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.

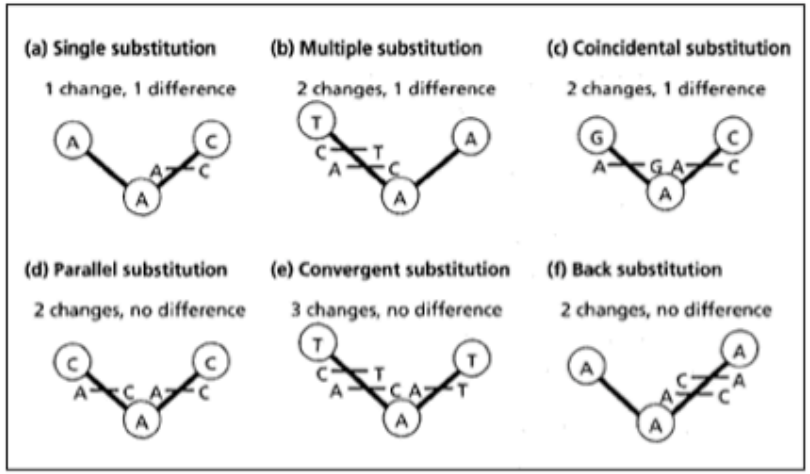


Fig. 5.9
Six kinds of nucleotide substitution. In each case the ancestral nucleotide was A. In all except the case of a single substitution, the number of substitutions that actually occurred is greater than would be counted if we just compared the two descendant sequences. In the lower three cases the 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

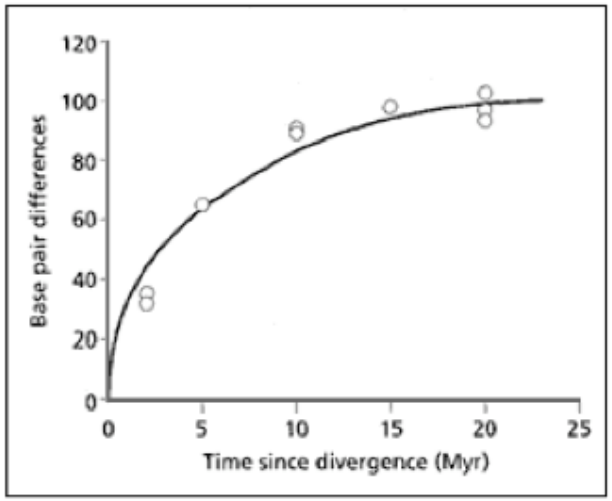


Fig. 5.11
Number of nucleotide substitutions between pairs of bovid mammal mitochondrial sequences (684 basepairs from the *COII* gene) against estimated time of divergence. Notice that the observed number of substitutions is not linear with time but curvilinear. Data from Janecek *et al*. (1996).

Διόρθωση της απόστασης μεταξύ 2 ακολουθιών

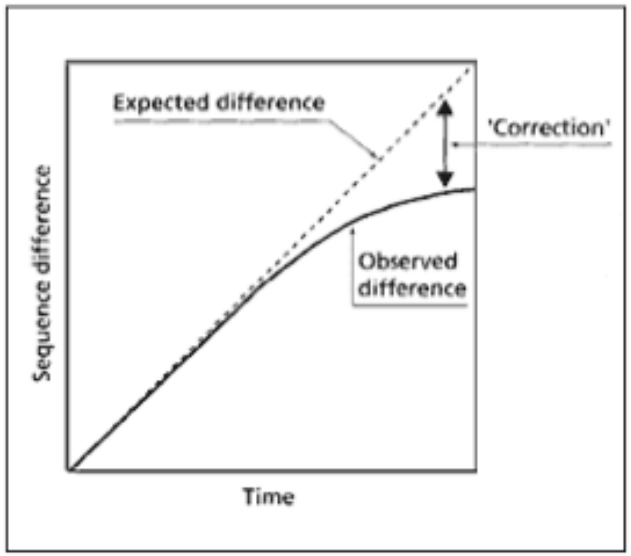
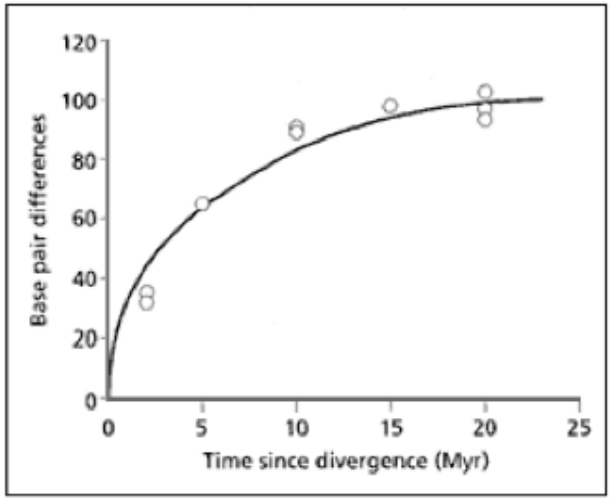


Fig. 5.12
The need to correct observed sequence differences. The extent of observed differences between two sequences is not linear with time (as we would expect if the rate of molecular evolution is approximately constant) but curvilinear due to multiple hits. The goal of distance correction methods 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 υπόψην 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

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

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.

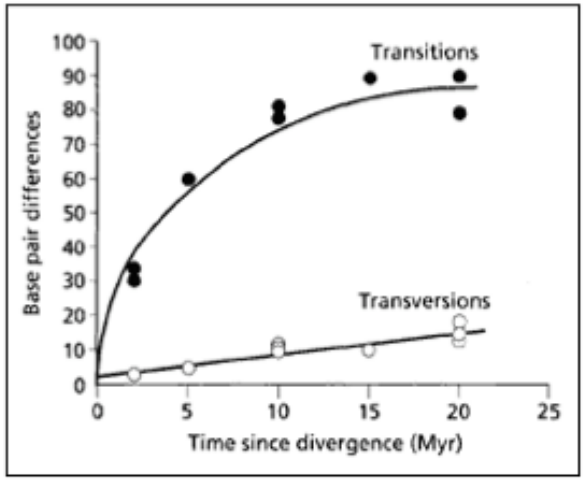


Fig. 5.13
The number of transitions and transversions between the same bovid mammal sequences used in Fig. 5.11. Transitions accumulate much more rapidly than transversions and become saturated, whereas transversions accumulate more slowly and show no evidence of saturation.

Μοντέλο αντικατάστασης Kimura

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

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

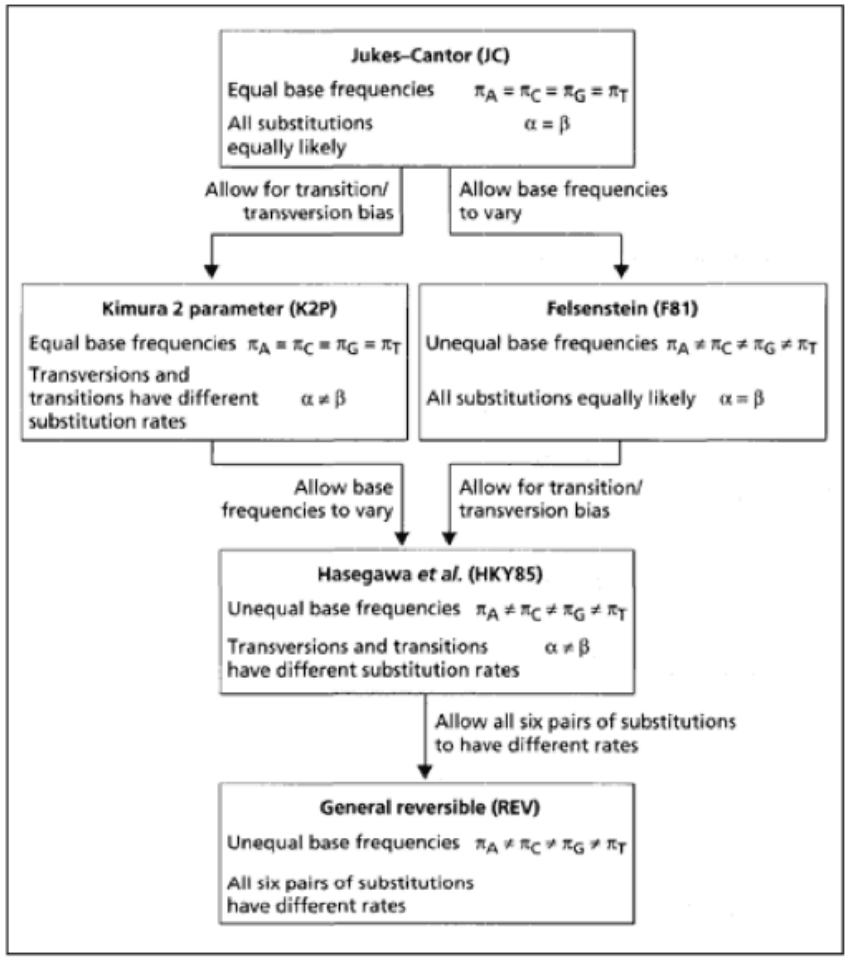


Fig. 5.14
Interrelationships among five models for estimating the number of nucleotide substitutions among a pair of DNA sequences. The JC, K2P, F81 and HKY85 models can all be generated by constrainingvarious parameters of the REV model.

Διόρθωση των παρατηρούμενων αποστάσεων για πρωτεΐνες

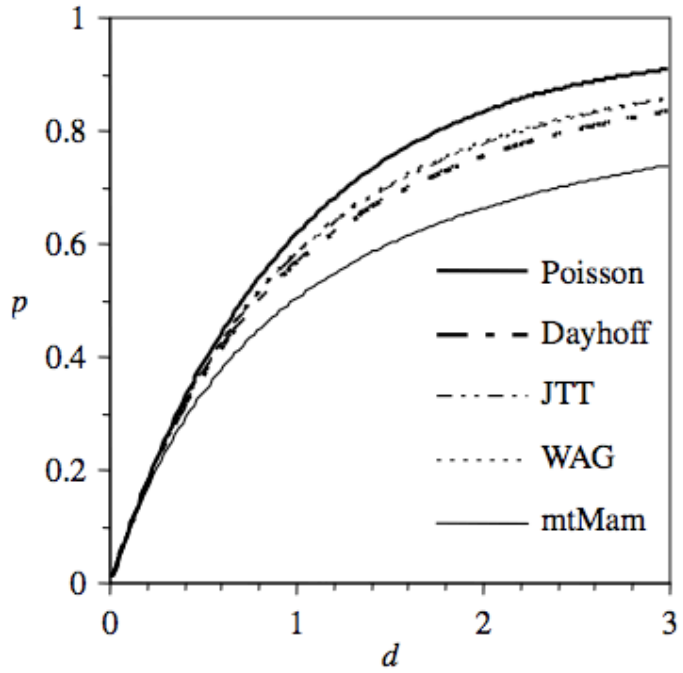


Fig. 2.2 The expected proportion of different sites (p) between two sequences separated by time or distance d under different models. The models are, from top to bottom, Poisson, WAG (Whelan and Goldman 2001), JTT (Jones *et al.* 1992), DAYHOFF (Dayhoff *et al.* 1978), and MTMAM (Yang *et al.* 1998).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

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

whereas p is the observed pairwise distance between two sequences.

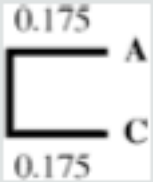
UPGMA

- Βασίζεται στην υπόθεση ότι όλες οι ακολουθίες εξελίσονται 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.



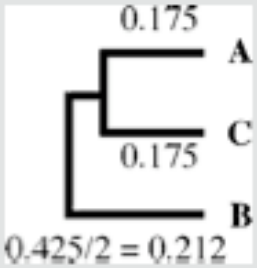
2. Because A and C are joined into a cluster, they are treated as one new composite taxon, which is used to create a reduced matrix. The distance of A-C cluster to every other taxa is one half of a taxon to A and C, respectively. That means that the distance of B to A-C is (AB + BC)/2; and that of D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

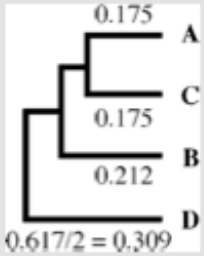


UPGMA

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.



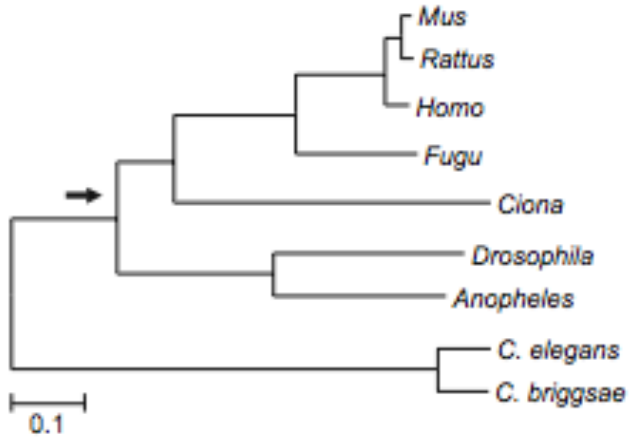
6. Because distance trees allow branches to be additive, the resulting distances between taxa from the tree path can be used to create a distance matrix. Obviously, the estimated distances do not match the actual evolutionary distances shown, which illustrates the failure of UPGMA to precisely reflect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

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

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:

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

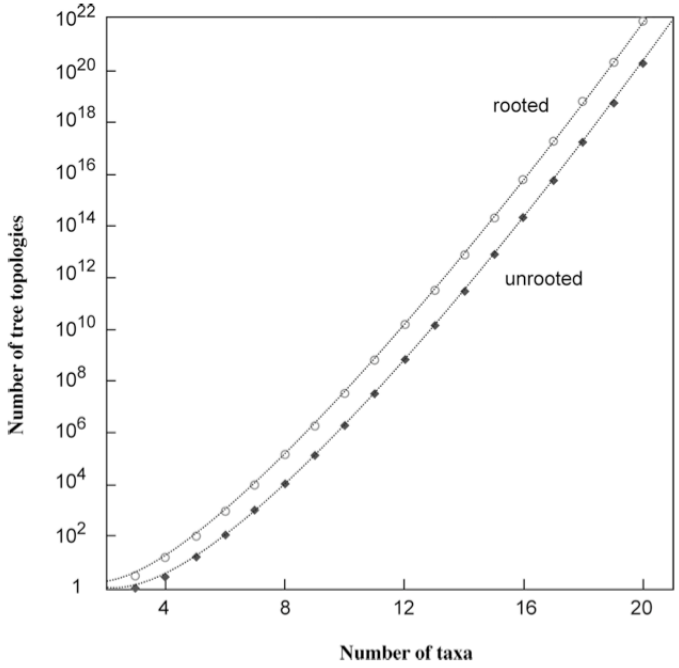
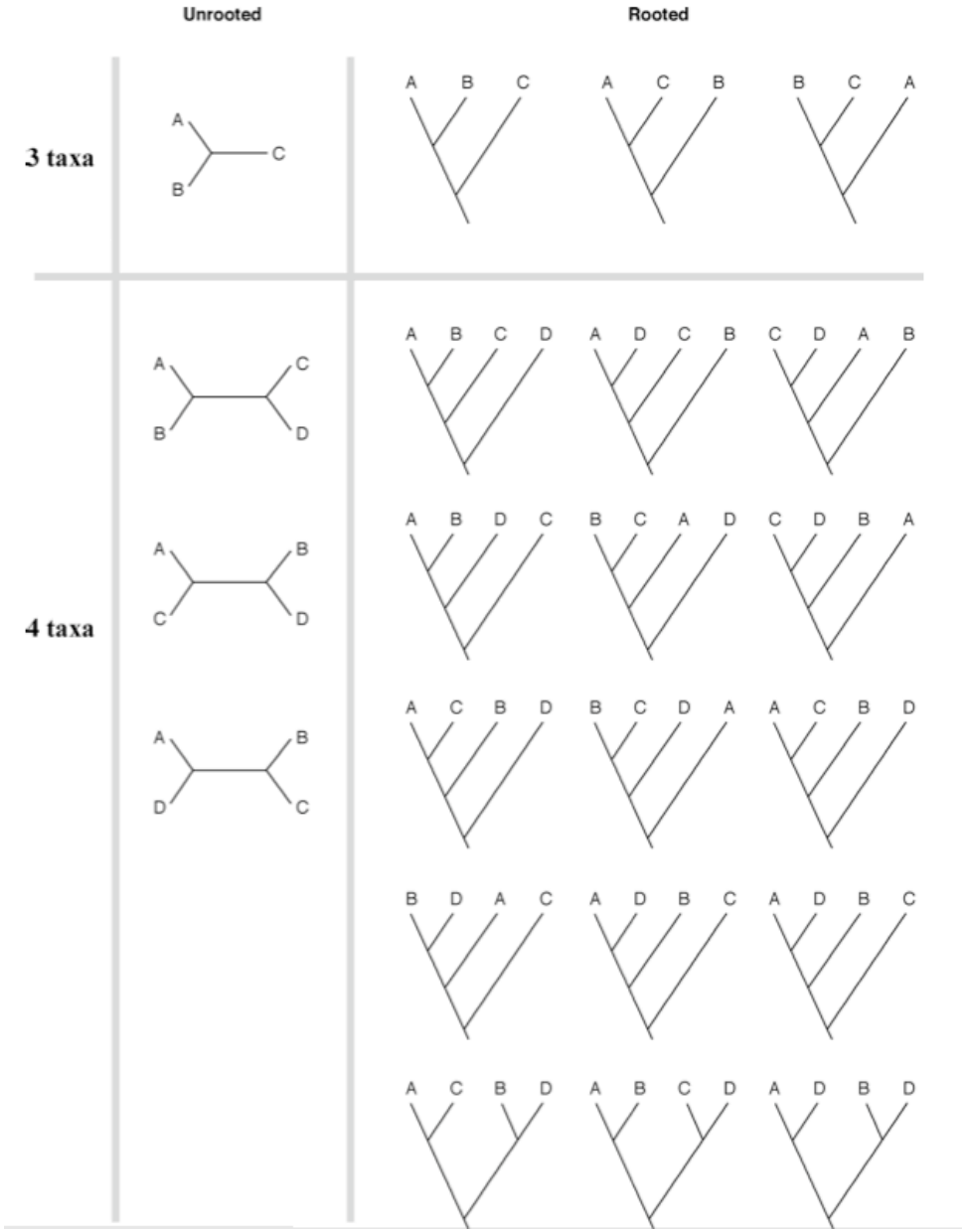


Figure 10.8: Total number of rooted (○) and unrooted (◆) tree topologies as a function of the number of taxa. The values in the *y*-axis are plotted in the log scale.

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.

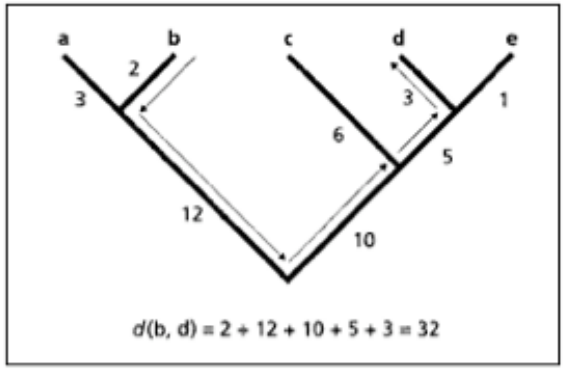


Fig. 5.21
The evolutionary distance between b and d is the sum of the edge lengths along the path in the tree between the two sequences.

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 υπόψην 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.

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.

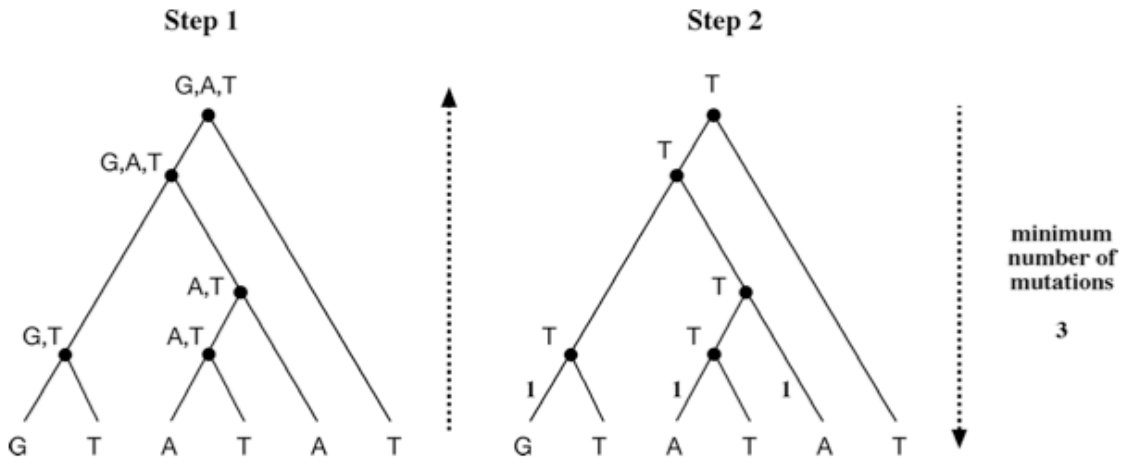


Figure 11.2: Using parsimony to infer ancestral characters at internal nodes involves a two-step procedure. The first step involves going from the leaves to the root and counting all possible ancestral characters at the internal nodes. The second step goes from the root to the leaves and assigns ancestral 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< taxa <20, τότε εφαρμόζεται το branch and bound.
- Όταν taxa > 20, εφαρμόζονται ευρετικές μέθοδοι.

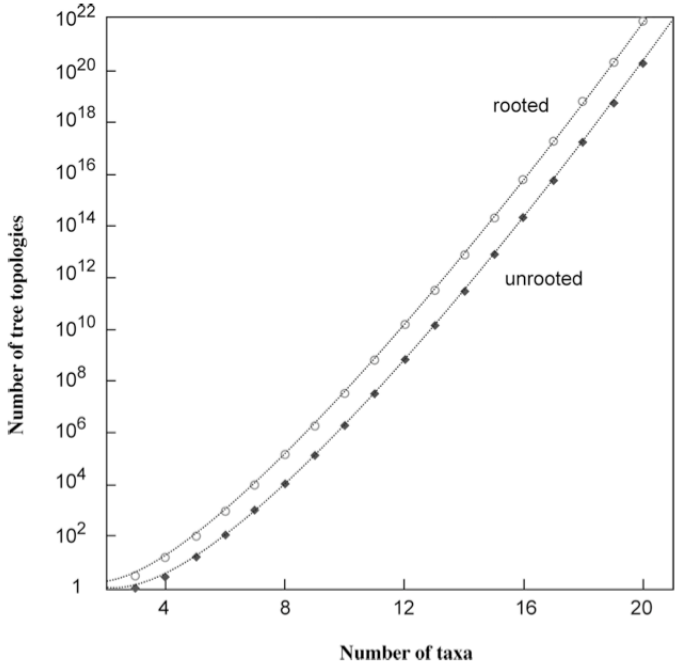


Figure 10.8: Total number of rooted (○) and unrooted (◆) tree topologies as a function of the number of taxa. The values in the *y*-axis are plotted in the log scale.

Αναζητώντας το καλύτερο δένδρο

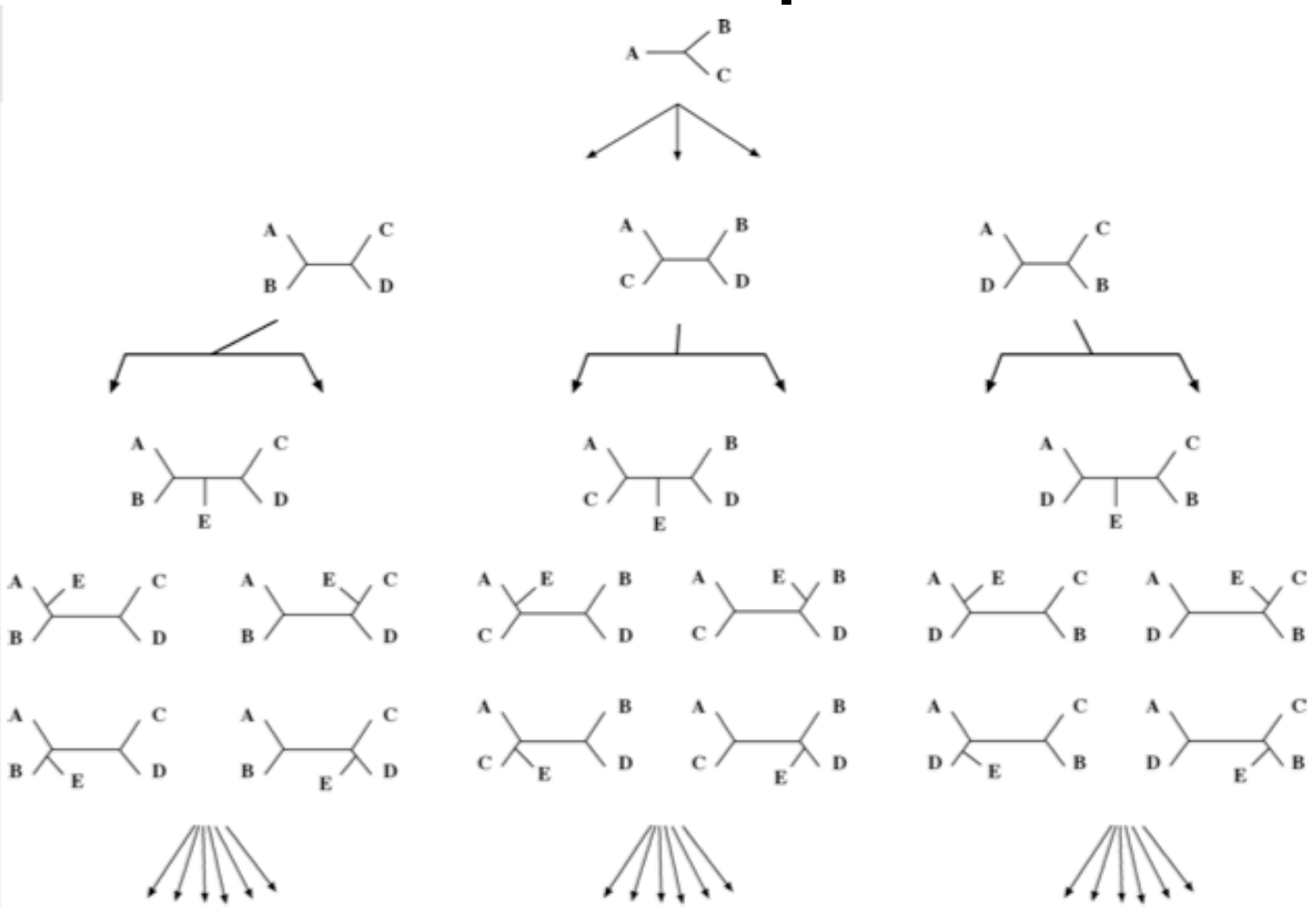


Figure 11.4: Schematic of exhaustive tree construction in the MP procedure. The tree starts with three taxa with one topology. One taxon is then added at a time in an progressive manner, during which the total branch lengths of all possible topologies are calculated.

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 συγκεκριμμένο μονοπάτι

Αναζητώντας το καλύτερο δένδρο

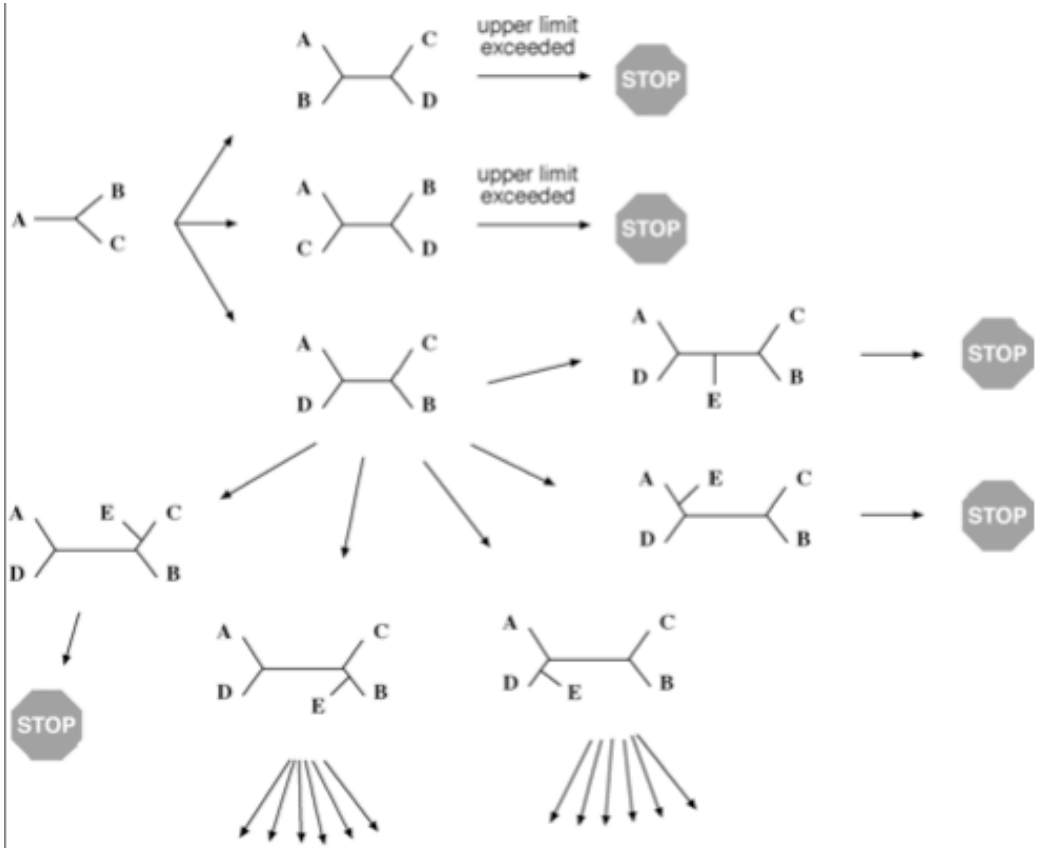


Figure 11.5: Schematic illustration of the branch-and-bound algorithm. Tree building starts with a stepwise addition of taxa of all possible topologies. Whenever the total branch length for a given topology exceeds the upper bound, the tree search in that direction stops, thereby reducing the total computing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

Αναζητώντας το καλύτερο δένδρο

- Ευρετικές μέθοδοι:
  - Δημιουργείται ένα δένδρο με neighbor joining και υπολογίζονται τα εξελικτικά βήματα για το συγκεκριμμένο 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

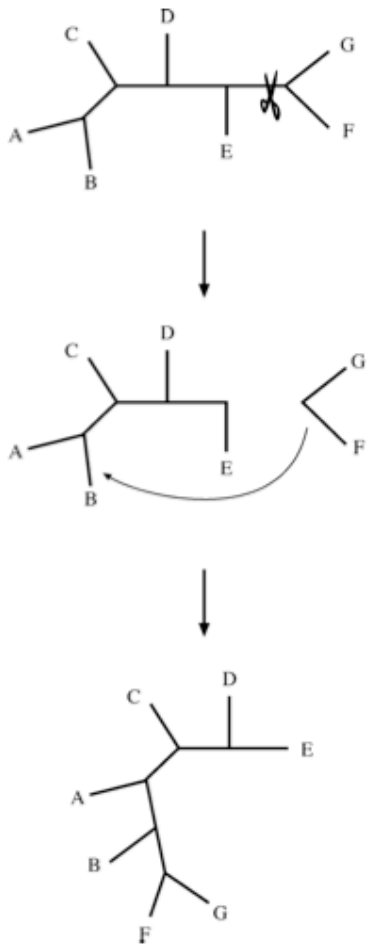


Figure 11.6: Schematic representation of a typical branch swapping process in which a branch is cut and moved to another part of the tree, generating a new topology.

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 εξελίσονται με γρήγορους ρυθμούς και επομένως έχουν μακρείς βραχίονες, έλκονται μεταξύ τους.

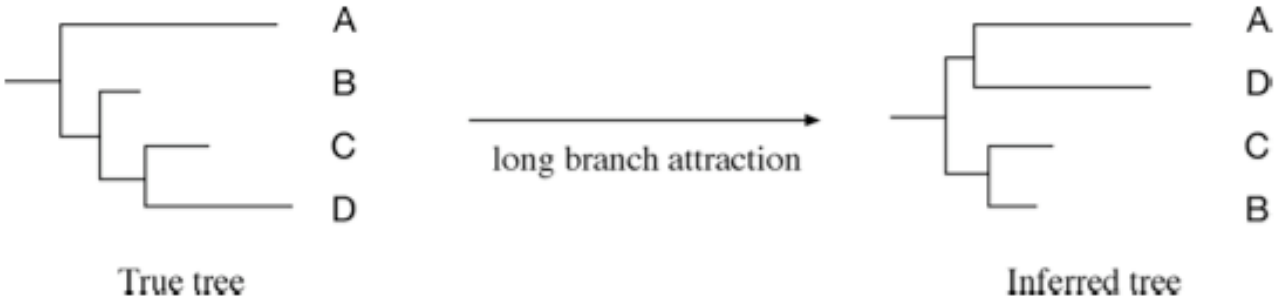


Figure 11.7: The LBA artifact showing taxa A and D are artifactually clustered during phylogenetic construction.

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 συγκεκριμμένο μοντέλο εξέλιξης (μοντέλο αντικατάστασης).
- Οι πιθανότητες μετατρέπονται σε log-likelihood scores.
- Δένδρο με το μεγαλύτερο log-likelihood score επιλέγεται.

Μέγιστη πιθανοφάνεια

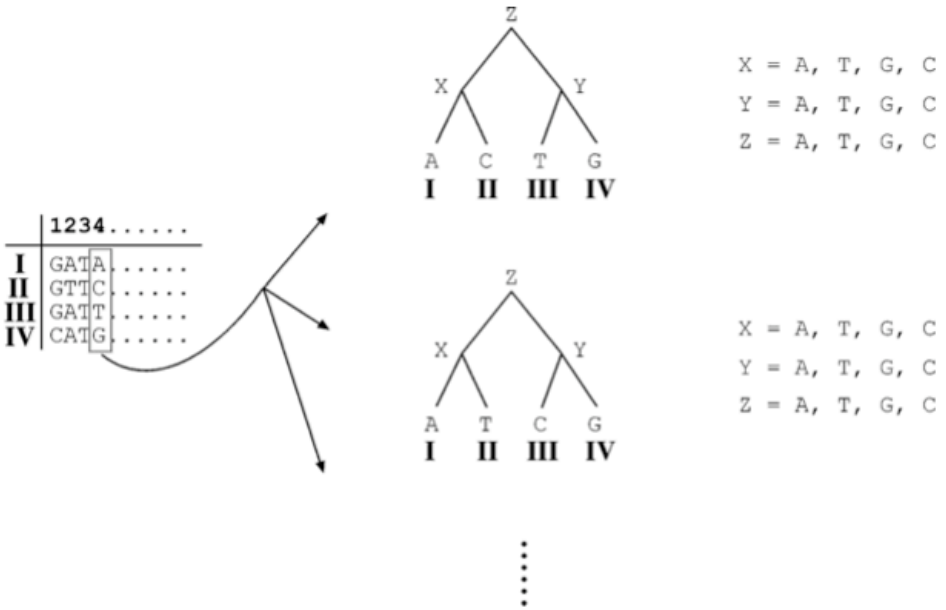


Figure 11.8: Schematic representation of the ML approach to build phylogenetic trees for four taxa, I, II, III, and IV. The ancestral character states at the internal nodes and root node are assigned X, Y, and Z, respectively. The example only shows some of the topologies derived from one of the sites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

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

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

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).
  - Boostrap -> 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

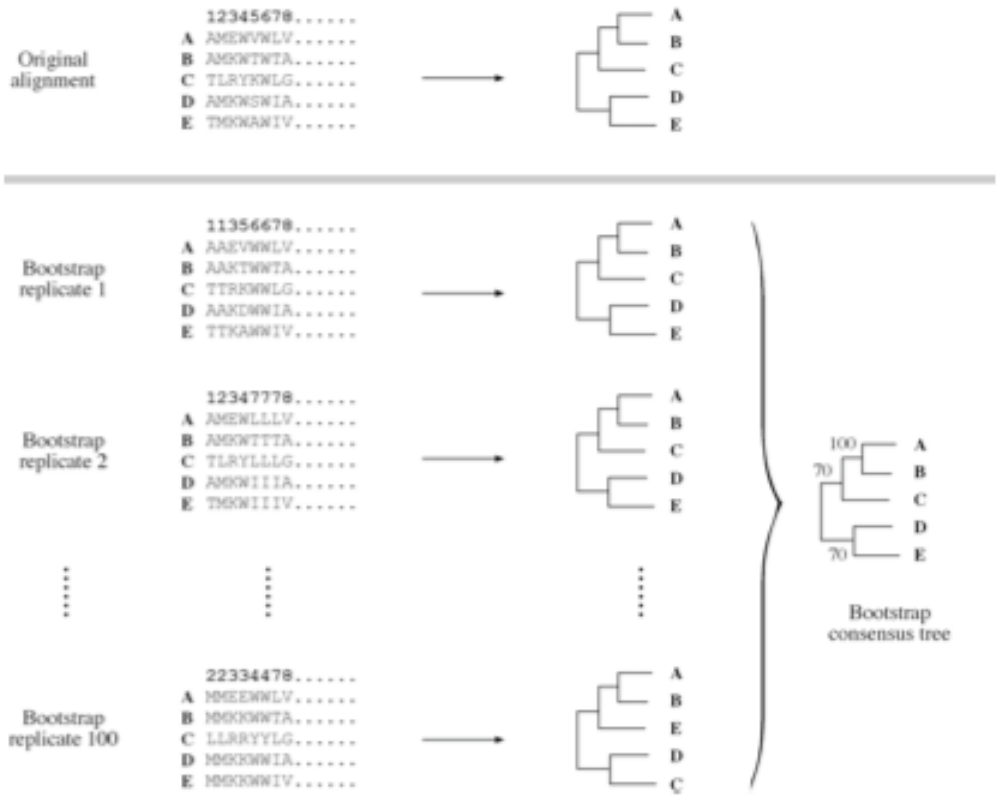


Figure 11.10: Schematic representation of a bootstrap analysis showing the original alignment and modified replicates in which certain sites are randomly replaced with other existing sites. The resulting altered replicates are used to building trees for statistical analysis at each node.

Jacknife

- Το Jacknife 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 Jacknife 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, Ν πληροφοριακές θέσεις. Για κάθε θέση, υπολογίζεται το μήκος βραχιόνων του καθένα από τα 2 δένδρα. Αυτό γίνεται και για τις Ν 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.