο δέντρο, ενώνοντας το πιο κοντινό από όλα τα δυνατά ζευγάρια των taxa. Ένας κλάδος ενώνει την ομάδα με τα υπόλοιπα taxa μέχρι να ολοκληρωθεί το δέντρο
- Δίνει καλύτερα αποτελέσματα από το UPGMA, αλλά έχει προβλήματα σε μεγαλύτερες αποστάσεις



# Unweighted Pair-Group Method with Arithmetic mean (UPGMA)



# Neighbor Joining Method

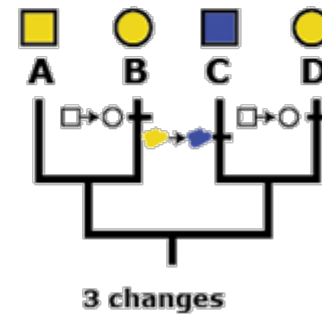
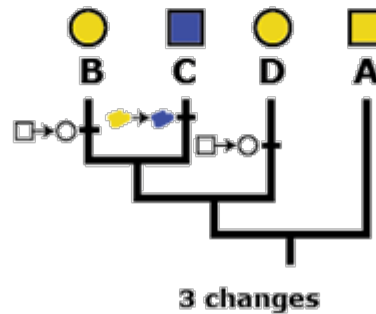
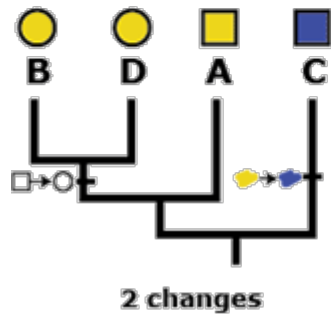


# Δέντρο μέσω χαρακτήρων

- Μέγιστη φειδωλότητα (maximum parsimony)
  - η πιο απλή εξήγηση είναι και η πιο πιθανή
  - Το πιο σωστό δέντρο είναι αυτό που περιλαμβάνει τα λιγότερα βήματα από τον κοινό πρόγονο
- Μέγιστη πιθανοφάνεια (maximum likelihood)
  - Όπως και η MP υπολογίζει μια τιμή για όλα τα πιθανά δέντρα και καταλήγει σε αυτό που είναι πιο πιθανό
  - Η πιθανοφάνεια υπολογίζεται για κάθε συνδυασμό taxa, ώστε το εξελικτικό μοντέλο που οδηγεί σε αυτό το αποτέλεσμα να είναι το πιο πιθανό, δηλαδή να εξηγεί καλύτερα τα δεδομένα που έχουμε
  - Η μέθοδος αυτή είναι η πιο αργή στην εκτέλεση αλλά συνήθως βγάζει τα πιο έγκυρα δέντρα με την πιο χρήσιμη εξελικτική πληροφορία

# Maximum parsimony

taxon	characters	
	shape	color
A	□	yellow
B	○	yellow
C	□	blue
D	○	yellow



# Maximum likelihood

- Θεωρούμε ένα εξελικτικό μοντέλο μεταλλάξεων
- Ξεκινούμε να προσθέτουμε taxa στο δέντρο και εξετάζουμε την πιθανοφάνεια της κάθε εναλλακτικής τοπολογίας του κάθε δέντρου σε κάθε βήμα
- Συνεχίζουμε στο επόμενο βήμα με το δέντρο που έχει την μέγιστη πιθανοφάνεια
  - Η πιθανοφάνεια είναι κρίνεται με βάση την πιθανότητα κάθε μέρους μιας τοπολογίας να οδηγεί τελικά στα δεδομένα που έχουμε
- Διαδικασία που απαιτεί χρόνο και πόρους
  - Δίνει τα πιο αξιόπιστα φυλογενετικά δέντρα

# Maximum Parsimony vs Maximum Likelihood

More Information Online [WWW.DIFFERENCEBETWEEN.COM](http://WWW.DIFFERENCEBETWEEN.COM)

	Maximum Parsimony	Maximum Likelihood
DEFINITION	Maximum parsimony is the technique of drawing a phylogenetic tree with the minimal number of character state changes	Maximum likelihood is the technique of drawing a phylogenetic tree with the maximum likelihood between genetic data
CHARACTERS CONSIDERED	Low in number	High in number
BRANCHES	Branches are long	Short trees with short branches
RELIABILITY	Less reliability	High reliability

# Εργαλεία υπολογισμού και κατασκευής ML δέντρων

- IQtree
  - <http://www.iqtree.org/>
- RaxML
  - <https://raxml-ng.vital-it.ch/#/>

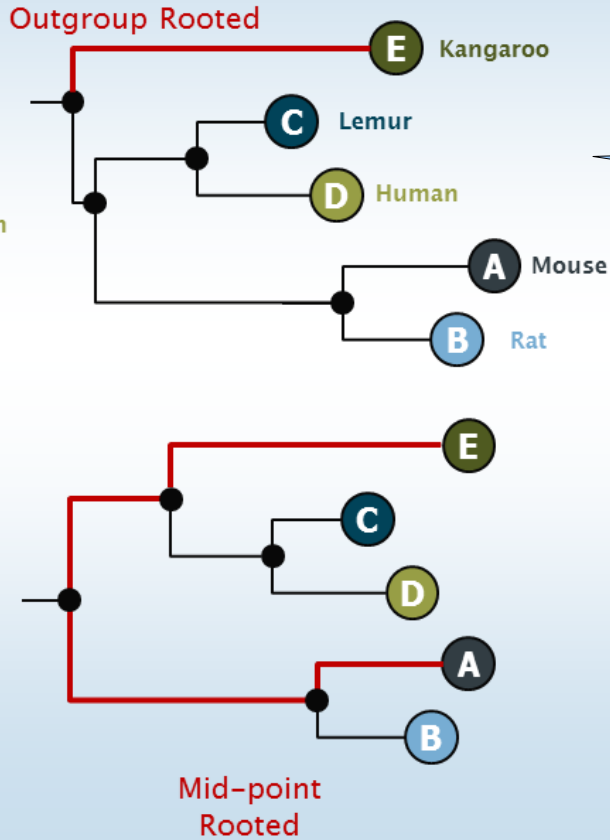
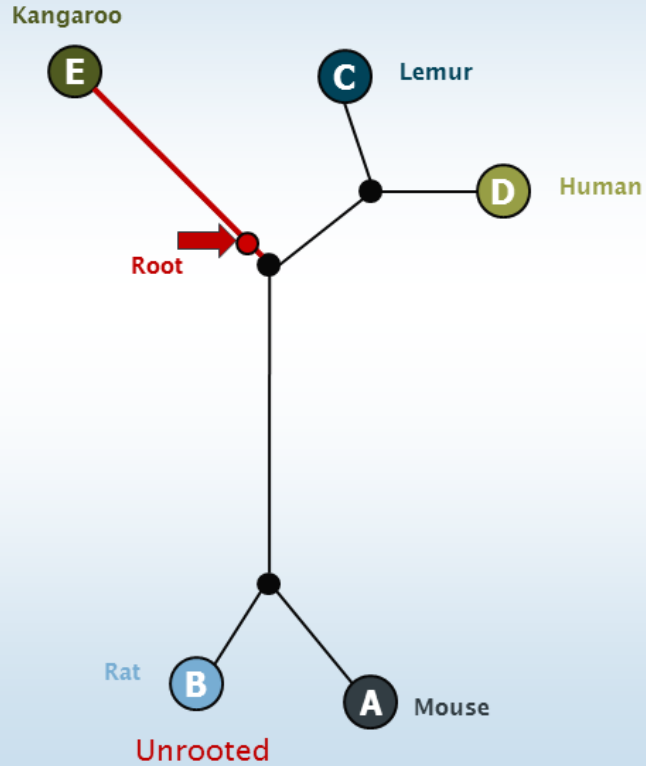
# Ρίζα και εξωομάδα



- Ως **ρίζα** (root) θεωρούμε τον κοινό πρόγονο των taxa ενός δέντρου
  - Συνήθως δεν γνωρίζουμε τη ρίζα και απλώς την τοποθετούμε στο μέσο των taxa με την μεγαλύτερη απόκλιση
    - Δεν είναι πάντα σωστό αυτό
- Η **εξωομάδα** (outgroup) είναι ένα taxon που γνωρίζουμε ότι βρίσκεται πολύ μακριά εξελικτικά από τα υπόλοιπα taxa του δέντρου
  - Η εξωομάδα μπορεί να χρησιμοποιηθεί για να βρεθεί η ρίζα σε ένα δέντρο
    - Κάποιες φορές η εξωομάδα δεν είναι τόσο μακρινή και δημιουργεί ασάφειες
    - Το ίδιο και αν η εξωομάδα είναι τόσο μακρινή ώστε να ομαδοποιούνται άσχετα είδη



# Outgroup Rooting

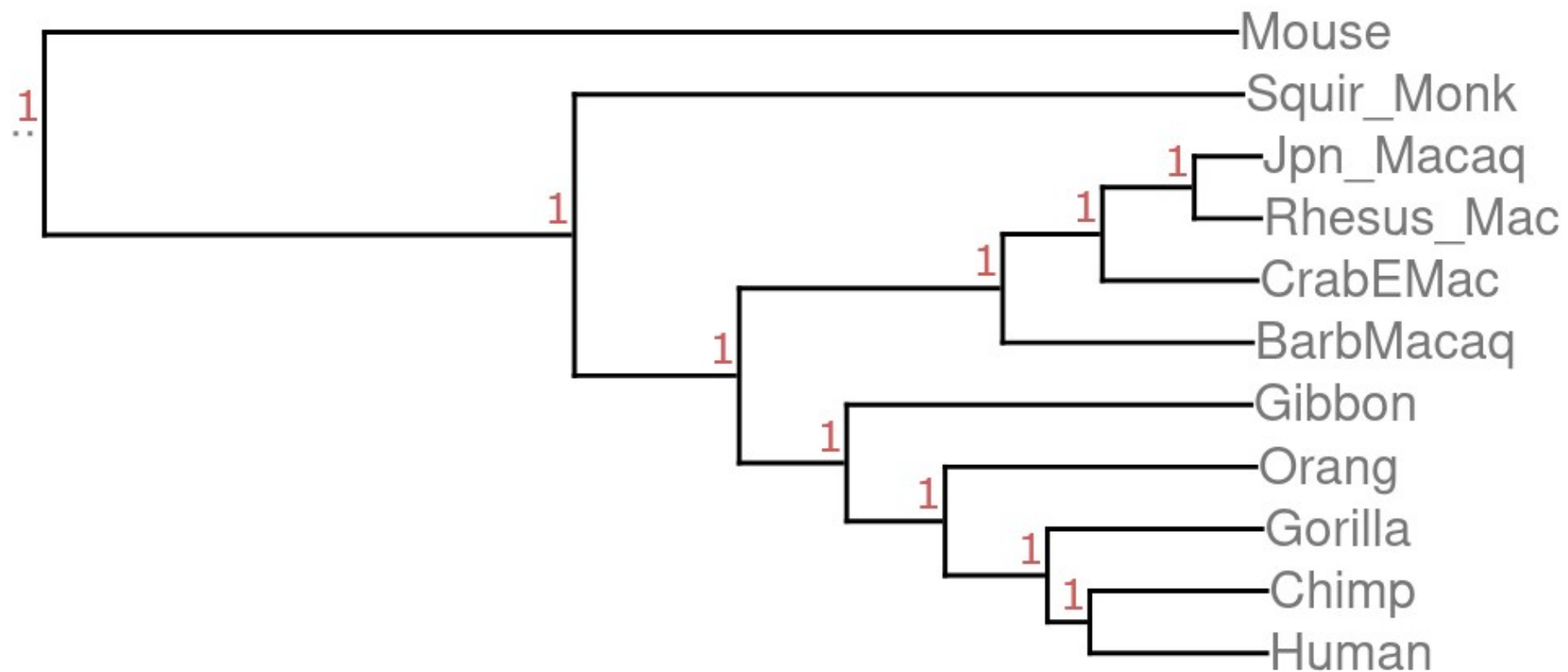


Σωστό

Λάθος

# Δέντρο χωρίς ρίζα

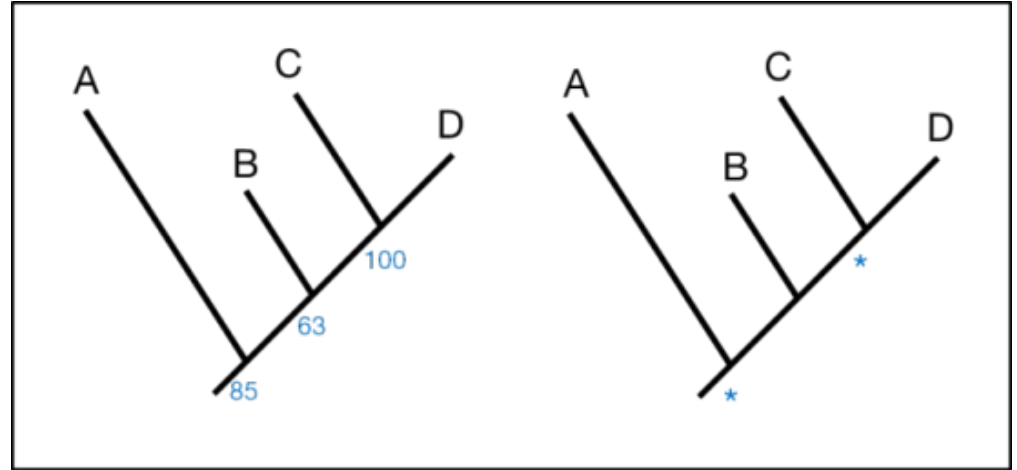




0.18

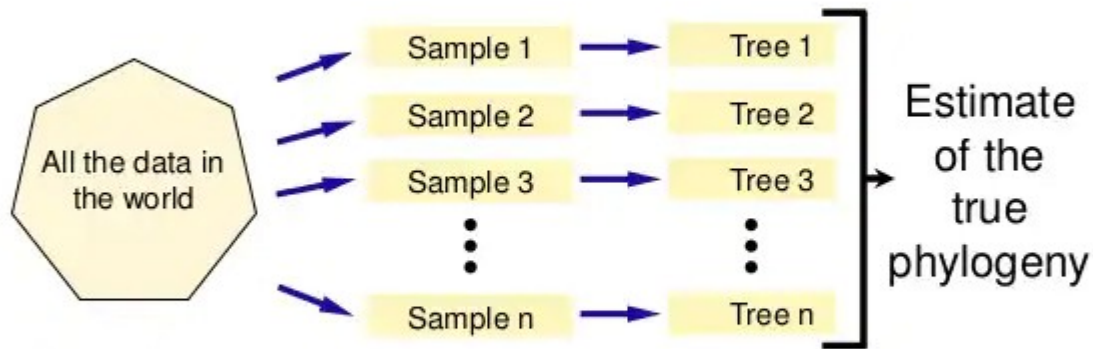
# Bootstrap μέθοδοι

- Έλεγχος αξιοπιστίας ενός φυλογενετικού δέντρου
- Επαναληπτικές δειγματοληψίες από τα δεδομένα και έλεγχος επανάληψης συγκεκριμένων κόμβων
- Οι πιο αξιόπιστοι κόμβοι είναι αυτοί που εμφανίζονται στις περισσότερες περιπτώσεις



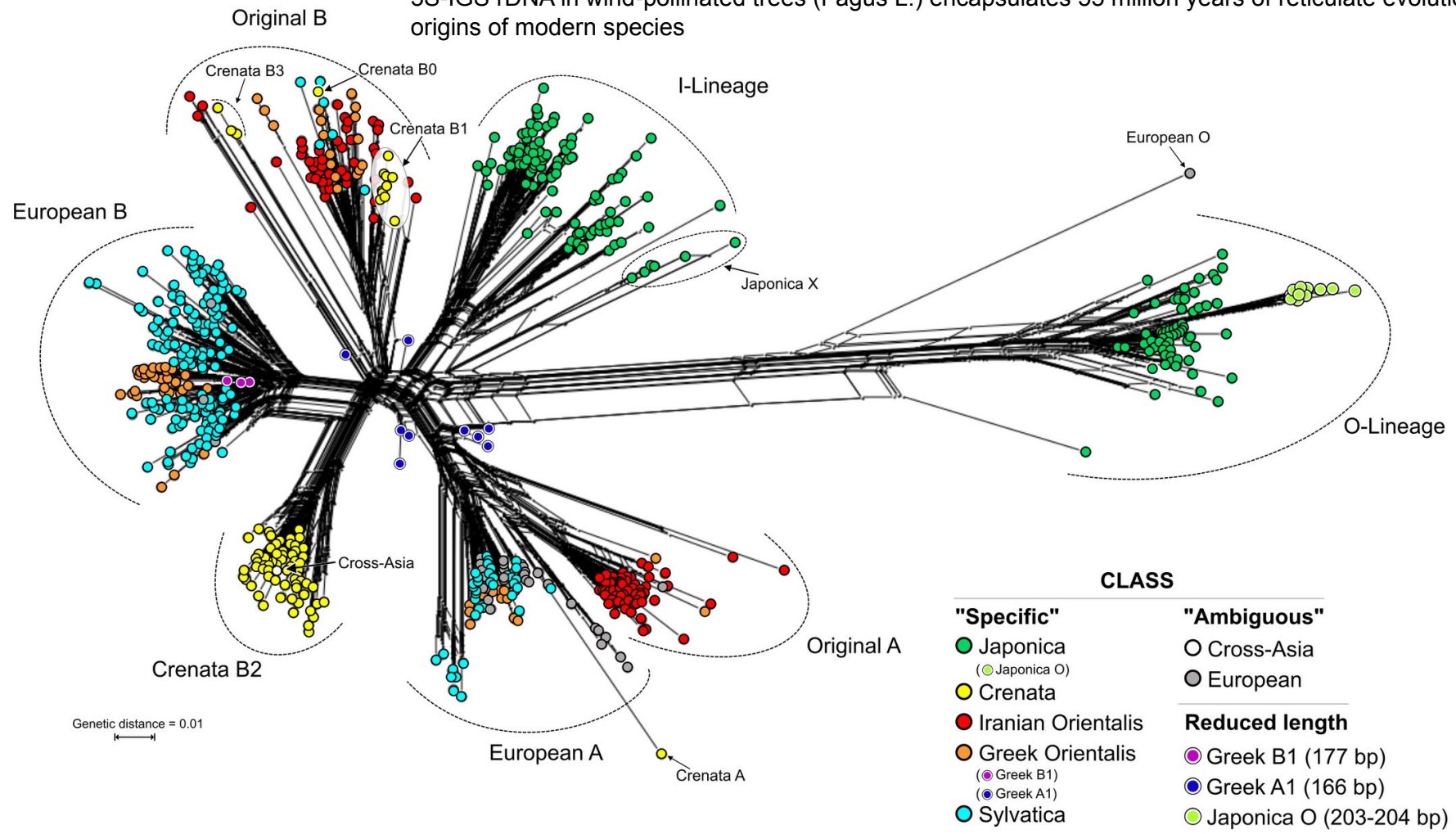
# Bootstrapping

the ideal world



Build replicates by resampling from unlimited data

5S-IGS rDNA in wind-pollinated trees (*Fagus L.*) encapsulates 55 million years of reticulate evolution and hybrid origins of modern species



# Κατασκευή δέντρων



<http://etetoolkit.org/treeview/>

A screenshot of the NCBI Tree Viewer website. The browser address bar shows "https://www.ncbi.nlm.nih.gov/projects/treeview/". The page title is "NCBI's Tree Viewer". Under the heading "Sample Links:", there are several links with descriptions: "tree\_txt.asn file shown as Circular Tree", "tree\_txt.asn file shown as Slanted Cladogram", "carbohydrate\_kinase\_FGGY\_midpoint.asn file shown as Rectangle Cladogram sorted by distance", "genometree167.asn shown as Circular Tree", "recA\_nj\_nw.asn file shown as Rectangle Cladogram", "sample\_tree.asn file shown as Radial Tree", "kmer\_9219\_tree.asn file shown as Rectangle Cladogram", "kmer.asn file shown as Rectangle Cladogram with vertical only zoom", "primates.asn file from ftp site shown as Radial Tree", and "NewickTree.nwk in Newick format file from ftp site shown as Rectangle Cladogram". At the bottom, there is a "Contact | Copyright | Disclaimer | Privacy | Accessibility" link and the address "National Center for Biotechnology Information, US National Library of Medicine, 8600 Rockville Pike, Rockville, MD USA 20894".

<https://www.ncbi.nlm.nih.gov/projects/treeview/>