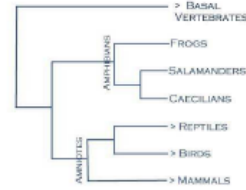


Μοριακή Εξέλιξη και Φυλογένεση



Το δένδρο της Ζωής

Γιατί είναι σημαντική η Φυλογένεση;

Για την κατανόηση και την κατάταξη της ποικιλότητας της ζωής στη Γη

Για τον έλεγχο εξελικτικών υποθέσεων:

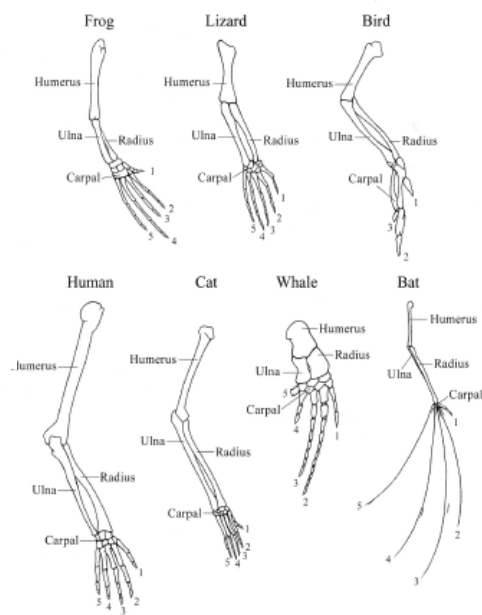
- εξέλιξη χαρακτήρων
- συν-εξέλιξη
- τρόποι και πρότυπα ειδογένεσης
- εξέλιξη συσχετισμένων χαρακτήρων
- βιογεωγραφία
- γεωγραφική προέλευση
- ηλικία ταξινομικών ομάδων
- φύση μοριακής εξέλιξης
- επιδημιολογία ασθενειών



...και πολλές άλλες εφαρμογές!

Ομολογία

Ομοιότητες
(μορφολογικές, μοριακές,
συμπεριφοράς) που
κληρονομούνται μέσω
κοινού προγόνου



Φειδωλότητα

- Βασίζεται στην αρχή του “ξυραφιού του Occam” δηλαδή «Η καλύτερη εξήγηση είναι η απλούστερη εξήγηση»
- Το καλύτερο δένδρο είναι επομένως αυτό το οποίο απαιτεί τον λιγότερο αριθμό εξελικτικών γεγονότων

Φυλογενετική ανάλυση

Αρχές:

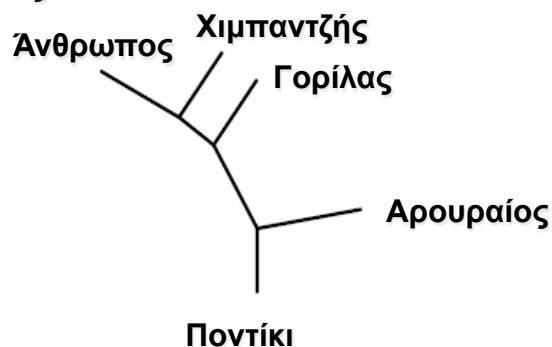
- Υποθέτει ότι όμοια χαρακτηριστικά είναι ομόλογα εκτός και αν αποδειχθεί το αντίθετο
- Χρησιμοποιεί **κοινούς αποκλίνοντες** χαρακτήρες και όχι κοινούς προγονικούς



Willi Hennig
(1913-1976)

Οπτικοποιώντας την Ιστορία – Χρησιμοποιώντας Φυλογενετικά δένδρα

Ο στόχος των φυλογενετικών δένδρων είναι να διευκρινίσει πως μια ομάδα αντικειμένων (συνήθως γονίδια ή οργανισμοί) συνδέονται μεταξύ τους και να οπτικοποιήσει τις εξελικτικές σχέσεις



Φυλογενετικά δένδρα

Κόμβος: ένα σημείο του κλαδιού σε ένα δένδρο (μια υποτιθέμενη προγονική ταξινομική ομάδα)

Κλαδί ή βραχιόνας: προσδιορίζει τη σχέση μεταξύ των ομάδων με όρους απογόνου και προγόνου

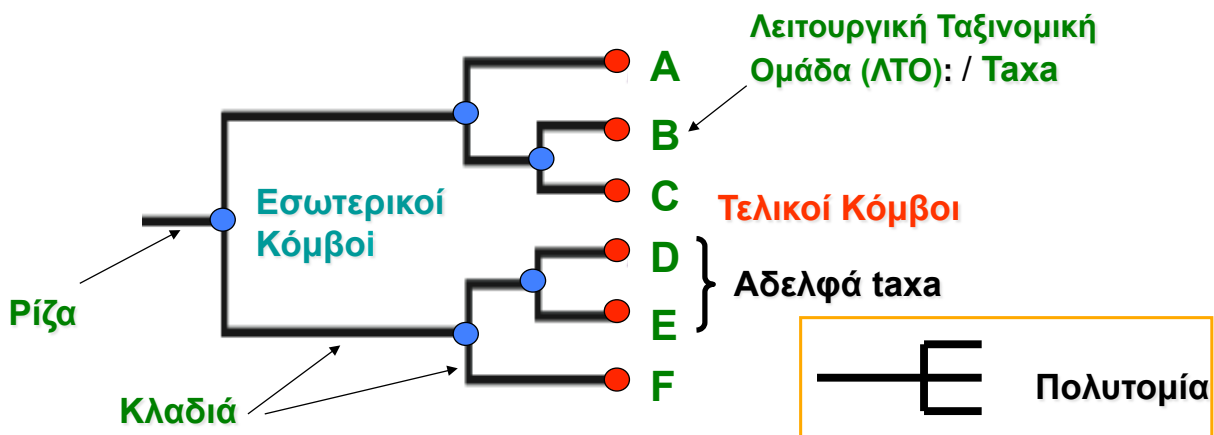
Τοπολογία: το πρότυπο διακλάδωσης του δένδρου

Μήκος κλαδιού: αντιπροσωπεύει τον αριθμό των αλλαγών που συνέβησαν στο κλαδί

Ρίζα: ο κοινός πρόγονος όλων των ομάδων του δένδρου

Κλάδος: μια ομάδα από δύο ή περισσότερες ομάδες ή DNA αλληλουχίες που περιλαμβάνει τον κοινό τους πρόγονο και όλους τους απογόνους τους

Λειτουργική Ταξινομική Ομάδα (ΛΤΟ): ταξινομικό επίπεδο του επιλεγμένου δείγματος που χρησιμοποιείται στη μελέτη, όπως άτομα, πληθυσμοί, είδη, γένη, ή βακτηριακά στελέχη



Λίγη ορολογία...

Προγονικός Χαρακτήρας

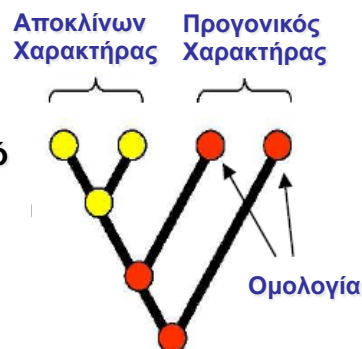
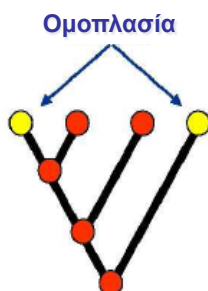
• Ο χαρακτήρας του κοινού προγόνου

Αποκλίνων Χαρακτήρας

• Ένας χαρακτήρας που είναι διαφορετικός από τον πρόγονο

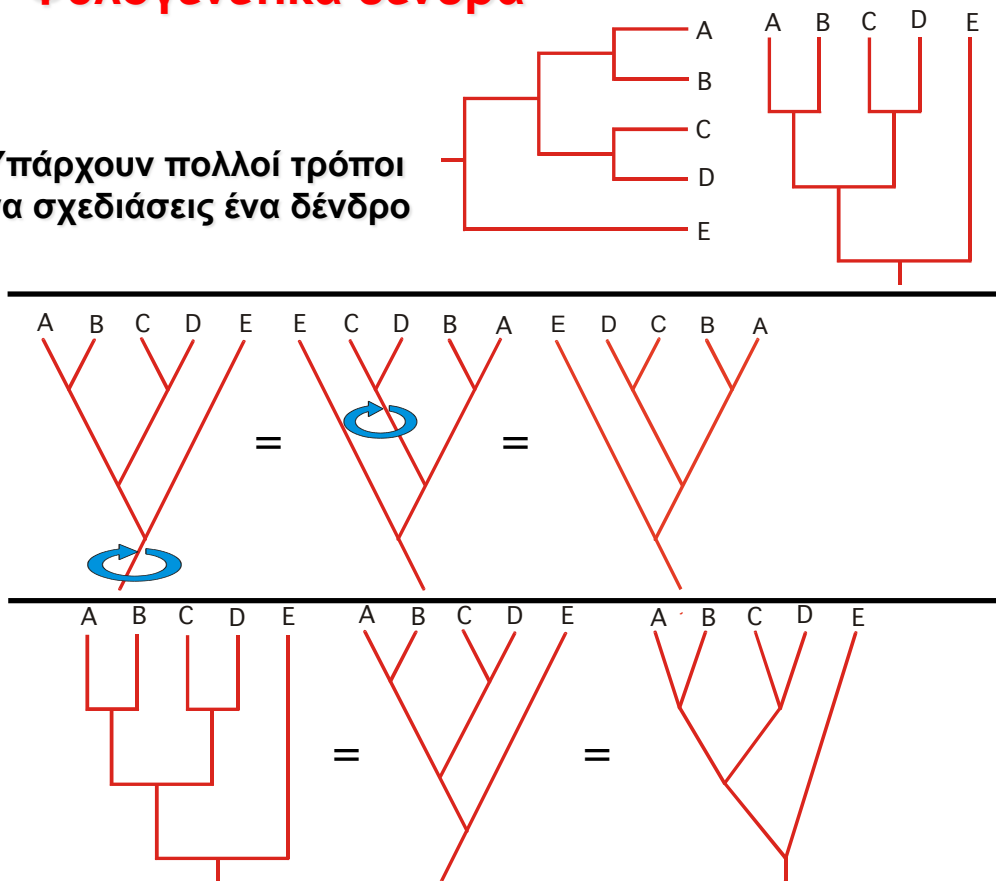
Ομοπλασία

• Ομοιότητα που οφείλεται σε παράλληλη ή συγκλίνουσα εξέλιξη



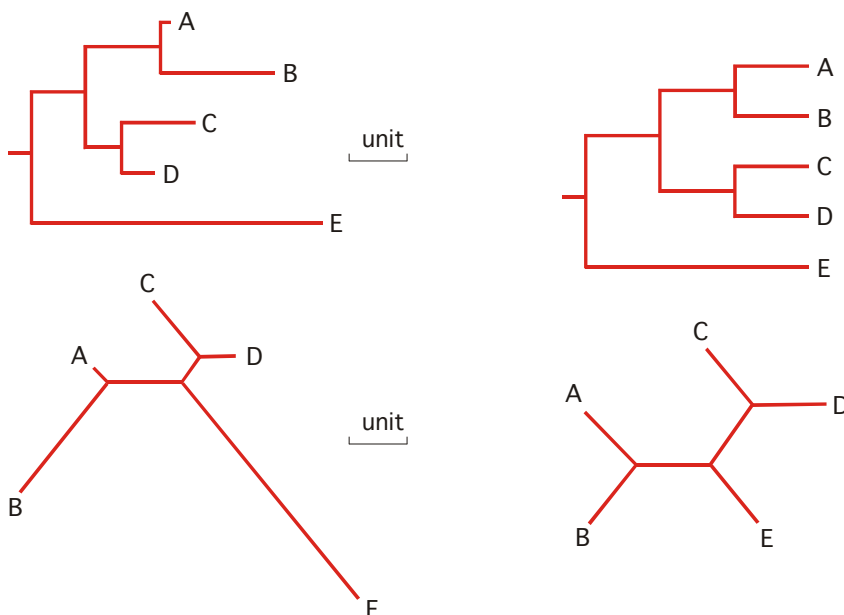
Φυλογενετικά δένδρα

Υπάρχουν πολλοί τρόποι να σχεδιάσεις ένα δένδρο

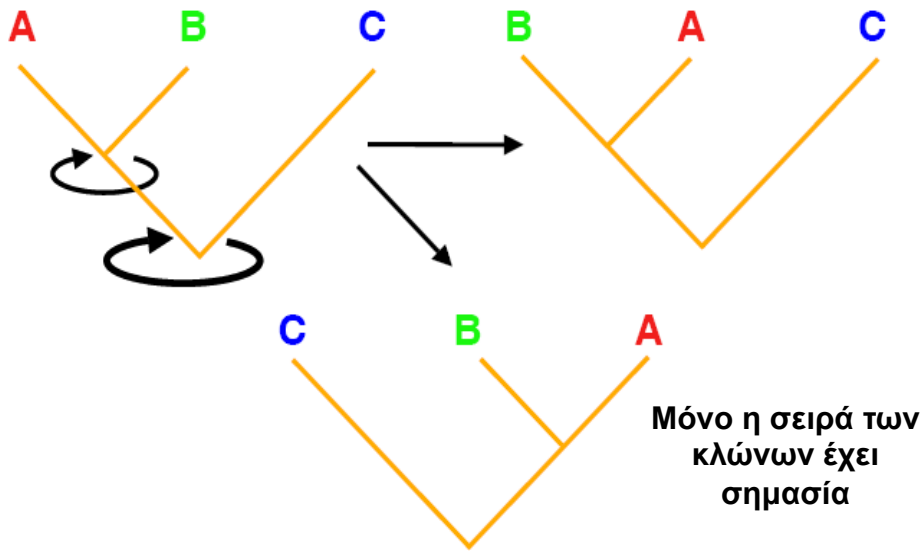


Φυλογενετικά δένδρα

Τα δένδρα μπορούν να είναι σε κλίμακα ή όχι (με ή χωρίς μήκος κλαδιών)



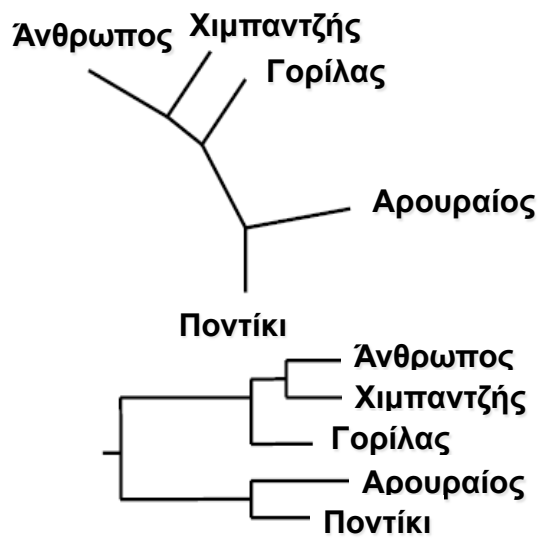
Τοπολογία Δένδρων



Φυλογενετικά δένδρα

Τα δένδρα μπορούν να είναι άρριζα...

... ή με ρίζα



Η πρόσθεση ρίζας στα δένδρα

Evolutionary Ecology 10 p

Rooting trees

The root is the node from which all other nodes descend

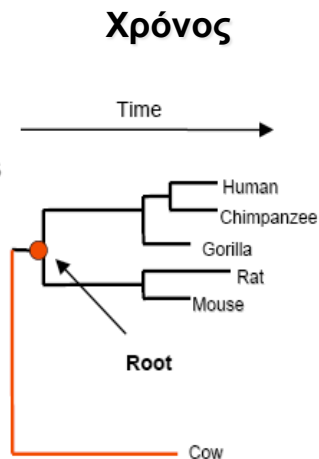
Rooted trees have a direction corresponding to evolutionary time and rooted trees thus allow us to define ancestor-descendent relationships

The place of the root for a tree can be determined using an **outgroup**

–The outgroup should be a species or sequence known to be less closely related to the rest of the species or sequences than they are to each other

–The outgroup should ideally be as closely related as possible to the rest of the species or sequences while still satisfying the condition above

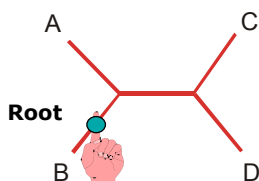
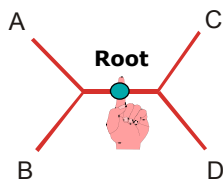
The root must then be placed somewhere between the outgroup and the rest (either on the node or in a branch)



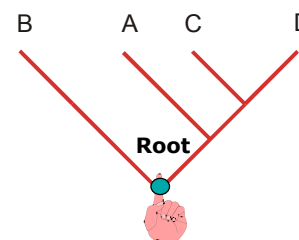
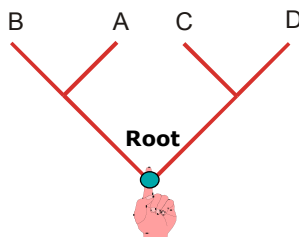
Φυλογενετικά δένδρα

Τα δένδρα μπορούν να είναι άρριζα ή έρριζα

Άρριζα δένδρα

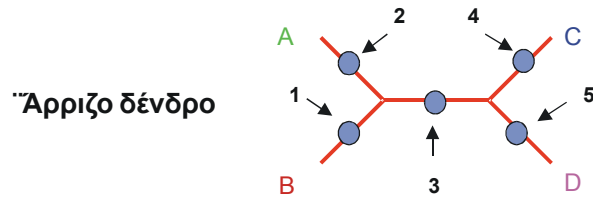


Έρριζα δένδρα

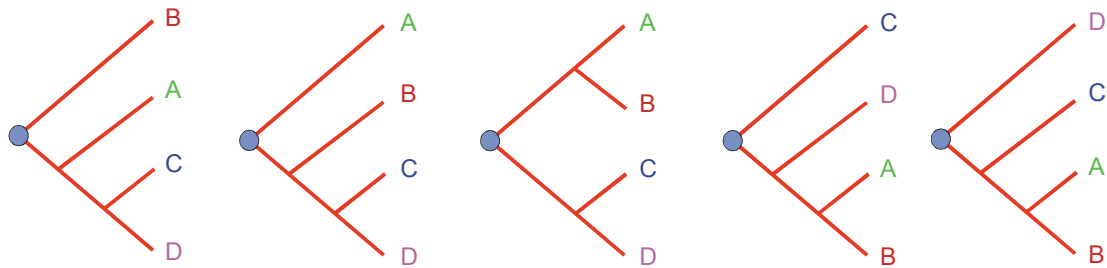


Φυλογενετικά δένδρα

Τα δένδρα μπορούν να είναι άρριζα ή έρριζα



Έρριζο δένδρο 1 Έρριζο δένδρο 2 Έρριζο δένδρο 3 Έρριζο δένδρο 4 Έρριζο δένδρο 5



! " # \$ % & ' () * + , - . / : ; < = > ? @ [\] ^ _ ` { | } ~ ¡ ¢ £ ¤ ¥ ¦ § ¨ © ª « ¬ ® ¯ ° ± ² ³ ´ µ ¶ · ¸ ¹ º » ¼ ½ ¾ ¿ À Á Â Ã Ä Å Æ Ç È É Ê Ë Ì Í Î Ï Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß à á â ã ä å æ ç è é ê ë ì í î ï ð ñ ò ó ô õ ö ÷ ø ù ú û ü ý þ ÿ

Φυλογενετικά δένδρα

Πιθανά φυλογενετικά δένδρα

Ταξινομικές ομάδες (n)	έρριζα $(2n-3)!/(2n-2(n-2)!)$	άρριζα $(2n-5)!/(2n-3(n-3)!)$
2	1	1
3	3	1
4	15	3
5	105	15
6	954	105
7	10,395	954
8	135,135	10,395
9	2,027,025	135,135
10	34,459,425	2,027,025

Πόσα πιθανά δένδρα;

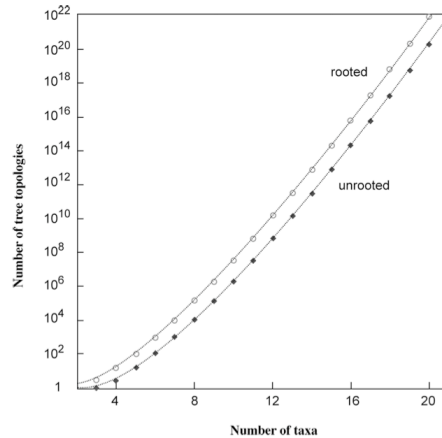
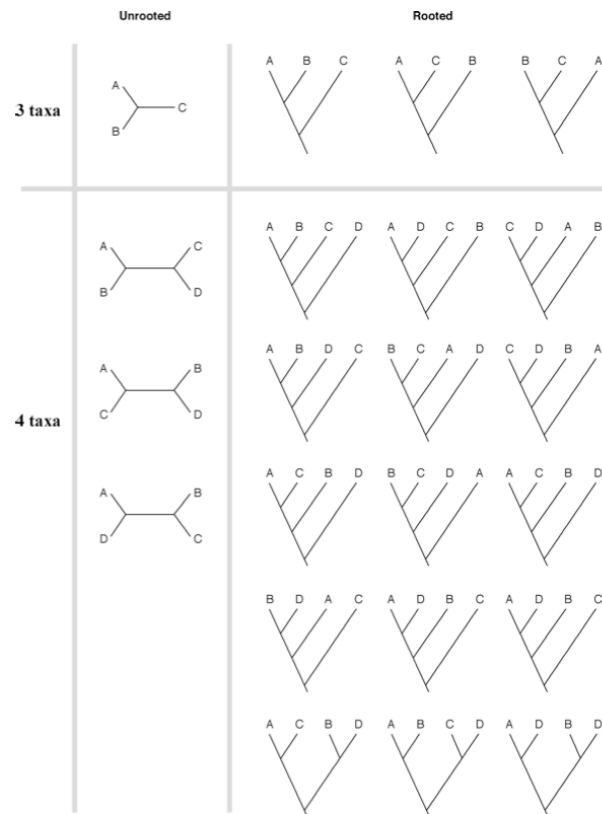
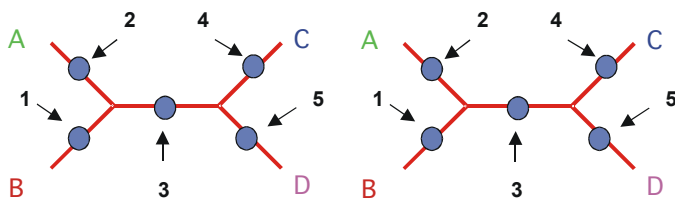


Figure 10.8: Total number of rooted (○) and unrooted (◆) tree topologies as a function of the number of taxa. The values in the y-axis are plotted in the log scale.

Φυλογενετικά δένδρα

Πως βάζουμε μια ρίζα;



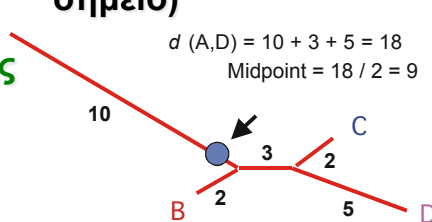
Χρησιμοποιώντας πληροφορίες από τους προγόνους

Στις περισσότερες περιπτώσεις όχι διαθέσιμος

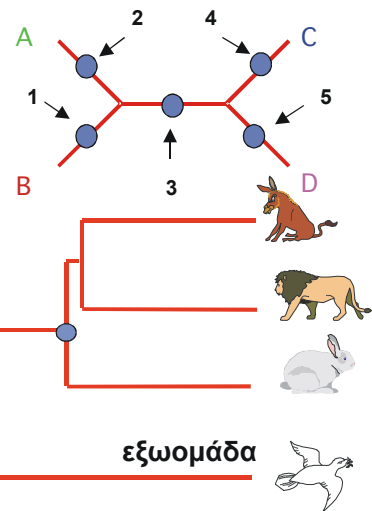
Με χρήση στατιστικών εργαλείων (π.χ. μέσο-σημείο)

$$d(A,D) = 10 + 3 + 5 = 18$$

$$\text{Midpoint} = 18 / 2 = 9$$



Χρειάζονται προϋποθέσεις. ΠΡΟΣΟΧΗ!

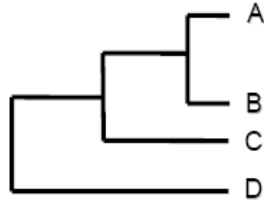


Η εξωμάδα θα πρέπει να είναι ένα είδος ή αλληλουχία λιγότερο συγγενικό με τα υπόλοιπα είδη ή αλληλουχίες από ότι αυτά μεταξύ τους

Η εξωμάδα θα πρέπει να είναι όσο το δυνατόν συγγενέστερη προς τα είδη του δένδρου αλλά και να ικανοποιεί και την προηγούμενη προϋπόθεση

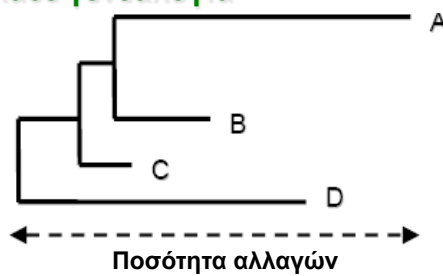
...και ακόμη περισσότερη ορολογία

Στα κλαδογράμματα ή δένδρογράμματα φαίνονται μόνο οι σχέσεις μεταξύ ειδών/αλληλουχιών. Το μήκος των κλαδιών δεν ενδιαφέρει



Κλαδογράμμα

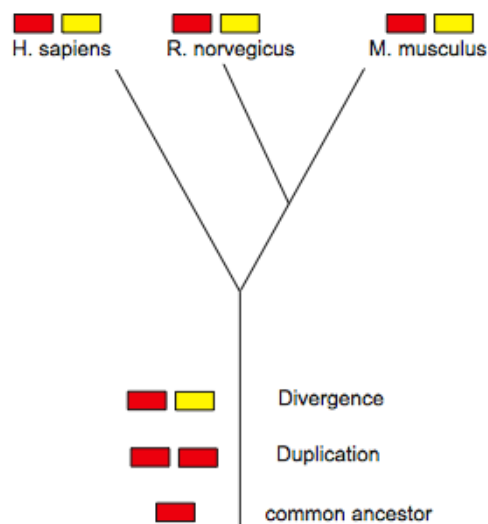
Στα φυλογράμματα και η ομαδοποίηση και τα μήκη των κλαδιών είναι σημαντικά. Τα μήκη είναι ανάλογα με την ποσότητα των αλλαγών για κάθε γενεαλογία



Φυλόγραμμα

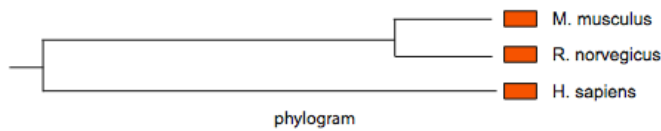
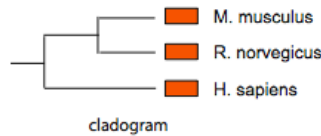
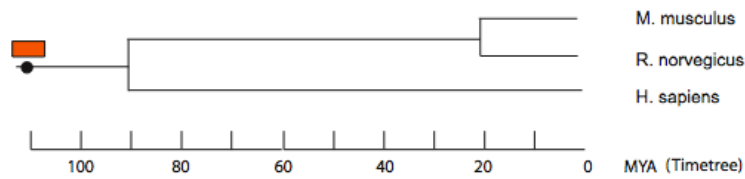
Φυλογένεση

Ορθολογία - Παραλογία



Φυλογένεση οργανισμών

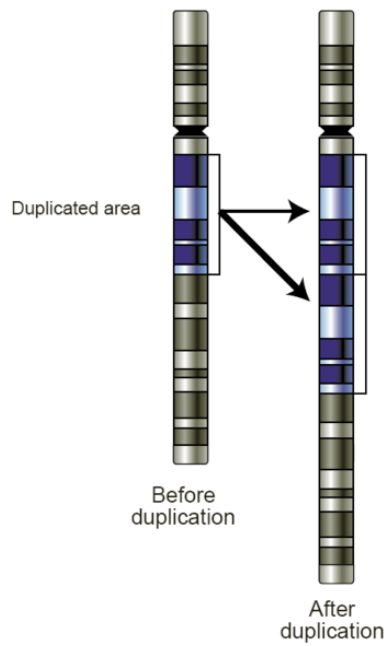
- Επιλέγουμε/βρίσκουμε το ορθόλογο γονίδιο-δείκτη στους οργανισμούς που μελετάμε και ακολουθεί φυλογένεση



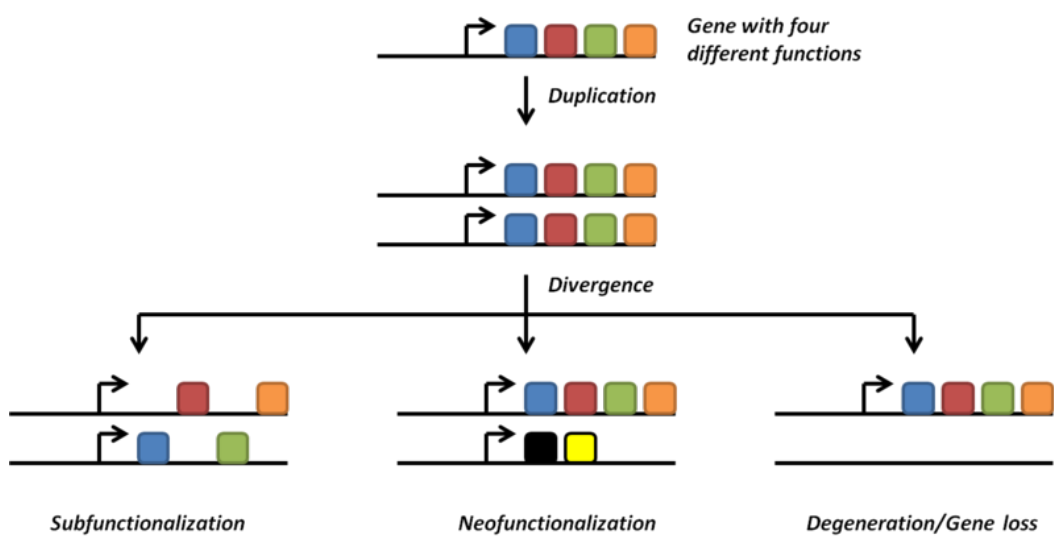
- Το ποντίκι και ο αρουραίος είχαν λιγότερο χρόνο να εξελιχθούν ξεχωριστά, από ότι ο άνθρωπος σε σχέση με το ποντίκι ή σε σχέση με τον αρουραίο. Οι μεταλλάξεις που συσσωρεύτηκαν σε κάθε ορθόλογη ακολουθία πρέπει να είναι ανάλογες του χρόνου απόκλισης των οργανισμών.

Γονιδιακός διπλασιασμός

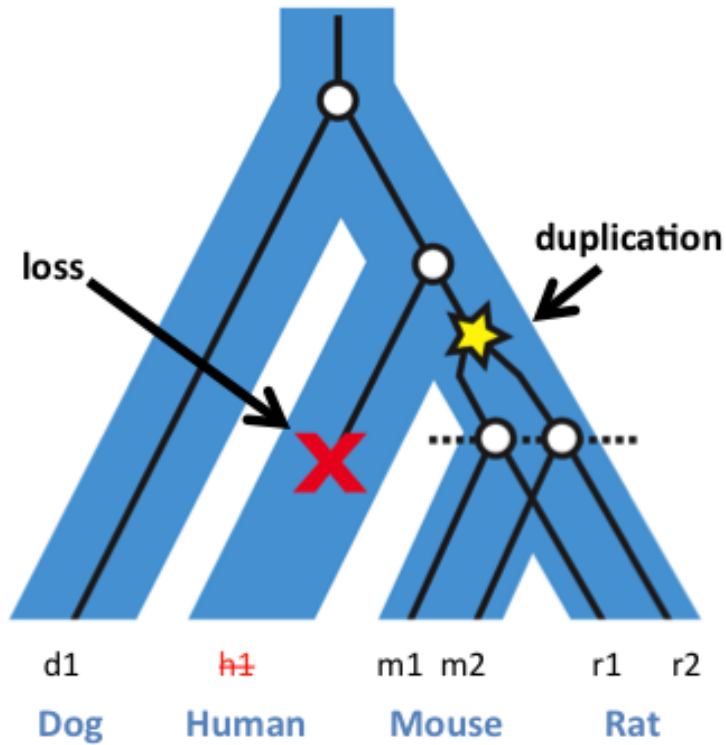
- Περιορισμένος (1 – μερικά γονίδια)
- Εκτεταμένος (block – χρωμόσωμα - πολυποειδία)



Η μοίρα των διπλασιασμένων γονιδίων



Διπλασιασμός και απώλεια

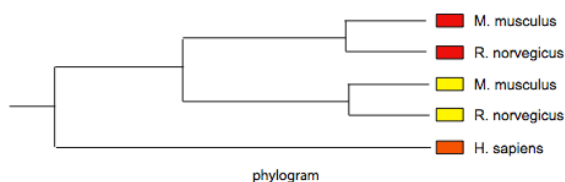
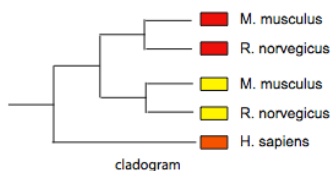
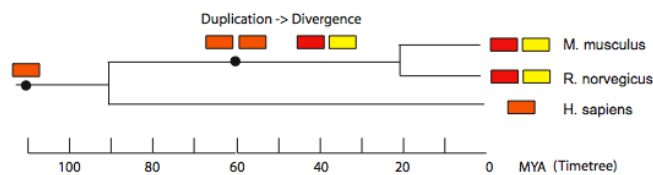


Φυλογένεση

Φυλογένεση γονιδίων

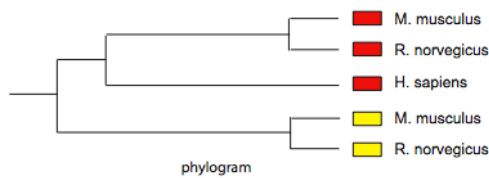
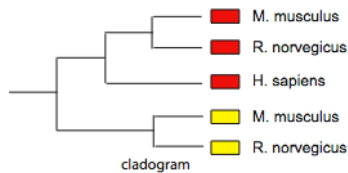
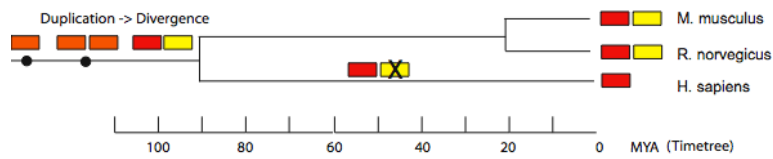
Βρίσκουμε τις ομόλογες ακολουθίες στους οργανισμούς που μας ενδιαφέρουν και ακολουθεί φυλογένεση, για να καταλάβουμε πότε συνέβησαν οι γονιδιακοί διπλασιασμοί, και ποιά ομόλογα είναι πιο κοντινά μεταξύ τους.

Πρέπει να γνωρίζουμε τις εξελικτικές σχέσεις των οργανισμών



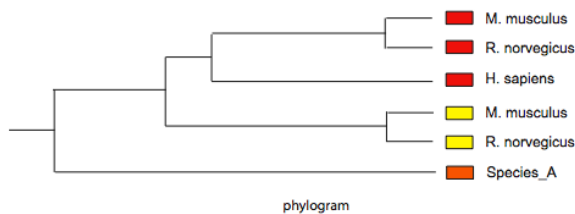
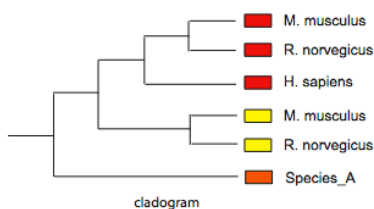
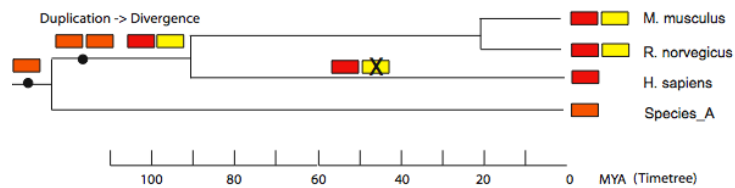
Φυλογένεση γονιδίων

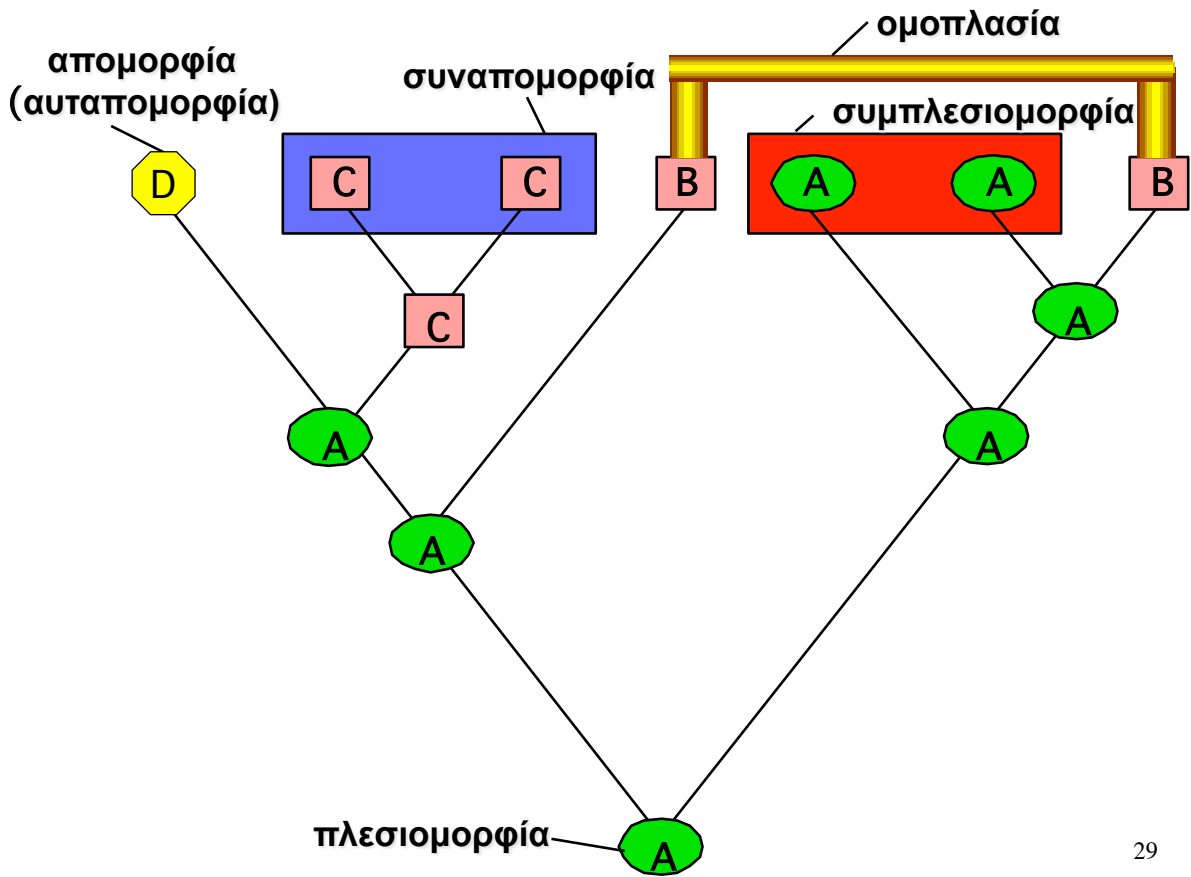
- Απώλεια αντίγραφου



Φυλογένεση γονιδίων

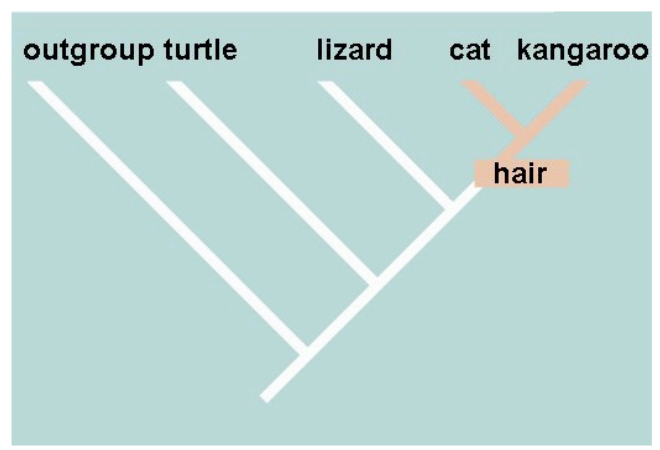
Δειγματοληψία ορθόλογων από πιο απομακρυσμένους οργανισμούς, μέχρι να εντοπιστεί ο χρόνος που συνέβη ο διπλασιασμός. Απολιθώματα βοηθούν στην χρονολόγηση





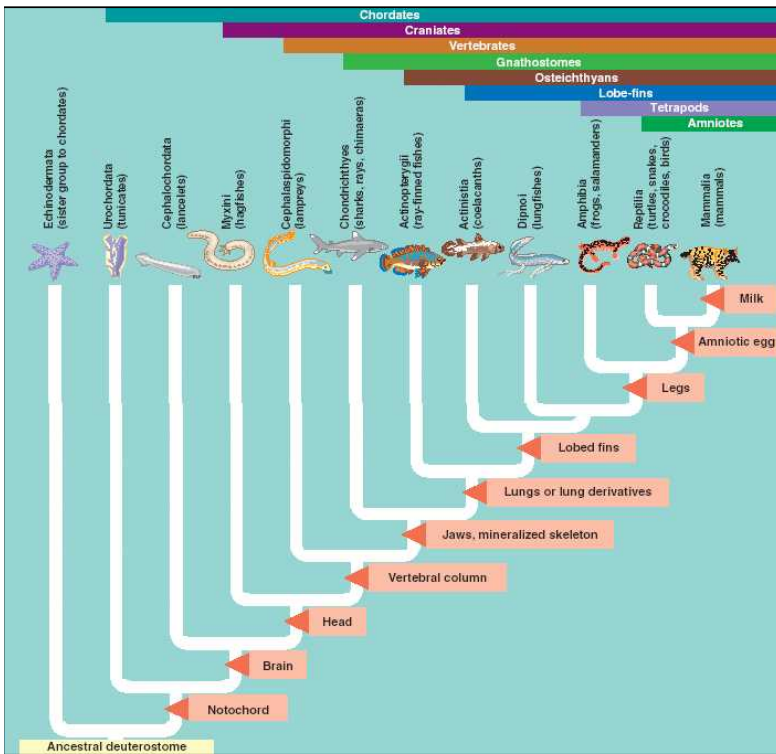
συναπομορφία

1- / (92 % - 14 x 7 (, % % 1 #) % 2



Within this taxon set, hair is a **shared derived character (synapomorphy)** of therians that indicates a close relationship of marsupial and placental mammals.

συναπομορφία



Κάθε χαρακτήρας με ροζ είναι μια **συναπομορφία**

Κοινός – για όλους τους απογόνους στον κλάδο

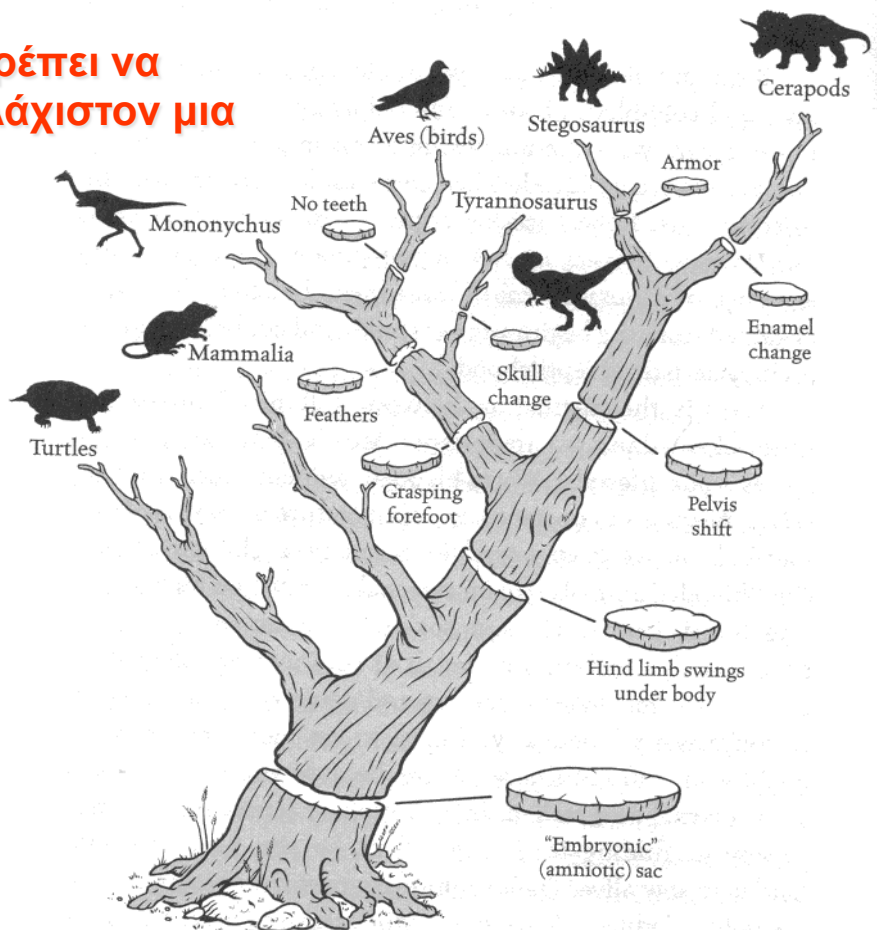
π.χ. όλα τα χορδωτά μοιράζονται μια νοτοχορδή

Αποκλίνων – όχι παρών σε προγονικές ομάδες

π.χ. τα προγονικά δευτεροστόμια δεν έχουν νοτοχορδή

;\$<* 14\$&- 2.)'.*/ (%
μ- /) \$=*#/% # "4\$, /5# (μ/%
5" (% - μ-) 0 #%

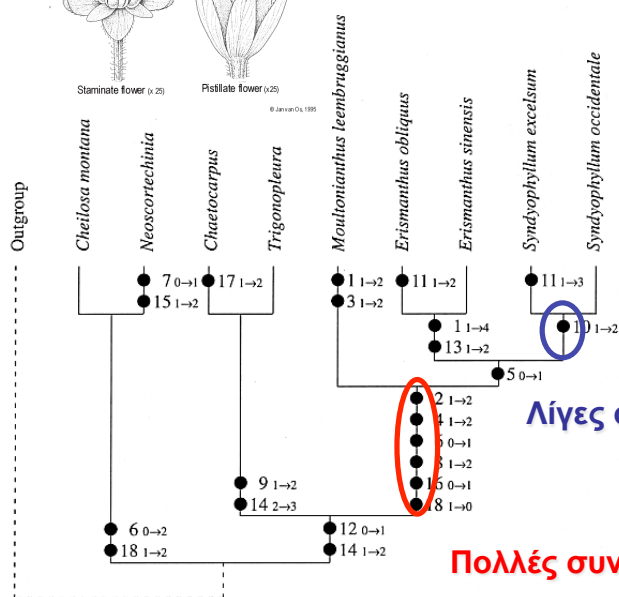
Κάθε κλάδος πρέπει να μοιράζεται τουλάχιστον μια συναπομορφία



συναπομορφία

Πως μπορούμε να καταλάβουμε πότε ένας κλάδος υποστηρίζεται;

Εν μέρει από τους αριθμούς των συναπομορφιών



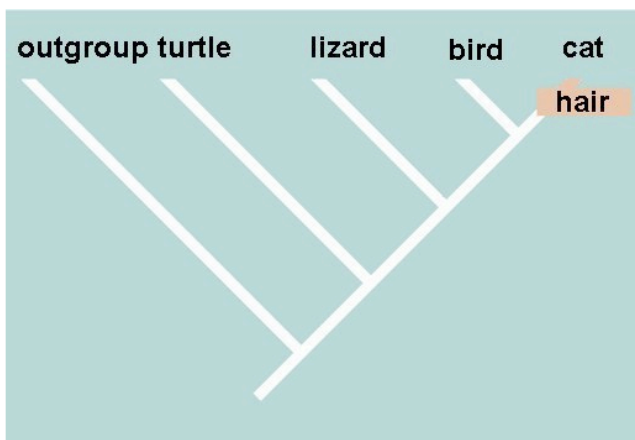
Λίγες συναπομορφίες = ασθενής υποστήριξη

Πολλές συναπομορφίες = ισχυρή υποστήριξη

Fig. 1. The phylogeny of the tribe Erismantheae (genera *Erismanthus*, *Moultonianthus*, and *Syndyphyllum*) in relation to the tribes Chaetocarpeae (*Chaetocarpus* and *Trigonoptera*) and Cheiloseae (*Cheilosia* and *Neoscortechinia*). Delimitation of tribes after Webster (1994).

αυταπομορφία

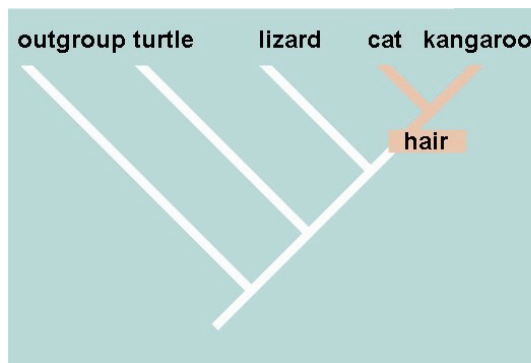
μ- (%&/192 % - 14χ 7 (, % %1#) %2



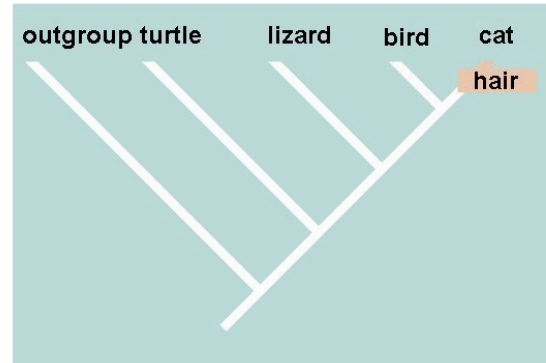
Within this taxon set, hair is a unique character (autapomorphy) of mammals

Συναπομορφία - αυταπομορφία

Ο όρος χρησιμοποιείται σε σχέση με την κλίμακα των taxa



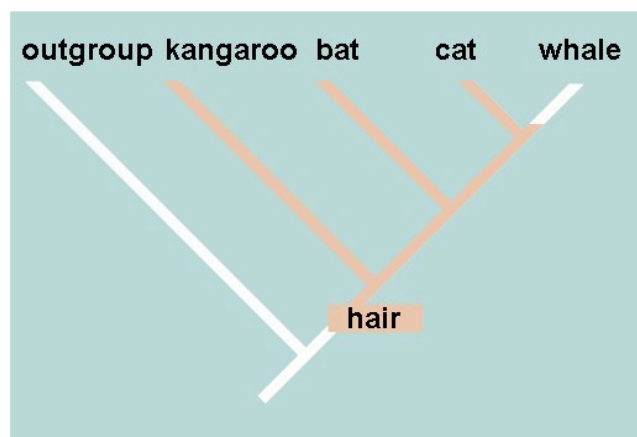
Within this taxon set, hair is a **shared derived character (synapomorphy)** of therians that indicates a close relationship of marsupial and placental mammals.



Within this taxon set, hair is a **unique character (autapomorphy)** of mammals

συνπλεσιομορφία

1- / (92 .) - > - (/ 192 , % 1 #) % 2



Within this taxon set, hair is a **shared ancestral character (symplesiomorphy)** of non-cetaceans. It would *not* indicate a close relationship of marsupial and placental mammals.

συναπομορφία

1-/(92 % - 14χ7 (, % %1#) %2

αυταπομορφία

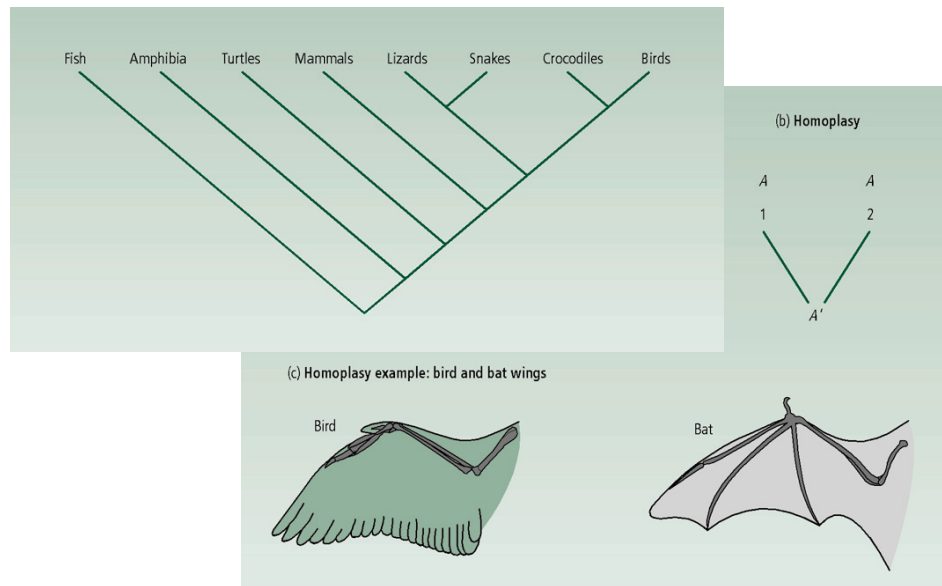
μ- (%&/192 % - 14χ7 (, % %1#) %2

συνπλεσιομορφία

1-/(92 .) - >- (/192 , % %1#) %2

Ομοπλασία

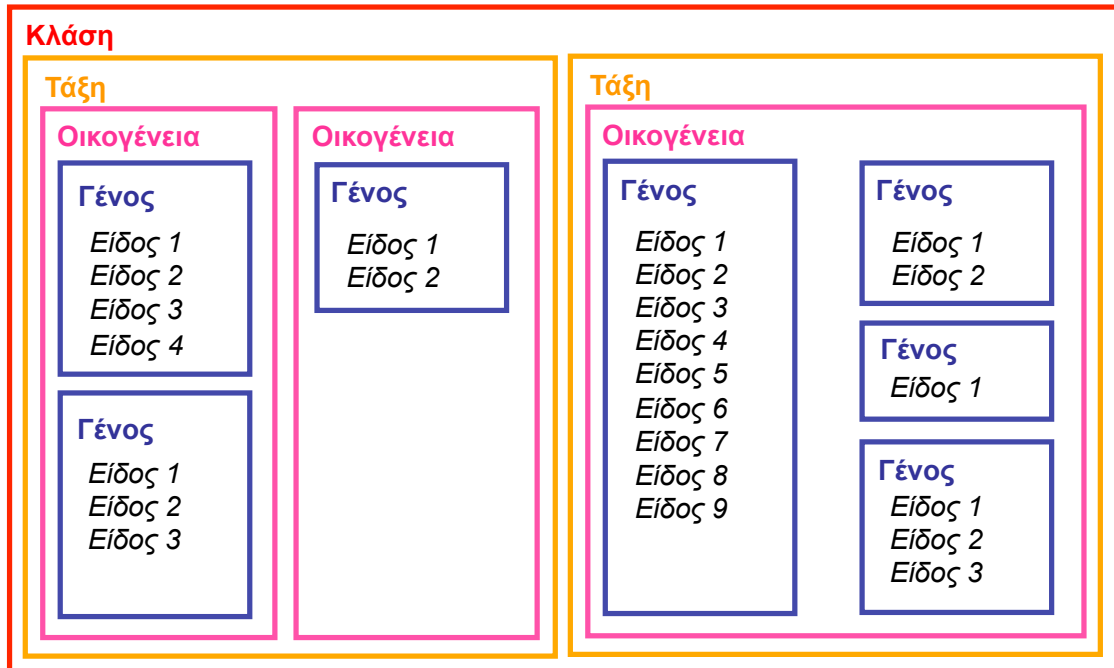
Κοινά χαρακτηριστικά που μοιράζονται κάποια taxa, τα οποία όμως δεν υπήρχαν στον κοντινότερο κοινό τους πρόγονο. Δηλαδή, οι συγκεκριμένοι κοινοί χαρακτήρες εμφανίστηκαν ανεξάρτητα στο κάθε taxon



Φυλογένεση και συστηματική κατάταξη

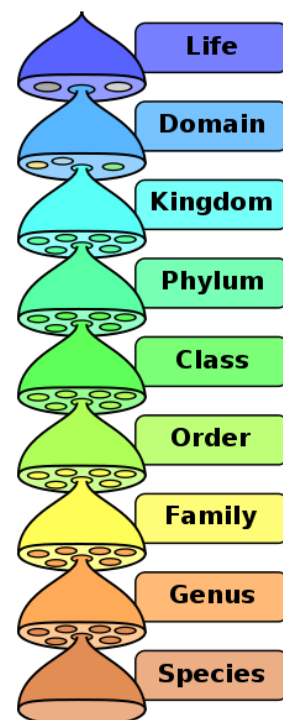
Ιεραρχία

Όλες οι ταξινομικές κατατάξεις είναι ιεραρχικές – πώς διαφέρει η φυλογένεση;

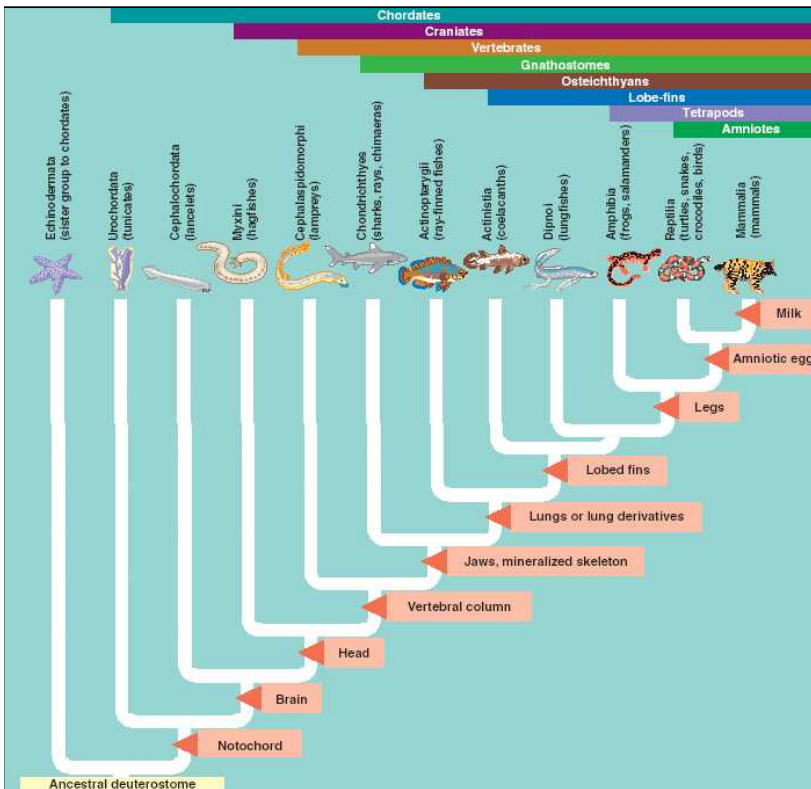


Ταξινόμηση οργανισμών

- Ιεραρχική κατηγοριοποίηση/ομαδοποίηση οργανισμών.
- Linnaeus (1707-1778) ομαδοποίησε οργανισμούς με βάση κοινούς χαρακτήρες.
- Αργότερα, η ταξινόμηση προσαρμόστηκε στην εξελικτική θεωρία του Δαρβίνου, ώστε να ομαδοποιούνται οι οργανισμοί με βάση την κοινή τους προέλευση.



Φυλογένεση και συστηματική κατάταξη



Κατάταξη

Σημειώστε ότι οι ομάδες καθορίζονται βάσει των κοινών προγόνων

π.χ. Όλα τα τετράποδα μοιράζονται έναν κοινό πρόγονο με πόδια, αλλά άλλα χορδοτά δεν μοιράζονται αυτόν τον κοινό πρόγονο

Οι χαρακτήρες που χρησιμοποιούνται στη φυλογένεση είναι **συναπομορφίες**

Φυλογένεση και συστηματική κατάταξη

Μονοφυλετική ομάδα

Περιλαμβάνει έναν πρόγονο και όλους τους απογόνους του

Παραφυλετική ομάδα

Περιλαμβάνει τον πρόγονο, αλλά όχι όλους τους απογόνους του

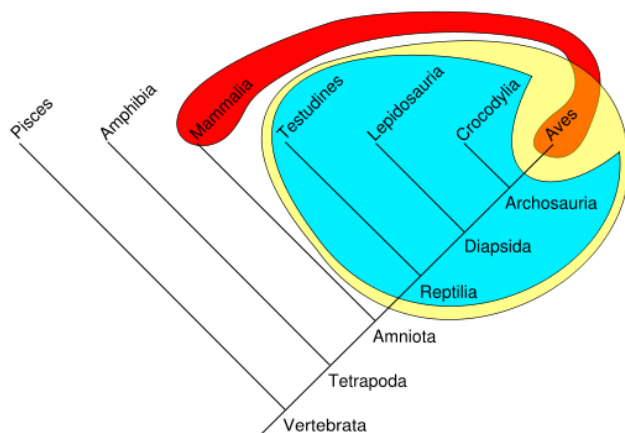
Πολυφυλετική ομάδα

Περιλαμβάνει δύο συγκλίνοντες απογόνους αλλά όχι τον κοινό τους πρόγονο

Μόνο μονοφυλετικές ομάδες αναγνωρίζονται στην κλαδιστική κατάταξη



Τα ομοιόθερμα ζώα, θηλαστικά και πτηνά αποτελούν μία πολυφυλετική ομάδα



Φυλογένεση και συστηματική κατάταξη

Μονοφυλία

Κάθε μία από τις έγχρωμες γενεαλογίες στη φυλογένεση αυτή των εχινόδερμων είναι μια καλή μονοφυλετική ομάδα

Asteroidea

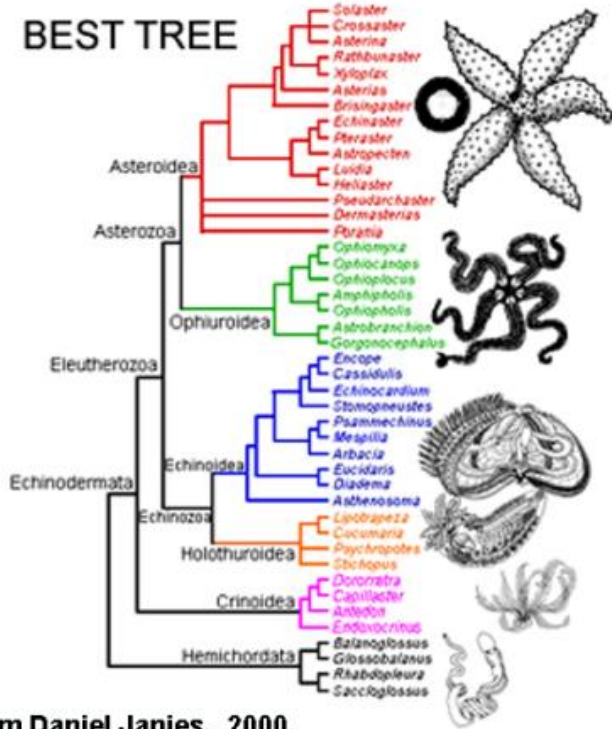
Ophiuroidea

Echinoidea

Holothuroidea

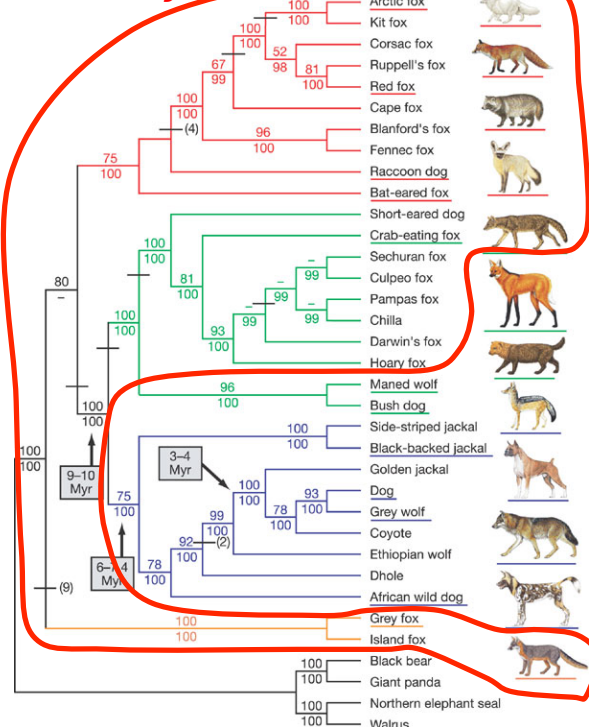
Crinoidea

Κάθε ομάδα μοιράζεται έναν κοινό πρόγονο που δεν είναι κοινός για άλλα μέλη μιας άλλης ομάδας



Φυλογένεση και συστηματική κατάταξη

Αλεπούδες



Lindblad-Toh et al. (2005) *Nature* 438: 803-819

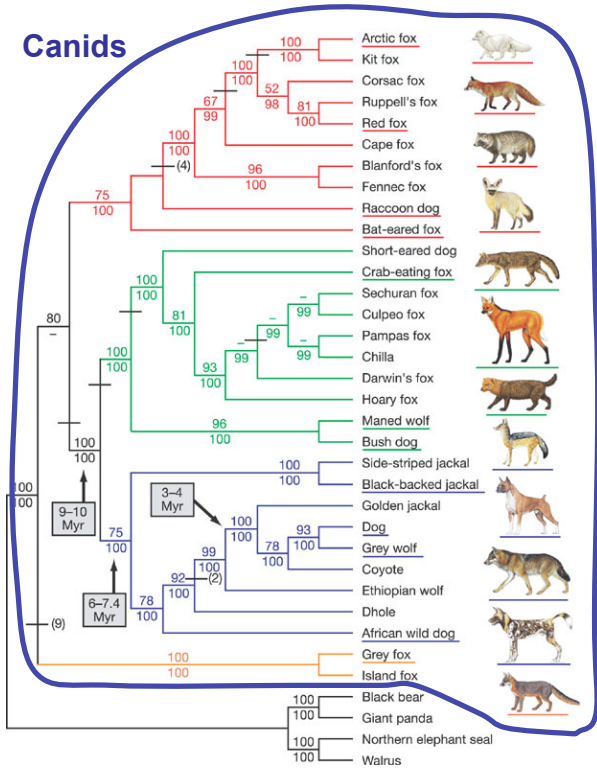
Παραφυλετικές ομάδες

Παραφυλία

Οι «αλεπούδες» είναι παραφυλετικές σε σχέση με τους σκύλους, λύκους, τσακάλια, coyotes, κλπ

Αυτό είναι ένα κλασικό παράδειγμα επειδή «αλεπούδες» και «σκύλοι» δεν αποτελούν τυπικές ταξινομικές ομάδες, αλλά δείχνει ότι ένας σκύλος ή ένας λύκος είναι απλά μια αποκλίνουσα αλεπού με τη φυλογενετική έννοια

Φυλογένεση και συστηματική κατάταξη



Lindblad-Toh et al. (2005) *Nature* 438: 803-819

Παραφυλετικές ομάδες

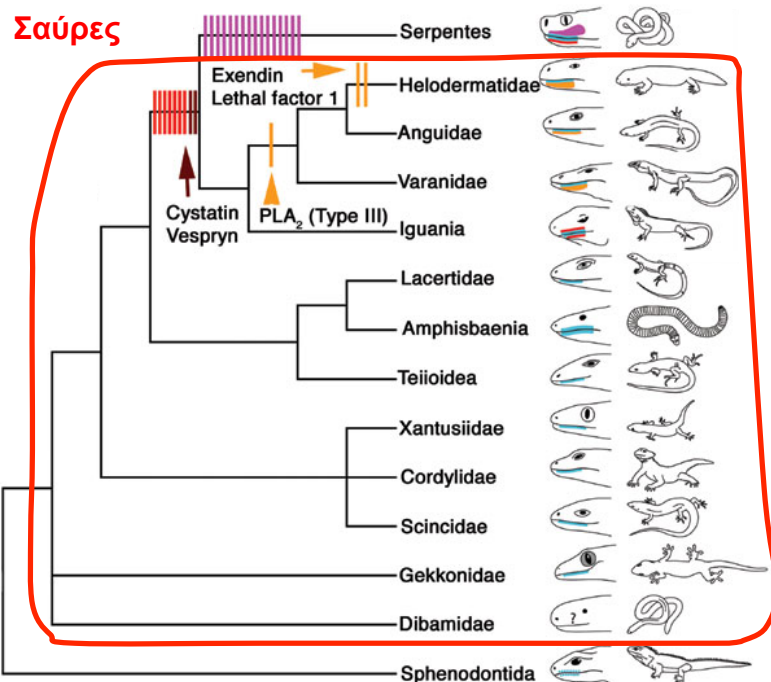
Μονοφυλία

Να σημειωθεί ότι τα canids παραμένουν μια καλή **μονοφυλετική** ομάδα μέσα στα Mammalia (Θηλαστικά)

Κάθε έγχρωμη γενεαλογία μέσα στα canids είναι επίσης **μονοφυλετικός** κλάδος

Φυλογένεση και συστηματική κατάταξη

Παραφυλετικές ομάδες



Fry et al. (2006) *Nature* 439: 584-588

Παραφυλία

Οι «Σαύρες» (Sauria) είναι **παραφυλετικές** σε σχέση με τα «Φίδια» (Serpentes)

Serpentes είναι ένας **μονοφυλετικός** κλάδος μέσα στις Σαύρες

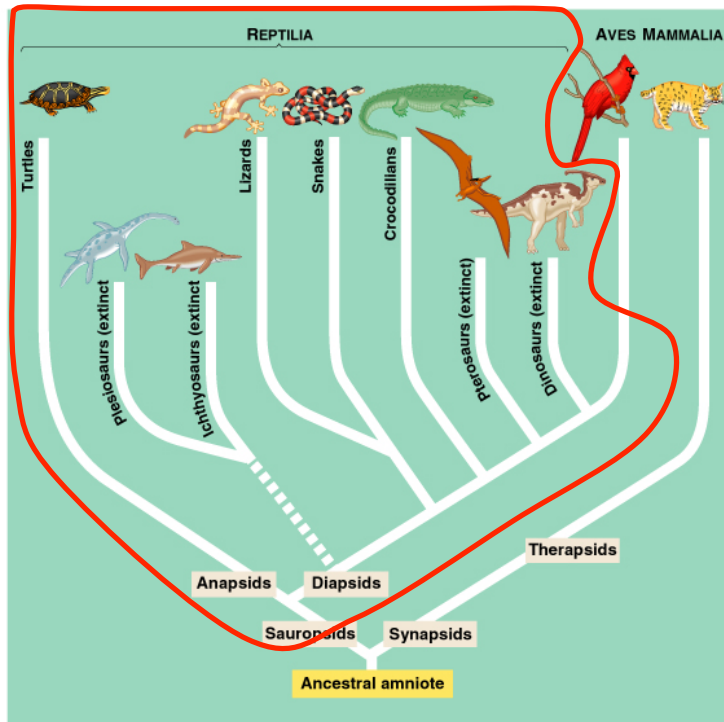
Squamata (σαύρες + φίδια) είναι ένας **μονοφυλετικός** αδελφός κλάδος με τα sphenodontida

??%0 +&/%* ψ %%% 48 2
% - 14ψ - "5*2 «, 7) +2
\$1) %ω 5%6) *2

Φυλογένεση και συστηματική κατάταξη

Παραφυλετικές ομάδες

Reptilia



© 1999 Addison Wesley Longman, Inc.

Παραφυλία

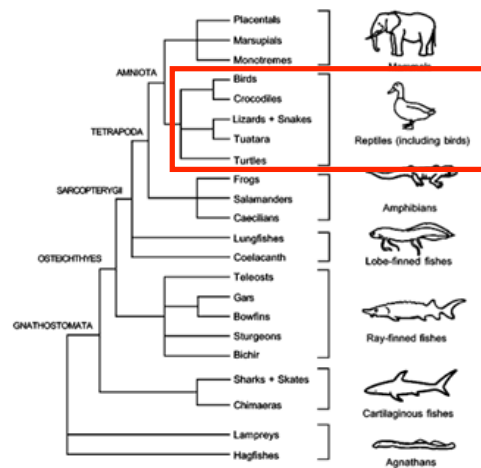
Τα πουλιά συγγενεύουν πολύ περισσότερο με τα κροκοδειλοειδή απ' ό τι με άλλα υπάρχοντα σπονδυλωτά

Archosauria = Πουλιά + Κροκόδειλοι

Σκεφτόμαστε τα ερπετά ως χελώνες, σαύρες, φίδια και κροκοδείλους

Αλλά τα Reptilia είναι μια παραφυλετική ομάδα εκτός και αν συμπεριληφθούν και τα Aves

Τι σημαίνει αυτό;



Αυτό που σημαίνει είναι ότι τα «ερπετά» είναι έγκυρος κλάδος μόνο εάν συμπεριλάβει και τα πουλιά

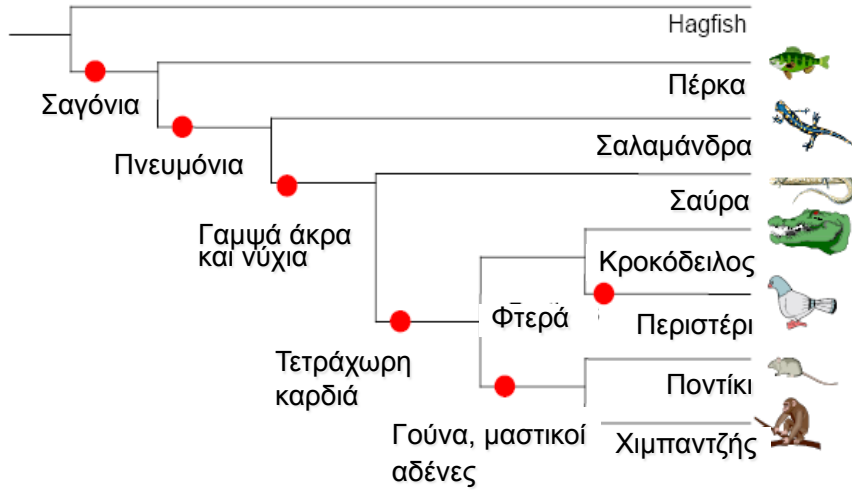
Τα πουλιά είναι ακόμη πουλιά αλλά τα Aves δεν μπορούν θεωρηθούν ως «Κλάση» ισοδύναμη με αυτή των Reptilia επειδή εξελικτικά είναι εμβόλιμα μέσα στα Reptilia

Ερπετά Aves (πουλιά)	Χελώνες
	Κροκόδειλοι
	Σαύρες και φίδια
	Tuataras

Έλεγχος εξελικτικών υποθέσεων

Χαρτογράφηση εξελικτικών μεταβάσεων

Τα δένδρα μπορούν να χρησιμοποιηθούν για τη μελέτη της εξέλιξης των χαρακτήρων



Παλαιά γεγονότα Σχετικός εξελικτικός χρόνος Πρόσφατα γεγονότα

Έλεγχος εξελικτικών υποθέσεων

Χαρτογράφηση εξελικτικών μεταβάσεων

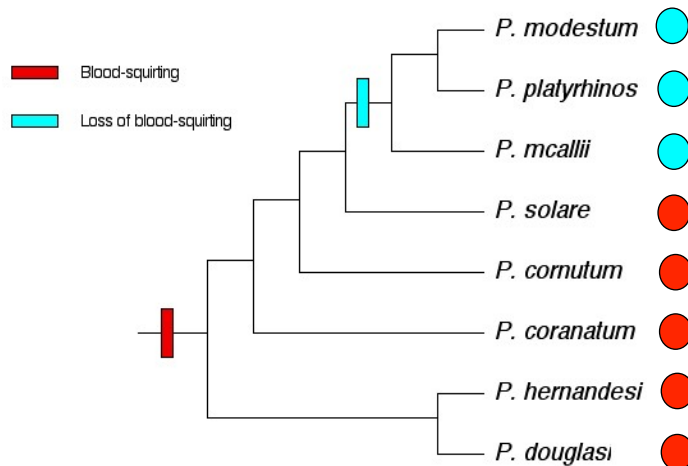
Μερικές κερασφόρες σαύρες εκτοξεύουν αίμα από τα μάτια όταν δεχτούν επίθεση από «σκύλους»

Πόσες φορές η εκτόξευση αίματος εξελίχθηκε;

! "# @ 0 " 4 - > ' (* 5 @ .) - # * % * / μ % * 3 * 4 / 1 # 1 : * μ 0 \$ (/ 5 @ 1 % μ / % * 3 * 4 / 1 # 1 : % 8 4 * / % # @ * 1 # 9 3 * " 5 @ % μ / % # 2



Εκτόξευση αίματος; ● Όχι ● Ναι

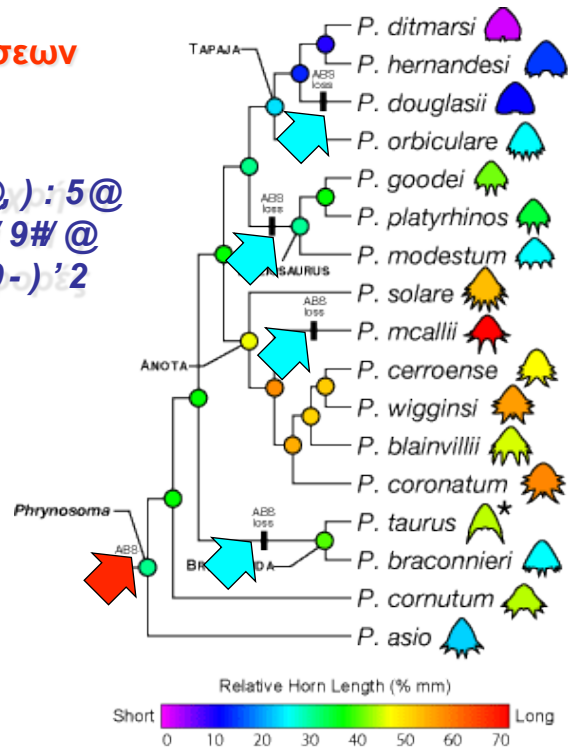


Έλεγχος εξελικτικών υποθέσεων

Χαρτογράφηση εξελικτικών μεταβάσεων

A5#95-, μ/%('%0"4->'(*5@μ*#@,):5@
 .-44%48(,%%#)7(,.)-#*#/9#@
 1#93"5@%μ%#-2,\$<@*.-44'20-)'2
 1%# \$#@*3'4/3@#@-μ\$&%2

Η ερμηνεία αυτών των εξελικτικών
 σεναρίων εξαρτάται από τη
 φυλογένεση

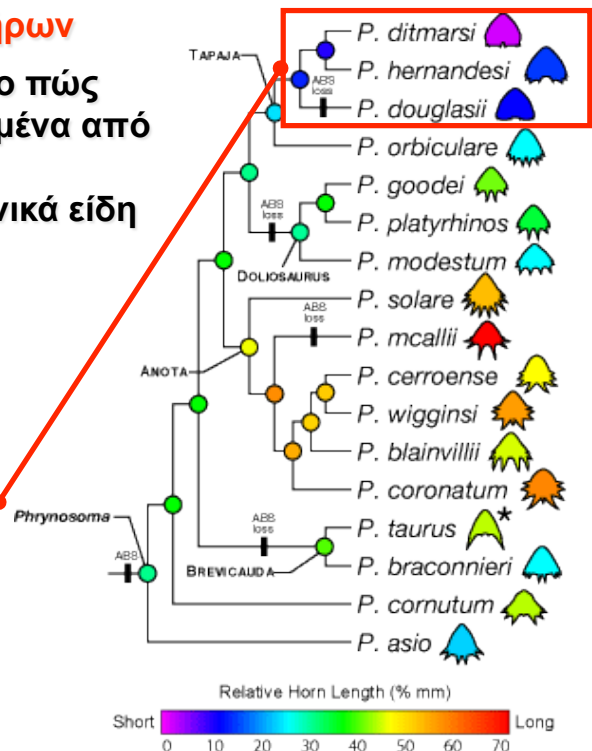
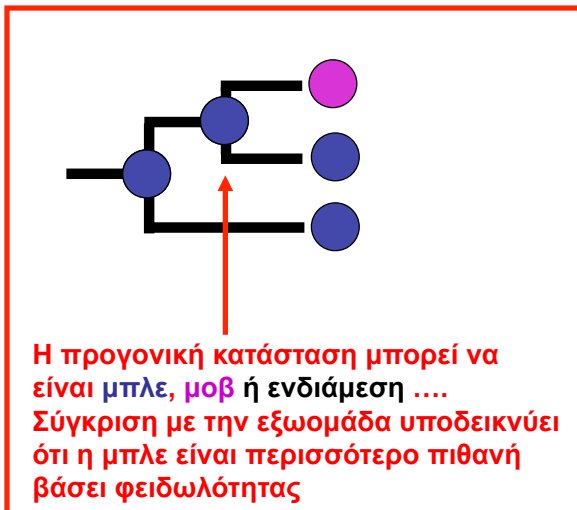


Leaché and McGuire. Molecular Phylogenetics and Evolution 39: 628-644

Έλεγχος εξελικτικών υποθέσεων

Ανακατασκευή προγονικών χαρακτήρων

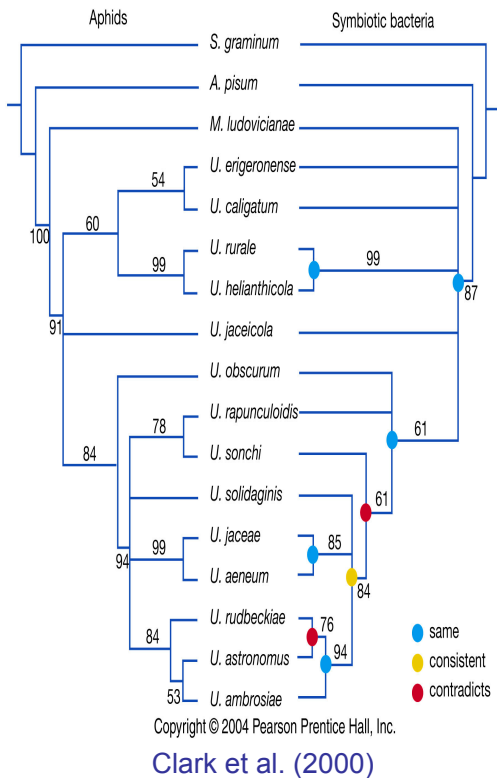
Αυτή η φυλογένεση δείχνει επίσης το πώς
 μπορούν να χρησιμοποιηθούν δεδομένα από
 ζωντανά είδη για να συμπεράνουμε
 καταστάσεις χαρακτήρων σε προγονικά είδη



Leaché and McGuire. Molecular Phylogenetics and Evolution 39: 628-644

Έλεγχος εξελικτικών υποθέσεων - συνεξέλιξη

(c)



Συν-εξέλιξη

Οι αφίδες και ορισμένα βακτήρια είναι συμβιοτικά

Δεδομένης της στενής σχέσης, θα περιμέναμε ότι η ειδογένεση στις αφίδες θα προκαλούσε παράλληλη ειδογένεση στα βακτήρια

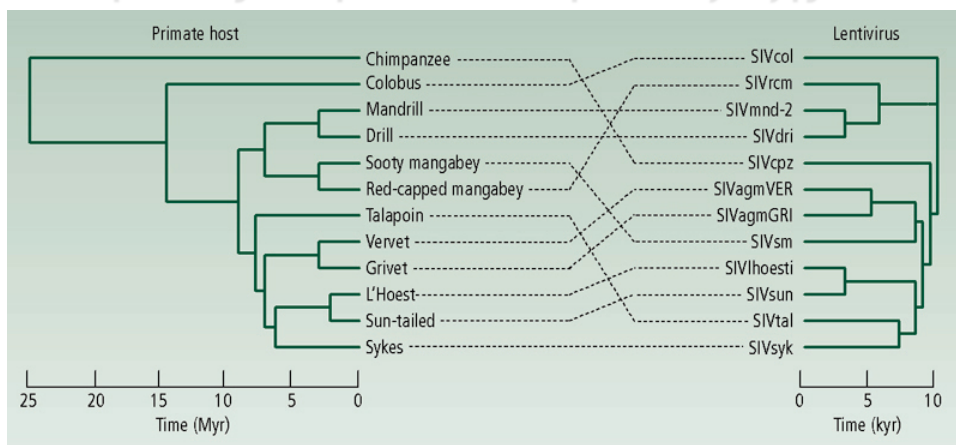
Όταν συγκρίνονται φυλογενέσεις από κάθε είδος υπάρχουν ενδείξεις για αμοιβαία κλαδογένεση (αλλά επίσης και αντιφάσεις)

Έλεγχος εξελικτικών υποθέσεων - Συνεξέλιξη

Πρωτεύοντα και lentiviruses (D.L Robertson)

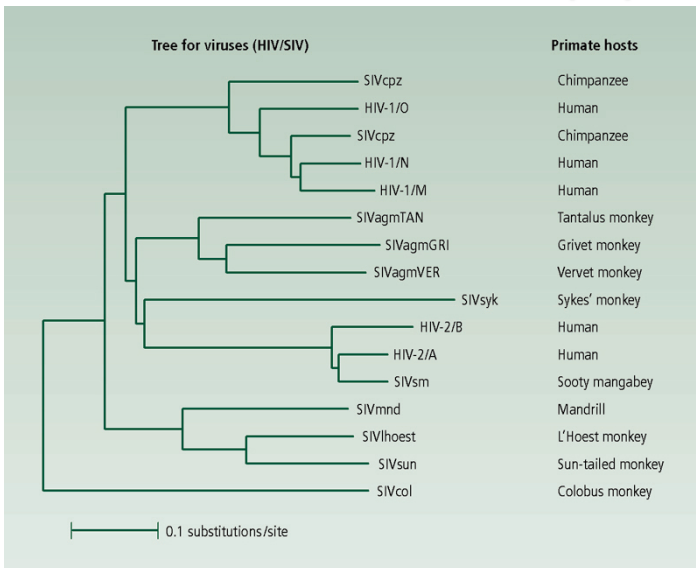
Τα δένδρα (επιφανειακά) δείχνουν συνεξέλιξη σε ικανοποιητικό βαθμό.

Ωστόσο, το μοριακό ρολόι (αν δεν είναι εντελώς λάθος!) δείχνει ότι οι ειδογένεσεις δεν ήταν αποτέλεσμα συνεξέλιξης



Έλεγχος εξελικτικών υποθέσεων

Από που προήλθε ο ιός HIV;



Πρωτοεμφανίστηκε μυστηριωδώς στις αρχές της δεκαετίας του 1980.

Ο τύπος HIV-1 εισήλθε στους ανθρώπους, ίσως περισσότερες από μια φορές, από τον χιμπατζή.

Ο τύπος HIV-2 εισήλθε στους ανθρώπους, από τους sooty mangabees



Έλεγχος εξελικτικών υποθέσεων

Από που προήλθε ο ιός HIV;

- HIV-1 εισήλθε στους ανθρώπους από τον χιμπατζή περίπου το 1908.
- HIV-2 εισήλθε στους ανθρώπους από τα sooty mangabees περίπου το 1933.
- Παρέμεινε σε χαμηλά επίπεδα στον ανθρώπινο πληθυσμό μέχρι το 1950.
- Η αστικοποίηση ίσως να ευθύνεται για την εξάπλωσή του.
- Ο SIV δεν είναι τόσο επικίνδυνος για τις μαϊμούδες όσο ο HIV για τους ανθρώπους.
- Αρχικά πίστευαν ότι ο SIV υπήρχε στις μαϊμούδες για πολλά χρόνια και έτσι υπήρχε χρόνος να προσαρμοστεί το ανοσοποιητικό σύστημα.
- Φυλογενετικές αναλύσεις έδειξαν ότι ο SIV εισήλθε στους χιμπατζήδες το 1492 και στα sooty mangabees το 1808!
- Άρα, δεν ισχύει η παραπάνω θεωρία γιατί οι μαϊμούδες δεν αρρωσταίνουν τόσο έντονα.

Έλεγχος εξελικτικών υποθέσεων

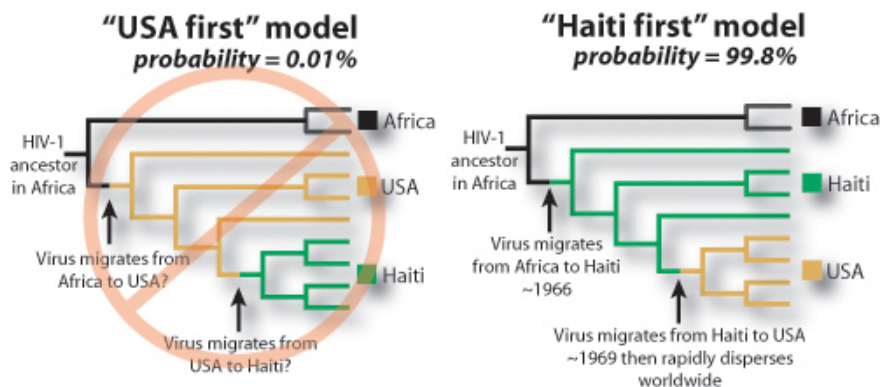
Από που προήλθε ο ιός HIV-1 subtype M; Προέλευση στην Κεντρική Αφρική.

Όταν πρωτοεντοπίστηκε, αρκετοί ασθενείς στην Αμερική ήταν πρόσφατοι Αϊτινοί μετανάστες.

Κάποιοι ισχυρίζονταν ότι πήγε από την Αμερική στην Αϊτή στα μέσα των 70s, λόγω σεξοτουρισμού.

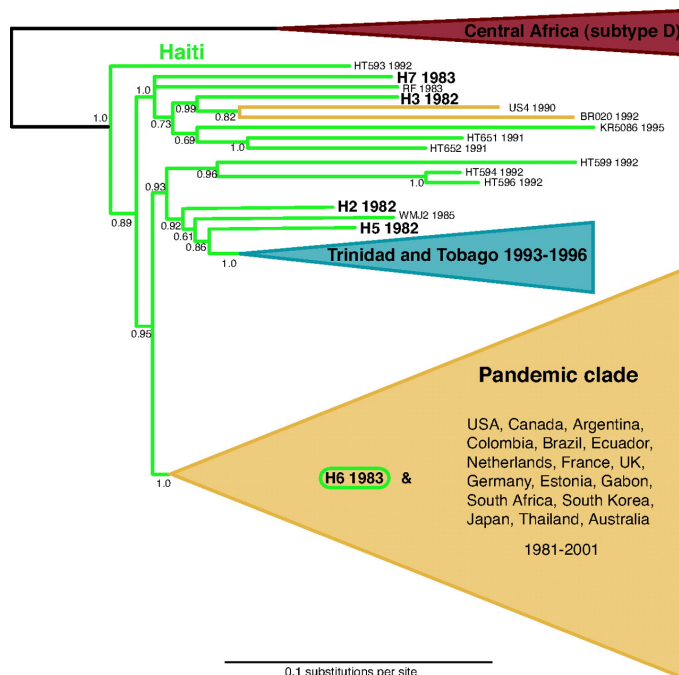
Από την Αϊτή στην Αμερική ή το αντίθετο;

Ο Worobey χρησιμοποίησε ακολουθίες HIV από συντηρημένα δείγματα Αϊτινών ασθενών (1983)



Έλεγχος εξελικτικών υποθέσεων

Η πορεία του HIV



Επιδημία χολέρας στην Αϊτή 2010

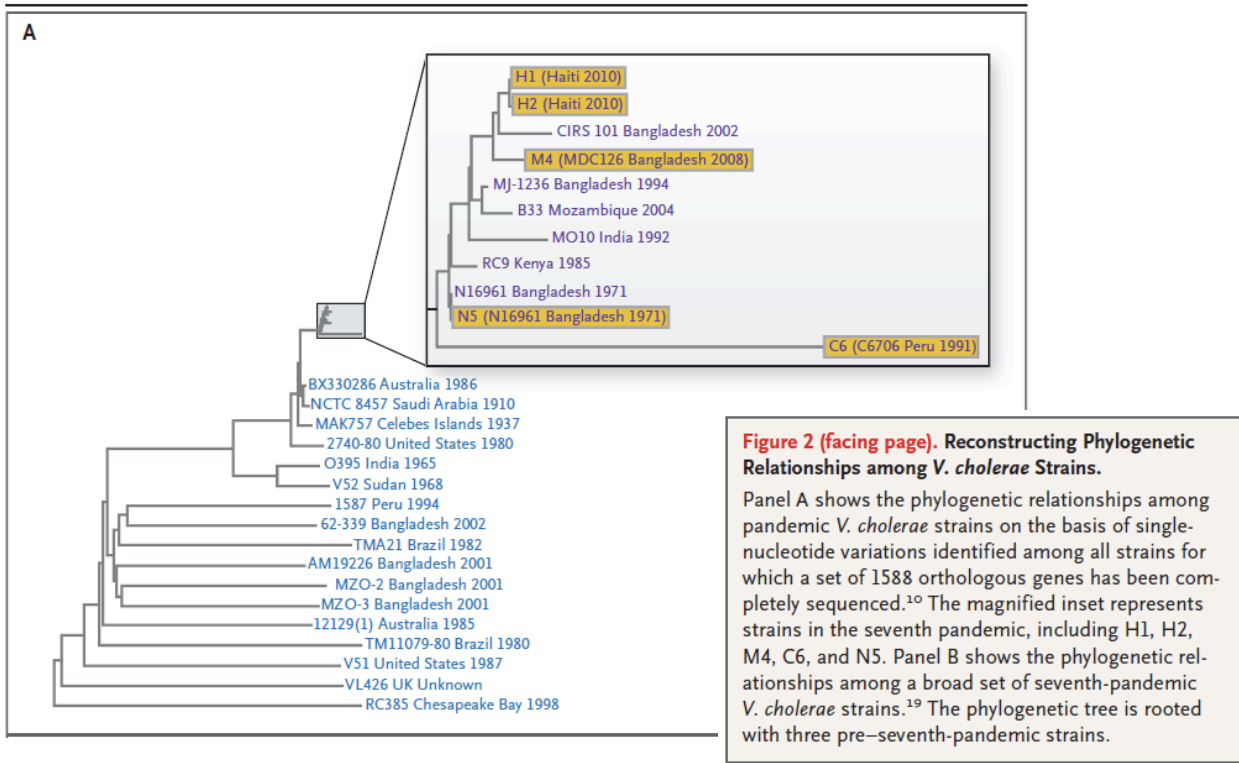
- Μετά τον σεισμό στην Αϊτή (Ιανουάριος 2010), ξέσπασε επιδημία χολέρας (Οκτώβριος 2010).
 - Το βακτήριο *Vibrio cholerae* ελευθερώνει μια τοξίνη που προκαλεί έντονες διάρροιες και αφυδάτωση, έως και θάνατο, εντός ολίγων ωρών, αν δεν αντιμετωπιστεί!
 - Η μετάδοση γίνεται όταν τα κόπρανα ενός μολυσμένου ατόμου έρθουν σε επαφή με πόσιμο νερό ή τροφή.
 - Τα άτομα που δεν παράγουν αρκετό γαστρικό υγρό στο στομάχι τους, ή τα άτομα με ομάδα αίματος O είναι πιο ευάλωτα.
 - Το *Vibrio cholerae* υπάρχει σε υδάτινα περιβάλλοντα ανά την υφήλιο και εάν οι συνθήκες είναι ευνοϊκές, μπορεί να ξεσπάσει επιδημία.
 - Η χολέρα είναι διαδεδομένη στην Ασία.
 - Τα πρώτα κρούσματα παρατηρήθηκαν σε κεντρικές περιοχές του νησιού, στην κοιλάδα Artibonite, μια εβδομάδα μετά την έλευση Νεπαλέζων κυανόκρανων, κοντά στο στρατόπεδό τους.
 - Λύμματα από το στρατόπεδο κατέληγαν σε γειτονικό ποταμό.
 - Οι κάτοικοι κατηγόρησαν τον ΟΗΕ ότι
 - οι κυανόκρανοι που ήρθαν να βοηθήσουν ευθύνονται για το ξέσπασμα της επιδημίας.
 - ότι ο ΟΗΕ προσπάθησε να αποκρύψει το γεγονός και να μην αναλάβει τις ευθύνες του
- Ξέσπασαν ταραχές.

Επιδημία χολέρας στην Αϊτή 2010

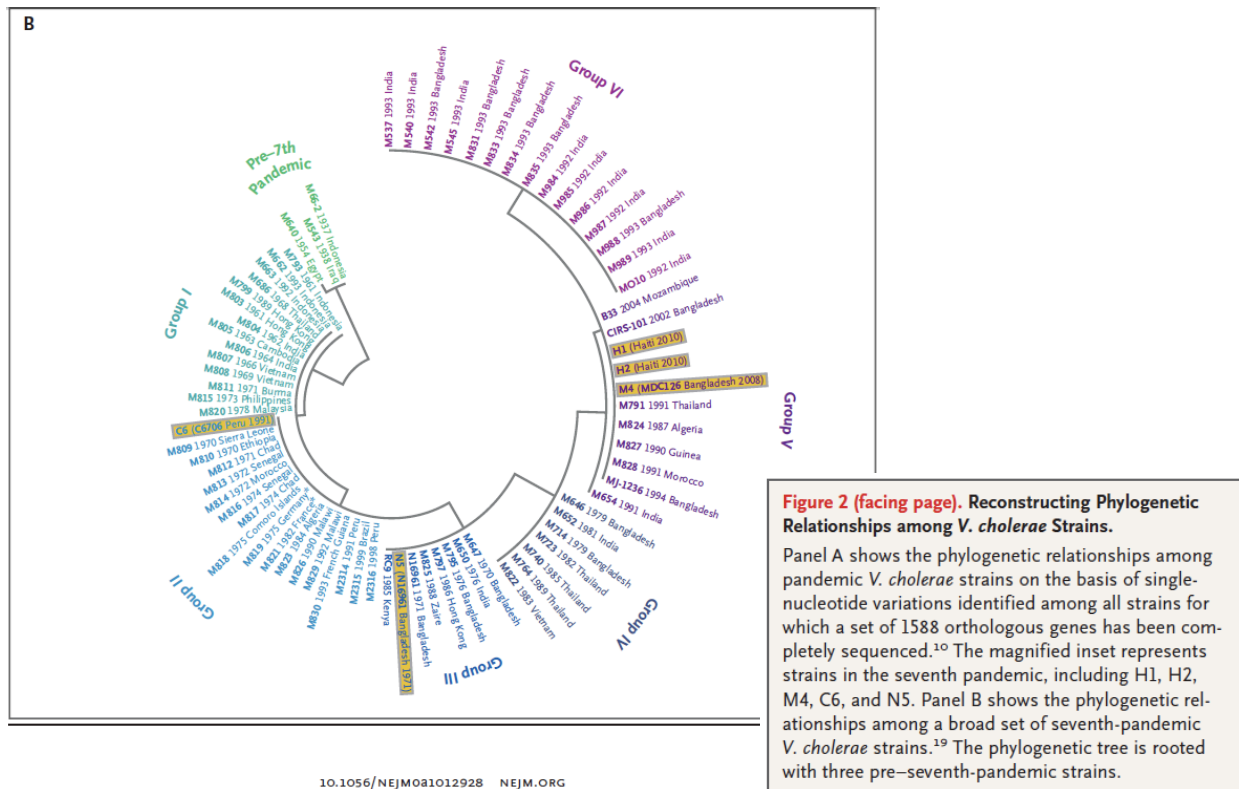
- Αλληλούχιση του γονιδιώματος:
 - 2 κλινικών στελεχών από την τωρινή επιδημία στην Αϊτή.
 - 1 κλινικό στέλεχος από την επιδημία του 1991 στη Νότια Αμερική.
 - 2 στέλεχη που απομονώθηκαν στη Νότια Ασία το 2002 και 2008.
- Επίσης χρησιμοποιήθηκαν οι μερικές αλληλουχίες από 23 άλλα στελέχη ανά την υφήλιο (τα τελευταία 98 χρόνια).
- 1588 συντηρημένα ορθόλογα γονίδια χρησιμοποιήθηκαν από το κάθε στέλεχος, για να γίνει το φυλογενετικό δένδρο.

Επιδημία χολέρας στην Αϊτή 2010

ORIGIN OF CHOLERA OUTBREAK STRAIN IN HAITI



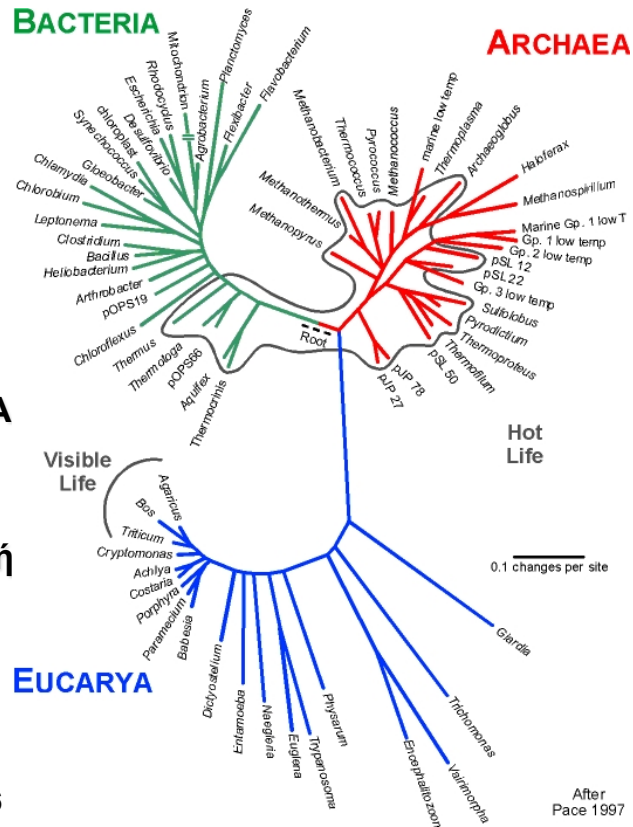
Επιδημία χολέρας στην Αϊτή 2010



Άλλες σημαντικές ανακαλύψεις βάσει φυλογένεσης

Το παγκόσμιο δένδρο της ζωής

Με τη χρήση αλληλουχιών rRNA
 Ικανότητα μελέτης των σχέσεων
 μη καλλιεργούμενων
 οργανισμών, από τη θερμή πηγή
 στο Εθνικό Πάρκο Yellowstone



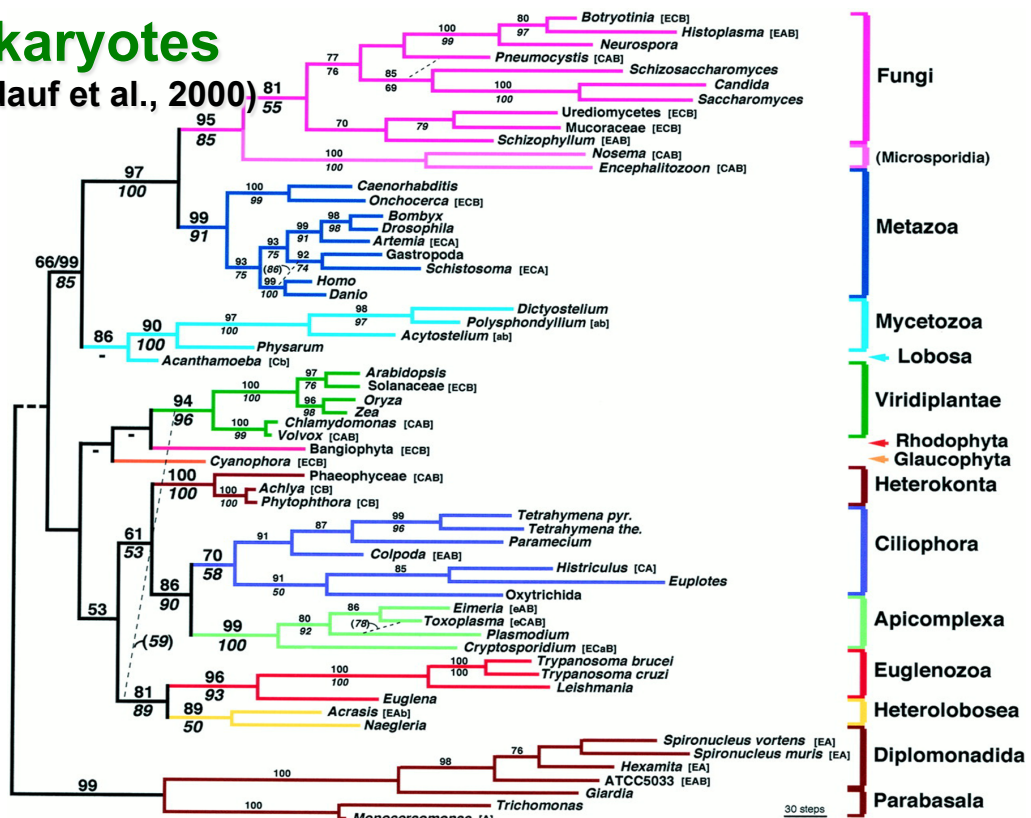
Barns et al., 1996

After Pace 1997

Άλλες σημαντικές ανακαλύψεις βάσει φυλογένεσης

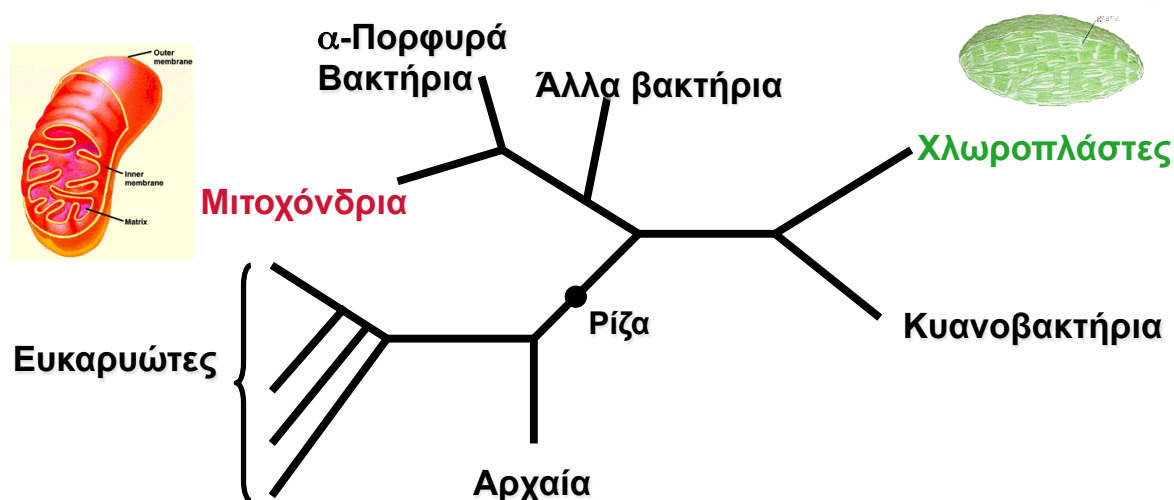
Eukaryotes

(Baldauf et al., 2000)



Άλλες σημαντικές ανακαλύψεις βάσει φυλογένεσης

Ενδοσυμβίωση: Προέλευση των Μιτοχονδρίων και των Χλωροπλάστων



Μιτοχόνδρια και **χλωροπλάστες** προήλθαν από τα α-πορφυρά βακτήρια και τα κυανοβακτήρια αντίστοιχα, μέσω ξεχωριστών ενδοσυμβιωτικών γεγονότων

Μέθοδοι Φυλογένεσης

- Μέθοδοι που βασίζονται σε αποστάσεις
 - UPGMA
 - Κοντινότερης γειτονίας (Neighbor joining)
 - Fitch-Margoliash
 - Ελάχιστης εξέλιξης
- Μέθοδοι που βασίζονται σε χαρακτήρες
 - Μέγιστη φειδωλότητα (Maximum Parsimony)
 - Μέγιστη πιθανοφάνεια (Maximum Likelihood)

Μέθοδοι αποστάσεων

- Αρχικά υπολογίζονται οι αποστάσεις ανάμεσα σε όλα τα πιθανά ζεύγη ακολουθιών.
- Δημιουργείται ένας πίνακας αποστάσεων.
- Με βάση τον πίνακα αυτό, δημιουργούνται δένδρα με μεθόδους που βασίζονται:
 - Στην ομαδοποίηση. Η ομαδοποίηση ξεκινάει από τις πιο κοντινές ακολουθίες και σταδιακά ενσωματώνει όλο και πιο απομακρυσμένες:
 - UPGMA
 - Neighbor joining
 - Στην βελτιστοποίηση. Ο αλγόριθμος συγκρίνει τις πιθανές τοπολογίες και επιλέγει αυτή που οι αποστάσεις πάνω στο δένδρο ταιριάζουν καλύτερα με τις αποστάσεις στον αρχικό πίνακα αποστάσεων:
 - Fitch-Margoliash
 - Ελάχιστη εξέλιξη

Υπολογισμός της απόστασης μεταξύ δύο ακολουθιών

- Παρατηρούμενη απόσταση: από την στοίχιση, μπορούμε να δούμε σε ποιές θέσεις δεν ταιριάζουν οι χαρακτήρες.
- Η παρατηρούμενη απόσταση δεν συμπίπτει με την πραγματική (εξελικτική) απόσταση, λόγω πολλαπλών αντικαταστάσεων στην ίδια θέση. Όσο μεγαλύτερη η απόσταση, τόσο πιο πολλές αντικαταστάσεις συνέβησαν στην ίδια θέση.

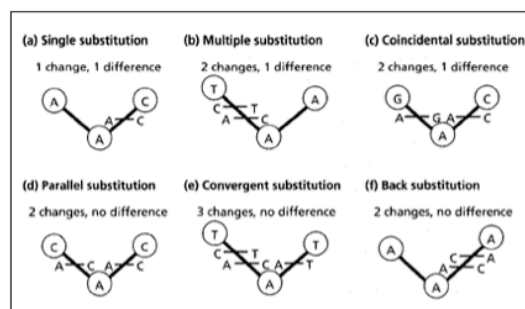


Fig. 5.9

Six kinds of nucleotide substitution. In each case the ancestral nucleotide was A. In all except the case of a single substitution, the number of substitutions that actually occurred is greater than would be counted if we just compared the two descendant sequences. In the lower three cases the nucleotides are identical in both descendant sequences, but this similarity has not been directly inherited from the ancestral sequence. Such similarity is termed 'homoplasy'.

Υπολογισμός της απόστασης μεταξύ δύο ακολουθιών

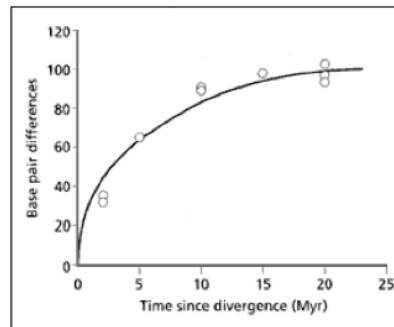


Fig. 5.11
Number of nucleotide substitutions between pairs of bovid mammal mitochondrial sequences (684 basepairs from the *COII* gene) against estimated time of divergence. Notice that the observed number of substitutions is not linear with time but curvilinear. Data from Janecek *et al.* (1996).

Διόρθωση της απόστασης μεταξύ 2 ακολουθιών

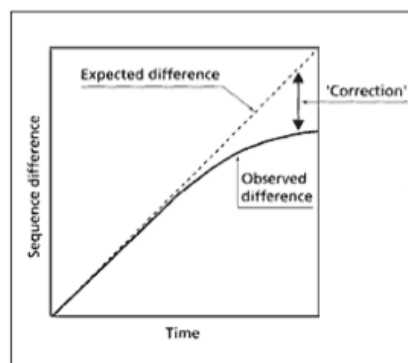
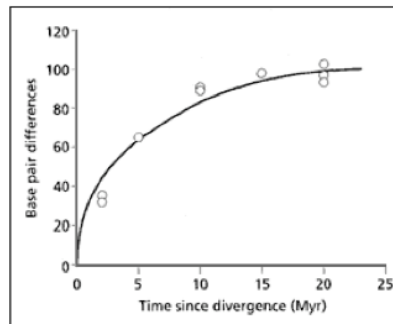


Fig. 5.12
The need to correct observed sequence differences. The extent of observed differences between two sequences is not linear with time (as we would expect if the rate of molecular evolution is approximately constant) but curvilinear due to multiple hits. The goal of distance correction methods is to recover the amount of evolutionary change that the multiple hits have overprinted and to 'correct' the distances for unobserved hits. In effect, the methods seek to 'straighten out' the line representing observed differences.

Μοντέλα αντικατάστασης

- Στατιστικά μοντέλα που λαμβάνουν υπόψη τις πολλαπλές αντικαταστάσεις (για την ίδια θέση) και διορθώνουν την παρατηρούμενη απόσταση, μετατρέποντας την σε εξελικτική.
- Αν η απόσταση είναι πολύ μεγάλη, τότε έχει επέλθει κορεσμός και δεν είναι δυνατόν να γίνει σωστή διόρθωση.



Μοντέλο αντικατάστασης Jukes - Cantor

- Είναι το απλούστερο μοντέλο για ακολουθίες DNA.
- κάθε νουκλεοτίδιο εμφανίζεται με την ίδια συχνότητα
- έχει την ίδια πιθανότητα να μεταλλαχθεί σε ένα από τα υπόλοιπα 3 νουκλεοτίδια

$$d_{AB} = -(3/4) \ln[1 - (4/3) p_{AB}] \quad (\text{Eq. 10.3})$$

where d_{AB} is the evolutionary distance between sequences A and B and p_{AB} is the observed sequence distance measured by the proportion of substitutions over the entire length of the alignment.

For example, if an alignment of sequences A and B is twenty nucleotides long and six pairs are found to be different, the sequences differ by 30%, or have an observed distance 0.3. To correct for multiple substitutions using the Jukes–Cantor model, the corrected evolutionary distance based on Equation 10.3 is:

$$d_{AB} = -3/4 \ln[1 - (4/3 \times 0.3)] = 0.38$$

Μοντέλο αντικατάστασης Kimura

- Πιο εξελιγμένο μοντέλο.
- κάθε νουκλεοτίδιο εμφανίζεται με την ίδια συχνότητα
- Θεωρεί ότι οι μεταπτώσεις έχουν άλλη πιθανότητα να συμβούν, από ότι οι μεταστροφές.

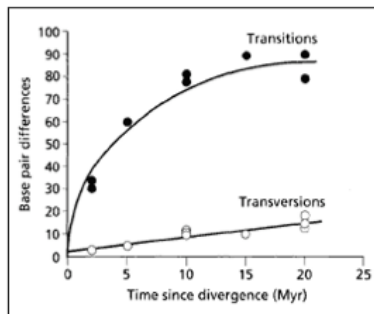


Fig. 5.13

The number of transitions and transversions between the same bovid mammal sequences used in Fig. 5.11. Transitions accumulate much more rapidly than transversions and become saturated, whereas transversions accumulate more slowly and show no evidence of saturation.

Μοντέλο αντικατάστασης Kimura

$$d_{AB} = -(1/2) \ln(1 - 2p_{ti} - p_{tv}) - (1/4) \ln(1 - 2p_{tv}) \quad (\text{Eq. 10.4})$$

An example of using the Kimura model can be illustrated by the comparison of sequences A and B that differ by 30%. If 20% of changes are a result of transitions and 10% of changes are a result of transversions, the evolutionary distance can be calculated using Equation 10.4:

$$d_{AB} = -1/2 \ln(1 - 2 \times 0.2 - 0.1) - 1/4 \ln(1 - 2 \times 0.1) = 0.40$$

Μοντέλα αντικατάστασης για DNA

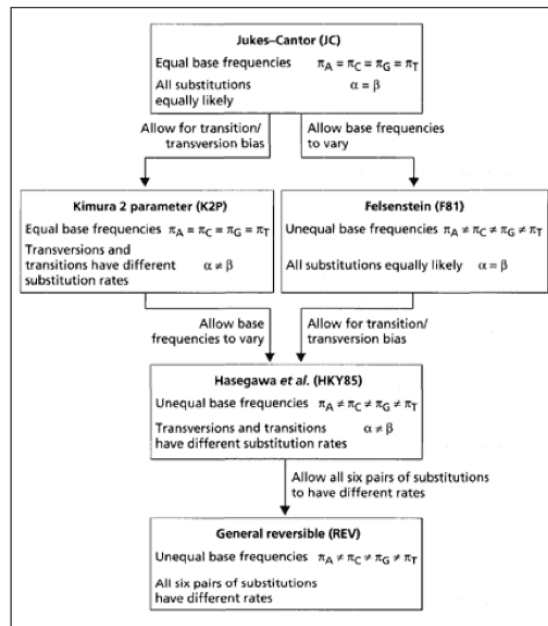


Fig. 5.14

Interrelationships among five models for estimating the number of nucleotide substitutions among a pair of DNA sequences. The JC, K2P, F81 and HKY85 models can all be generated by constraining various parameters of the REV model.

Διόρθωση των παρατηρούμενων αποστάσεων για πρωτεΐνες

2.3 Estimation of distance between two protein sequences • 47

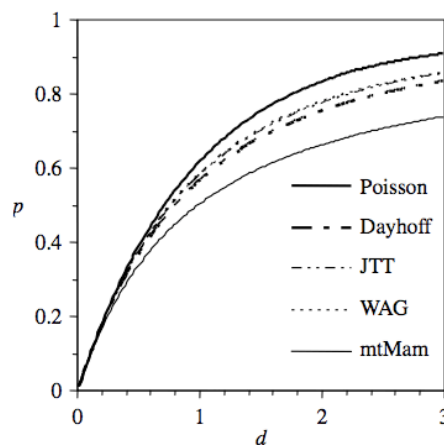


Fig. 2.2 The expected proportion of different sites (p) between two sequences separated by time or distance d under different models. The models are, from top to bottom, Poisson, WAG (Whelan and Goldman 2001), JTT (Jones *et al.* 1992), DAYHOFF (Dayhoff *et al.* 1978), and MTMAM (Yang *et al.* 1998). Note that the results for WAG, JTT, and DAYHOFF are almost identical.

Διόρθωση των παρατηρούμενων αποστάσεων για πρωτεΐνες

- Διόρθωση με πίνακες αντικατάστασης:
 - PAM
 - JTT (Jones-Taylor-Thornton)
- Διόρθωση με αντίστοιχες μεθόδους Jukes-Cantor ή Kimura, προσαρμοσμένες για πρωτεΐνες.

distances. For example, the Kimura model for correcting multiple substitutions in protein distances is:

$$d = -\ln(1 - p - 0.2p^2) \quad (\text{Eq. 10.5})$$

whereas p is the observed pairwise distance between two sequences.

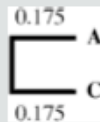
UPGMA

- Βασίζεται στην υπόθεση ότι όλες οι ακολουθίες εξελίσσονται με ένα σταθερό ρυθμό και ότι όλες απέχουν το ίδιο από την ρίζα (κοινό πρόγονο).
- Το τελευταίο τάξον που ενσωματώνεται αποτελεί και την εξωμάδα. Ουσιαστικά, δημιουργείται δένδρο με ρίζα.
- Αποδέχεται την ύπαρξη ενός μοριακού ρολογιού με σταθερή ταχύτητα.
- Στην πραγματικότητα, αυτό δεν ισχύει.
- Σήμερα, το UPGMA χρησιμοποιείται περισσότερο για την ομαδοποίηση δεδομένων από μικροσυστοιχίες και όχι για φυλογένεση.
- Είναι ένας γρήγορος αλγόριθμος κατασκευής δένδρων.

UPGMA

	A	B	C
B	0.40		
C	0.35	0.45	
D	0.60	0.70	0.55

1. Using a distance matrix involving four taxa, A, B, C, and D, the UPGMA method first joins two closest taxa together which are A and C (0.35 in grey). Because all taxa are equidistant from the node, the branch length for A to the node is $AC/2 = 0.35/2 = 0.175$.



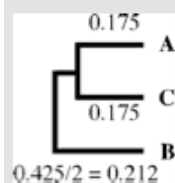
2. Because A and C are joined into a cluster, they are treated as one new composite taxon, which is used to create a reduced matrix. The distance of A-C cluster to every other taxa is one half of a taxon to A and C, respectively. That means that the distance of B to A-C is $(AB + BC)/2$; and that of D to A-C is $(AD + CD)/2$.

	A-C	B
B	$\frac{0.4 + 0.45}{2} = 0.425$	
D	$\frac{0.55 + 0.6}{2} = 0.575$	0.70

UPGMA

	A-C	B
B	$\frac{0.4 + 0.45}{2} = 0.425$	
D	$\frac{0.55 + 0.6}{2} = 0.575$	0.70

3. In the newly reduced-distance matrix, the smallest distance is between B and A-C (in grey), which allows the grouping of B and A-C to create a three-taxon cluster. The branch length for the B is one half of B to the A-C cluster.

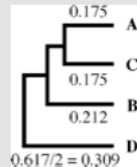


UPGMA

4. When B and A-C are grouped and treated as a single taxon, this allows the matrix to reduce further into only two taxa, D and B-A-C. The distance of D to the composite taxon is the average of D to every single component which is $(BD + AD + CD)/3$.

	B-A-C
D	$\frac{0.7 + 0.6 + 0.55}{3} = 0.617$

5. D is the last branch to add to the tree, whose branch length is one half of D to B-A-C.



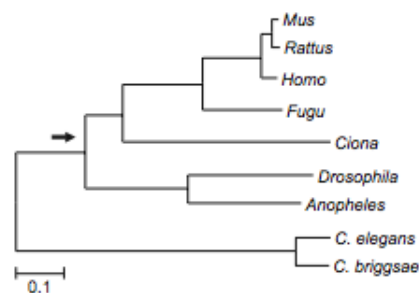
6. Because distance trees allow branches to be additive, the resulting distances between taxa from the tree path can be used to create a distance matrix. Obviously, the estimated distances do not match the actual evolutionary distances shown, which illustrates the failure of UPGMA to precisely reflect the experimental observation.

	A	B	C
B	0.42		
C	0.35	0.42	
D	0.62	0.62	0.62

	A	B	C
B	0.40		
C	0.35	0.45	
D	0.60	0.70	0.55

Μέθοδος σύνδεσης γειτονίας neighbor joining

- Είναι παρόμοια μέθοδος με το UPGMA.
- Οστόσο, δεν θεωρεί ότι όλες οι ακολουθίες εξελίσσονται με τον ίδιο ρυθμό.
- Το δένδρο που παράγεται είναι άρριζο και πρέπει εμείς να επιλέξουμε που είναι η ρίζα.



Μέθοδοι βελτιστοποίησης

- Οι μέθοδοι που βασίζονται σε ομαδοποίηση παράγουν ένα δένδρο.
- Δεν γνωρίζουμε πόσο καλύτερο είναι αυτό το δένδρο από άλλα εναλλακτικά δένδρα.
- Οι μέθοδοι βελτιστοποίησης ελέγχουν τα διάφορα πιθανά δένδρα και βρίσκουν αυτό που ταιριάζει καλύτερα στον αρχικό πίνακα αποστάσεων.

Πόσα πιθανά δένδρα;

- Το σύνολο των πιθανών διαφορετικών δένδρων για ένα αριθμό taxa αυξάνει εκθετικά

$$N_R = (2n - 3)!/2^{n-2}(n - 2)! \quad (\text{Eq. 10.1})$$

In this formula, $(2n - 3)!$ is a mathematical expression of factorial, which is the product of positive integers from 1 to $2n - 3$. For example, $5! = 1 \times 2 \times 3 \times 4 \times 5 = 120$.

For unrooted trees, the number of unrooted tree topologies (N_U) is:

$$N_U = (2n - 5)!/2^{n-3}(n - 3)! \quad (\text{Eq. 10.2})$$

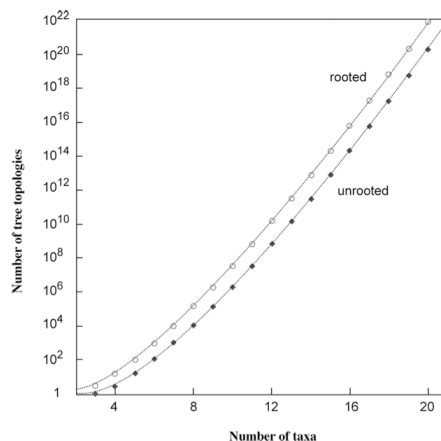
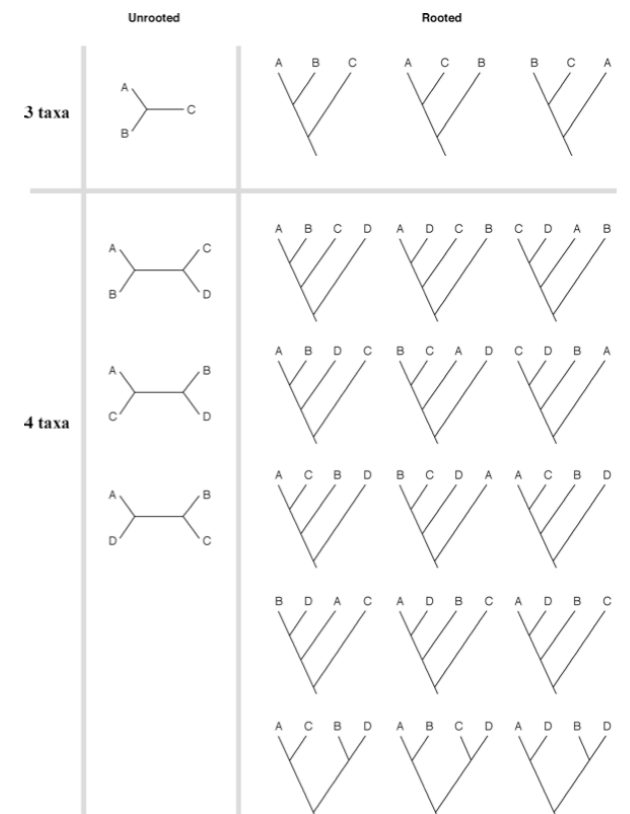


Figure 10.8: Total number of rooted (○) and unrooted (◆) tree topologies as a function of the number of taxa. The values in the y-axis are plotted in the log scale.

Πόσα πιθανά δένδρα;



Fitch-Margoliash

- Διερευνά για το κάθε πιθανό δένδρο ποιές είναι οι αποστάσεις με βάσει αυτό και στην συνέχεια επιλέγει το δένδρο που η υπολογισμένες του αποστάσεις αποκλίνουν το λιγότερο δυνατό από τον αρχικό πίνακα αποστάσεων.

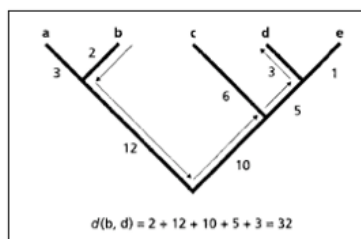


Fig. 5.21

The evolutionary distance between b and d is the sum of the edge lengths along the path in the tree between the two sequences.

Ελάχιστη εξέλιξη

- Παρόμοιο με το Fitch-Margoliash.
- Διερευνά τα πιθανά δένδρα.
- Επιλέγει το δένδρο που το συνολικό μήκος των βραχιόνων του είναι το ελάχιστο δυνατό, για τα υπάρχοντα δεδομένα αποστάσεων.
- Η μέθοδος αυτή είναι λίγο καλύτερη από την Fitch-Margoliash.

Υπέρ και κατά μεθόδων βασισμένων σε αποστάσεις

- Οι μέθοδοι βελτιστοποίησης δίνουν καλύτερα αποτελέσματα από τις μεθόδους ομαδοποίησης, αλλά είναι πιο αργές.
- Αν τα δεδομένα είναι πολλά, τότε προτιμάται μια μέθοδος ομαδοποίησης.
- Οι μέθοδοι αποστάσεων διορθώνουν τις παρατηρούμενες αποστάσεις. Όταν οι ακολουθίες είναι απομακρυσμένες, αυτή η διόρθωση έχει μεγάλες επιπτώσεις και πρέπει να γίνεται.
- Με τις μεθόδους αποστάσεων χάνεται πληροφορία και δεν είναι δυνατόν να ανακατασκευαστεί μια προγονική ακολουθία.

Μέθοδοι που βασίζονται σε χαρακτήρες

Μέγιστη φειδωλότητα (Maximum Parsimony)

Μέγιστη πιθανοφάνεια (Maximum Likelihood)

Βασίζονται στους χαρακτήρες των ακολουθιών και όχι στις αποστάσεις μεταξύ των ακολουθιών.

Είναι δυνατή η ανακατασκευή των προγονικών ακολουθιών.

Μέγιστη φειδωλότητα (Maximum Parsimony)

- Διερευνά τα πιθανά δένδρα και επιλέγει το/τα δένδρο/α που εξηγεί τα δεδομένα με τα λιγότερα δυνατά εξελικτικά βήματα / αντικαταστάσεις.
- Επιτρέπει την ανακατασκευή προγονικών ακολουθιών.
- Βασίζεται στο ξυράφι του Όκαμ (13ος αιώνας), όπου η πιο σύντομη/ απλή εξήγηση είναι μάλλον και η πραγματική.
- Δεν λαμβάνει υπόψη το γεγονός ότι περισσότερες από μια αντικαταστάσεις συνέβησαν στην ίδια θέση.
- Επομένως, για κοντινές ακολουθίες λειτουργεί καλά, για απομακρυσμένες ακολουθίες, που αυξάνεται η πιθανότητα πολλαπλών αντικαταστάσεων στην ίδια θέση, είναι προβληματική μέθοδος.

Μέγιστη φειδωλότητα (Maximum Parsimony)

- Δεν χρησιμοποιεί όλες τις θέσεις μια πολλαπλής στοίχισης, άλλα μόνο εκείνες που έχουν αρκετή πληροφορία για να επιτραπεί ο διαχωρισμός/ομαδοποίηση των ακολουθιών.
- Τέτοιες θέσεις πρέπει να έχουν τουλάχιστον 2 ειδών διαφορετικούς χαρακτήρες και ο κάθε ένας από αυτούς να υπάρχει τουλάχιστον σε 2 ακολουθίες.

Figure 11.1: Example of identification of informative sites that are used in parsimony analysis. Sites 2, 5, and 8 (grey boxes) are informative sites. Other sites are noninformative sites, which are either constant or having characters occurring only once.

taxa \ sites	1	2	3	4	5	6	7	8
I	A	A	T	T	A	G	C	T
II	G	G	T	C	G	T	A	G
III	A	A	T	G	C	G	C	T
IV	A	G	T	A	A	G	C	A
V	A	C	T	T	C	G	C	G
VI	A	C	A	T	G	G	C	A

Μέγιστη φειδωλότητα (Maximum Parsimony)

- Για μια πιθανή τοπολογία δένδρου, υπολογίζεται πόσα συνολικά εξελικτικά βήματα / αντικαταστάσεις χρειάζονται (στο σύνολο των θέσεων που χρησιμοποιούνται).
- Επιλέγεται το δένδρο με τα λιγότερα εξελικτικά βήματα.
- Συχνά, υπάρχουν περισσότερες από μια βέλτιστες λύσεις/δένδρα, γιατί δεν γνωρίζουμε ποιοί ήταν πραγματικά οι χαρακτήρες στις προγονικές ακολουθίες. Τότε δημιουργείται ένα δένδρο συναίνεσης από τα εξίσου βέλτιστα δένδρα.

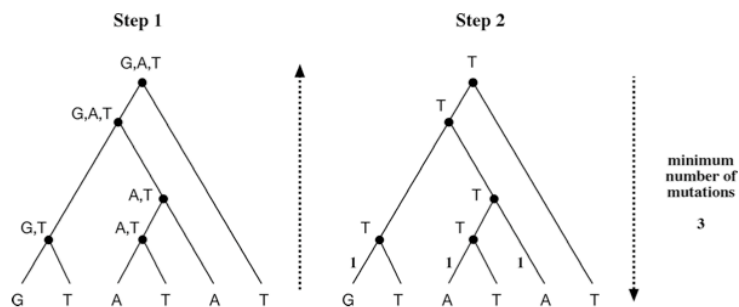


Figure 11.2: Using parsimony to infer ancestral characters at internal nodes involves a two-step procedure. The first step involves going from the leaves to the root and counting all possible ancestral characters at the internal nodes. The second step goes from the root to the leaves and assigns ancestral characters that involve minimum number of mutations. In this example, the total number of mutations is three if T is at the root, whereas other possible character states increase that number.

Αναζητώντας το καλύτερο δένδρο

- Όταν ο αριθμός των taxa είναι μικρός, τότε μπορούν να υπολογιστούν όλα τα δυνατά δένδρα (brute force).
- Όταν $10 < \text{taxa} < 20$, τότε εφαρμόζεται το branch and bound.
- Όταν $\text{taxa} > 20$, εφαρμόζονται ευρετικές μέθοδοι.

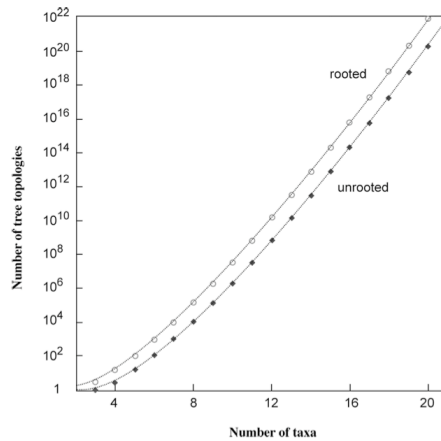


Figure 10.8: Total number of rooted (○) and unrooted (◆) tree topologies as a function of the number of taxa. The values in the y-axis are plotted in the log scale.

Αναζητώντας το καλύτερο δένδρο

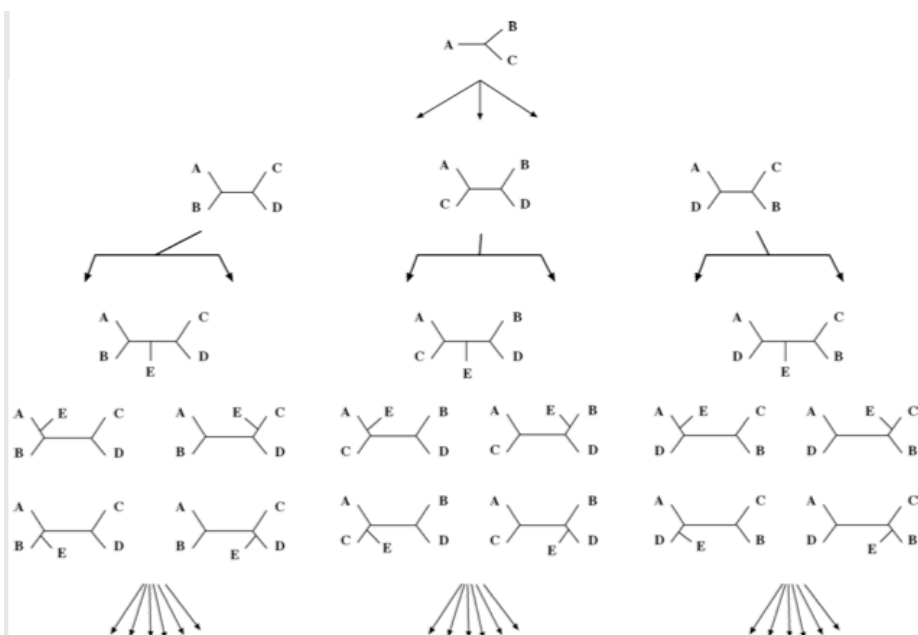


Figure 11.4: Schematic of exhaustive tree construction in the MP procedure. The tree starts with three taxa with one topology. One taxon is then added at a time in an progressive manner, during which the total branch lengths of all possible topologies are calculated.

Αναζητώντας το καλύτερο δένδρο

- Branch and bound.
- Δημιουργείται το δένδρο με UPGMA ή neighbor joining.
- Υπολογίζονται τα εξελικτικά βήματα για αυτό το δένδρο.
- Ο αριθμός αυτός αποτελεί την 'οροφή'. Ένα δένδρο μέγιστης φειδωλότητας θα πρέπει να έχει τον ίδιο αριθμό βημάτων ή και μικρότερο.
- Καθώς χτίζεται σταδιακά το δένδρο φειδωλότητας, αν σε κάποιο στάδιο κάποιες επιλογές καταλήγουν σε βήματα που ξεπερνούν την οροφή, τότε απορρίπτεται το συγκεκριμένο μονοπάτι

Αναζητώντας το καλύτερο δένδρο

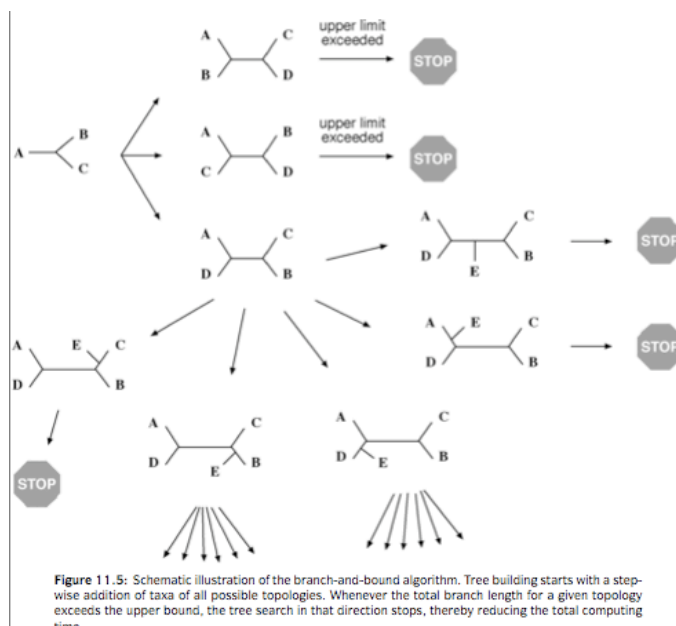


Figure 11.5: Schematic illustration of the branch-and-bound algorithm. Tree building starts with a step-wise addition of taxa of all possible topologies. Whenever the total branch length for a given topology exceeds the upper bound, the tree search in that direction stops, thereby reducing the total computing time.

Αναζητώντας το καλύτερο δένδρο

- Ευρετικές μέθοδοι:
 - Δημιουργείται ένα δένδρο με neighbor joining και υπολογίζονται τα εξελικτικά βήματα για το συγκεκριμένο δένδρο.
 - Δοκιμάζονται τροποποιήσεις πάνω στο δένδρο αυτό. Αν βρεθεί ένα τροποποιημένο δένδρο με μικρότερο αριθμό εξελικτικών βημάτων, τότε επιλέγεται αυτό και οι τροποποιήσεις γίνονται πάνω του, έως ότου βρεθεί ένα ακόμα καλύτερο δένδρο. Η διαδικασία συνεχίζεται έως ότου να μην βρίσκεται καλύτερο δένδρο.
- Ευρετικές μέθοδοι είναι γρήγορες, όμως δεν δίνουν πάντοτε την καλύτερη λύση.

Αναζητώντας το καλύτερο δένδρο

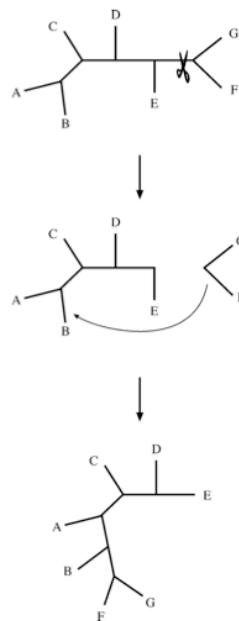


Figure 11.6: Schematic representation of a typical branch swapping process in which a branch is cut and moved to another part of the tree, generating a new topology.

Μέγιστη φειδωλότητα (Maximum Parsimony)

- Δεν διορθώνει για πολλαπλές αντικαταστάσεις πάνω στην ίδια θέση, άρα είναι προβληματική όταν μελετάμε απομακρυσμένες ακολουθίες.
- Δεν χρησιμοποιεί όλες τις θέσεις μιας πολλαπλής στοίχισης.
- Η λύση επηρεάζεται από τον αλγόριθμο αναζήτησης του καλύτερου δένδρου.
- Είναι επιρρεπής στην έλξη μεταξύ μακρινών βραχιόνων (long branch attraction).

Έλξη μεταξύ μακρινών βραχιόνων (long branch attraction).

- Τάξα που εξελίσσονται με γρήγορους ρυθμούς και επομένως έχουν μακρείς βραχίονες, έλκονται μεταξύ τους.

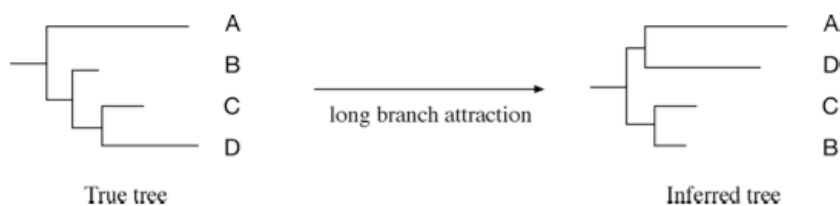


Figure 11.7: The LBA artifact showing taxa A and D are artifactually clustered during phylogenetic construction.

Μέγιστη πιθανοφάνεια

- Βασίζεται σε χαρακτήρες.
- Χρησιμοποιεί όλες τις θέσεις μια πολλαπλής στοίχισης.
- Χρησιμοποιεί πιθανότητες και μοντέλα αντικατάστασης.
- Υπολογίζονται οι χαρακτήρες σε κάθε προγονική ακολουθία.
- Υπολογίζει για το κάθε πιθανό εξελικτικό μονοπάτι (προγονικές ακολουθίες και δένδρο) την πιθανότητα του, με βάση τα παρατηρούμενα σημερινά δεδομένα και ένα συγκεκριμένο μοντέλο εξέλιξης (μοντέλο αντικατάστασης).
- Οι πιθανότητες μετατρέπονται σε log-likelihood scores.
- Δένδρο με το μεγαλύτερο log-likelihood score επιλέγεται.

Μέγιστη πιθανοφάνεια

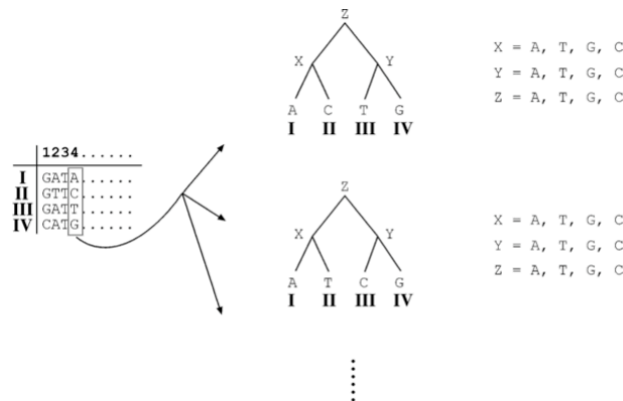


Figure 11.8: Schematic representation of the ML approach to build phylogenetic trees for four taxa, I, II, III, and IV. The ancestral character states at the internal nodes and root node are assigned X, Y, and Z, respectively. The example only shows some of the topologies derived from one of the sites in the original alignment. The method actually uses all the sites in probability calculation for all possible trees with all combinations of possible ancestral sequences at internal nodes according to a predefined substitution model.

$$L_{(4)} = \Pr(Z \rightarrow X) * \Pr(Z \rightarrow Y) * \Pr(X \rightarrow A) * \Pr(X \rightarrow C) * \Pr(Y \rightarrow T) * \Pr(Y \rightarrow G)$$

$$\ln L_{(4)} = \ln \Pr(Z \rightarrow X) + \ln \Pr(Z \rightarrow Y) + \ln \Pr(X \rightarrow A) + \ln \Pr(X \rightarrow C)$$

$$+ \ln \Pr(Y \rightarrow T) + \ln \Pr(Y \rightarrow G)$$

Αξιολόγηση του δένδρου

- Bootstrap:
 - Τυχαία δειγματοληψία θέσεων της πολλαπλής στοίχισης.
 - Μια θέση μπορεί να επιλεγεί περισσότερες από μια φορές ή και καμία.
 - Δημιουργία μιας νέας αλλαγμένης πολλαπλής στοίχισης
 - Η διαδικασία επαναλαμβάνεται 100-1000 φορές.
 - Για κάθε νέα πολλαπλή στοίχιση, υπολογίζεται το δένδρο.
 - Τα νέα δένδρα συγχωνεύονται σε ένα νέο δένδρο (consensus tree).
 - Bootstrap -> συχνότητα εμφάνισης ενός κόμβου.
 - Bootstrap 70% -> 95% εμπιστοσύνη.
 - Αν η μεθοδολογία δημιουργίας του δένδρου είναι λάθος, μπορεί να πάρουμε υψηλές τιμές bootstrap για το λάθος δένδρο.

bootstrap

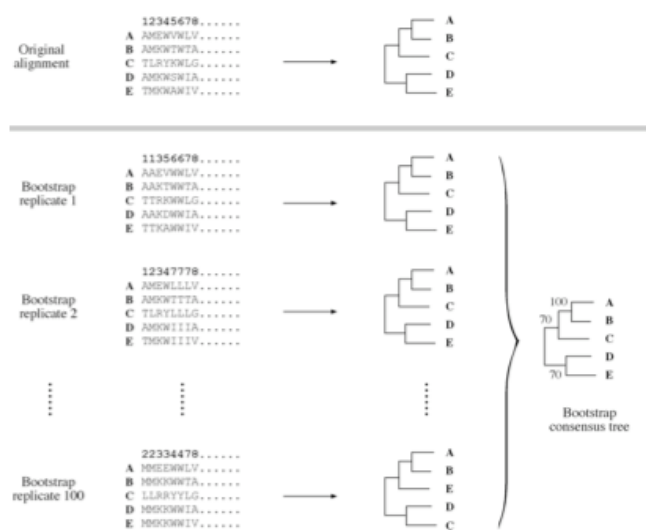


Figure 11.10: Schematic representation of a bootstrap analysis showing the original alignment and modified replicates in which certain sites are randomly replaced with other existing sites. The resulting altered replicates are used for building trees for statistical analysis at each node.