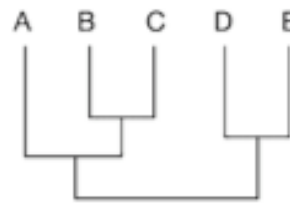
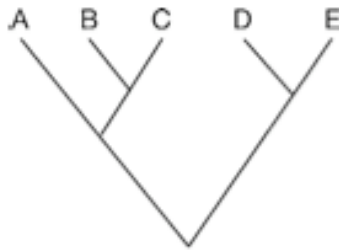


Φυλογένεση

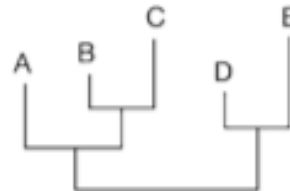
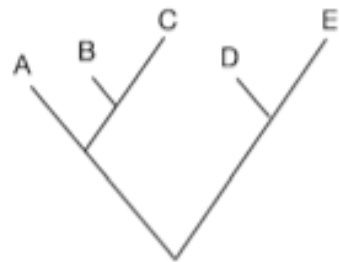
Γρηγόριος Αμούτζιας
Καθηγητής Βιοπληροφορικής με έμφαση στη Μικροβιολογία
Τμήμα Βιοχημείας και Βιοτεχνολογίας
Πανεπιστήμιο Θεσσαλίας

Φυλογένεση

- Η εκτίμηση της εξελικτικής ιστορίας γονιδίων/πρωτεϊνών ή οργανισμών.
- Η απεικόνιση αυτής της ιστορίας γίνεται με φυλογράμματα/κλαδογράμματα



Cladogram



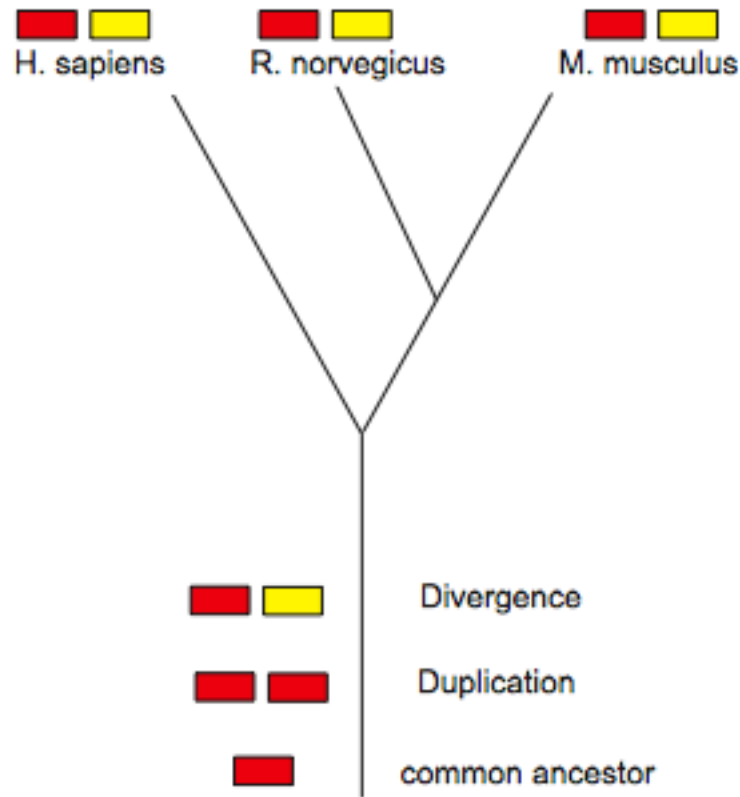
Phylogram

Figure 10.4: Phylogenetic trees drawn as cladograms (*top*) and phylograms (*bottom*). The branch lengths are unscaled in the cladograms and scaled in the phylograms. The trees can be drawn as angled form (*left*) or squared form (*right*).

Λίγη εξέλιξη: ομολογία

- Ομόλογα γονίδια: κοινός εξελικτικός πρόγονος.
- Ορθόλογα γονίδια: προέρχονται από ειδογένεση. Ουσιαστικά, ένα γονίδιο α (μεταλλαγμένο) σε δύο διαφορετικούς οργανισμούς. Συχνά έχουν την ίδια λειτουργία
- Παράλογα γονίδια: προέρχονται από γονιδιακό διπλασιασμό. Ανήκουν στην ίδια οικογένεια
- Ξενόλογα γονίδια: από οριζόντια μεταφορά

Λίγη εξέλιξη: ομολογία (II)



Λίγη εξέλιξη: Ορθόλογα, παράλογα και εξελικτική απόσταση

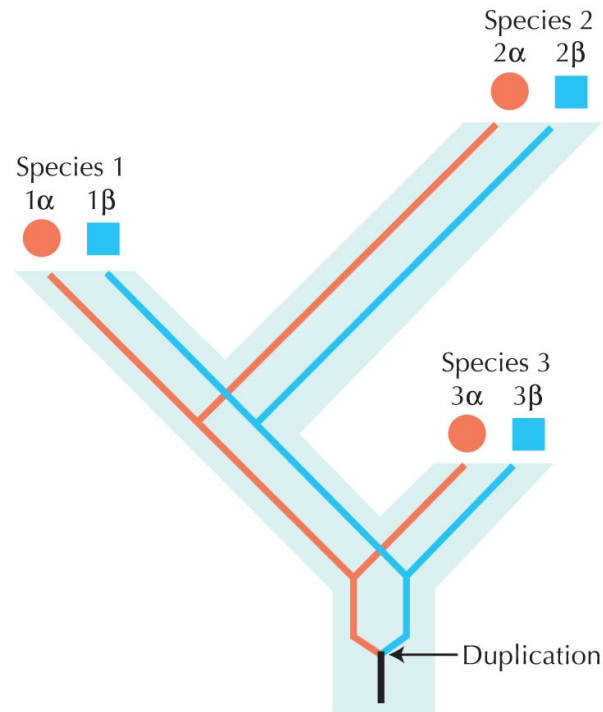
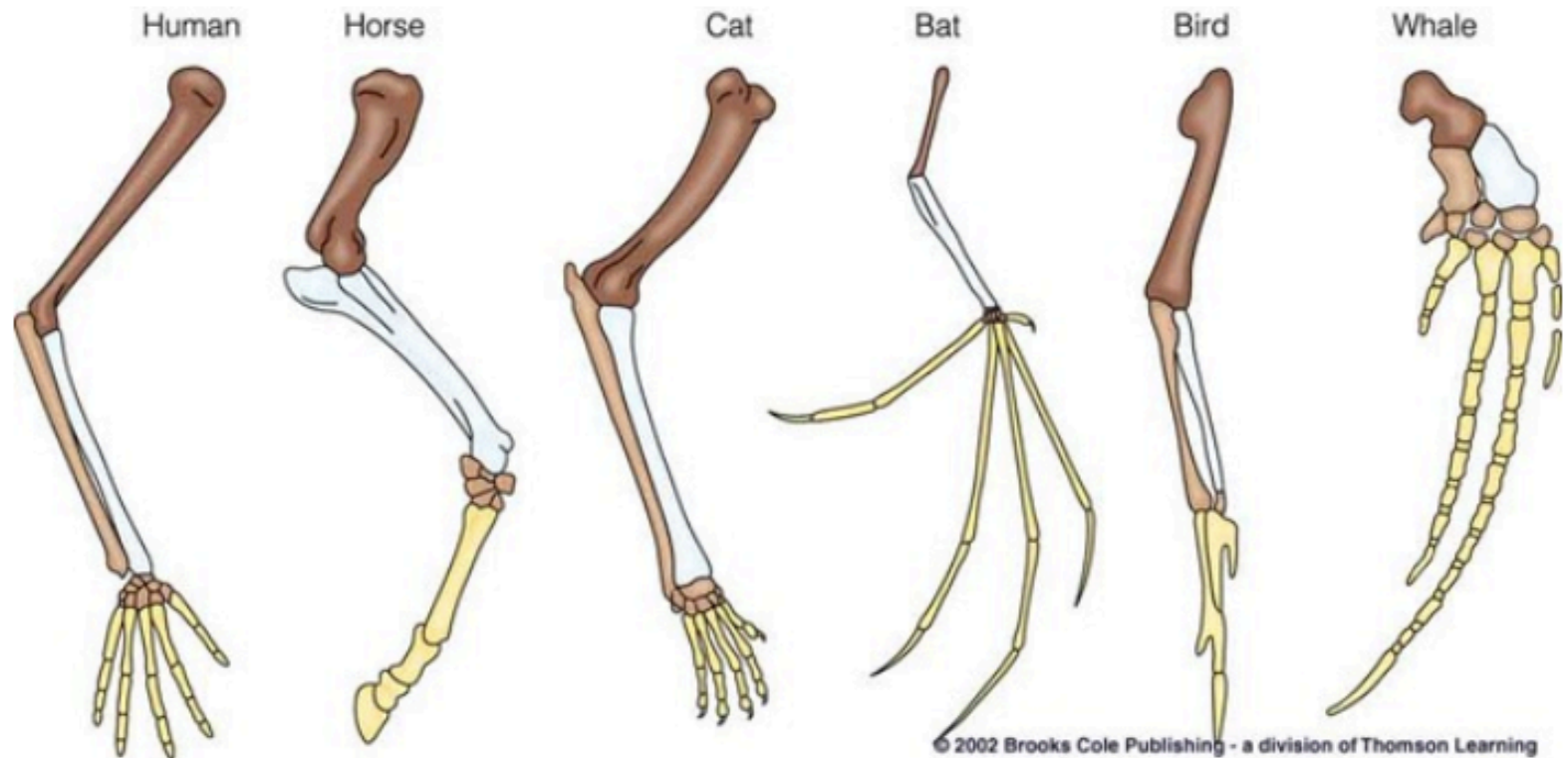


FIGURE 27.33. Sequence similarity does not always accurately represent relatedness. In this hypothetical tree, the evolution of three species is indicated by *thick light blue bars* and the evolution of specific genes is indicated by *thin red and blue lines* within the bars. An early gene duplication gave rise to the α and β genes; thus all of the α genes are paralogs of the β genes. Over time, the α and β genes diverged such that their functions now differ (indicated by *red* and *blue*). The multiple forms of the α genes arose through speciation events, and thus all are orthologs of each other (and likewise for all the β genes). If species 3 had a slower rate of evolution than the other two, then the α and β paralogs within species 3 will be more similar to each other (as measured by summing the branch lengths connecting them) than either is to their orthologs in other species. (Redrawn from Eisen J.A. *Genome Res.* **8**: 163–167, Table 4, © 1998 CSHLP.)

Vertebrate limbs are homologous.



Ομοπλασία

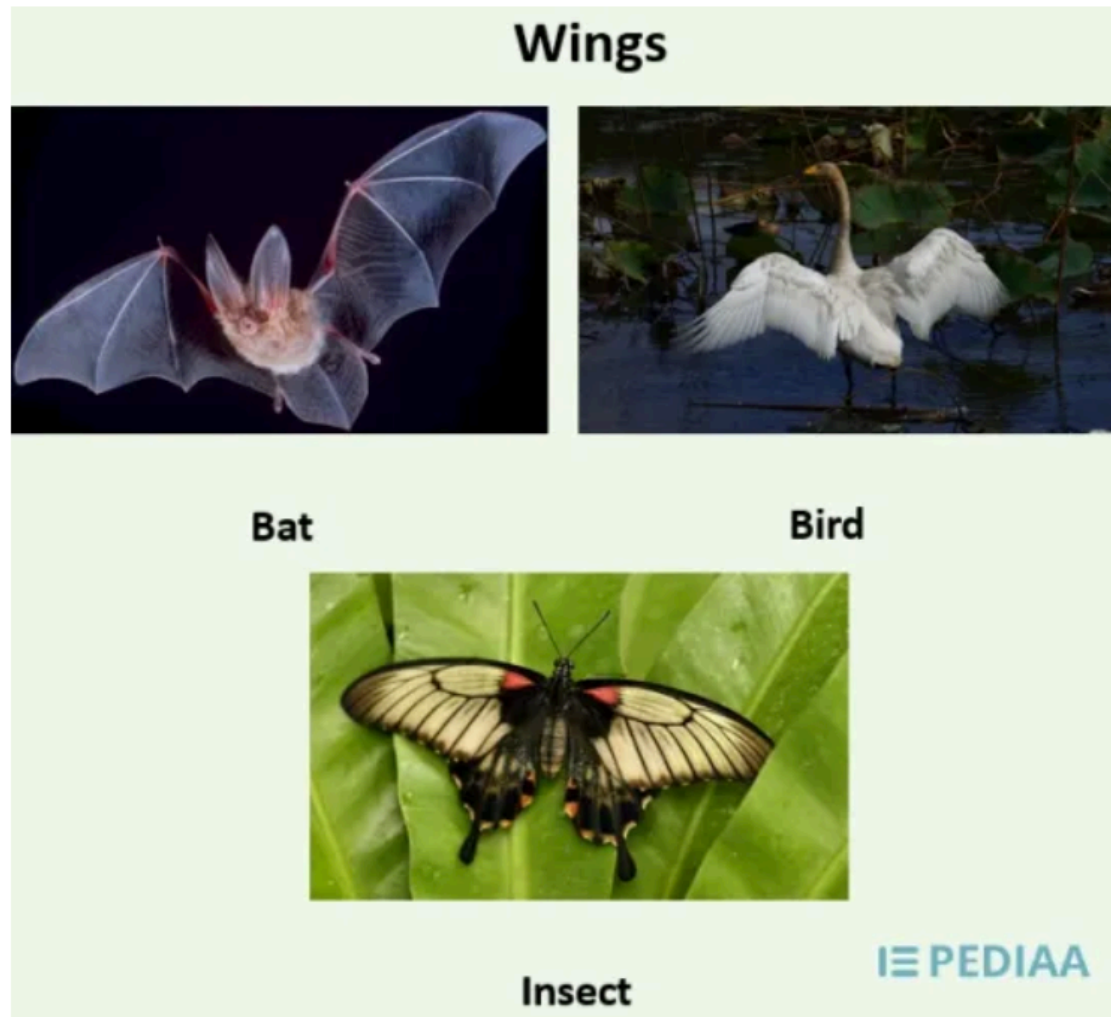


Figure 2: Homoplasious Wings

HOMOLOGY

VERSUS

HOMOPLASY

Homology refers to a similarity between two or more species that emerged from a common ancestor	Homoplasy refers to a shared character between two or more species that did not arise from a common ancestor	Basic structure is similar	Basic structures are different
Arises from a common ancestor	Does not arise from a common ancestor	Homologous characters perform different functions	Homoplasious characters perform similar functions
Results from divergent evolution	Results from convergent evolution	Homologous characters share a high degree of genetic similarity	Homoplasious characters do not share a genetic similarity
		Occurs as a result of the evolutionary relationship	Occurs as a result of an adaptation to the environmental conditions
		Can be used to describe the evolutionary relationship of different species	Cannot be used to describe the evolutionary relationship between species
		Vertebrate limbs are an example	Wings of birds, bats, and insects are an example
			Visit www.pediaa.com

Ομοπλασία ακολουθιών

A

Sequence 1 G A C G A G

Sequence 2 G C C G A C

B

Sequence 1

Sequence 2

G A C G **A** G

G C C G **A** C

↑ G → A at
position 5

↑ C → A at
position 5

G A C G G G

G C C G C C

↑ G → A at
position 2

↑ G → C at
position 5
G → C at
position 6

G G C G G G

G C C G G G } Ancestral sequences

FIGURE 27.20. Homoplasy in sequences. Homoplasy refers to the presence of identical character states that did not arise through shared descent but rather by some other process, such as convergent evolution. This can occur with any character, including molecular sequences. For example, consider these two aligned DNA sequences (*A*), which both have an A at position 5. If the ancestor of Sequence 1 was GACGGG and the ancestor of Sequence 2 was GCCGCC (*B*), then the shared As at position 5 are not a reflection of common ancestry and are instead an example of homoplasy.

Ομοπλασία ακολουθιών

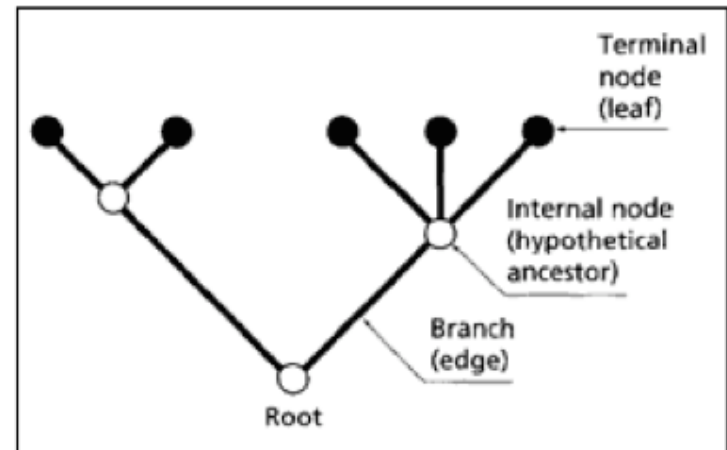
- Δεν είναι σπάνιο δύο φαινοτυπικοί χαρακτήρες να μοιάζουν λόγω ομοπλασίας/συγκλίνουσας εξέλιξης.
- Γενικά είναι πολύ δύσκολο δύο μοριακές ακολουθίες να μοιάζουν λόγω ομοπλασίας (εξαίρεση οι περιοχές χαμηλής πολυπλοκότητας).
- Για τους παραπάνω δύο λόγους προτιμούμε την φυλογένεση με μοριακές ακολουθίες για να δούμε την εξέλιξη των οργανισμών.

Στάδια φυλογενετικής ανάλυσης

- Επιλογή ακολουθιών:
 - Επιλογή μοριακών δεικτών
 - Εντοπισμός ομόλογων ακολουθιών
 - Π.χ. Blast, HMMs
- Πολλαπλή στοίχιση
 - Διορθώσεις στην στοίχιση
- Υπολογισμός φυλογενετικού δένδρου
 - Επιλογή εξελικτικού μοντέλου
 - Επιλογή μεθόδου δημιουργίας του δένδρου
 - Αξιολόγηση/αξιοπιστία του δένδρου

Στοιχεία ενός φυλογενετικού δένδρου

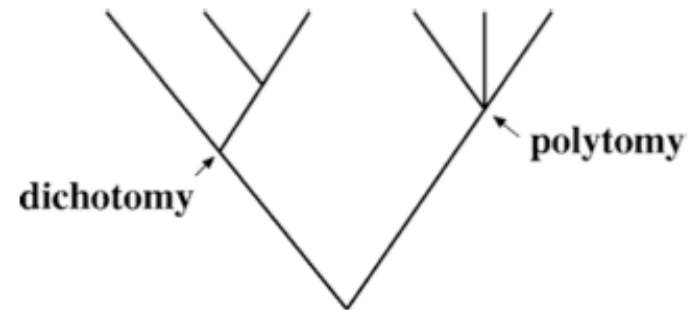
- Φύλλα (leafs)
 - Taxon
 - Operational taxonomic units (OTUs)
- Βραχίονες (branches)
- Κόμβοι (nodes)
- Κλάδοι (clades)
 - Μονοφυλετικά group
- Ρίζα (root)



Στοιχεία ενός φυλογενετικού δένδρου

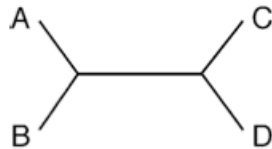
- Τοπολογία του δένδρου
 - Διχοτόμιση (dichotomy)
 - Πολυτόμιση (polytomy)
 - Radiation
 - Unresolved phylogeny

Figure 10.2: A phylogenetic tree showing an example of bifurcation and multifurcation. Multifurcation is normally a result of insufficient evidence to fully resolve the tree or a result of an evolutionary process known as *radiation*.

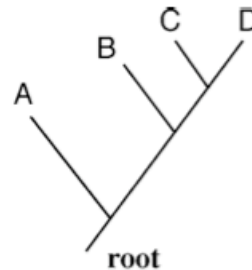


Δένδρα με/χωρίς ρίζα

- Χωρίς ρίζα
 - Δεν γνωρίζουμε τον κοινό πρόγονο.
 - Απεικονίζονται μόνο οι σχετικές θέσεις των taxa.
 - Δεν φαίνεται η εξελικτική πορεία.
- Με ρίζα
 - Γνωρίζουμε τον κοινό πρόγονο.
 - Φαίνεται η εξελικτική πορεία.
 - Χρησιμοποιούνται:
 - Outgroup
 - Midpoint rooting approach (υποθέτει την ύπαρξη μοριακού ρολογιού - σταθερού ρυθμού εξέλιξης για όλες τις ακολουθίες).



Unrooted



Rooted

Figure 10.3: An illustration of rooted versus unrooted trees. A phylogenetic tree without definition of a root is unrooted (*left*). The tree with a root is rooted (*right*).

Δένδρα με/χωρίς ρίζα

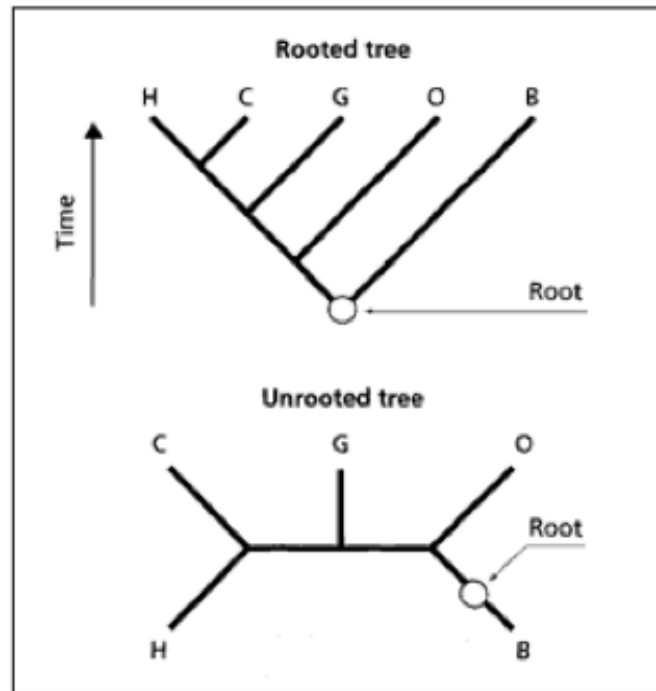


Fig. 2.6

Rooted and unrooted trees for human (H), chimp (C), gorilla (G), orang-utan (O), and gibbon (B). The rooted tree (top) corresponds to the unrooted tree below.

Δένδρα με/χωρίς ρίζα

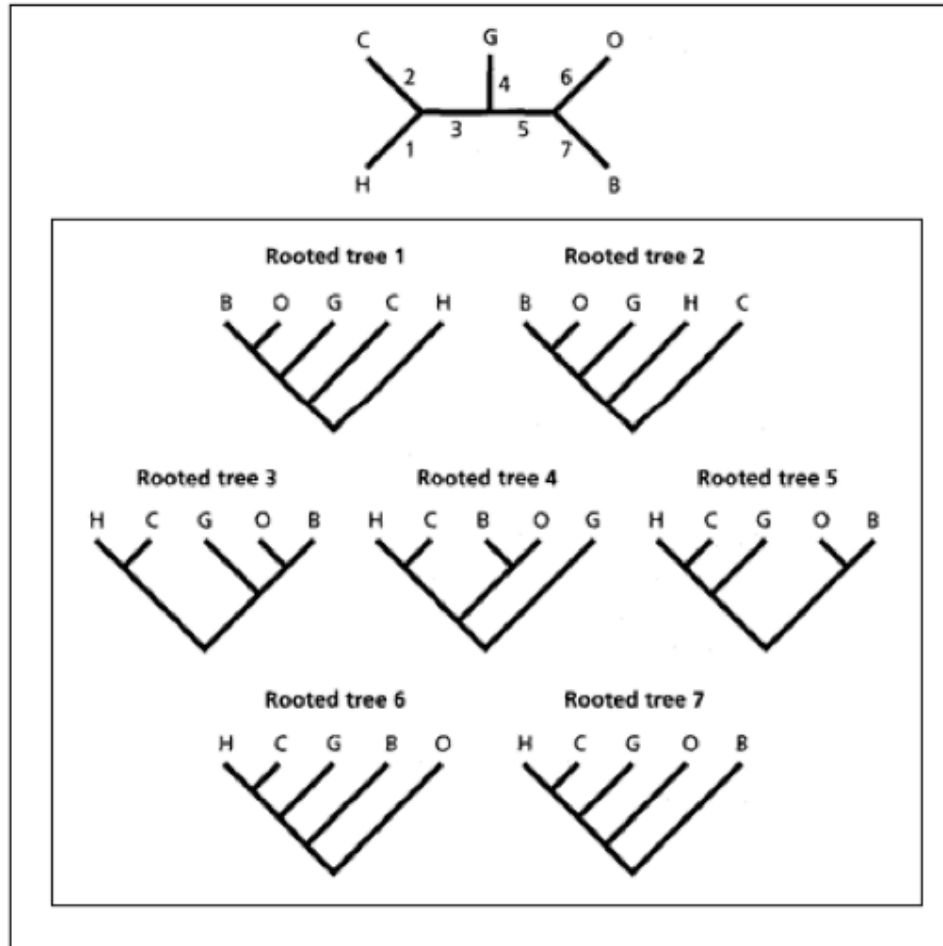
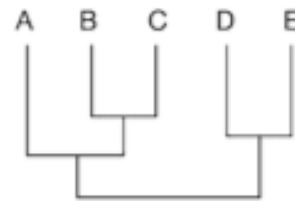
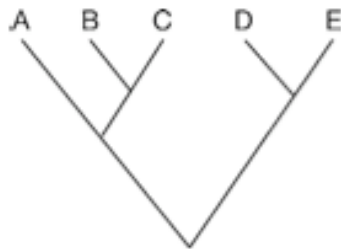


Fig. 2.7

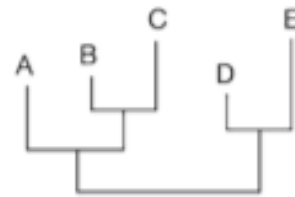
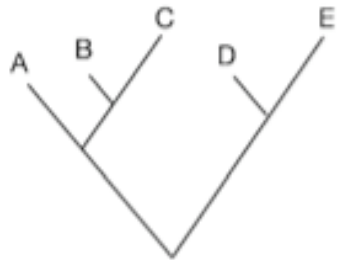
The seven rooted trees that can be derived from an unrooted tree for five sequences. Each rooted tree 17 corresponds to placing the root on the corresponding numbered branch of the unrooted tree.

(Sequence labels as for Fig. 2.6.)

Κλαδόγραμμα/φυλόγραμμα



Cladogram

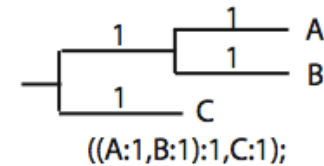
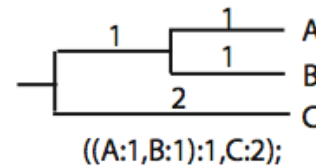
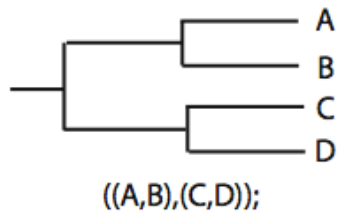
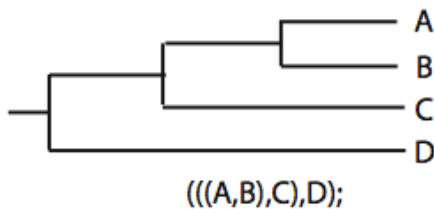
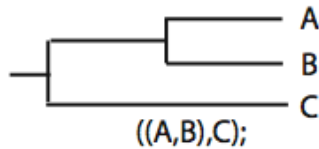


Phylogram

Figure 10.4: Phylogenetic trees drawn as cladograms (*top*) and phylograms (*bottom*). The branch lengths are unscaled in the cladograms and scaled in the phylograms. The trees can be drawn as angled form (*left*) or squared form (*right*).

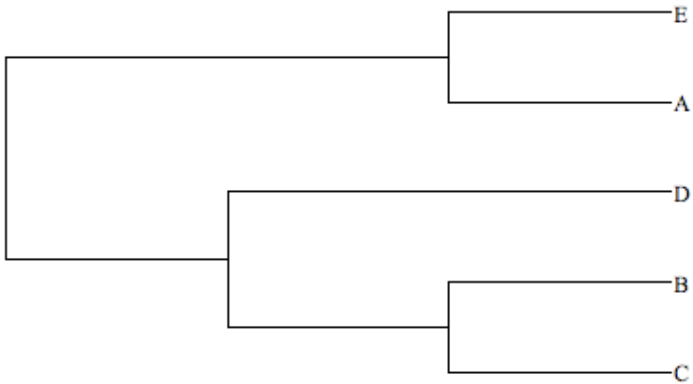
Newick format

- Δένδρα αποθηκεύονται σε μορφή Newick ή Nexus (παραλλαγή του Newick).



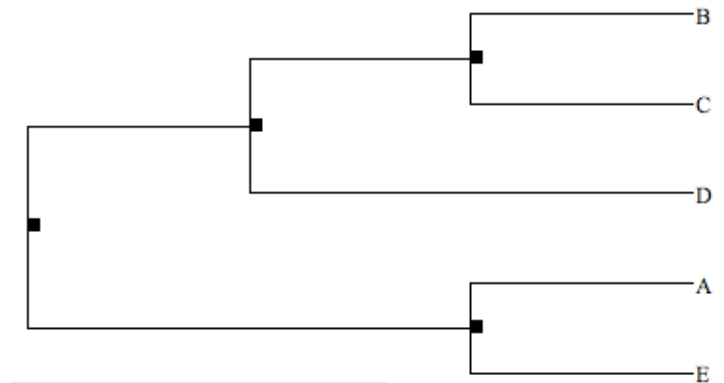
- Ποιό είναι το δένδρο: $(((C,B),D),(A,E));$

Newick format



`((C,B),D),(A,E));`

Είναι το ίδιο δένδρο;



Newick format

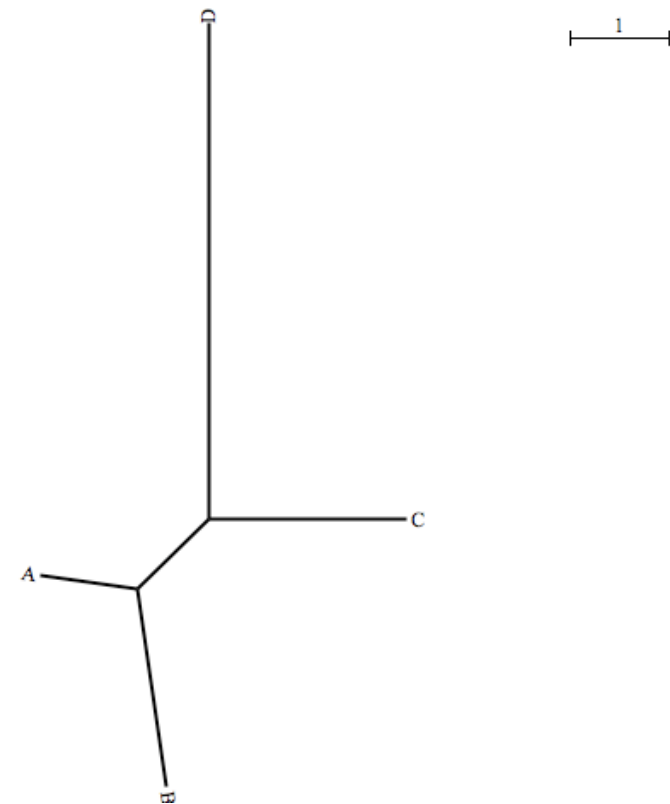
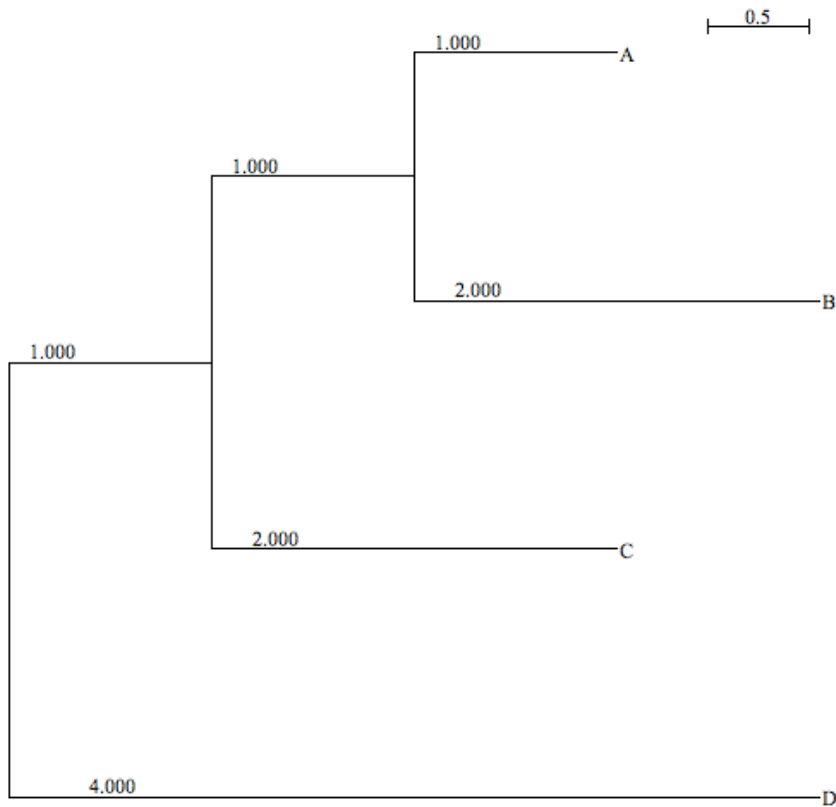
Ποιό είναι το δένδρο;

```
((A:1,B:2):1,C:2):1,D:4);
```

Newick format

$((A:1,B:2):1,C:2):1,D:4);$

distanceAC=1+1+2



Φυλογένεση γονιδίων/ πρωτεϊνών

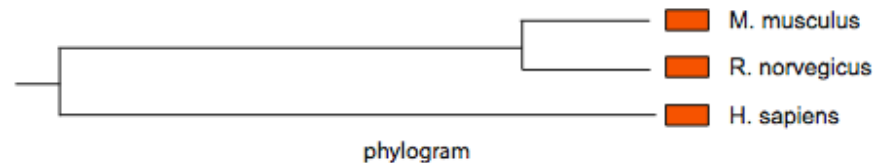
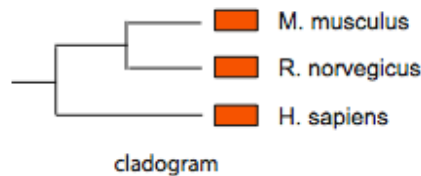
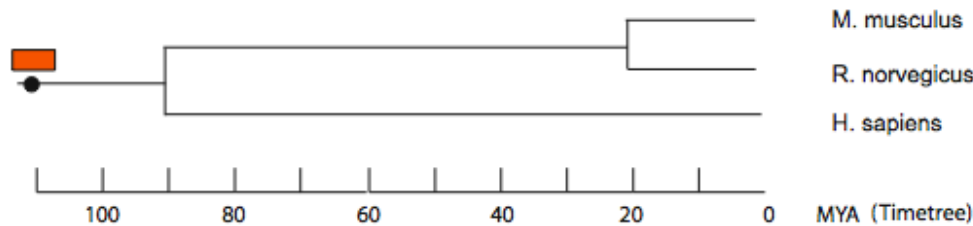
- Φυλογένεση γονιδίων ή πρωτεϊνών.
 - Δείχνει την εξελικτική πορεία μιας οικογένειας γονιδίων.
 - Κάθε κόμβος (node) στο δένδρο είναι ένας γονιδιακός διπλασιασμός ή ειδογένεση.
 - Το κάθε γονίδιο/πρωτεΐνη μπορεί να έχει διαφορετική εξελικτική πορεία (π.χ. Οριζόντια μεταφορά) ή ρυθμό εξέλιξης από τα υπόλοιπα γονίδια ενός οργανισμού.
 - Άρα, η εξελικτική πορεία ενός μόνο γονιδίου/πρωτεΐνης ενδέχεται να μην αντανακλά την εξελικτική πορεία ενός οργανισμού

Φυλογένεση οργανισμών

- Δείχνει την εξελικτική πορεία μιας ομάδας οργανισμών.
- Οι κόμβοι (nodes) στο δένδρο απεικονίζουν γεγονότα ειδογένεσης.
- Η φυλογένεση μπορεί να γίνει από:
 - μια σειρά φαινοτυπικών χαρακτήρων
 - Ένα γονίδιο μοριακό δείκτη (π.χ. 16S rRNA)
 - Μια σειρά γονιδίων
 - Από την πλειψηφία των γονιδίων του κάθε γενώματος

Φυλογένεση οργανισμών

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



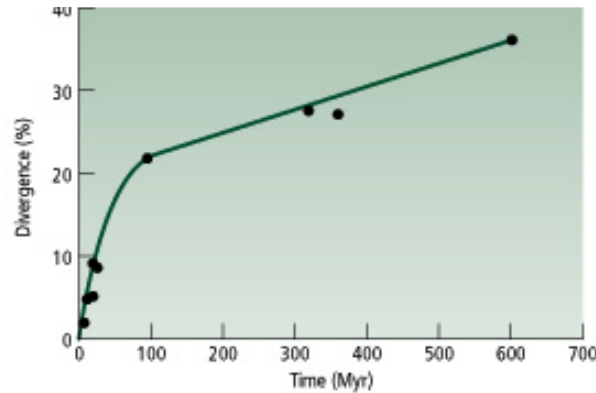
- Το ποντίκι και ο αρουραίος είχαν λιγότερο χρόνο να εξελιχθούν ξεχωριστά, από ότι ο άνθρωπος σε σχέση με το ποντίκι ή σε σχέση με τον αρουραίο. Οι μεταλλάξεις που συσσωρεύτηκαν σε κάθε ορθόλογη ακολουθία πρέπει να είναι ανάλογες του χρόνου απόκλισης των οργανισμών.
- Αν υποθέσουμε ότι ο ρυθμός μετάλλαξης είναι 1/1.000.000 χρόνια, πόσες μεταλλάξεις έχουν συσσωρευθεί σε κάθε ακολουθία, σε σχέση με τον κοινό πρόγονο;

Μοριακοί δείκτες για φυλογένεση οργανισμών

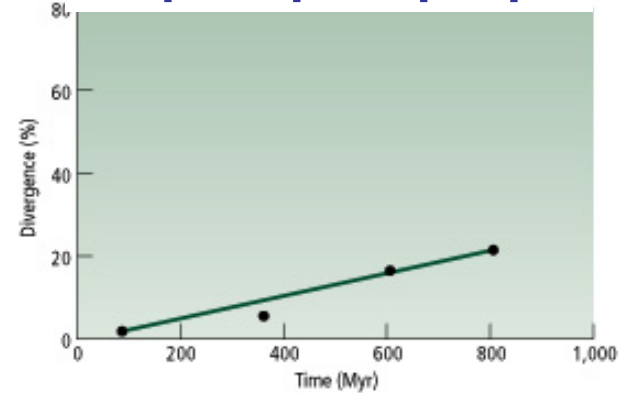
- DNA ή πρωτεΐνη, ανάλογα με την εξελικτική απόσταση των οργανισμών.
- Για πολύ 'κοντινούς' οργανισμούς:
 - Περιοχές του DNA που εξελίσσονται γρήγορα.
 - Π.χ. Για άτομα ενός ή περισσότερων πληθυσμών του ίδιου είδους, χρησιμοποιείται mtDNA που δεν κωδικοποιεί πρωτεΐνες.
- Για μέτρια αποκλίνοντες οργανισμούς:
 - rRNA ή πρωτεΐνες.
 - Mt-rRNA 10-100 MY
 - Nuc-rRNA 100-800 MY
- Για βαθιά αποκλείοντες οργανισμούς:
 - Βαθιά συντηρημένες πρωτεΐνες.

Διαφορετικά γονίδια για διαφορετικά ερωτήματα

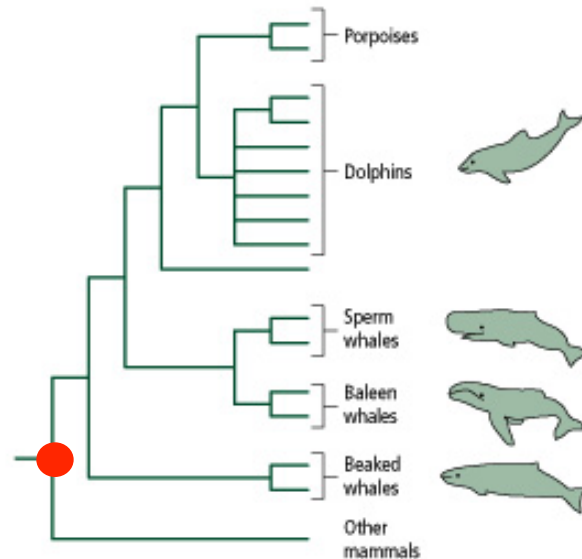
Μοριακό χρονόμετρο



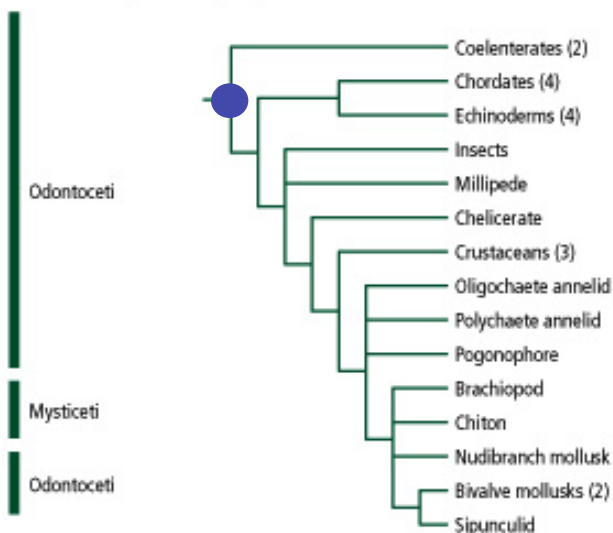
Μοριακή κλεψύδρα



(c) Cetaceans



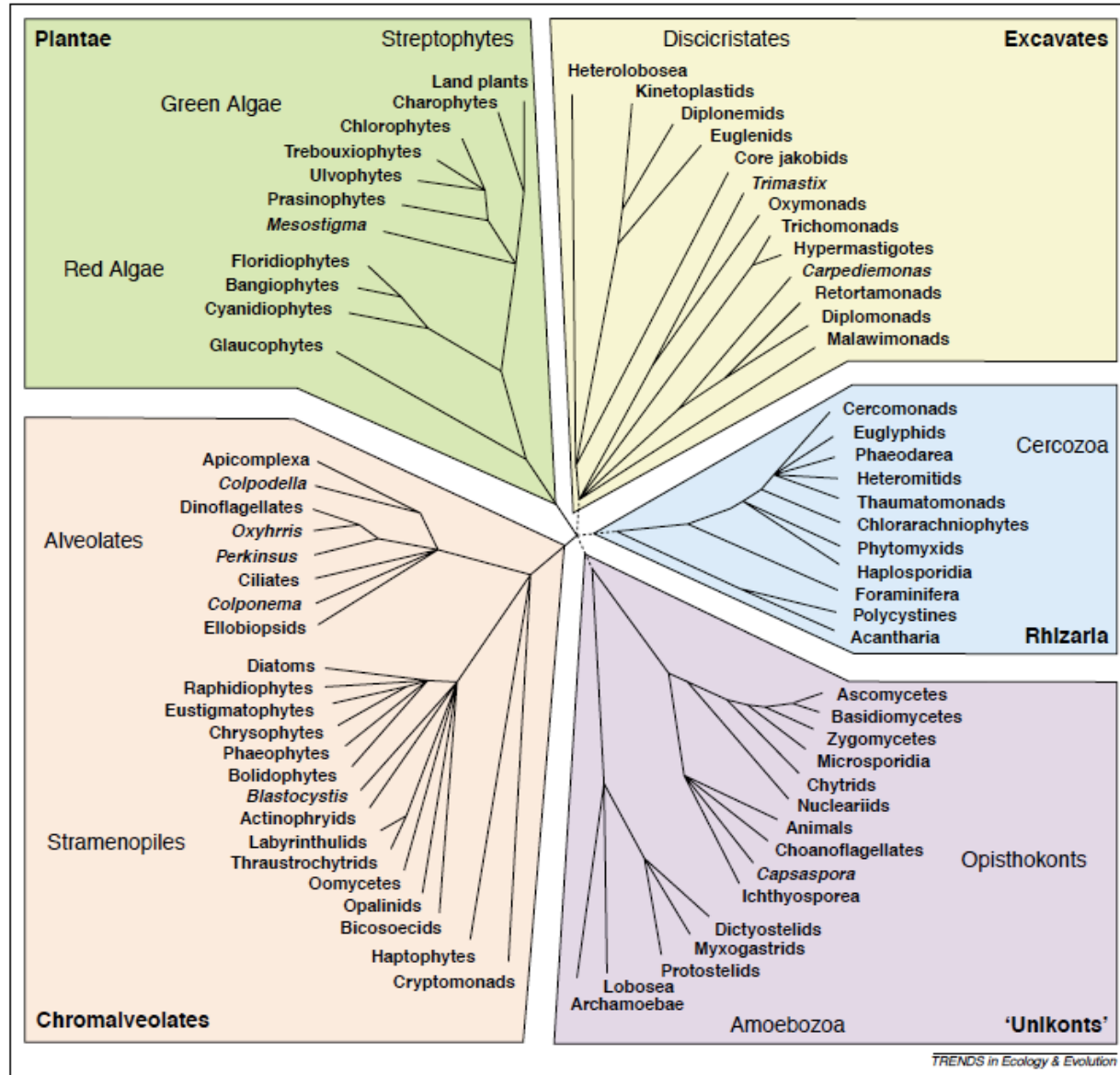
(d) Major animal groups



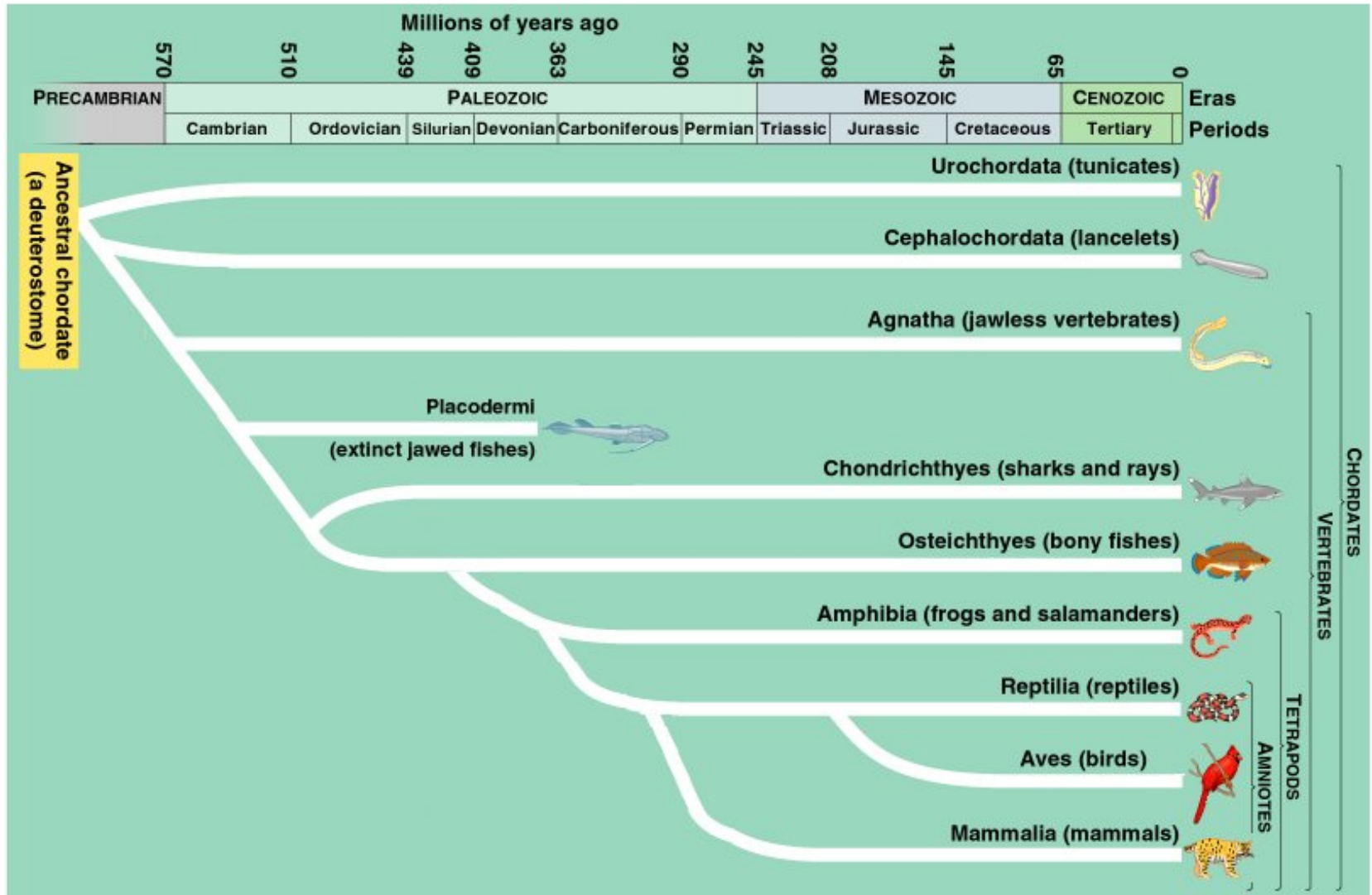
Βαθύτερη ρίζα: **35 mya (με mtRNA)**

600 mya (με πυρηνικό rRNA)

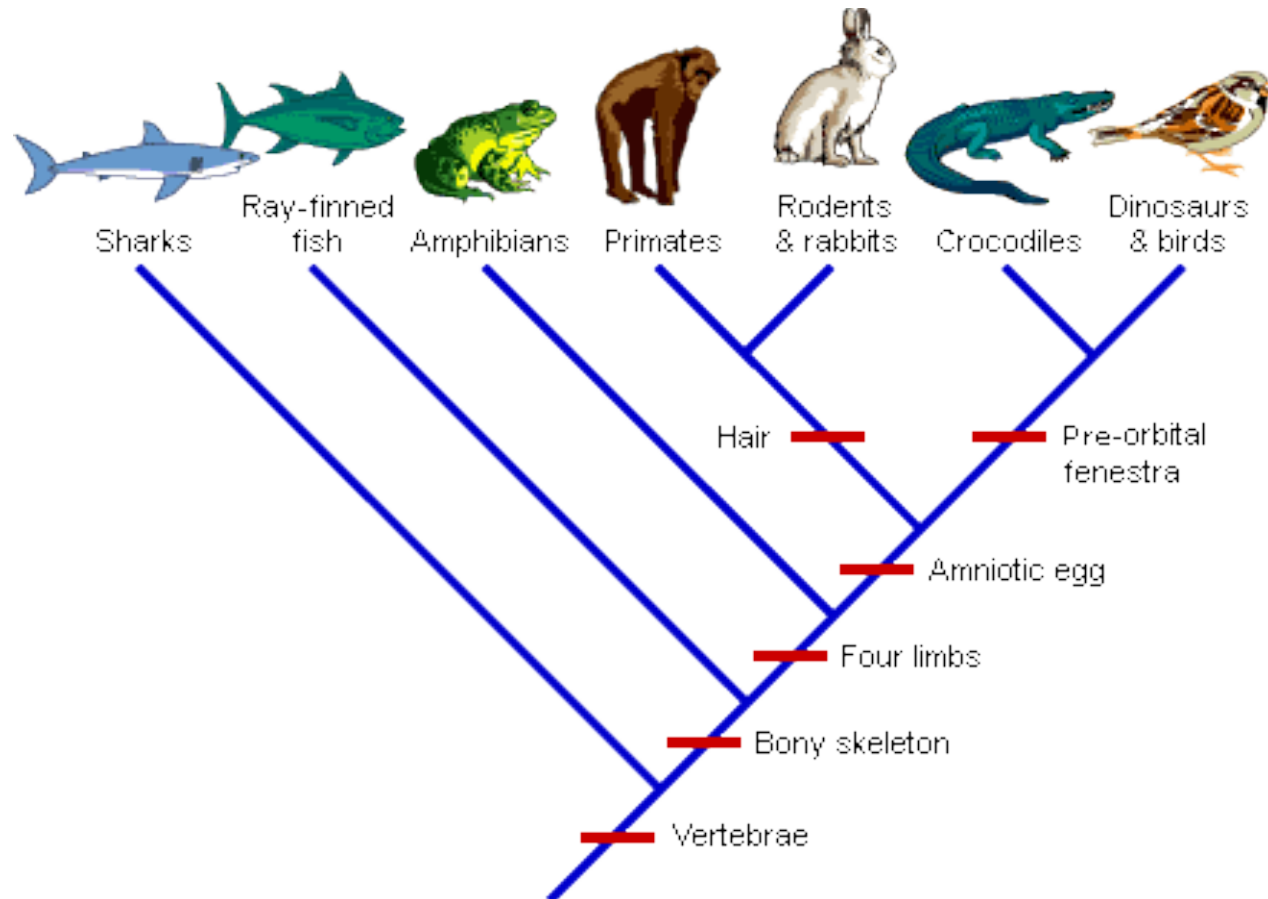
Φυλογένεση οργανισμών



Φυλογένεση χορδωτών

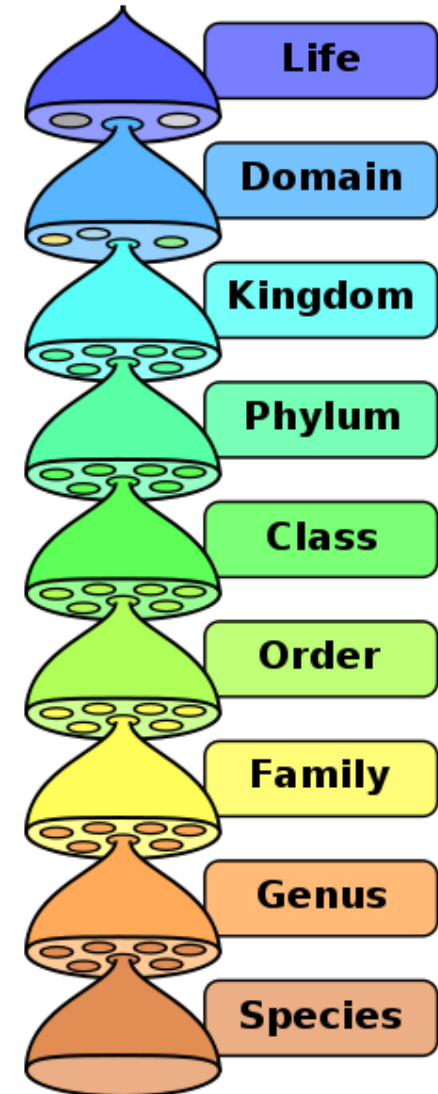


Φυλογένεση σπονδυλωτών





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.



NCBI taxonomy

Entrez
PubMed
Nucleotide
Protein
Genome
Structure
PMC
Taxonomy
BioCollections

Search for complete name lock

Display levels using filter:

Homo sapiens

Taxonomy ID: 9606 (for references in articles please use NCBI:txid9606)

current name

Homo sapiens Linnaeus, 1758

Genbank common name: **human**
 NCBI BLAST name: **primates**
 Rank: **species**
 Genetic code: [Translation table 1 \(Standard\)](#)
 Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)

Lineage (full)
[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Dipnotetrapodomorpha](#); [Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Boreoeutheria](#); [Euarchontoglires](#); [Primates](#); [Haplorrhini](#); [Simiiformes](#); [Catarrhini](#); [Hominoidea](#); [Hominidae](#); [Homininae](#); [Homo](#)

Entrez records		
Database name	Subtree links	Direct links
Nucleotide	28,545,975	28,544,573
Protein	1,970,019	1,969,460
Structure	65,719	65,715
Genome	1	1
Popset	25,639	25,638
SNP	1,116,839,409	1,116,839,409
Conserved Domains	83	83
GEO Datasets	3,852,164	3,852,164
PubMed Central	59,963	59,882
Gene	359,883	359,810
HomoloGene	18,713	18,713
SRA Experiments	4,889,983	4,888,879
GEO Profiles	61,958,910	61,958,910
Protein Clusters	13	13
Identical Protein Groups	1,895,122	1,895,029
BioProject	108,744	108,706
BioSample	9,192,551	9,192,289
Assembly	1,279	1,279
dbVar	8,129,185	8,127,852
Genetic Testing Registry	73,886	73,886
PubChem BioAssay	761,628	761,621
Taxonomy	3	1

Timetree

<http://www.timetree.org/>



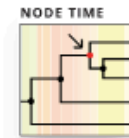
Timetree of Life

TimeTree is a public knowledge-base for information on the evolutionary timescale of life. Data from thousands of published studies are assembled into a searchable tree of life scaled to time.

Timepanels showing events in geological time and astronomical history are provided for comparison with timelines and timetrees. Results can be exported in different formats for additional analyses and publication.

Search

Get Divergence Time



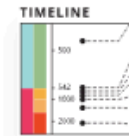
Specify 2 Taxon Names [?]

Taxon 1:

Taxon 2:

Search

Get an Evolutionary Timeline

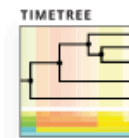


Specify a Taxon Name [?]

Taxon:

Search

Build a Timetree



Specify a Group of Taxa [?]

Group:

Search

or

Load a List of Species [?]

Upload:

No file selected.

Upload

Timetree

<http://www.timetree.org/>

- How It Works
- Download CSV

Specify 2 Taxon Names

Taxon 1:

Taxon 2:

Search

Resolve Ambiguity

Taxon 1:

Taxon 2:

Show Time

Climate Change

- Solar Luminosity
- CO₂ Levels
- O₂ Levels

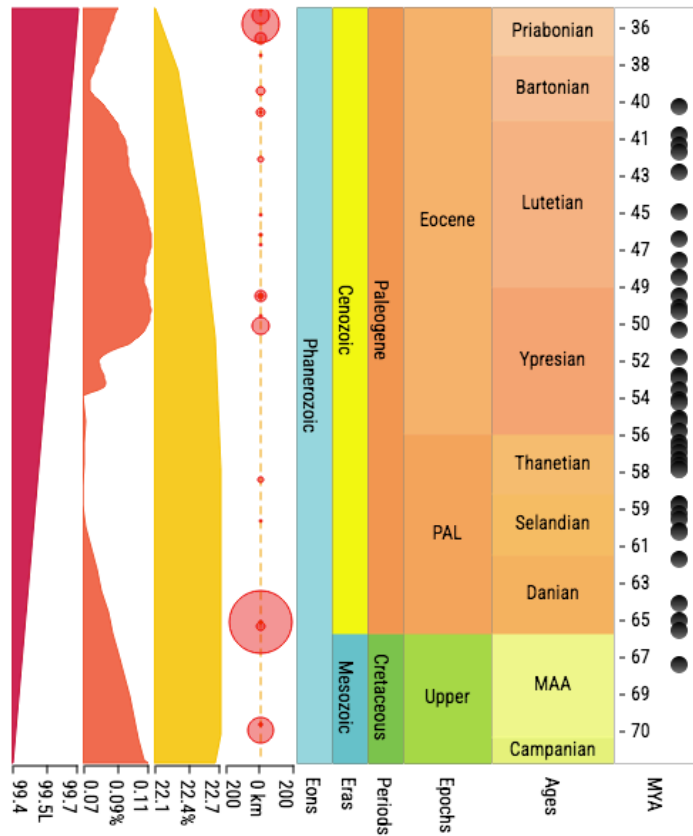
Display

-
-
-
-
-

Export Results

Select Image Format

To Image File



domestic cat

Felis catus

Versus

dog

Canis lupus familiaris

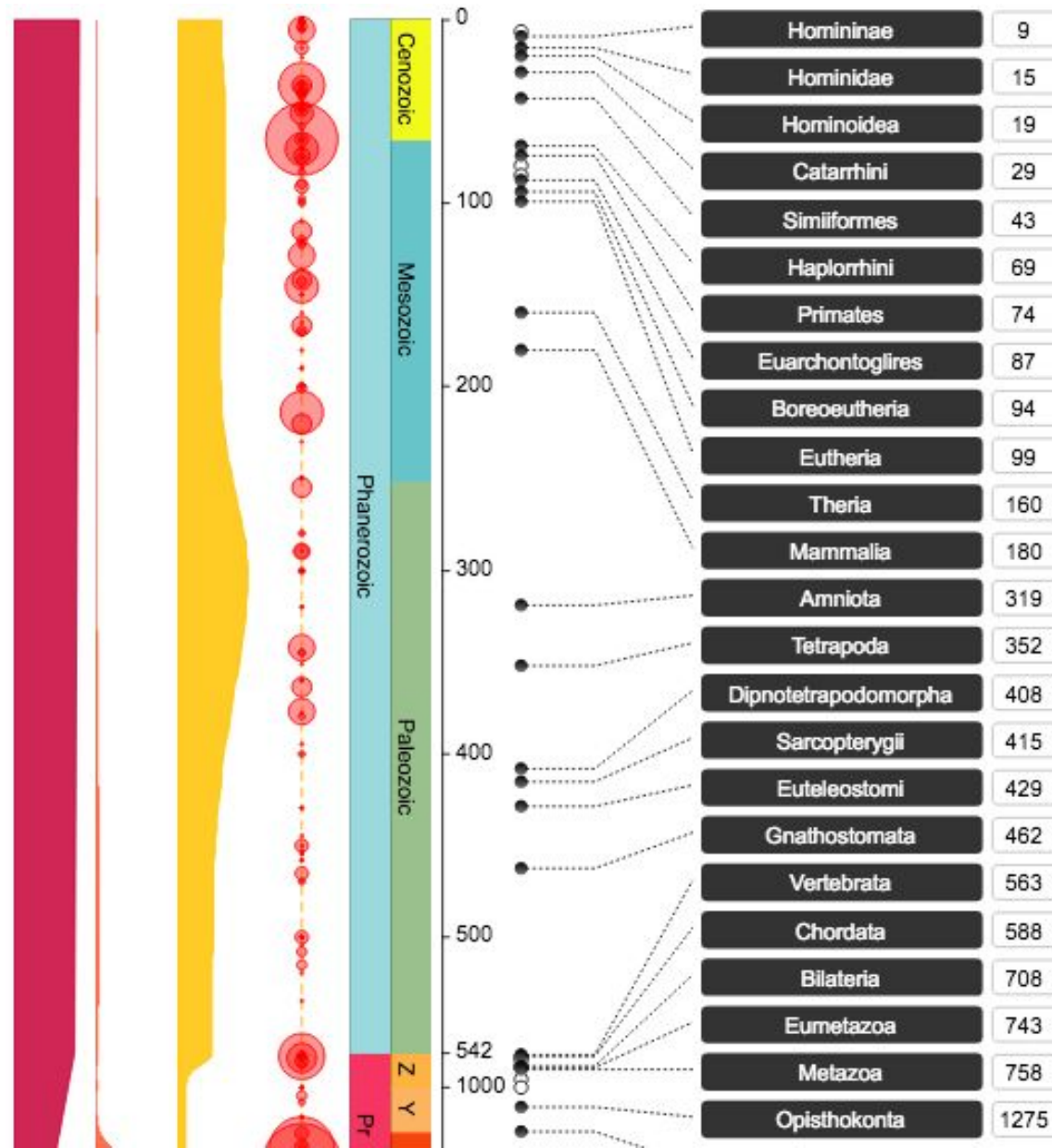
Median Time:
55 MYA
CI: (52.9 - 57.3 MYA)

Adjusted Time:
55 MYA

citation

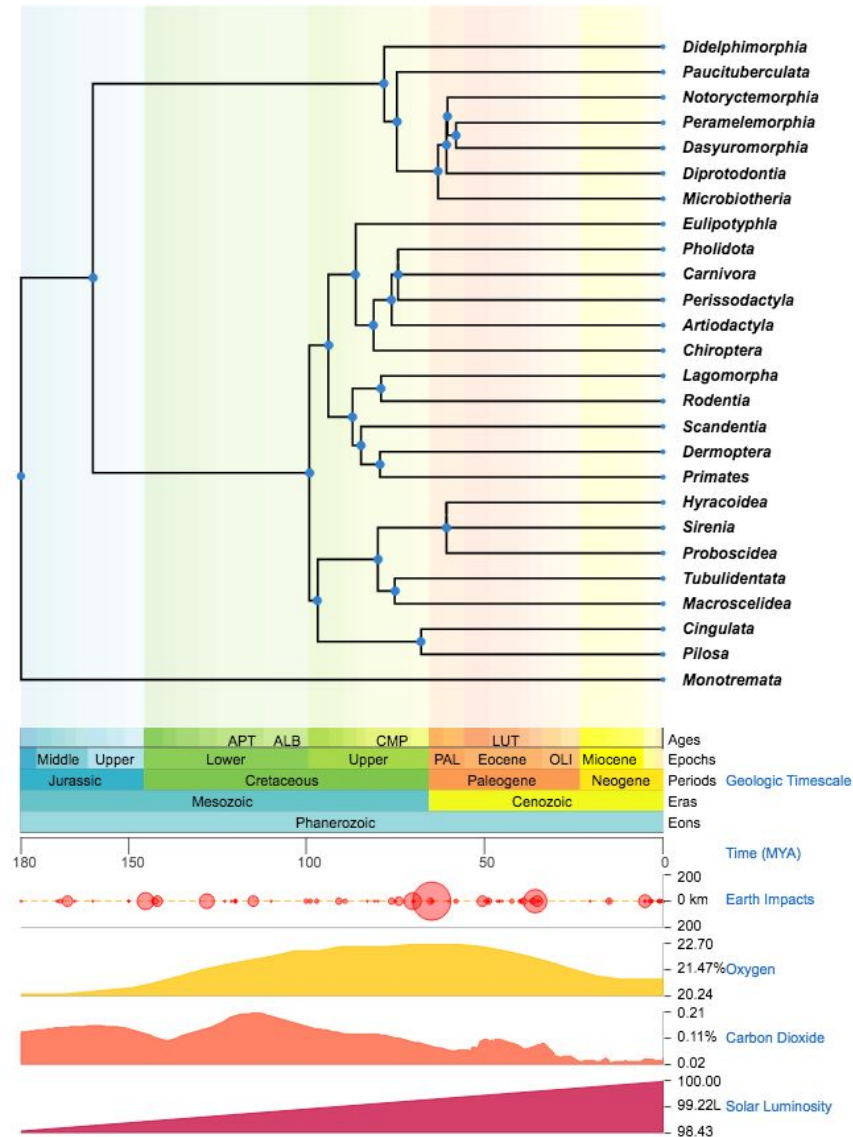
(Median and adjusted times were derived from 42 studies)

Timetree Evolutionary Timeline for "*Homo sapiens*"



Timetree of order mammalia

<http://www.timetree.org/>



Timetree



TIMETREE is a public resource for knowledge on the timescale and evolutionary history of life. Search the database below or go to the TIMETREE OF LIFE for other resources



TIMETREE
OF LIFE
BOOK

ABOUT SEARCH BOOK RESOURCES NEWS FAQs CONTACT

CONSORTIUM

CAUTION: These results are sensitive to both the classification (taxonomic names) and the "guide tree" used. For both, TimeTree uses [NCBI's Taxonomy Browser](#), which will yield a reliable result in most cases. However, NCBI may not use the classification you believe to be correct, or that was used by a study being queried, in which case the time estimates presented here may be unusable. Other options will be available in the future. Until then, check the individual study results at the bottom of the page for proper taxonomy before using summary information. If in doubt, consult [The Timetree of Life](#) or original literature.

Summary Information

Query Taxa: [Rattus/Mus musculus](#)

Result Comparison [Rattus/Mus](#)

Study	Weighted Average (#genes)	Simple Average
All (43)	36.8 Mya	26.1 Mya
Nuclear (25)	36.0 Mya	27.2 Mya
Mitochondrial (13)	35.8 Mya	26.7 Mya
Mixed (5)	19.4 Mya	19.0 Mya

Molecular Time Estimates

(values for estimates of standard error/deviation or confidence interval can be seen by clicking on individual times)

*Note: If you would like to suggest a publication for inclusion please [send us](#) a PubMed ID or PDF file.

Gene type	Timings	Publication Year	Source	Publication Title	PubMed
Mitochondrial	2	2007		Phylogenetic analyses of complete mi ...	Horer et al.
Time	# Genes	Taxon A	Taxon B	Data Type	Source
15.1	12	Rattus	Mus	Amino	table 3
15.9	12	Rattus	Mus	Nucleotide	table 3
Mixed	1	2007		Multiple molecular evidences for a l ...	Huchon et al.
Time	# Genes	Taxon A	Taxon B	Data Type	Source
16.4	6	Rattus	Mus	Nucleotide	figure s1
Mixed	1	2007		Using genomic data to unravel the ro ...	Murphy et al.
Time	# Genes	Taxon A	Taxon B	Data Type	Source
15.9	19	Rattus	Mus	Nucleotide	Supp Table 2
Nuclear	1	2007		Calibration choice, rate smoothing, ...	Hugall et al.
Time	# Genes	Taxon A	Taxon B	Data Type	Source
27.0	1	Rattus	Mus	Nucleotide	table 3
Nuclear	1	2006		The pattern and timing of diversific ...	Jansa et al.
Time	# Genes	Taxon A	Taxon B	Data Type	Source
18.7	1	Rattus	Mus	Nucleotide	Table 5
Mitochondrial	1	2006		A mitogenomic timescale for birds de ...	Pereira et al.
Time	# Genes	Taxon A	Taxon B	Data Type	Source
53.6	37	Rattus	Mus	Nucleotide	table 1
Nuclear	1	2006		Housekeeping genes for phylogenetic ...	Kullberg et al.
Time	# Genes	Taxon A	Taxon B	Data Type	Source
43.0	8	Rattus norvegicus	Mus musculus	Amino	table 5
Mixed	1	2006		Molecular phylogeny of the Cricetina ...	Neumann et al.
Time	# Genes	Taxon A	Taxon B	Data Type	Source
12.9	3	Rattus	Mus	Nucleotide	Table 3

Ιστότοποι/ΒΔ για βιοποικιλότητα

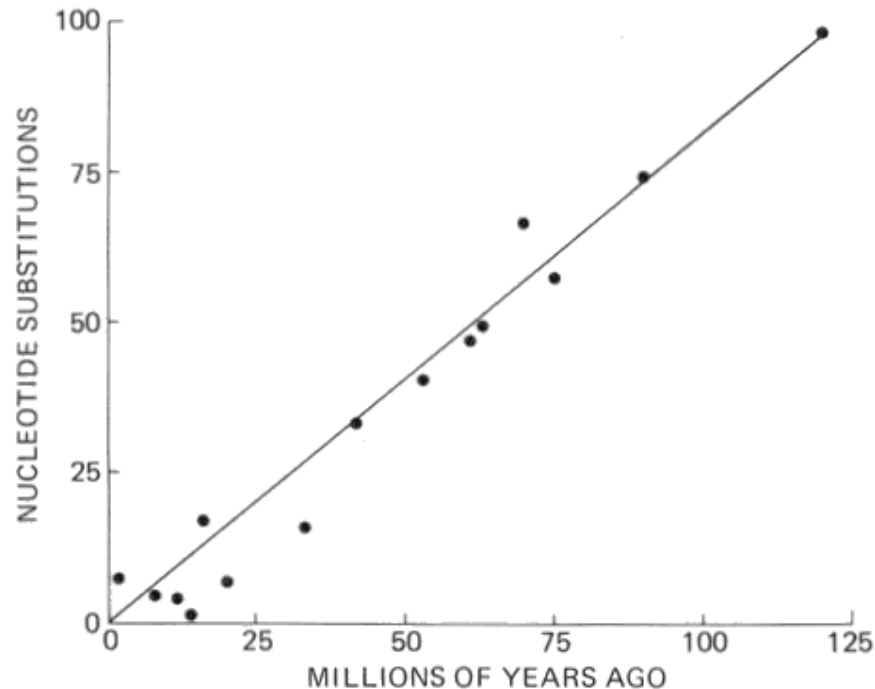
<http://www.timetree.org/resources.php>

Web Resources

Biodiversity Sites

- | | |
|--|--|
| ➤ AmphibiaWeb | An online system to search and retrieve information relating to amphibian biology and conservation. |
| ➤ Animal Diversity | An online database of animal natural history, distribution, classification, and conservation biology. |
| ➤ AntWeb | An advanced biodiversity information system at species level dedicated to ants. |
| ➤ Arkive | A unique collection of thousands of videos, images and fact-files illustrating the world's species. |
| ➤ Catalogue of Life | An informal partnership dedicated to creating an index of the world's organisms. |
| ➤ Encyclopedia of Life | Information about all known species, including their taxonomy, geographic distribution, collections, genetics, evolutionary history, morphology, behavior, ecological relationships, etc.. |
| ➤ FishBase | A global information system about fishes. |
| ➤ Global Biodiversity Information Facility | A source for information on biological specimen and observational data with access to more than 135 million data records from around the world. |
| ➤ IUCN Redlist | A comprehensive inventory of the global conservation status of plant and animal species. |
| ➤ Micro*scope | Descriptive information and images about all kinds of microbes. |
| ➤ Taxonomy Browser | The names of all organisms that are represented in the genetic databases with at least one nucleotide or protein sequence, arranged hierarchically. |

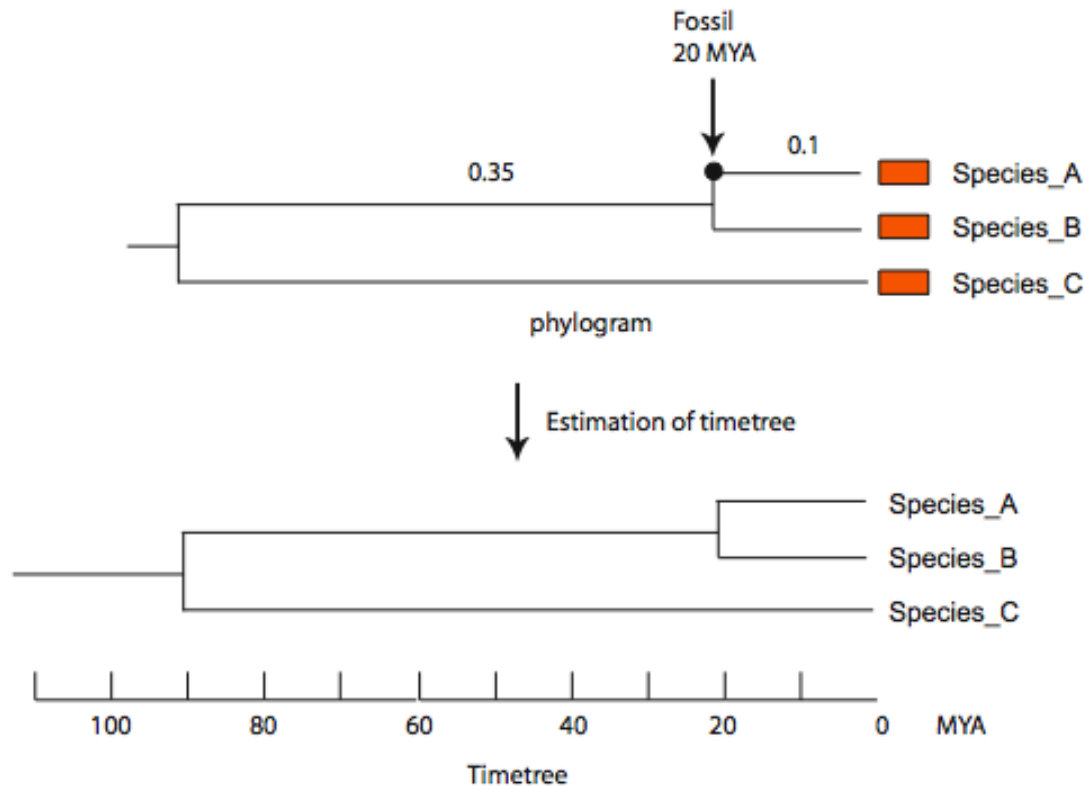
Μοριακά ρολόγια



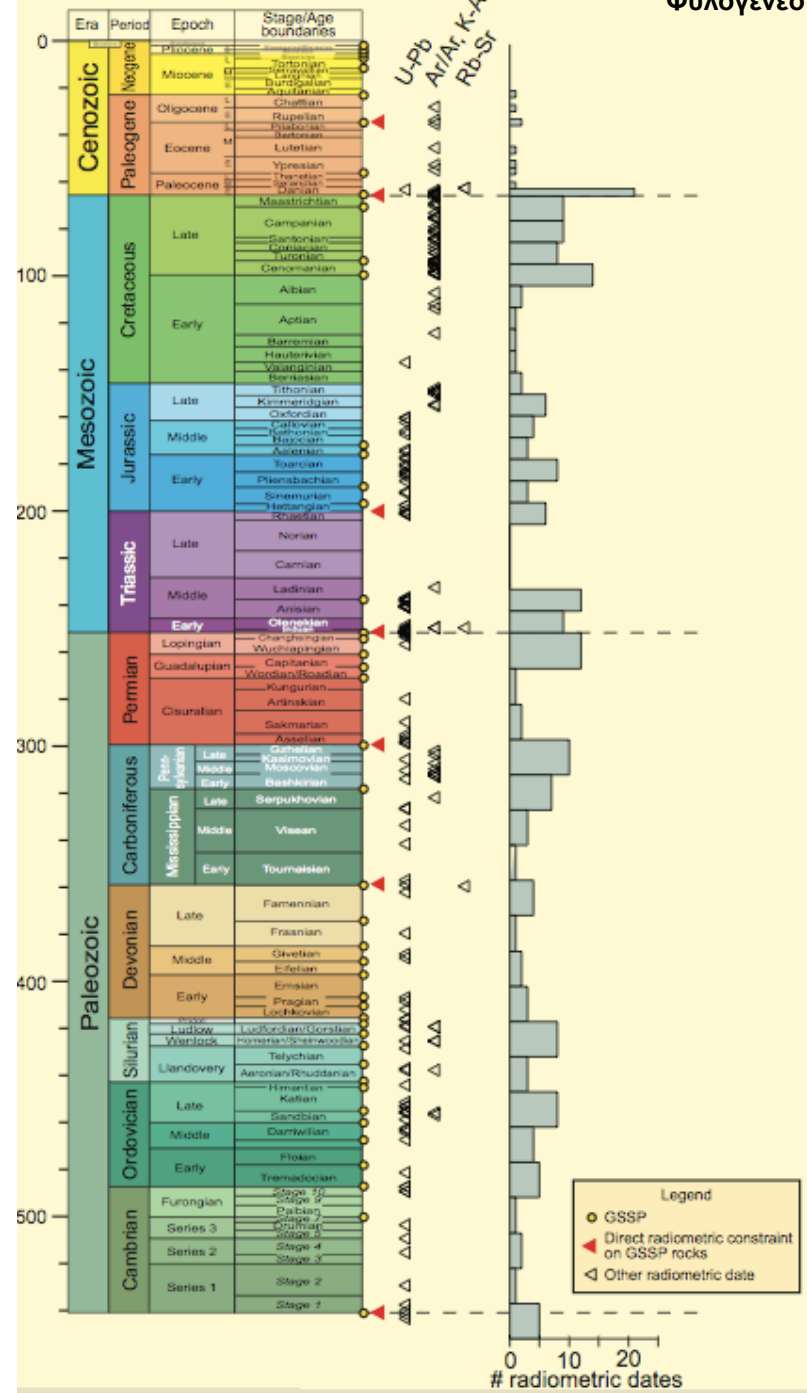
- Inferred pairwise nucleotide substitutions among 17 mammal species from seven gene products, as estimated from protein studies, plotted against date of divergence, as estimated from the fossil record. The line is drawn from the origin through the oldest comparison (the marsupial / placental divergence at 125 MYBP). The strong linear relationship suggests that **molecular differences between pairs of species are proportional to the time of their separation**, rather than the degree of organismal difference. Therefore, measures of genetic divergence can be used to date the time of divergence for species pairs for which no fossil data are available: genes function as **Molecular Clocks**. (from **A. C. Wilson** 1976).

Μοριακά ρολόγια

- Υποθέτουν ένα σταθερό ρυθμό μεταλλάξεων για μια εξελικτική γραμμή.
- Χρονολογημένα απολιθώματα επιτρέπουν την 'ζυγοστάθμιση' (calibration) του μοριακού ρολογιού.
- Ρυθμός εξέλιξης διαφέρει μεταξύ γονιδίων
- Στην πραγματικότητα, ο ρυθμός εξέλιξης μπορεί να μεταβάλλεται (επιβράδυνση ή επιτάχυνση) κατά τη διάρκεια εξέλιξης ενός κλάδου



Γεωλογικές περίοδοι



Απολιθώματα που χρησιμοποιούνται για ζυγοστάθμιση του μοριακού ρολογιού

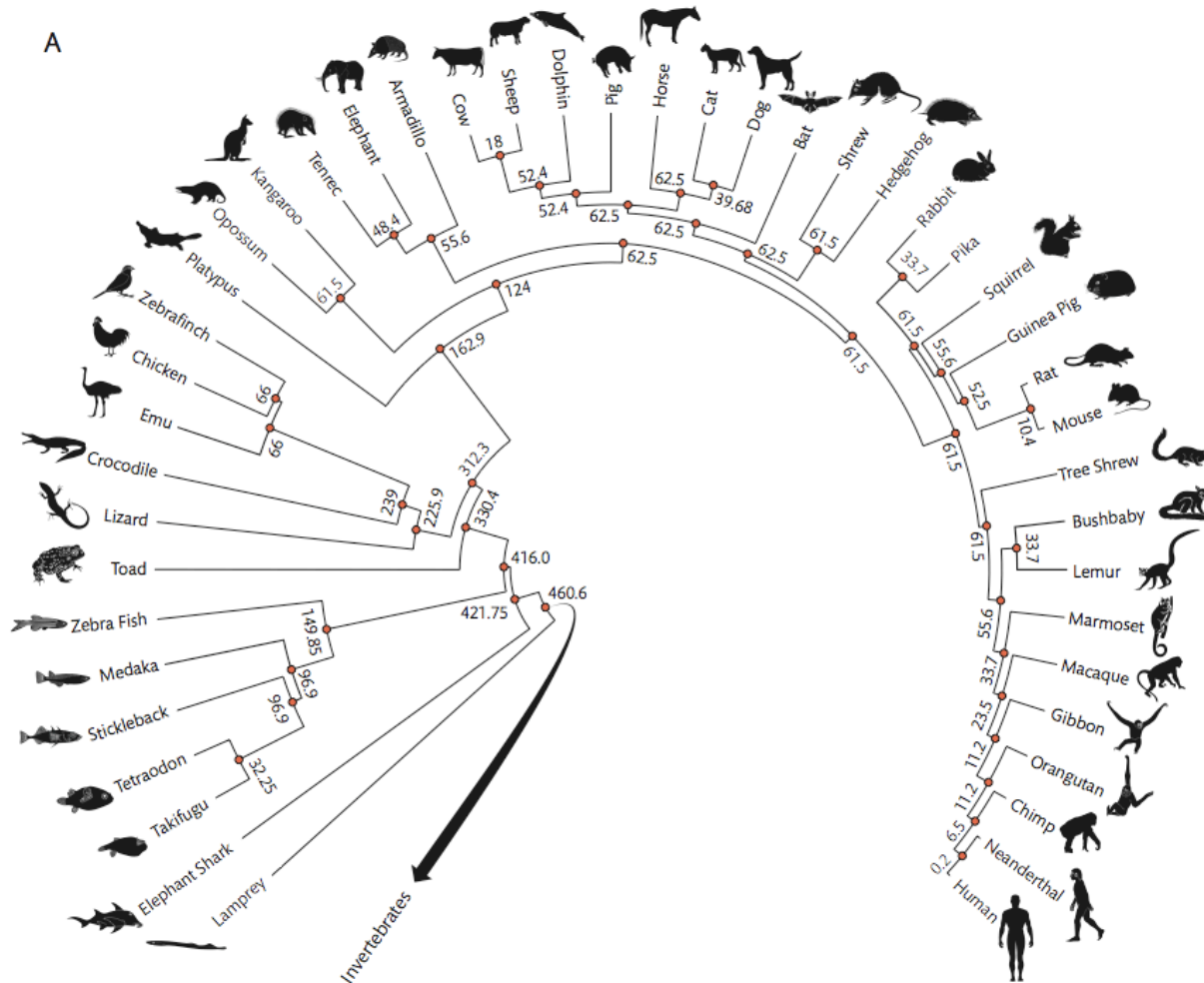


Fig. 3 Continues

Απολιθώματα που χρησιμοποιούνται για ζυγοστάθμιση του μοριακού ρολογιού

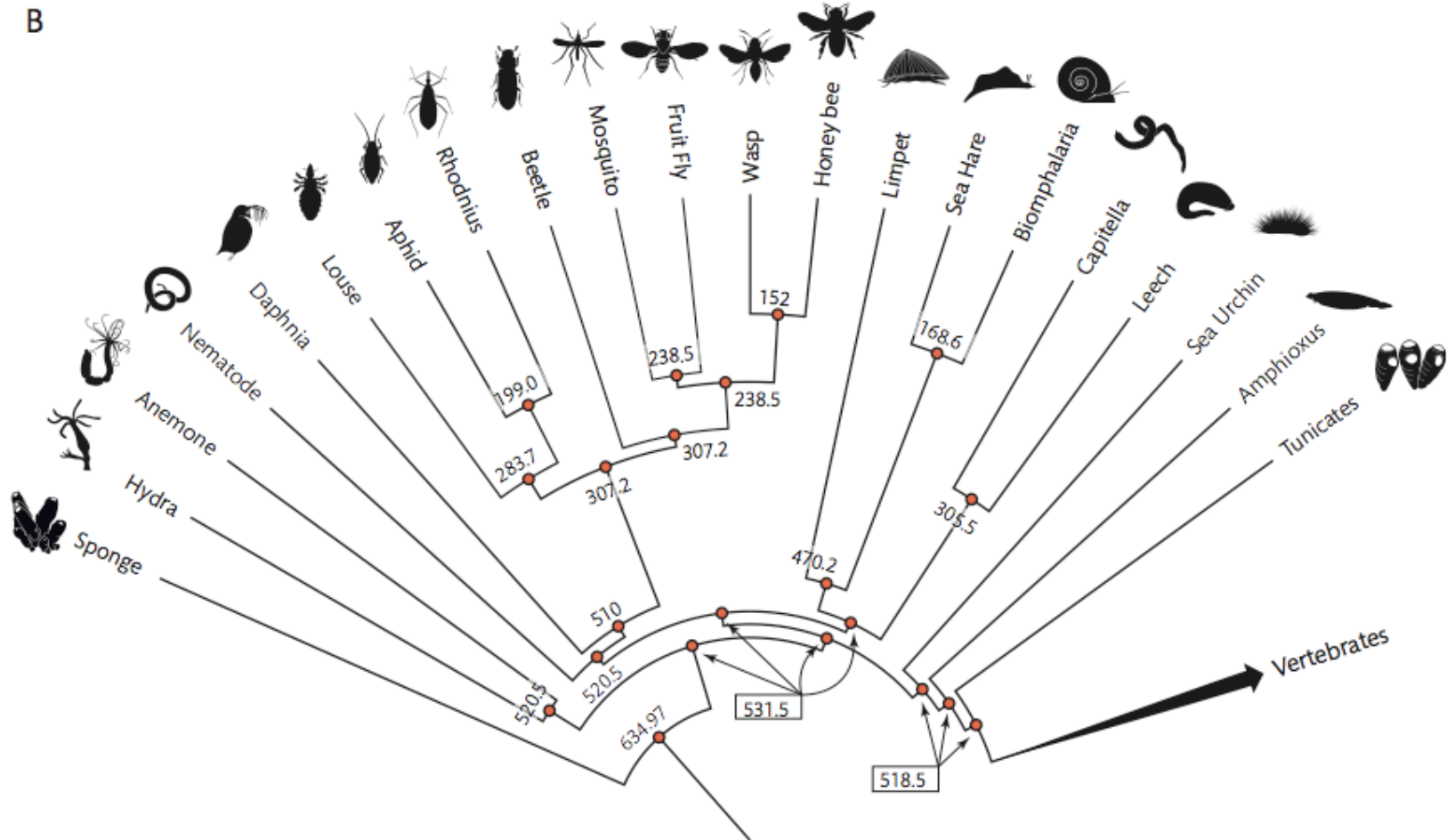
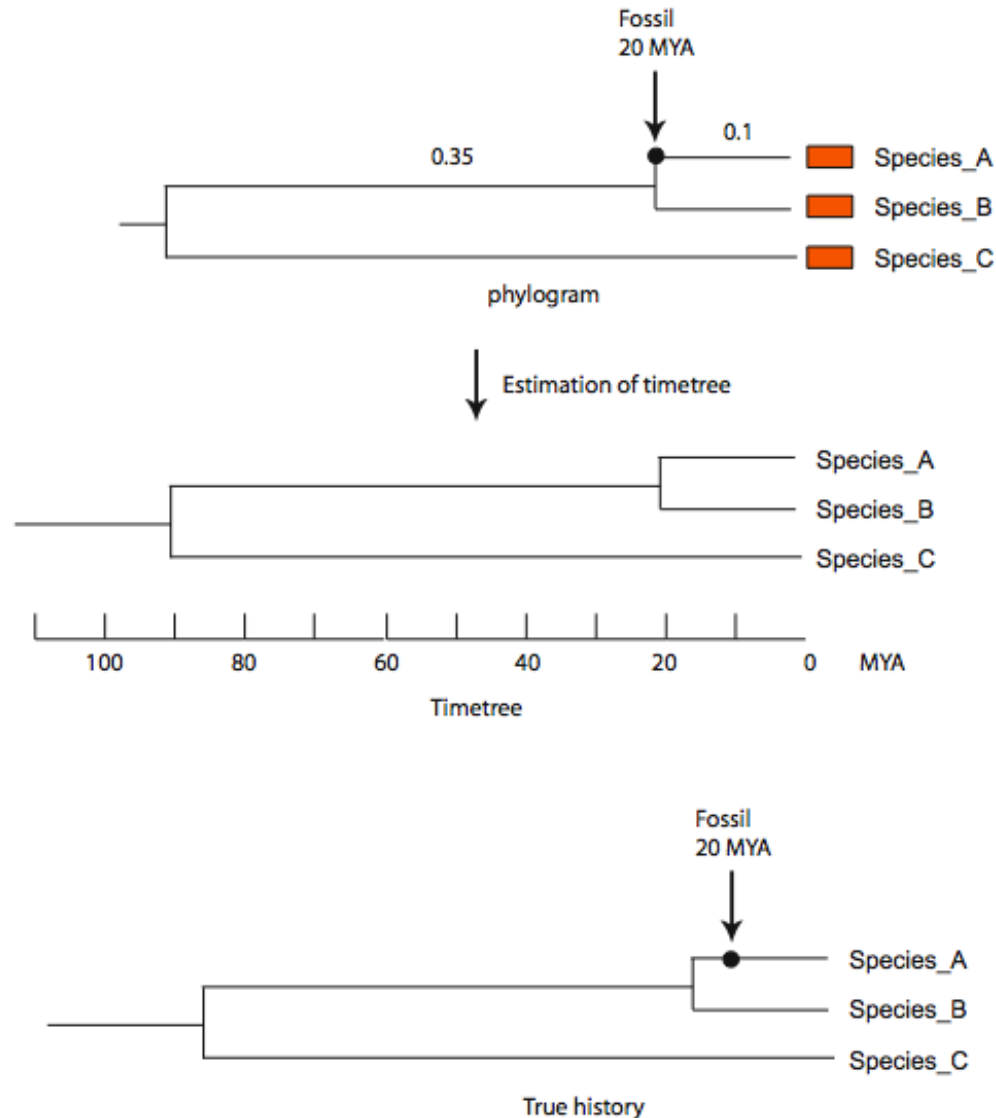


Fig. 3 Summary of paleontological constraints (in myr) on metazoan phylogeny. A. Vertebrates. B. Invertebrates. The dates are minimum constraints, and soft maximum constraints are given in the text, and at <http://www.fossilrecord.net/>.

Προβλήματα με τα απολιθώματα

- Χρονολογημένα απολιθώματα υποτιμούν τον χρόνο εμφάνισης μιας εξελικτικής γραμμής:
 - Όταν συνέβη η απολίθωση, ο κοινός πρόγονος ήταν ήδη παρόν για κάποιο χρονικό διάστημα.
 - Ο κοινός πρόγονος χρειάστηκε κάποιο χρόνο από την στιγμή της εμφάνισης του μέχρι να εξαπλωθεί γεωγραφικά και να βρίσκεται σε αφθονία, ώστε ένα τυχαίο γεγονός να οδηγήσει στην απολίθωση κάποιων ατόμων.
 - Τα απολιθώματα αποτελούν το κατώτατο χρονικό όριο εμφάνισης μιας εξελικτικής γραμμής

Χρονολογημένα απολιθώματα υποτιμούν τον χρόνο εμφάνισης μιας εξελικτικής γραμμής



Το πρόβλημα της Κάμβριας περιόδου

- Από τα απολιθώματα συμπεραίνουμε ότι οι περισσότερες και κυριότερες εξελικτικές γραμμές των μετάζωων εμφανίστηκαν ξαφνικά, σε μια σύντομη περίοδο πριν ~550 εκ. χρόνια, μεταξύ της προ-κάμβριας και κάμβριας περιόδου (Κάμβρια έκρηξη).
- Μελέτες με μοριακά ρολόγια δείχνουν ότι αυτές οι εξελικτικές γραμμές εμφανίστηκαν πολλές εκατοντάδες εκατομύρια χρόνια πιο πριν!

Το πρόβλημα της Κάμβριας περιόδου

- <http://www.sciencedirect.com/science/article/pii/S0960982213009160>
- Από τα απολιθώματα συμπεραίνουμε ότι τα περισσότερα εξελικτικά φύλα (κλάσεις) που συναντάμε σήμερα εμφανίστηκαν μέσα σε ένα εξαιρετικά σύντομο χρονικό διάστημα (~542 - 521 εκατομύρια χρόνια πριν), στην γεωλογική περίοδο του Κάμβριου. Το εξελικτικό αυτό γεγονός ονομάστηκε «έκρηξη του Κάμβριου» (Cambrian Explosion).
- Αυτή η φαινομενικά ταχύτατη εμφάνιση τόσων πολλών και διαφορετικών φύλων αποτελεί ένα παράδοξο στην εξελικτική θεωρία, που βασίζεται στην ιδέα των σταδιακών αλλαγών διά της φυσικής επιλογής (Ο Δαρβίνος προβληματίστηκε - χρησιμοποιείται και ως επιχείρημα από τους πολέμιους της εξέλιξης).
- Φυλογενετικές αναλύσεις από γενετικά και μορφολογικά δεδομένα με την μέθοδο του χαλαρού μοριακού ρολογιού δείχνουν ότι την περίοδο εκείνη ο ρυθμός εξέλιξης ήταν 4-5 φορές ταχύτερος από άλλες περιόδους.

Απολιθώματα και μοριακή χρονολόγηση

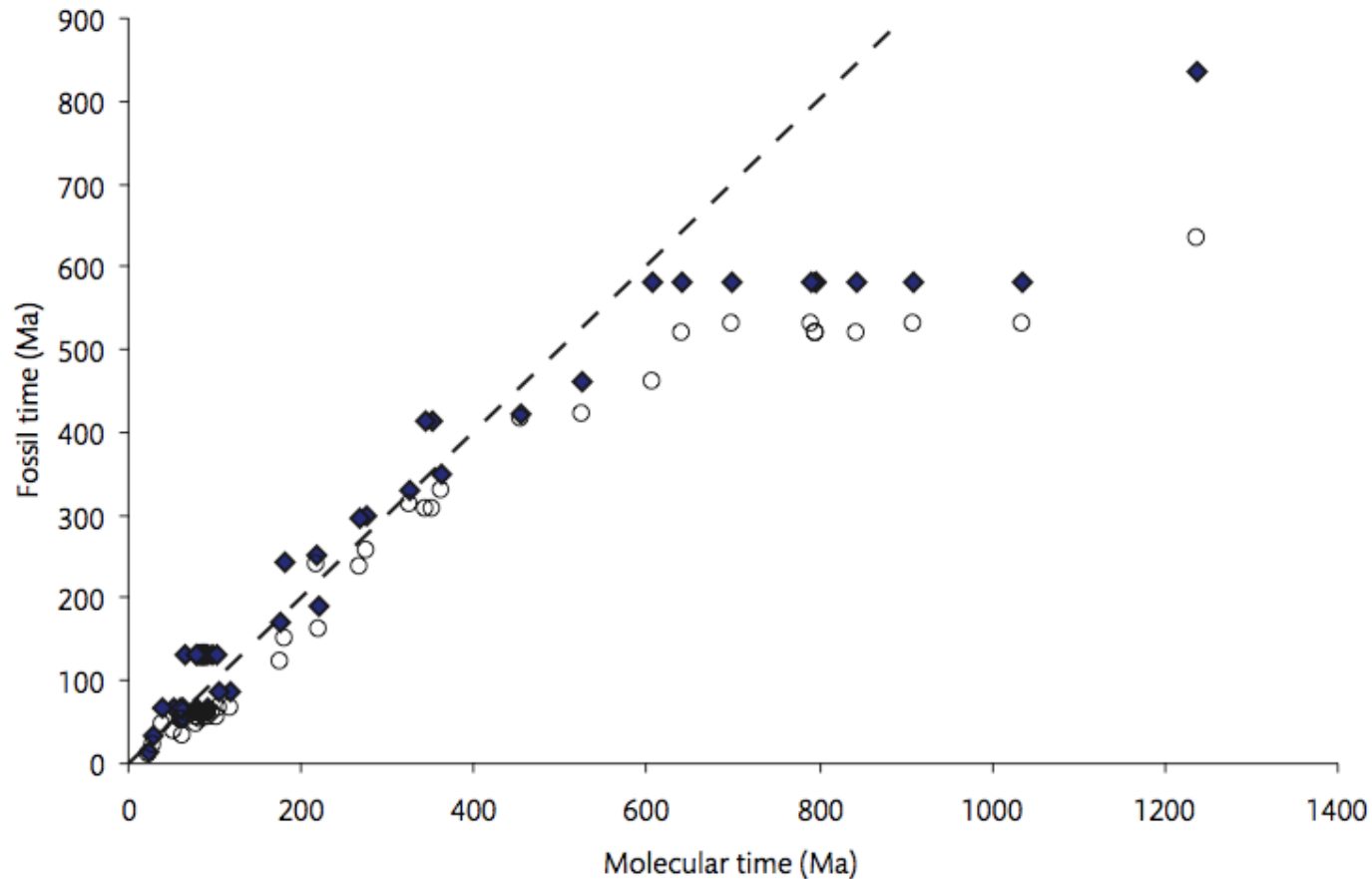


Fig. 3 Relationship between molecular time estimates (x-axis), based on data from *The Timetree of Life* (78) and fossil time estimates (y-axis) (55). Open circles = fossil minimum times; closed diamonds = fossil soft maximum times; dashed line = 1:1 relationship; Ma = million years ago.

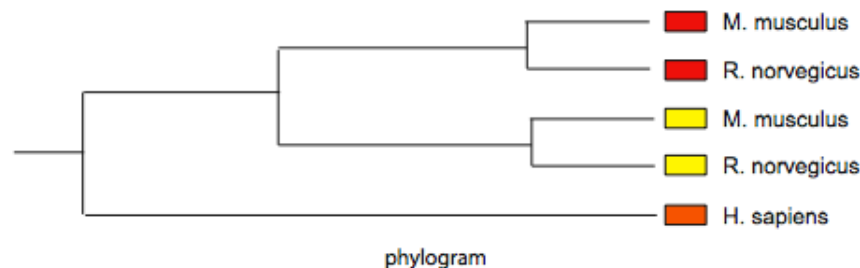
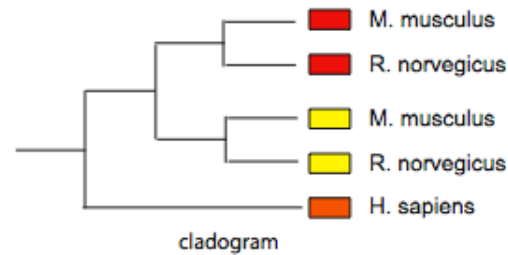
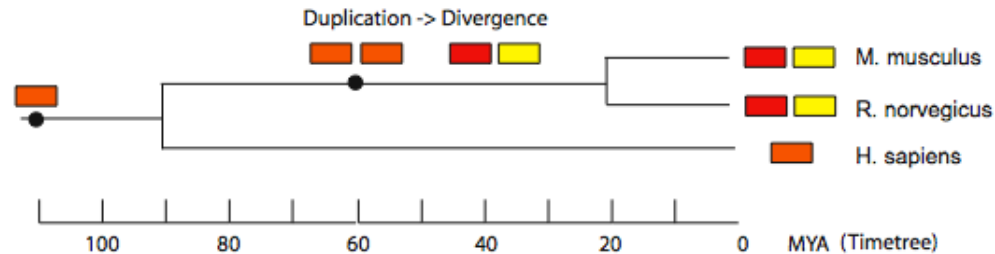
Το πρόβλημα της Κάμβριας περιόδου

- Είναι δυνατόν αβιοτικοί παράγοντες να επηρέασαν τη διαδικασία απολίθωσης.
- Το οξυγόνο αυξήθηκε δραματικά και πλησίασε στα σημερινά του επίπεδα την περίοδο του Κάμβριου (580-542 ΜΥΑ).
- Μεγαλύτερη συγκέντρωση οξυγόνου επέτρεψε τους οργανισμούς να αυξήσουν το μέγεθός τους.
- Μεγαλύτερο μέγεθος χρειάστηκε δομές που να συγκρατούν το βάρος. Αυτές οι νέες δομές μπορεί να ήταν πιο εύκολο να διατηρηθούν ως απολιθώματα.
- Πιθανόν, η μορφολογία των οργανισμών στην προ-κάμβρια περίοδο να μην επέτρεπε την απολίθωσή τους.

Φυλογένεση γονιδίων

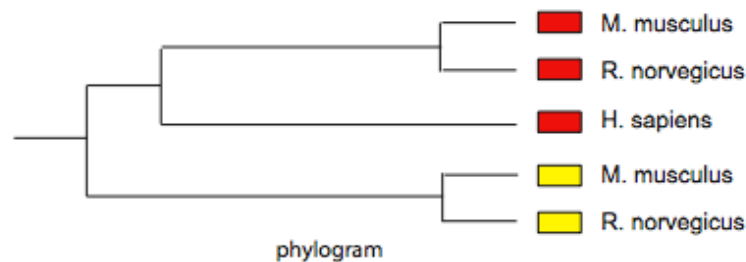
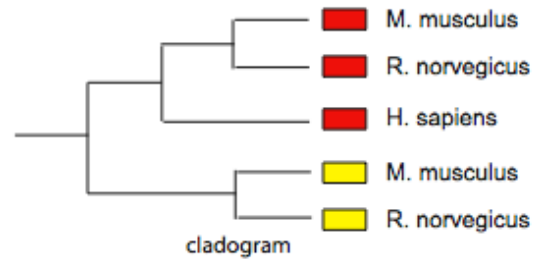
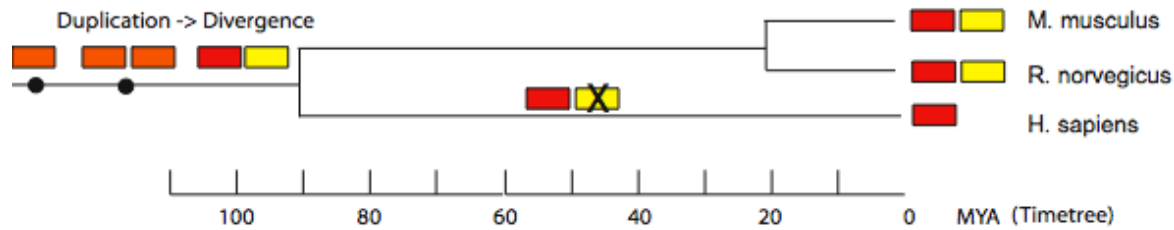
Βρίσκουμε τις ομόλογες ακολουθίες στους οργανισμούς που μας ενδιαφέρουν και ακολουθεί φυλογένεση, για να καταλάβουμε πότε συνέβησαν οι γονιδιακοί διπλασιασμοί, και ποιά ομόλογα είναι πιο κοντινά μεταξύ τους.

Πρέπει να γνωρίζουμε τις εξελικτικές σχέσεις των οργανισμών



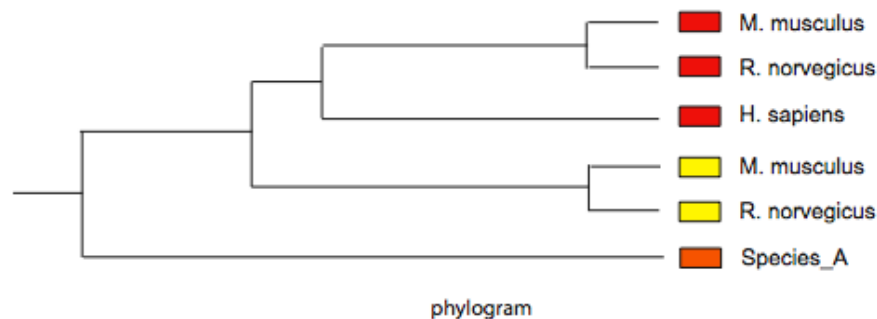
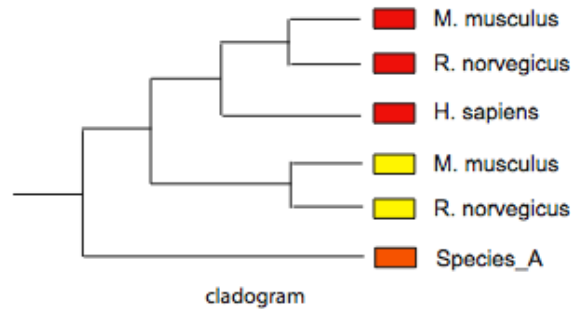
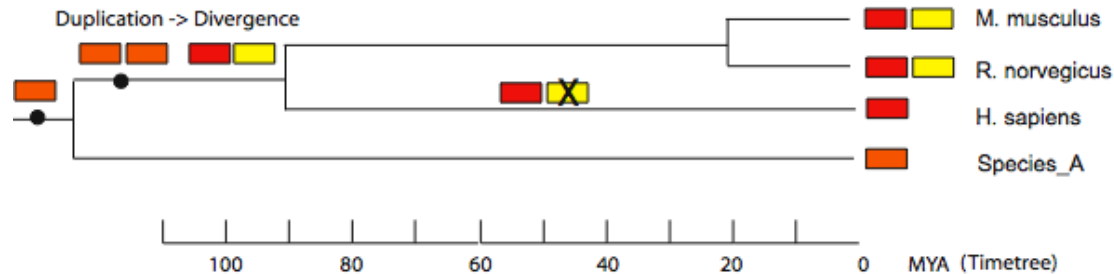
Φυλογένεση γονιδίων

- Απώλεια αντίγραφου



Φυλογένεση γονιδίων

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



Δένδρα συναίνεσης

- Μια μέθοδος μπορεί να οδηγήσει σε περισσότερα από ένα εξίσου καλά δένδρα.
- Ή, από τα ίδια δεδομένα, δημιουργούνται δένδρα με διαφορετικές μεθόδους.
- Το δένδρο συναίνεσης δείχνει ποιοί κόμβοι είναι κοινοί μεταξύ των διαφόρων δένδρων.
- Για κόμβους που δεν παρατηρείται συμφωνία, εμφανίζονται ως πολυτομημένοι.
- Μέθοδοι δημιουργίας δένδρου συναίνεσης:
 - απόλυτη συναίνεση (strict consensus) (100%)
 - Μέθοδος πλειοψηφίας (majority rule) (>50%)

Δένδρα συναίνεσης

- Το παράδειγμα της φυλογενετικής σχέσης ανθρώπου-χιμπατζή-γορίλα

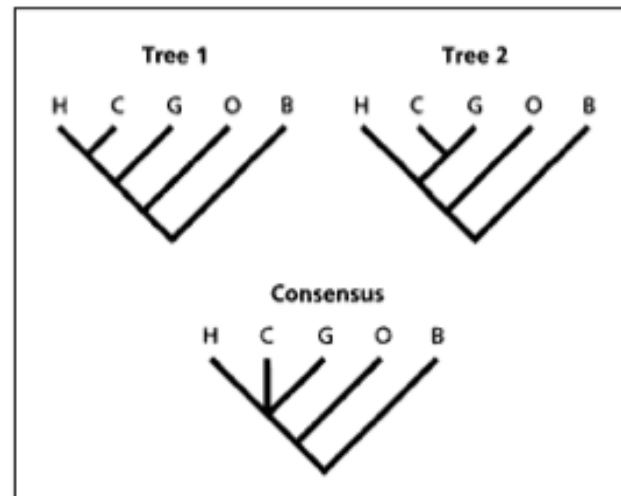
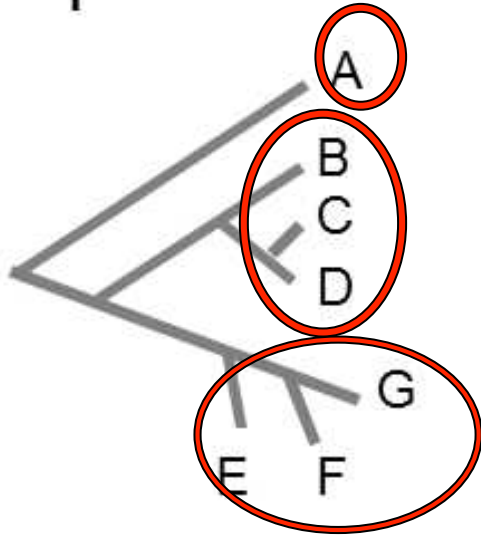


Fig. 2.26

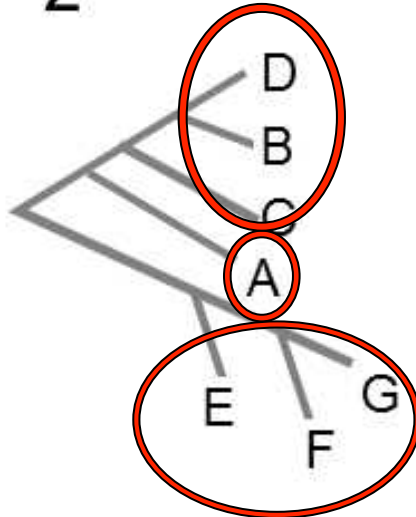
Two different trees for humans (H), chimps (C), gorillas (G), orang-utans (O) and gibbons (B), and their consensus tree.

Δένδρα συναίνεσης

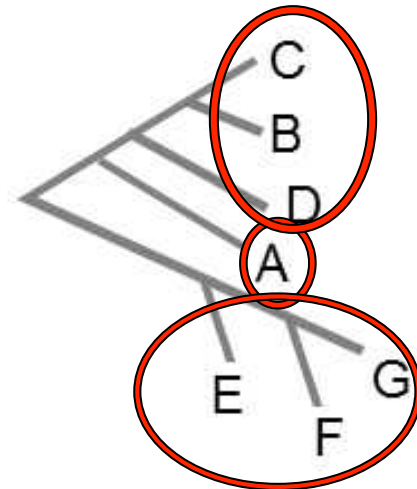
1



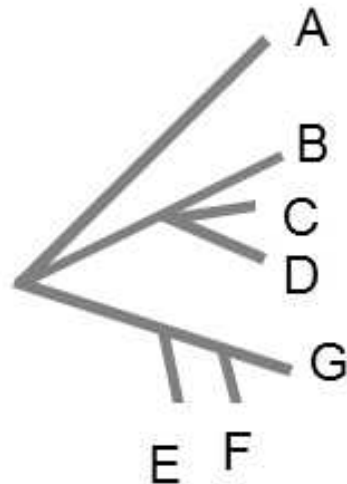
2



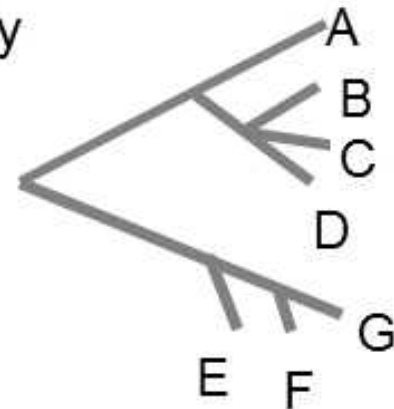
3



strictly
consensual
tree



50% majority
consensual
tree



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:

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

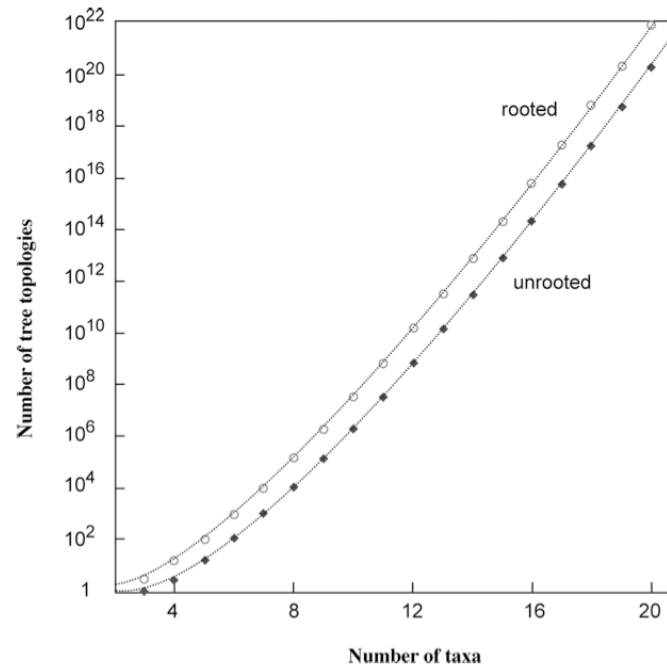
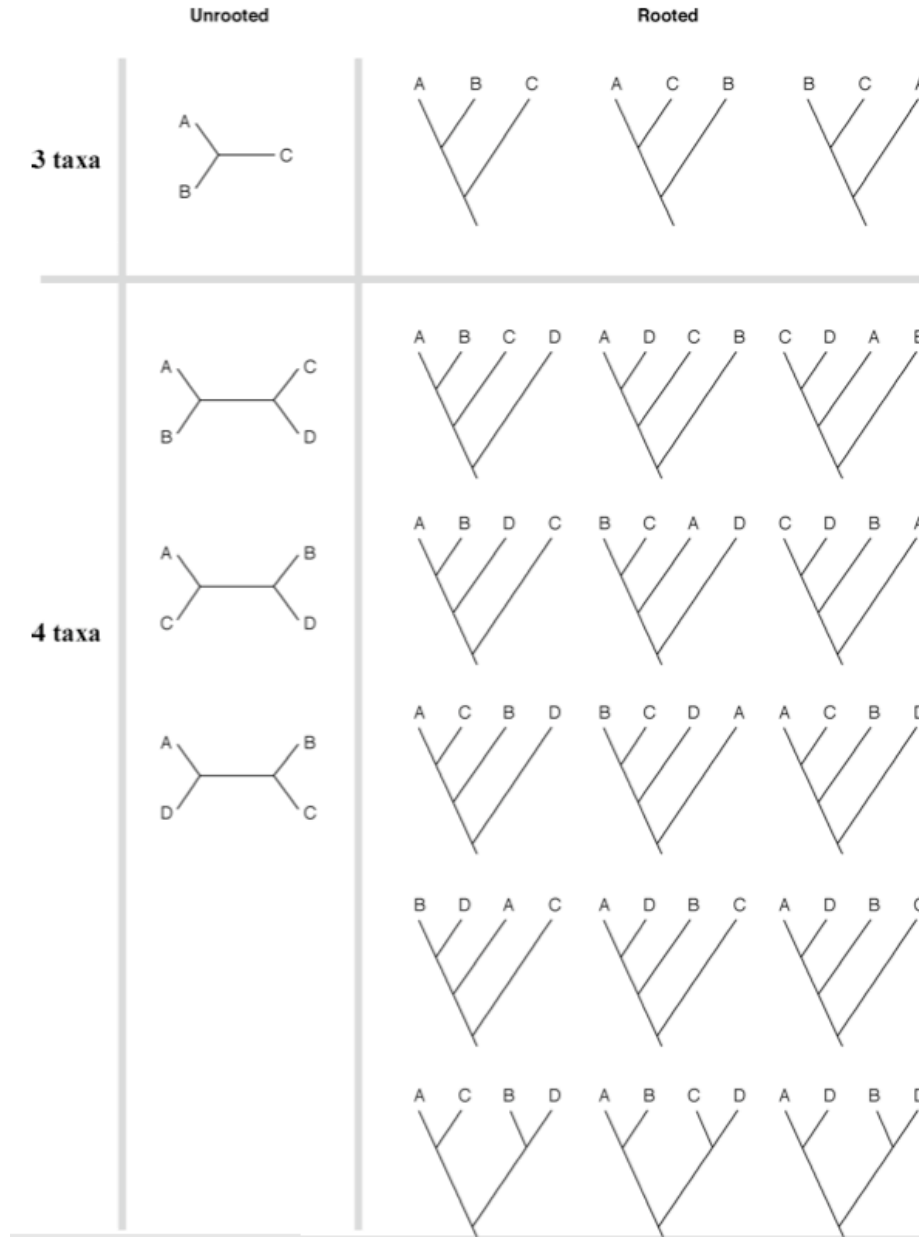


Figure 10.8: Total number of rooted (○) and unrooted (◆) tree topologies as a function of the number of taxa. The values in the y -axis are plotted in the log scale.

Πόσα πιθανά δένδρα;



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.

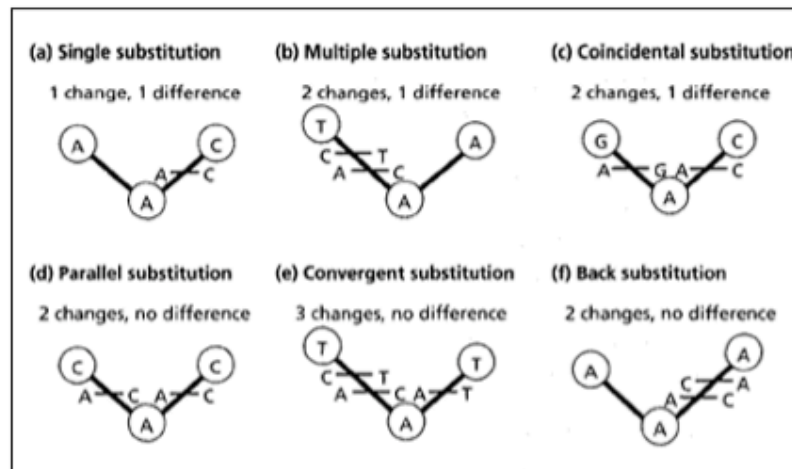


Fig. 5.9

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 'homoplasious'.

Υπολογισμός της απόστασης μεταξύ δύο ακολουθιών

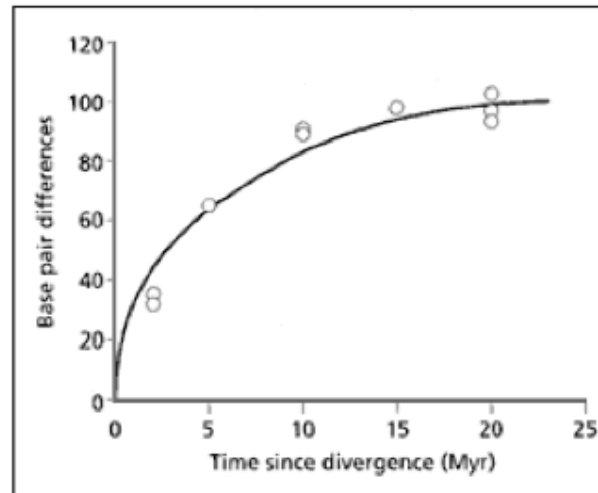


Fig. 5.11

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

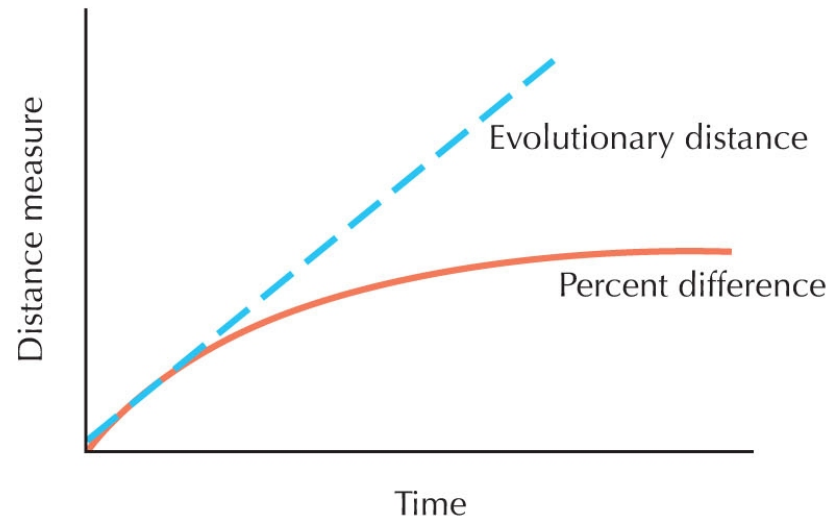
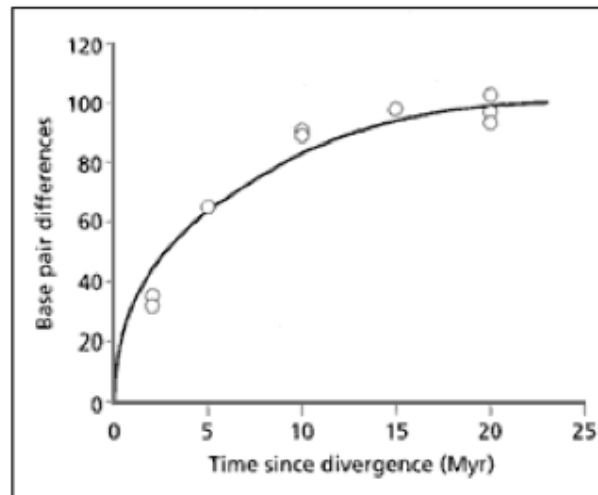


FIGURE 27.18. Distance corrections. If evolution occurs at a constant rate, then the percent difference between two OTUs should increase linearly with time (*blue dashed line*). However, this only applies if there are an infinite number of traits and states being compared. This does not occur with real data. For example, with DNA sequence data, there are only four possible states at any one character. This means that as two DNA sequences diverge, there is a limit to their maximum divergence. In addition, if multiple mutations occur at one site, the difference between the two DNA sequences will be less than the “evolutionary distance” between the organisms. The percent difference versus time will look something like the *red solid line* in this graph. For this reason, distance correction methods have been developed that allow the conversion of the *solid line* to the *dashed line* so that one can estimate the evolutionary distance from measures of percent difference.

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.

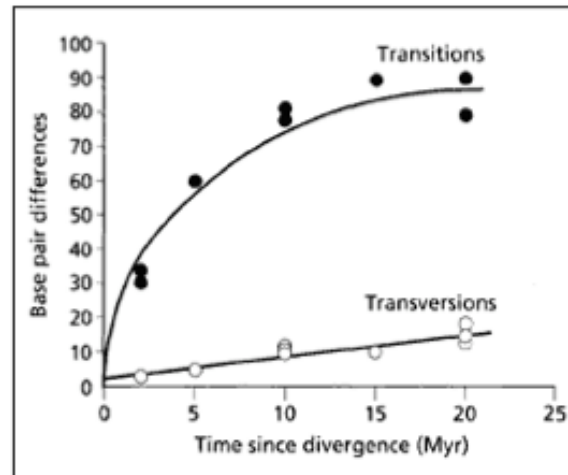


Fig. 5.13

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

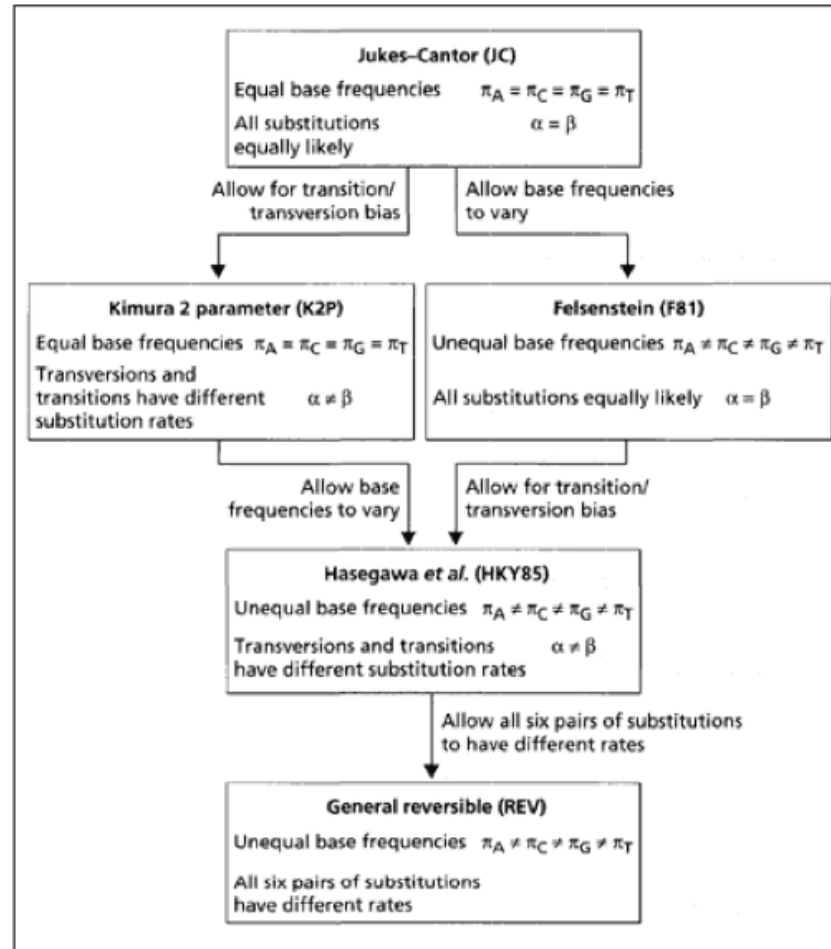


Fig. 5.14

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

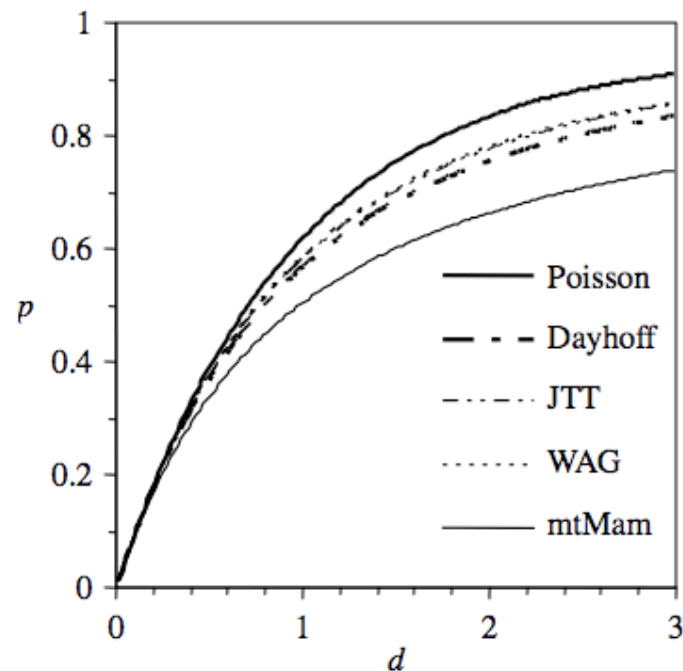


Fig. 2.2 The expected proportion of different sites (p) between two sequences separated by time or distance d under different models. The models are, from top to bottom, Poisson, WAG (Whelan and Goldman 2001), JTT (Jones *et al.* 1992), DAYHOFF (Dayhoff *et al.* 1978), and MTMAM (Yang *et al.* 1998).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

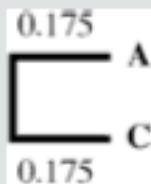
UPGMA

- 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 εξωομάδα. 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 $AC/2 = 0.35/2 = 0.175$.



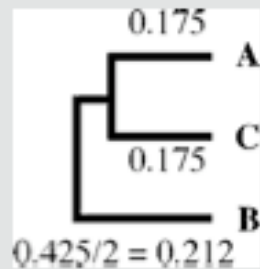
2. Because A and C are joined into a cluster, they are treated as one new composite taxon, which is used to create a reduced matrix. The distance of A-C cluster to every other taxa is one half of a taxon to A and C, respectively. That means that the distance of B to A-C is $(AB + BC)/2$; and that of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

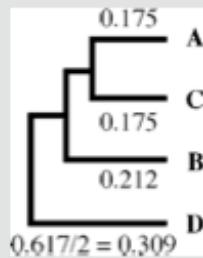


UPGMA

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.



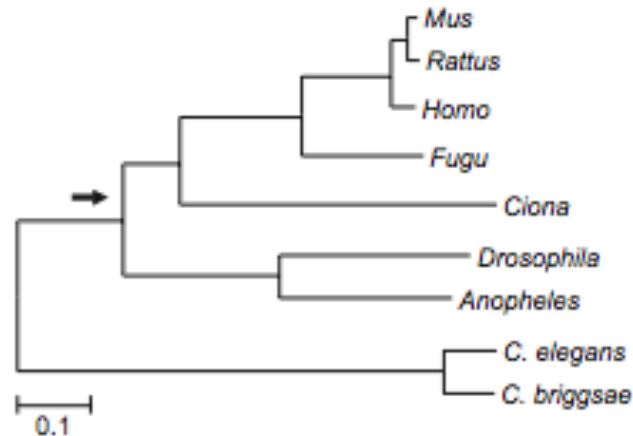
6. Because distance trees allow branches to be additive, the resulting distances between taxa from the tree path can be used to create a distance matrix. Obviously, the estimated distances do not match the actual evolutionary distances shown, which illustrates the failure of UPGMA to precisely reflect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

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

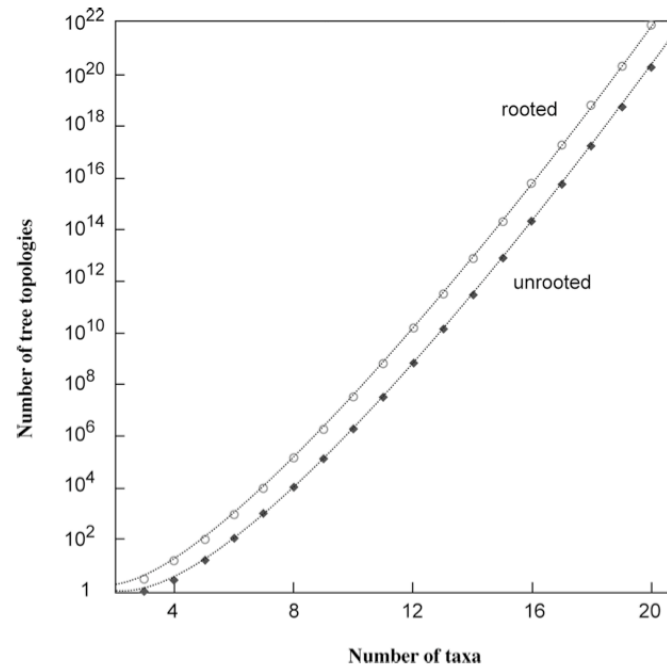
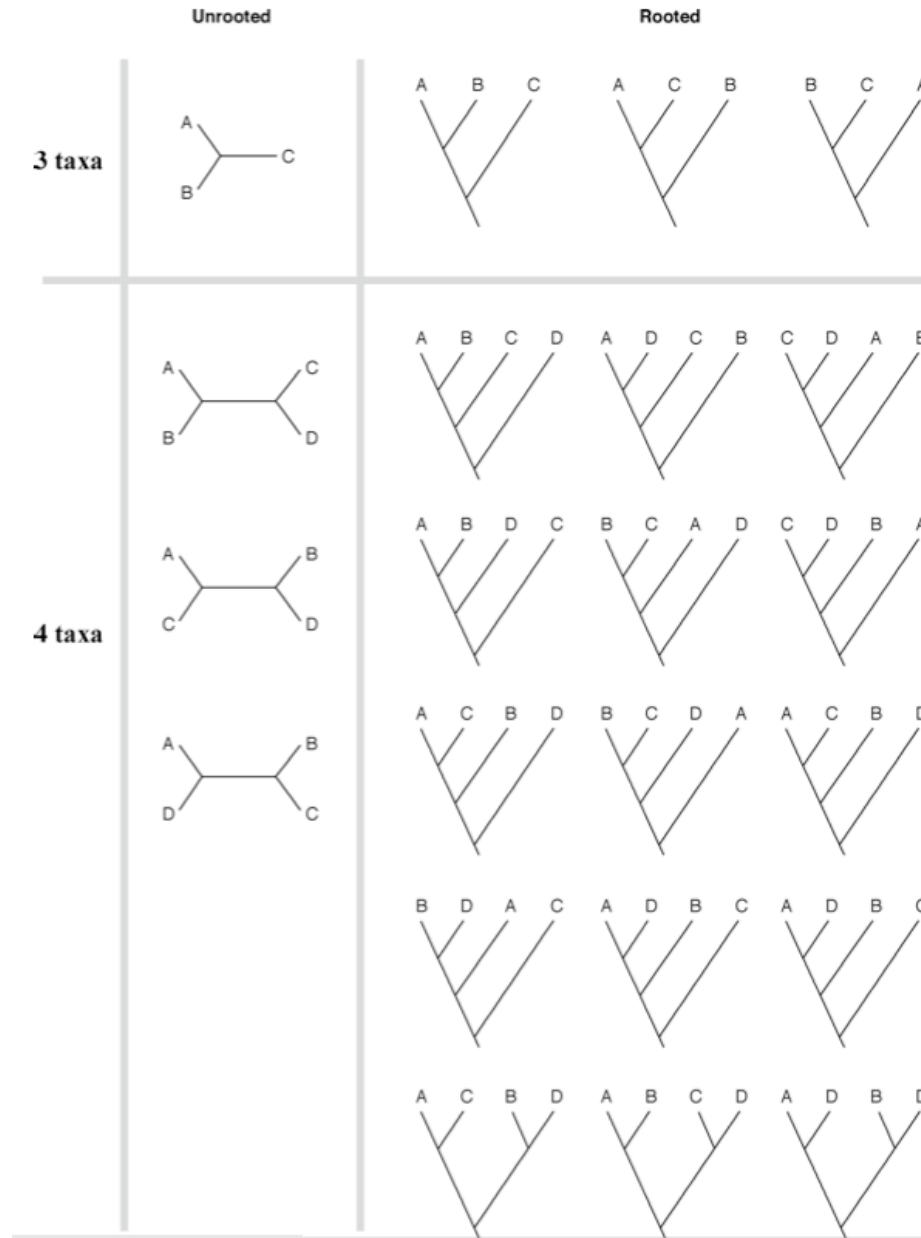


Figure 10.8: Total number of rooted (○) and unrooted (◆) tree topologies as a function of the number of taxa. The values in the y -axis are plotted in the log scale.

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.

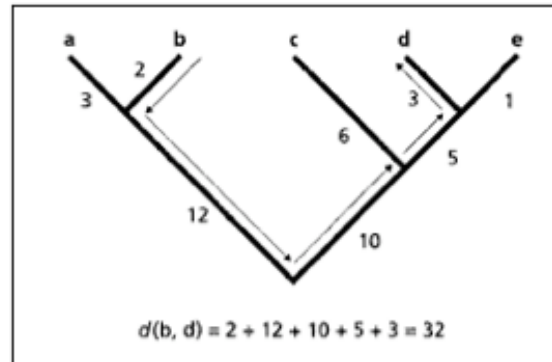


Fig. 5.21

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, αλλά 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.

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

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.

Μέγιστη φειδωλότητα (Maximum Parsimony)

- Για την κάθε 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.

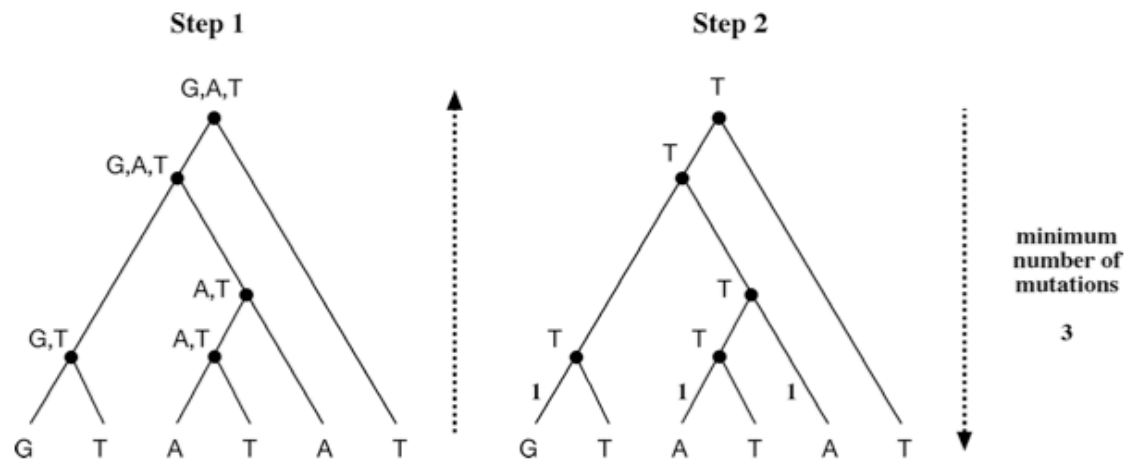


Figure 11.2: Using parsimony to infer ancestral characters at internal nodes involves a two-step procedure. The first step involves going from the leaves to the root and counting all possible ancestral characters at the internal nodes. The second step goes from the root to the leaves and assigns ancestral 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.

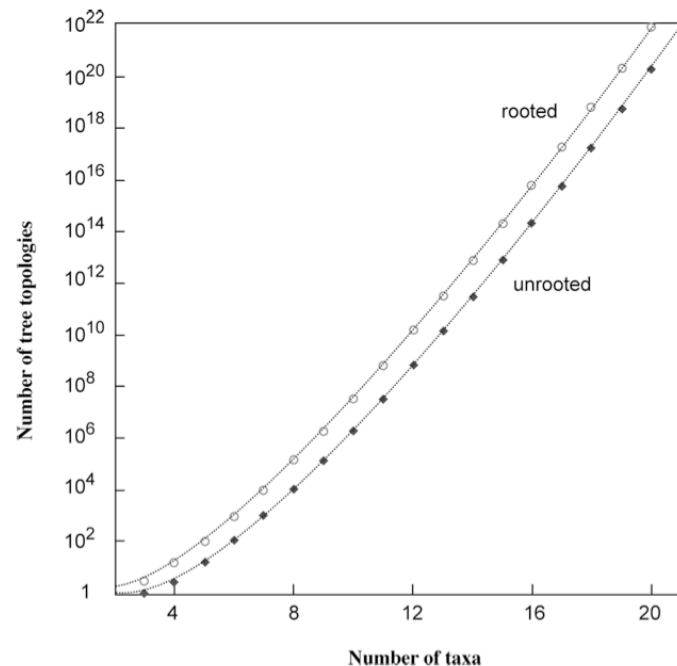


Figure 10.8: Total number of rooted (○) and unrooted (◆) tree topologies as a function of the number of taxa. The values in the y -axis are plotted in the log scale.

Αναζητώντας το καλύτερο δένδρο

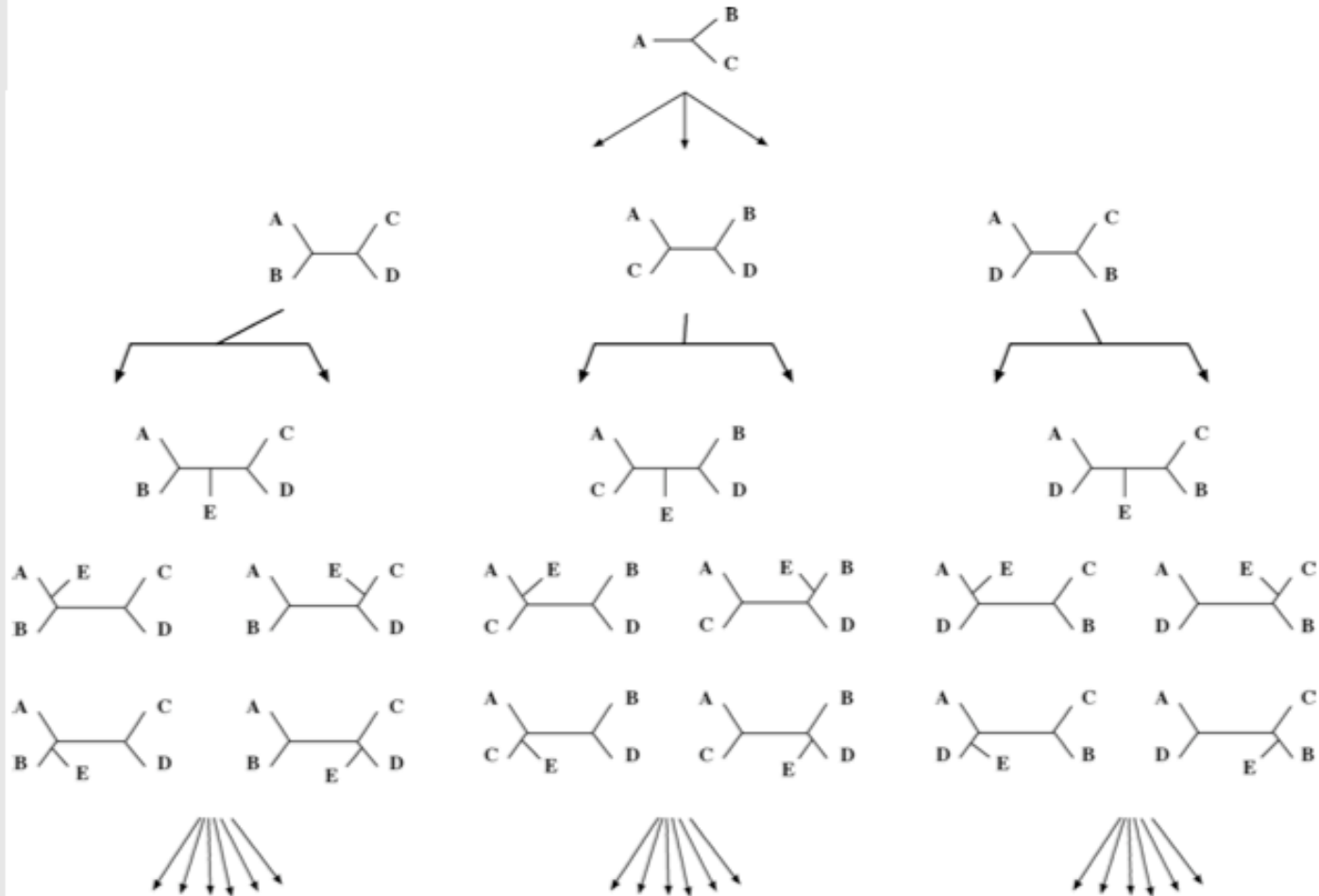


Figure 11.4: Schematic of exhaustive tree construction in the MP procedure. The tree starts with three taxa with one topology. One taxon is then added at a time in an progressive manner, during which the total branch lengths of all possible topologies are calculated.

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

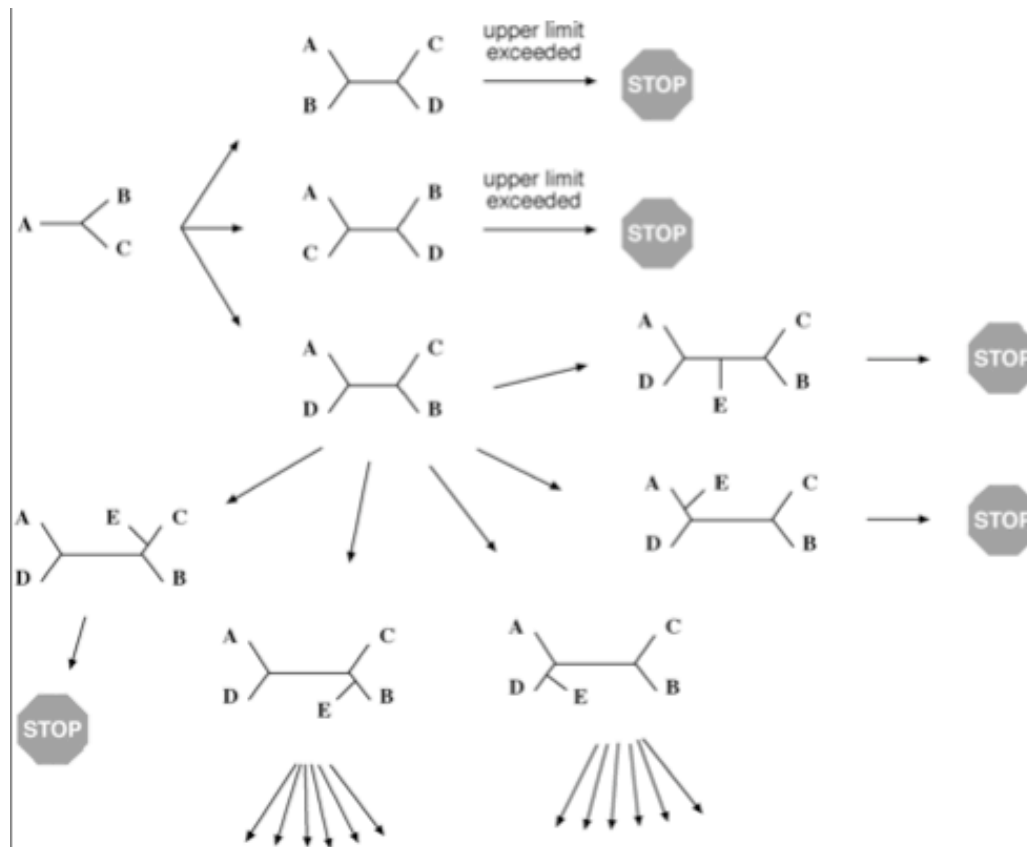


Figure 11.5: Schematic illustration of the branch-and-bound algorithm. Tree building starts with a step-wise addition of taxa of all possible topologies. Whenever the total branch length for a given topology exceeds the upper bound, the tree search in that direction stops, thereby reducing the total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

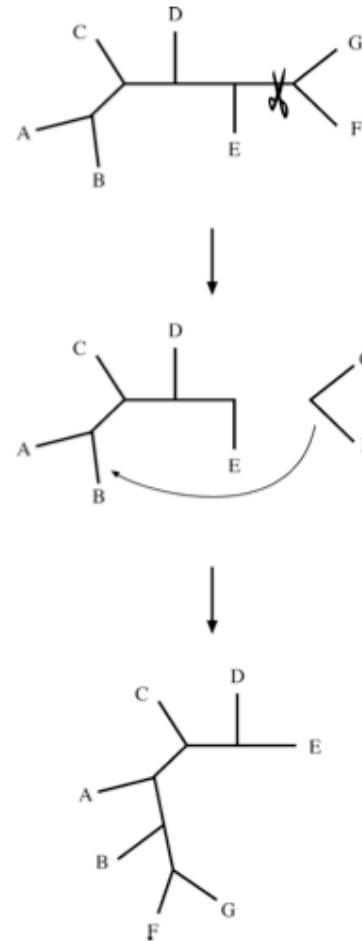


Figure 11.6: Schematic representation of a typical branch swapping process in which a branch is cut and moved to another part of the tree, generating a new topology.

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.

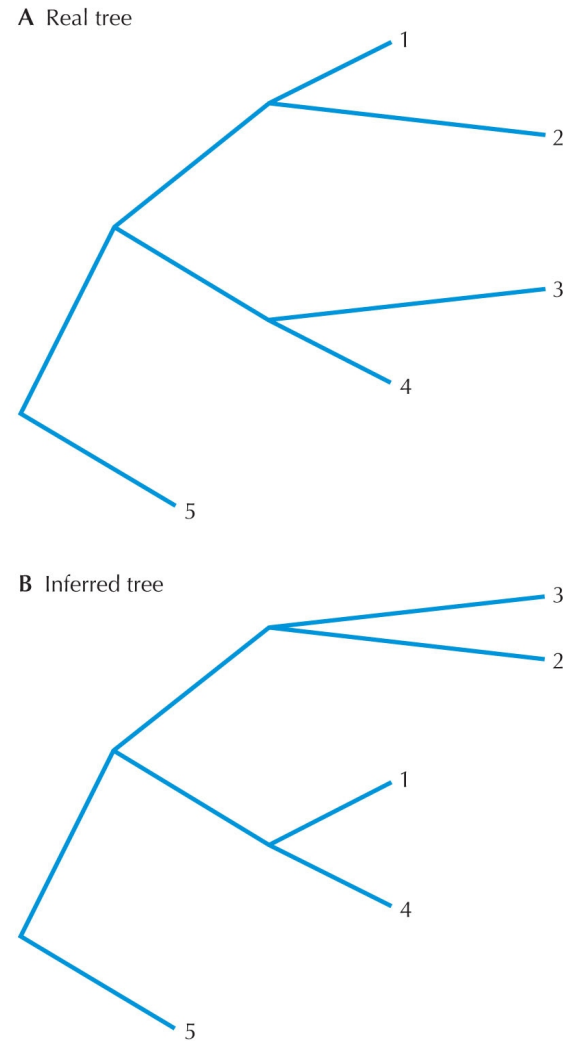


FIGURE 27.19. A simulated example of long-branch attraction. (A) The real tree of the relationships among five taxa, with two taxa (2 and 3) having long evolutionary branches. (B) An inferred tree of the taxa in which 2 and 3 are artificially grouped together because of the phenomenon of long-branch attraction.

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

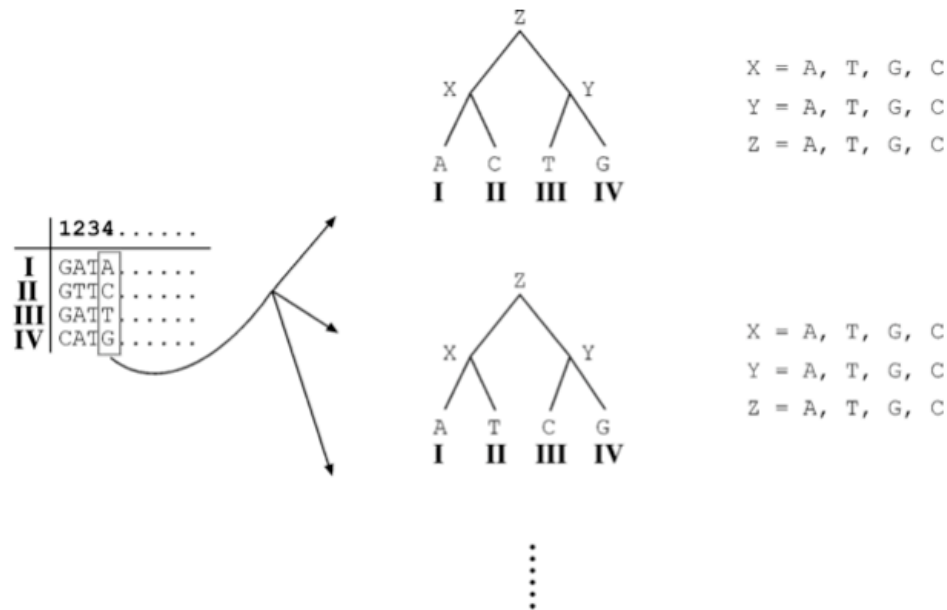


Figure 11.8: Schematic representation of the ML approach to build phylogenetic trees for four taxa, I, II, III, and IV. The ancestral character states at the internal nodes and root node are assigned X, Y, and Z, respectively. The example only shows some of the topologies derived from one of the sites 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

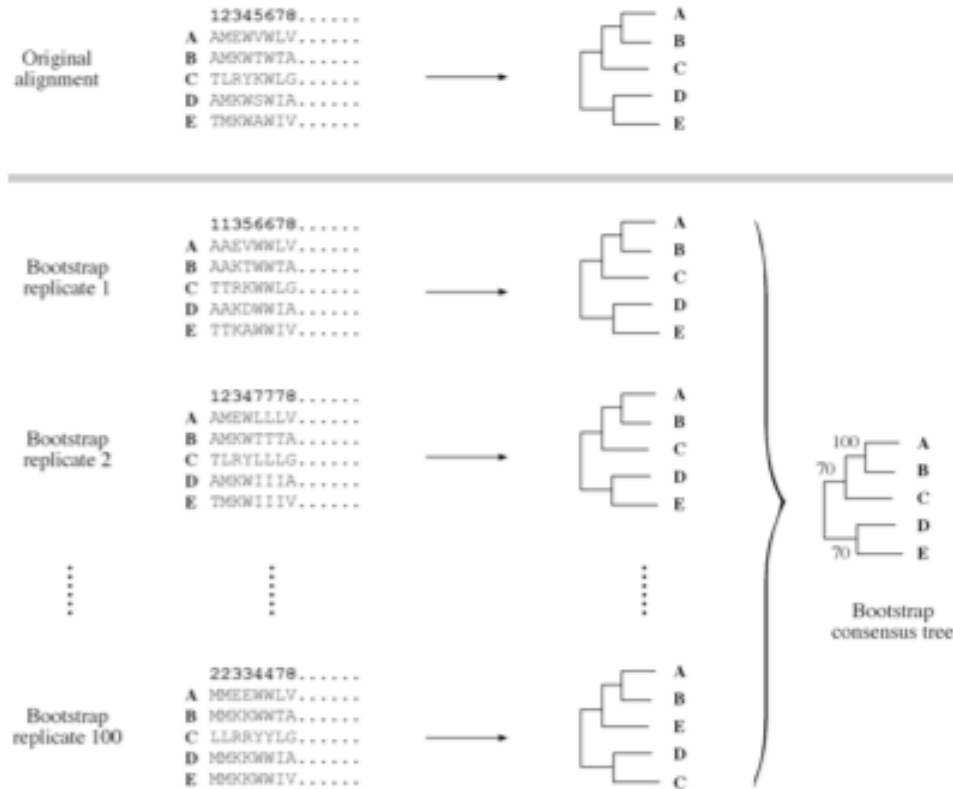


Figure 11.10: Schematic representation of a bootstrap analysis showing the original alignment and modified replicates in which certain sites are randomly replaced with other existing sites. The resulting altered replicates are used to building trees for statistical analysis at each node.

Jackknife

- Το Jackknife είναι παρόμοιο με το bootstrap.
- Επιλέγονται τυχαία (δίχως αντικατάσταση) οι μισές στήλες της πολλαπλής στοίχισης.
- Πρόβλημα: τα νέα δένδρα δημιουργούνται από λιγότερα δεδομένα.

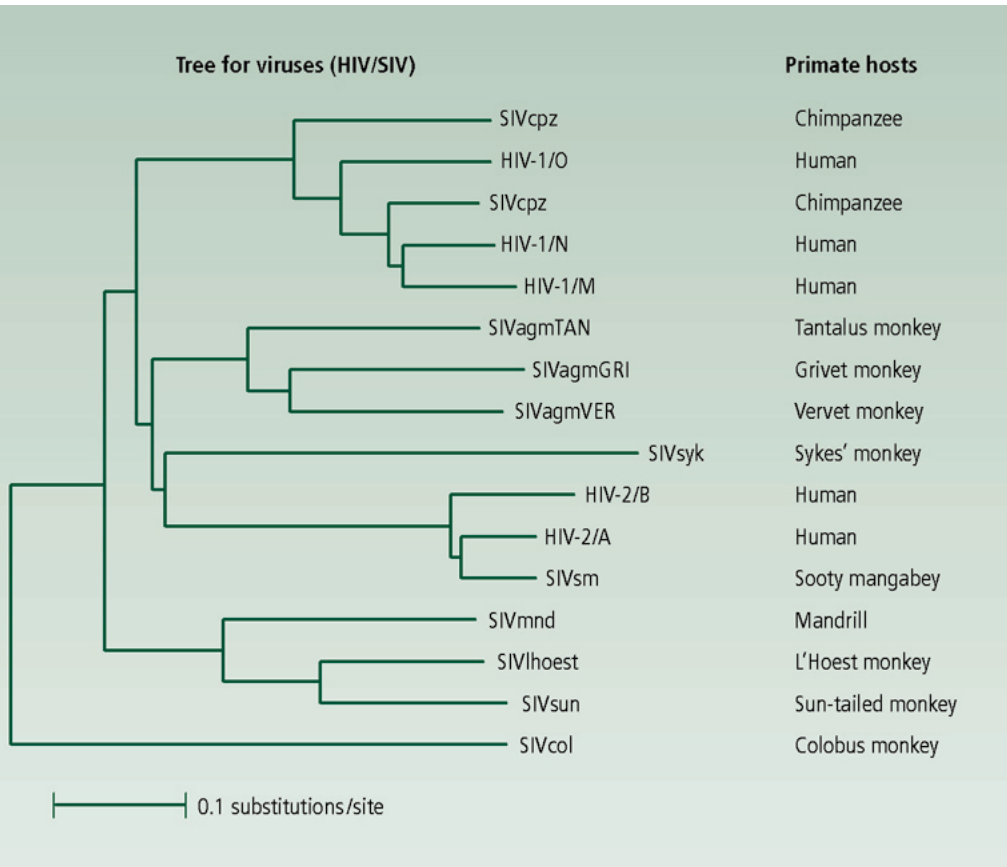
Tests που ελέγχουν αν ένα δένδρο είναι καλύτερο από ένα άλλο

- Συγκρίνονται 2 δένδρα στο σύνολό τους, με στατιστικές μεθόδους π.χ. Paired t-test ή χ^2 .
- Το bootstrap ή το Jackknife ελέγχει την αξιοπιστία του κάθε επιμέρους κλάδου.
- Για κάθε μέθοδο κατασκευής δένδρων χρησιμοποιείται και το αντίστοιχο τεστ.
- Για μέγιστη φειδωλότητα:
 - Kishino-Hasegawa test. 2 δένδρα, N πληροφοριακές θέσεις. Για κάθε θέση, υπολογίζεται το μήκος βραχιόνων του καθένα από τα 2 δένδρα. Αυτό γίνεται και για τις N θέσεις. Οι τιμές χρησιμοποιούνται σε paired t-test, για να φανεί αν η διαφορά μεταξύ των 2 δένδρων είναι στατιστικά σημαντική.
- Για μέγιστη πιθανοφάνεια:
 - Shimodaira-Hasegawa test. Αρχικά υπολογίζονται τα log-likelihood scores για τα 2 δένδρα. Οι βαθμοί ελευθερίας εξαρτώνται από το μοντέλο εξέλιξης που χρησιμοποιείται. Χρησιμοποιείται το χ^2 .

ΕΦΑΡΜΟΓΕΣ

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.

Η θεωρία του εμβολίου της πολιομυελίτιδας - OPV/ AIDS theory

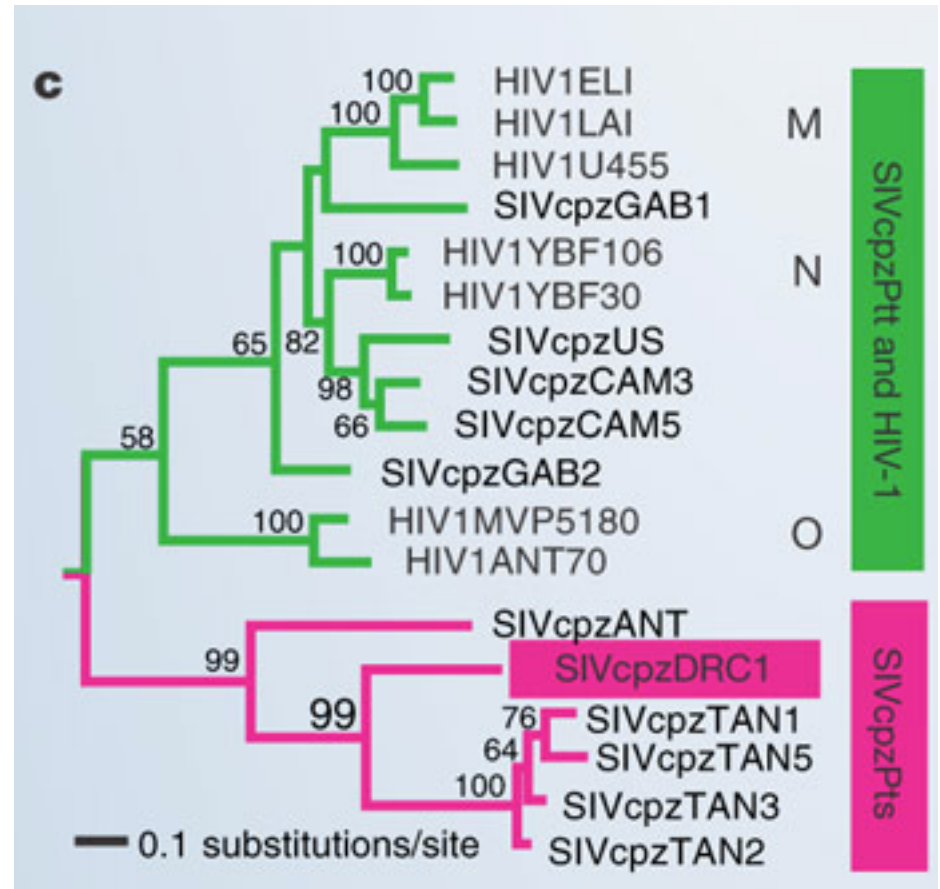
Η θεωρία αυτή υποστήριξε ότι το SIV πέρασε στους ανθρώπους και μεταλλάχθηκε σε HIV κατά την διαδικασία κατασκευής εμβολίων πολιομυελίτιδας σε ένα κέντρο παρασκευής εμβολίων στην περιοχή του Stanleyville (γνωστό πλέον ως Kisangani) στο Κονγκό (Democratic Republic of Congo). Η θεωρία υποστηρίζει ότι κατά την διαδικασία κατασκευής του εμβολίου χρησιμοποιήθηκε βιολογικό υλικό από χιμπατζήδες (μολυσμένους με SIV) της περιοχής.

Ο Worobey και οι συνεργάτες του μάζεψαν βιολογικό υλικό (μολυσμένων με SIV χιμπατζήδων) από την περιοχή της Kisangani και έκαναν φυλογενετική ανάλυση, χρησιμοποιώντας και ακολουθίες SIV από άλλους πληθυσμούς χιμπατζήδων και από ακολουθίες HIV.

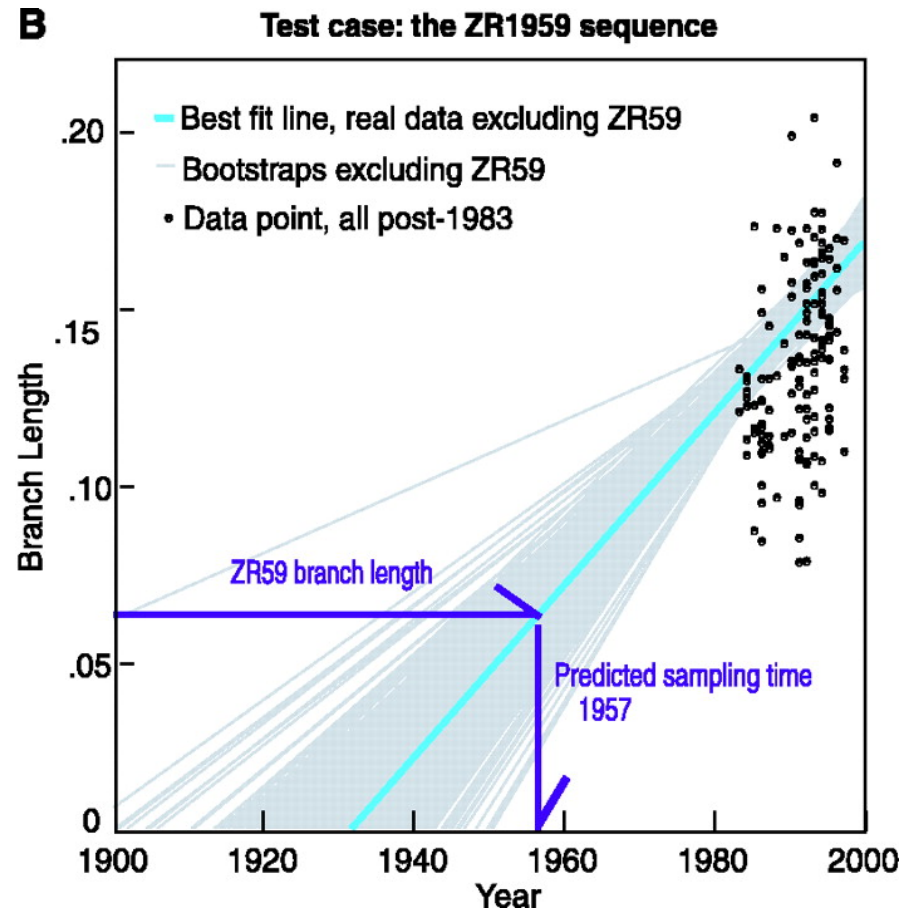
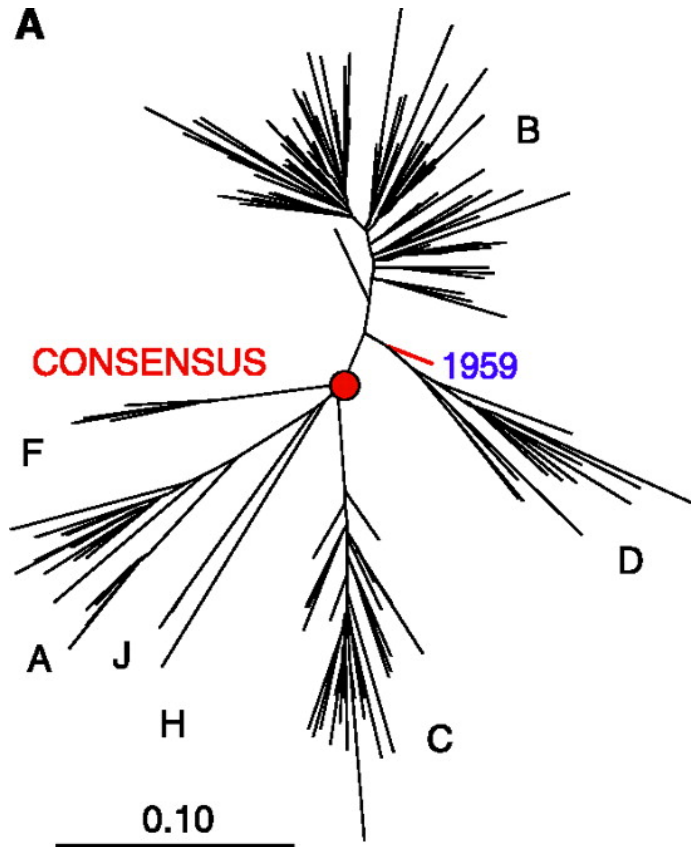
Αν ίσχυε η θεωρία του εμβολίου, τότε θα έπρεπε οι ακολουθίες SIV από την περιοχή του Kisangani να είναι οι πιο κοντινές SIV ακολουθίες σε αυτές του HIV.

<http://www.nature.com/nature/journal/v428/n6985/pdf/428820a.pdf>

Τα φυλογενετικά δέντρα όμως δεν το υποστηρίζουν αυτό το σενάριο. Επιπλέον, τα μοριακά ρολόγια υποστηρίζουν ότι ο HIV πέρασε στους ανθρώπους περίπου 30 χρόνια νωρίτερα από ότι υποστηρίζει η θεωρία του εμβολίου της πολιομυελίτιδας.



Timing the Ancestor of the HIV-1 Pandemic Strains



Molecular epidemiology of HIV transmission in a dental practice.

Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, Korber BT, Mullins JI, Schochetman G, Berkelman RL, Economou AN, et al.

Division of HIV/AIDS, Centers for Disease Control, Atlanta, GA 30333.

Abstract

Human immunodeficiency virus type 1 (HIV-1) transmission from infected patients to health-care workers has been well documented, but transmission from an infected health-care worker to a patient has not been reported. After identification of an acquired immunodeficiency syndrome (AIDS) patient who had no known risk factors for HIV infection but who had undergone an invasive procedure performed by a dentist with AIDS, six other patients of this dentist were found to be HIV-infected. Molecular biologic studies were conducted to complement the epidemiologic investigation. Portions of the HIV proviral envelope gene from each of the seven patients, the dentist, and 35 HIV-infected persons from the local geographic area were amplified by polymerase chain reaction and sequenced. Three separate comparative genetic analyses—genetic distance measurements, phylogenetic tree analysis, and amino acid signature pattern analysis—showed that the viruses from the dentist and five dental patients were closely related. These data, together with the epidemiologic investigation, indicated that these patients became infected with HIV while receiving care from a dentist with AIDS.

<http://www.ncbi.nlm.nih.gov/pubmed/1589796>

- Επτά ασθενείς ενός οδοντίατρου που έπασχε από AIDS προσβλήθηκαν επίσης από AIDS.
- Τον κατηγόρησαν ότι τους μετέδωσε τον ιό.
- Έγινε φυλογένεση ακολουθιών από τον οδοντίατρο, τους ασθενείς και από άλλους 35 ασθενείς AIDS από την τοπική κοινωνία.

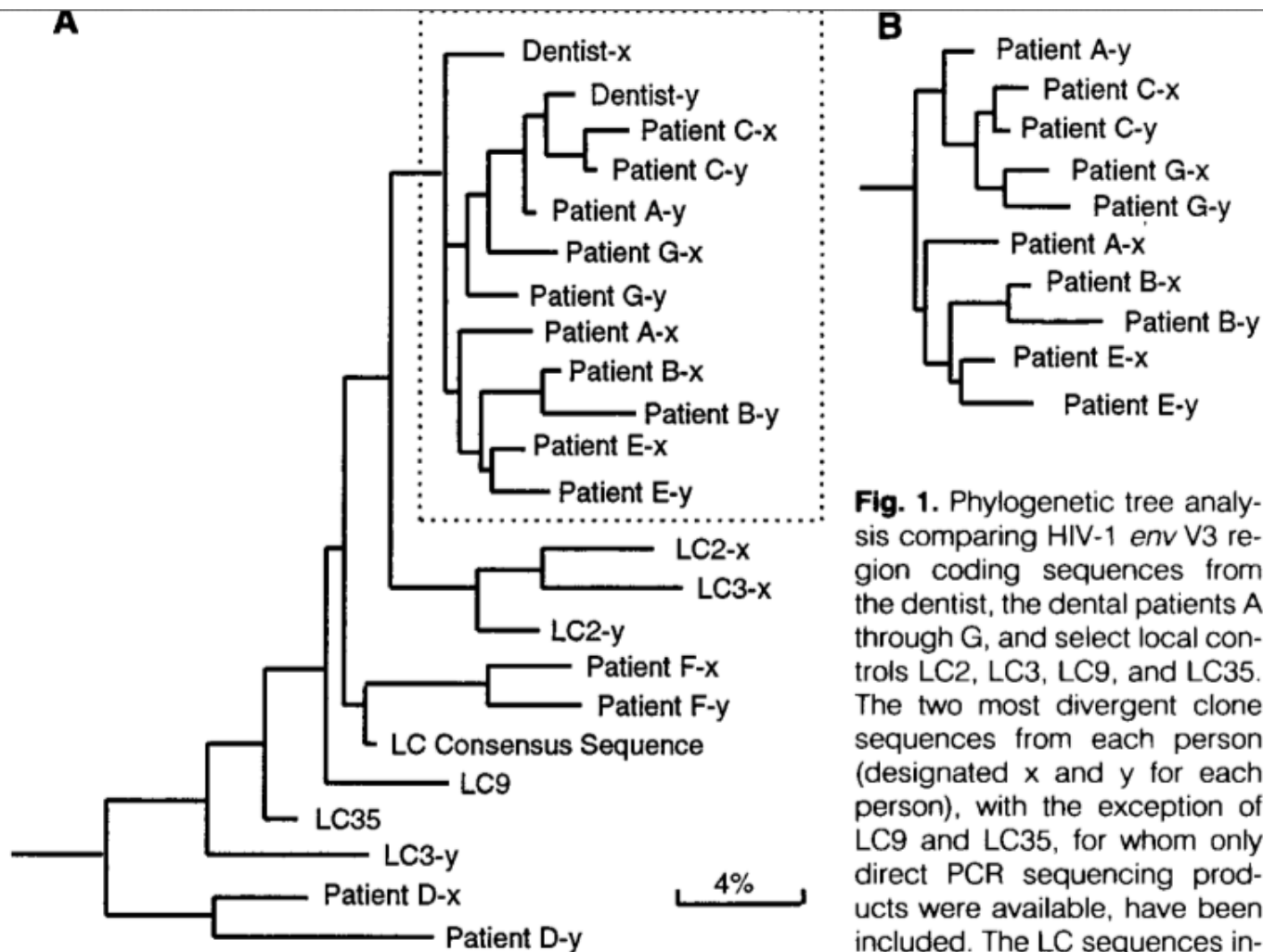


Fig. 1. Phylogenetic tree analysis comparing HIV-1 *env* V3 region coding sequences from the dentist, the dental patients A through G, and select local controls LC2, LC3, LC9, and LC35. The two most divergent clone sequences from each person (designated x and y for each person), with the exception of LC9 and LC35, for whom only direct PCR sequencing products were available, have been included. The LC sequences included were those found by

pairwise distance measurement (Table 1) and signature pattern analysis (Table 2) to be the closest control sequences to the dental group sequences, which are enclosed by a box in (A). An LC consensus sequence has also been included, and the tree was rooted upon the African sample ELI. The PAUP parsimony algorithm was used to analyze 279 aligned sites (25), of which 146 sites were varied. When the dentist's viral sequences were withdrawn from the analysis, or required to cluster with the LC consensus sequence (25), the dental clade remained otherwise unaffected, as shown in (B). Vertical distances are for clarity only; the lengths of the horizontal branches are proportional to the single base changes and can be read as percentage differences with the scale bar.

Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 pol and gag genes.

Albert J, Wahlberg J, Leitner T, Escanilla D, Uhlén M.

Department of Clinical Virology, Swedish Institute for Infectious Disease Control, Karolinska Institute, Stockholm.

Abstract

Transmission of human immunodeficiency virus type 1 (HIV-1) from a male accused of rape and deliberate transmission of HIV-1 was investigated by sequencing of the HIV-1 pol and gag genes from virus obtained from the male and from the female victim. Parts of the reverse transcriptase and p17gag genes were amplified and directly sequenced from uncultured peripheral blood mononuclear cells. The sequences were compared with sequences from 21 unrelated HIV-1-infected controls from the same geographic area (Stockholm, Sweden). Bootstrap analysis of phylogenetic trees demonstrated that the sequences from the female were significantly more closely related to the sequences from the male than to sequences from the controls. Furthermore, we found that the male and female shared two distinct genetic variants of HIV-1. In p17gag the major variant had an unusual, out-of-frame deletion of 3 nucleotides which the minor variant lacked. These results indicated that the male had transmitted more than one infectious unit to the female. From this study we concluded that it was highly likely that the HIV-1 strains carried by the male and female were closely epidemiologically linked.

PMID: 7520096 [PubMed - indexed for MEDLINE] PMID: PMC236997 **Free PMC Article**

<http://www.ncbi.nlm.nih.gov/pubmed/7520096>

- Θύμα βιασμού κατηγόρησε τον βιαστή ότι της μετέδωσε AIDS.
- Χρησιμοποιήθηκαν ακολουθίες, από το θύμα, τον βιαστή και από 21 ασθενείς AIDS της τοπικής κοινωνίας.
- Η φυλογένεση έδειξε ότι οι ακολουθίες του βιαστή και του θύματος ήταν οι πιο κοντινές.

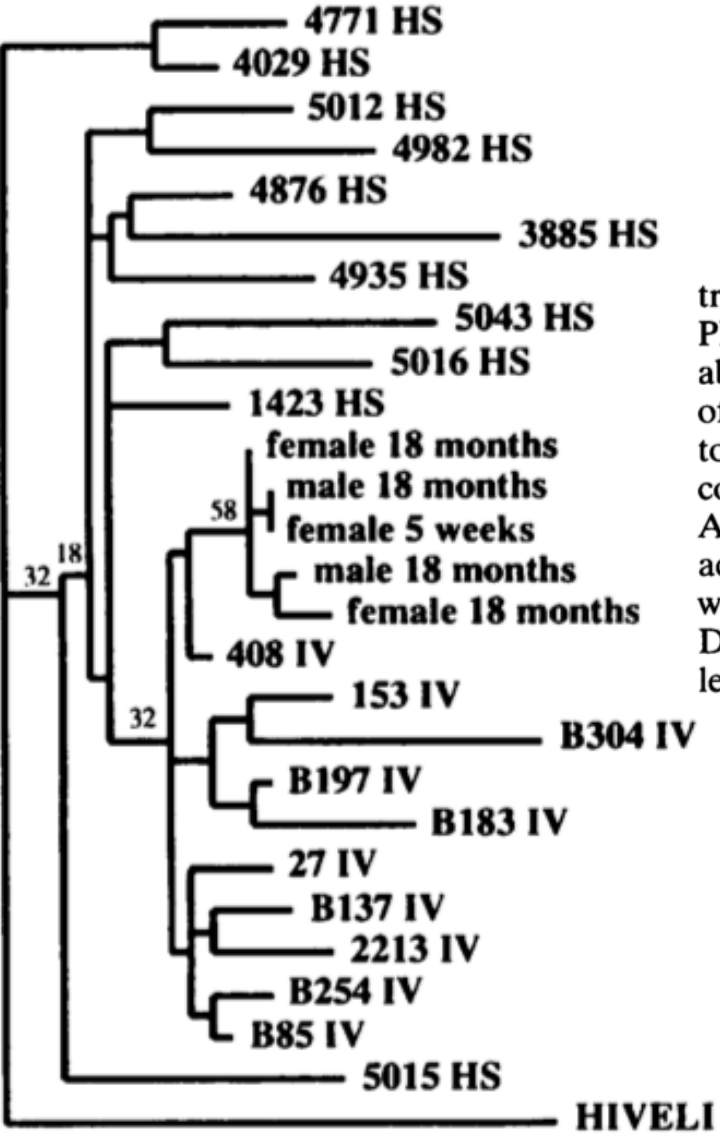
A

FIG. 5. (A) Phylogenetic tree analysis comparing HIV-1 reverse transcriptase nucleotide sequences obtained by direct sequencing of PBMC from the male, the female, 10 unrelated intravenous-drug abusers (IV), and 11 unrelated homosexual men (HS). The sequence of the African HIV-1 isolate ELI (HIVELI) was used as an outgroup to root the tree. The analysis was performed with DNA sequences corresponding to amino acids 8 to 222. Some controls had received AZT treatment; therefore, the nucleotides corresponding to amino acids 41, 67, 70, 215, and 219, which confer AZT resistance (16, 26), were excluded from the analysis. The tree was generated with the DNAML program from the PHYLIP package, version 3.4 (6). Selected bootstrap values are given. (B) Phylogenetic tree analysis

Έλεγχος εξελικτικών υποθέσεων

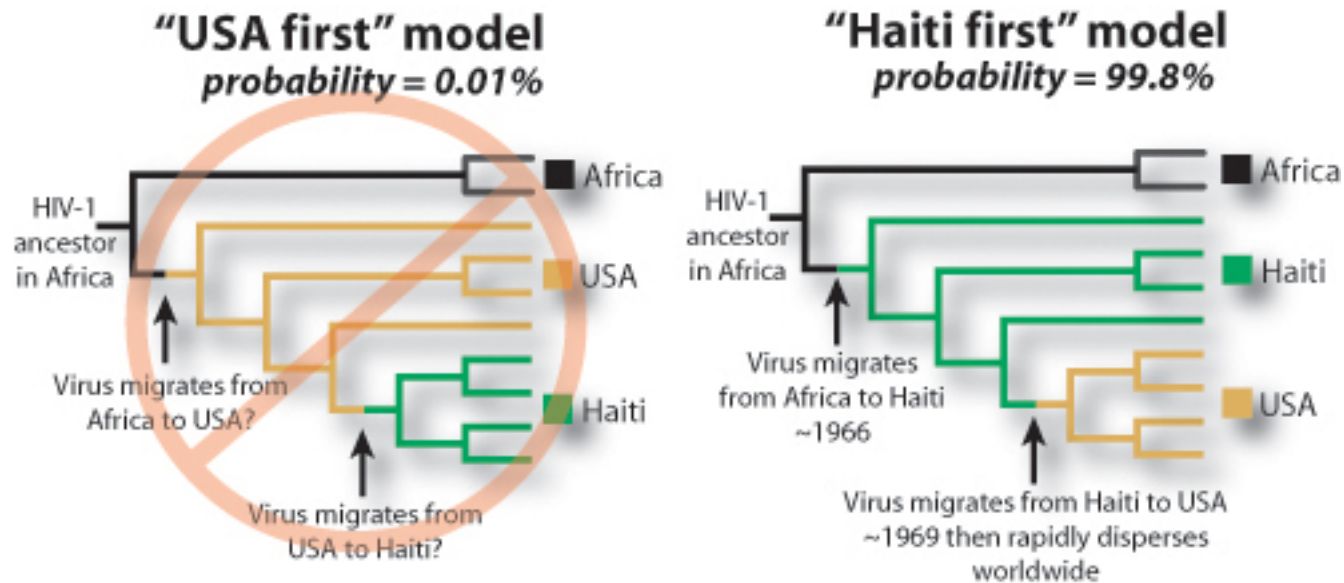
Από που προήλθε ο ιός HIV-1 subtype M; Προέλευση στην Κεντρική Αφρική.

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.

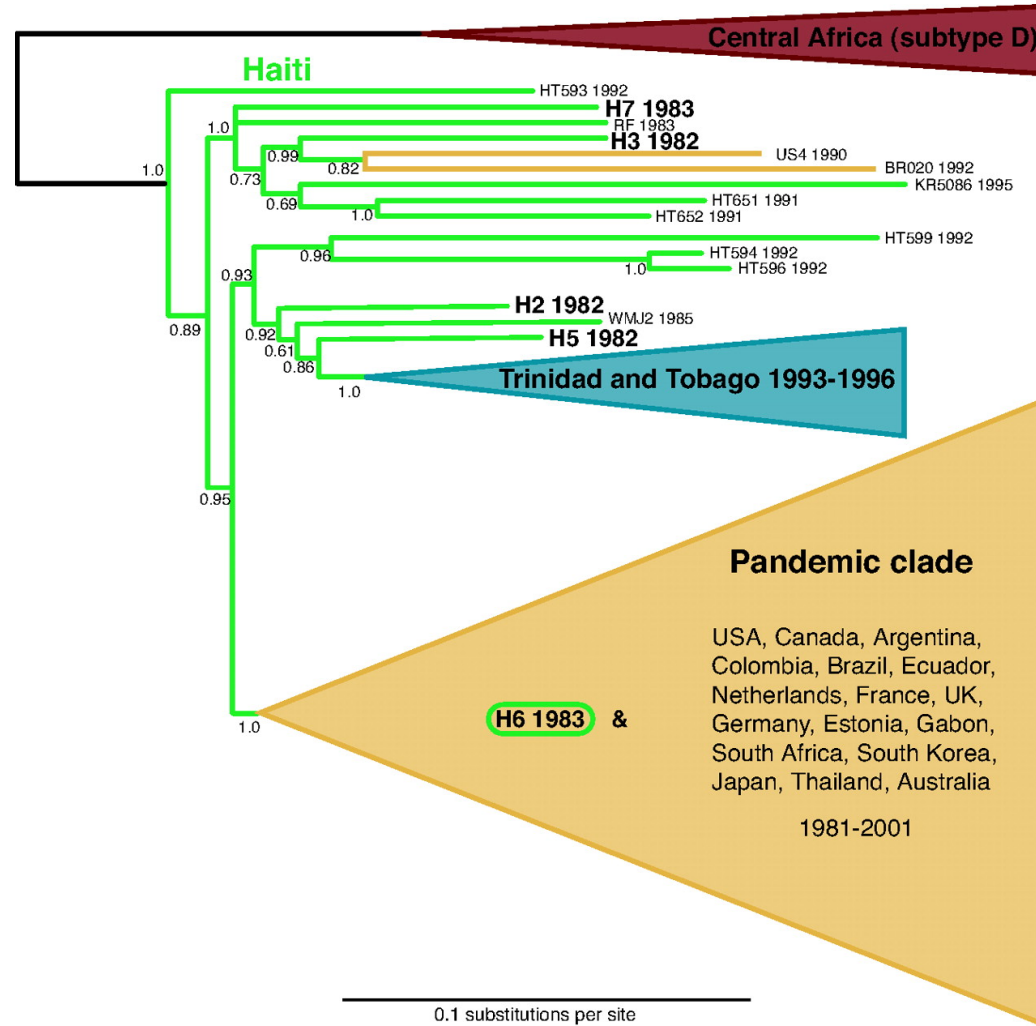
Από την Αϊτή στην Αμερική ή το αντίθετο;

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 (Ιανουαριος 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

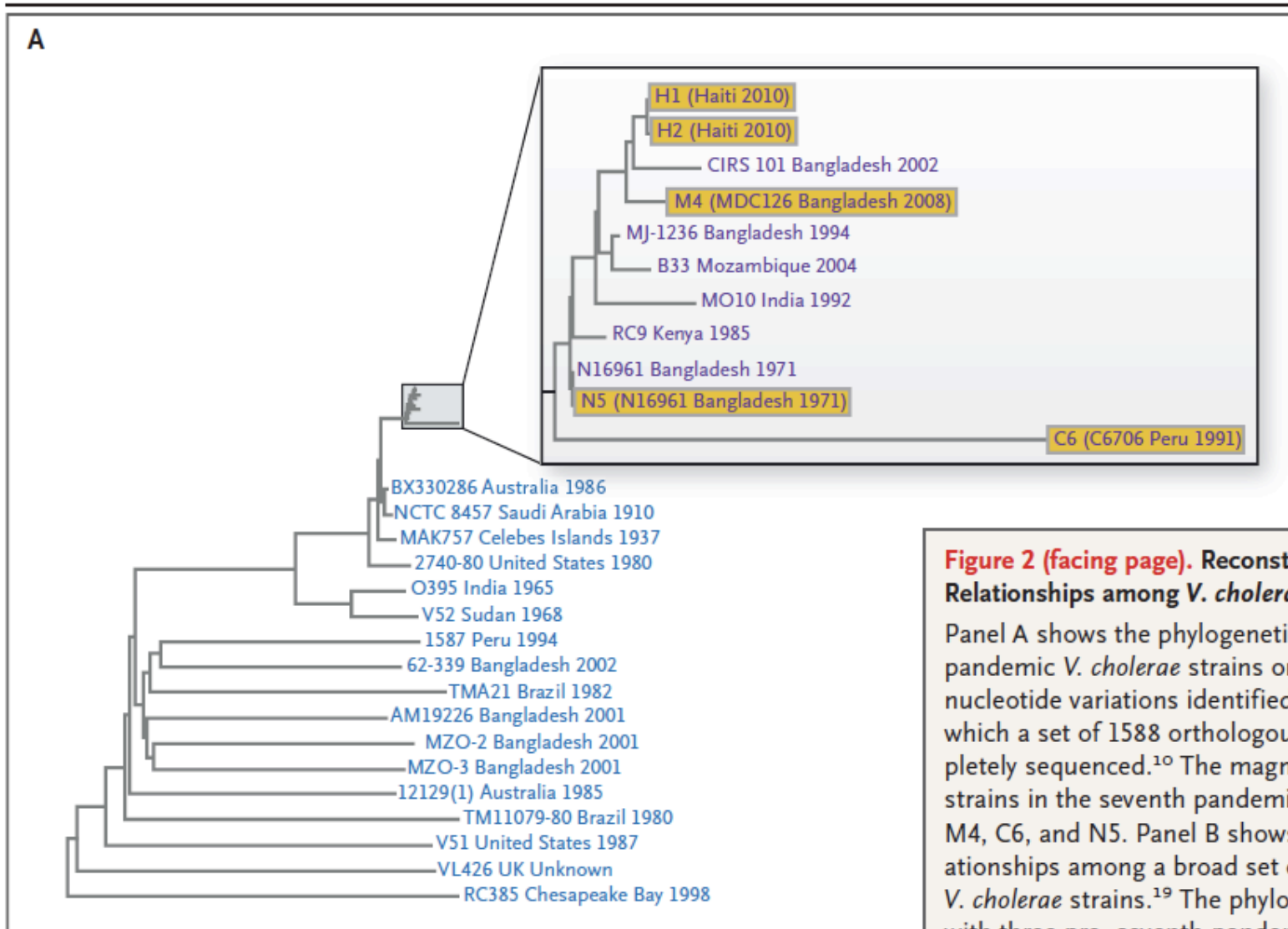


Figure 2 (facing page). Reconstructing Phylogenetic Relationships among *V. cholerae* Strains.

Panel A shows the phylogenetic relationships among pandemic *V. cholerae* strains on the basis of single-nucleotide variations identified among all strains for which a set of 1588 orthologous genes has been completely sequenced.¹⁰ The magnified inset represents strains in the seventh pandemic, including H1, H2, M4, C6, and N5. Panel B shows the phylogenetic relationships among a broad set of seventh-pandemic *V. cholerae* strains.¹⁹ The phylogenetic tree is rooted with three pre-seventh-pandemic strains.