



# Βιοπληροφορική

## Ενότητα 16:

## Μεθοδολογίες (Ανα-) Κατασκευής, 2 ΔΩ

Τμήμα: **Βιοτεχνολογίας**

Όνομα καθηγητή: **Τ. Θηραίου**



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης





# Μαθησιακοί Στόχοι

- Επεξήγηση των μεθόδων (ανα-)κατασκευής φυλογενετικών δέντρων.
- Παρουσίαση των πλεονεκτημάτων και μειονεκτημάτων των τεχνικών αυτών.



# Λέξεις Κλειδιά

- Λέξεις κλειδιά: Ανακατασκευή βάσει αποστάσεων, Ανακατασκευή βάσει χαρακτήρων.
- Key words: Distance-Based Methods, Character-Based Methods, UPGMA, Neighbor joining, Maximum Parsimony, Maximum Likelihood, Long branch attraction.



# Μεθοδολογίες (Ανα-)Κατασκευής 1/2

## COMPUTATIONAL METHOD

Optimality criterion

Clustering algorithm

DATA TYPE

Characters

PARSIMONY

MAXIMUM LIKELIHOOD

Distances

MINIMUM EVOLUTION

LEAST SQUARES

UPGMA

NEIGHBOR-JOINING



# Μεθοδολογίες (Ανα-)Κατασκευής 2/2

- Βάσει αποστάσεων

- UPGMA (Unweighted pair group method with Arithmetic Mean)
- Neighbor joining
- Fitch-Margoliash
- Minimal Evolution

- Βάσει χαρακτήρων

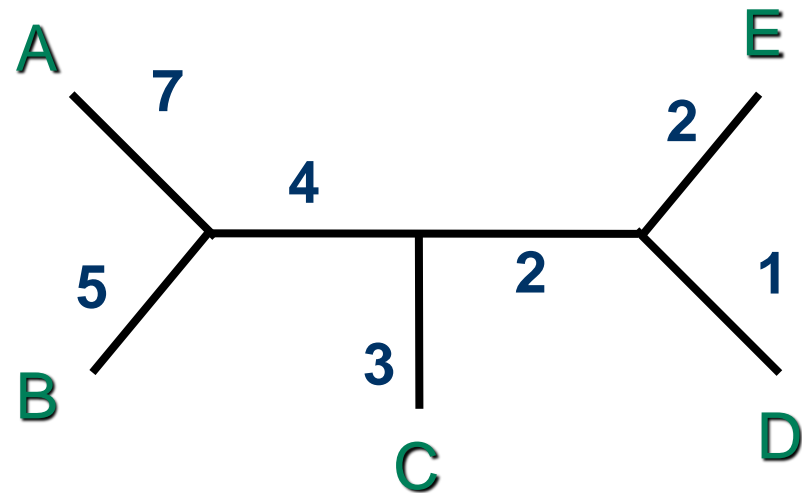
- Μέγιστη φειδωλότητα (Maximum Parsimony)
- Μέγιστη πιθανοφάνεια (Maximum Likelihood)



# Μεθοδολογία Αποστάσεων 1/4

- Δημιουργία **πίνακα αποστάσεων**  $M$  όπου  $M_{ij}$  είναι η απόσταση μεταξύ των ακολουθιών  $i$  και  $j$  και  $M_{ij} \geq 0$
- Δημιουργία ενός δένδρου όπου κάθε φύλλο του αντιστοιχεί σε ένα αντικείμενο του  $M$ , έτσι ώστε η **απόσταση** μεταξύ των **φύλλων  $i$  και  $j$**  να είναι  $M_{ij}$

	A	B	C	D	E
A	0				
B	12	0			
C	14	12	0		
D	14	12	6	0	
E	15	13	7	3	0

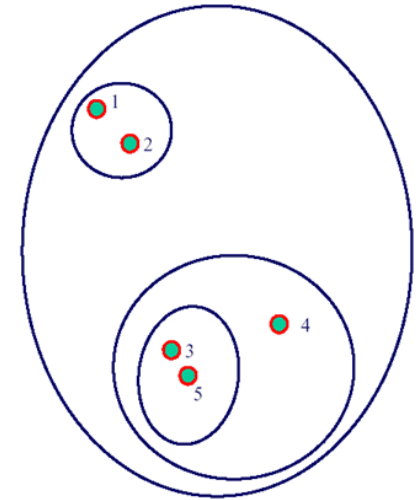




# Μεθοδολογία Αποστάσεων 2/4

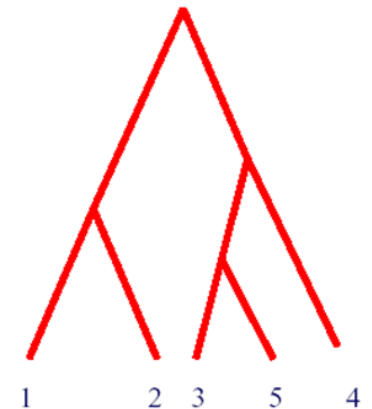
## ● Ομαδοποίηση

- αρχίζοντας από τις κοντινότερες ακολουθίες, σταδιακά ενσωματώνονται και οι πιο απομακρυσμένες
- UPGMA, Neighbor joining



## ● Βελτιστοποίηση

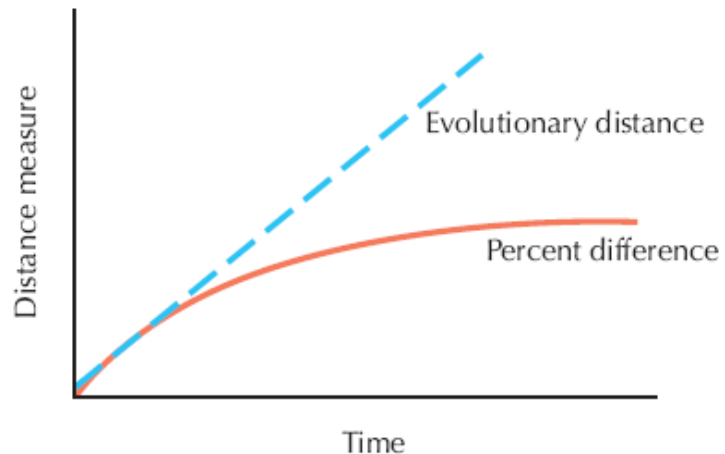
- σύγκριση των πιθανών τοπολογιών και επιλογή αυτής όπου οι αποστάσεις πάνω στο δένδρο ταιριάζουν καλύτερα με τα δεδομένα του πίνακα αποστάσεων
- Fitch-Margoliash, Minimal Evolution





# Μεθοδολογία Αποστάσεων 3/4

- Η παρατηρούμενη απόσταση **διαφέρει** από την εξελικτική απόσταση
  - **Πολλαπλές αντικαταστάσεις**
  - **Κορεσμός** όταν η εξελικτική απόσταση είναι πολύ μεγάλη
- Απαραίτητη η χρήση **μοντέλων αντικατάστασης** για τη **διόρθωση** της παρατηρούμενης απόστασης







# Μεθοδολογία Αποστάσεων 4/4

- **Δεν** είναι δυνατή η ανακατασκευή μιας **προγονικής ακολουθίας**
- μέθοδοι βελτιστοποίησης
  - καλύτερα αποτελέσματα
- μέθοδοι ομαδοποίησης
  - πιο γρήγορες
  - προτιμότερες όταν τα δεδομένα είναι πολλά



# Μέγιστη φειδωλότητα (Maximum Parsimony) 1/8

- Principle of Occam's Razor
  - When you have two competing theories that make exactly the same predictions, the simpler one is the better.
- Επιλογή του δένδρου που **ελαχιστοποιεί το πλήθος των εξελικτικών βημάτων** που απαιτούνται για την επεξήγηση των δεδομένων.



# Μέγιστη φειδωλότητα (Maximum Parsimony) 2/8

## ● Μέθοδος

- **Πολλαπλή στοίχιση** των ακολουθιών.
- Για **κάθε θέση/στήλη  $i$**  και για **κάθε δέντρο  $T$**  μετράμε το πλήθος των εξελικτικών βημάτων που απαιτούνται για να εξηγήσει το  $T$  τα δεδομένα της  $i$ .
- **"Βέλτιστο"** φυλογενετικό δέντρο  $\rightarrow$  το  $T$  με τις **λιγότερες αλλαγές** για όλες τις θέσεις.



# Μέγιστη φειδωλότητα (Maximum Parsimony) 3/8

- **non informative sites**

- θέσεις της πολλαπλής στοίχισης που δεν δίνουν πληροφορίες για την επιλογή του δέντρου

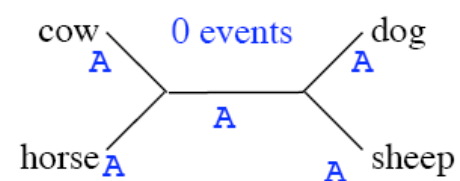
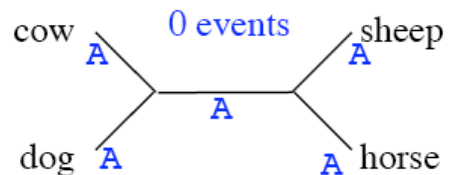
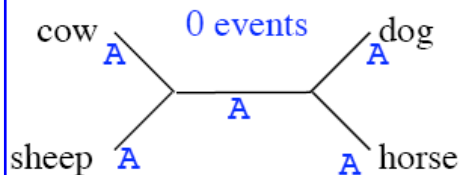
- **informative sites**

- θέσεις της πολλαπλής στοίχισης που περιέχουν τουλάχιστον **δύο διαφορετικούς χαρακτήρες** σε τουλάχιστον **δύο ακολουθίες ο καθένας**

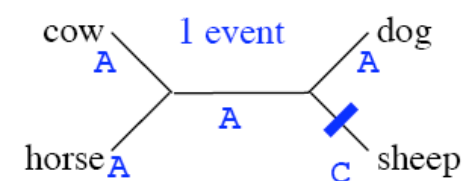
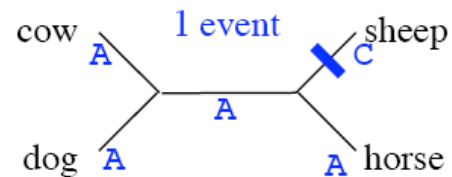
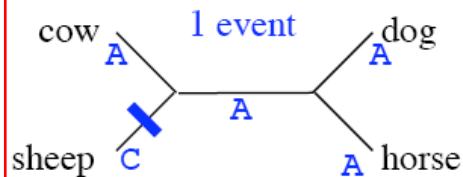


# Μέγιστη φειδωλότητα (Maximum Parsimony) 4/8

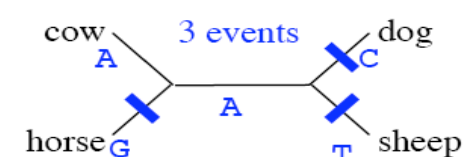
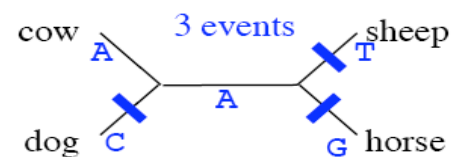
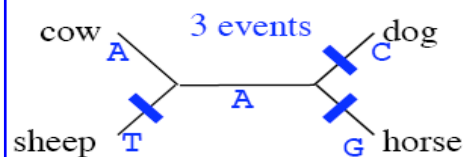
cow AAAAAACCA  
 sheep ACTACCAA  
 dog AACCCACC  
 horse AAGCAAAC



cow AAAAAACCA  
 sheep ACTACCAA  
 dog AACCCACC  
 horse AAGCAAAC

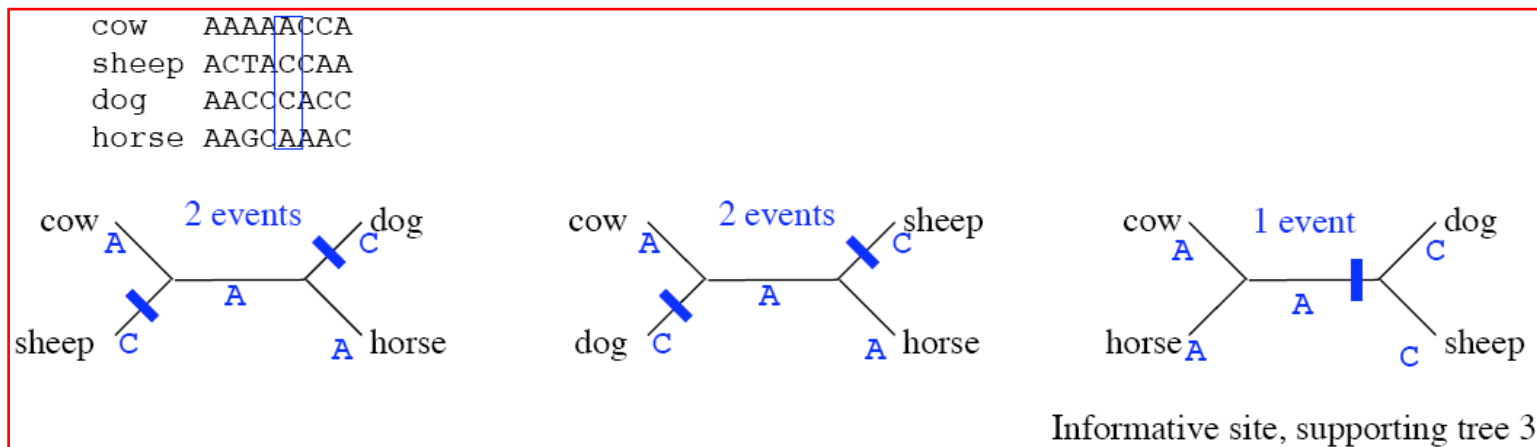
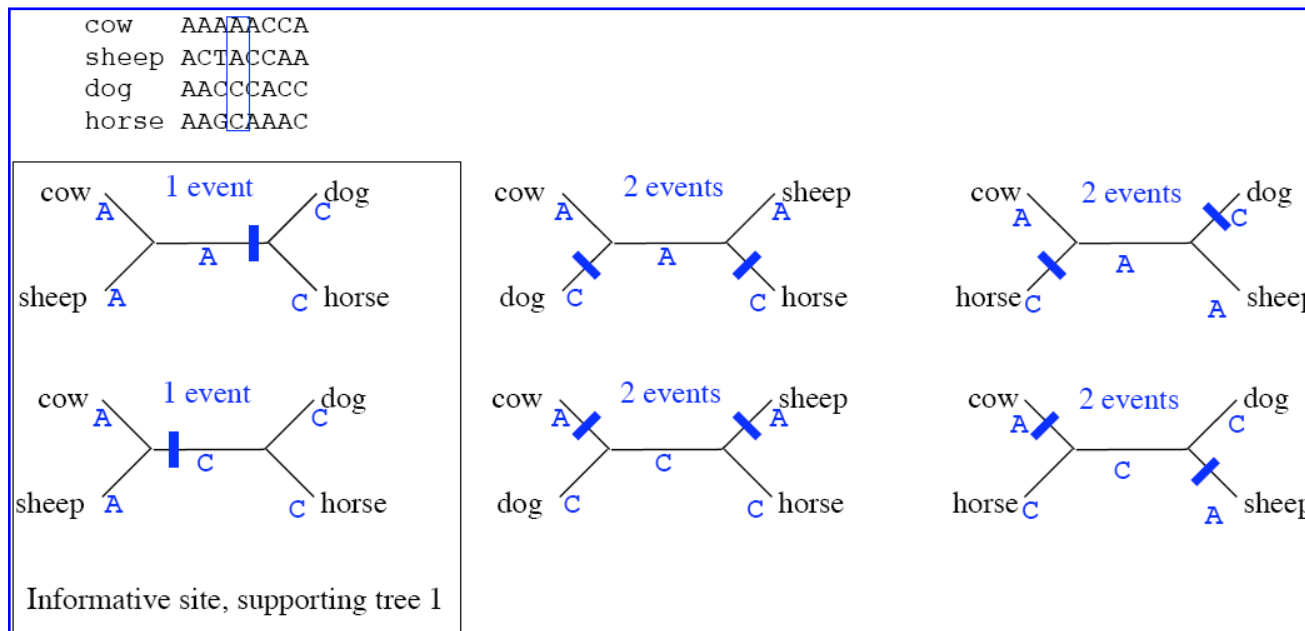


cow AAAA**A**ACCA  
 sheep ACTACCAA  
 dog AACCCACC  
 horse AAGCAAAC





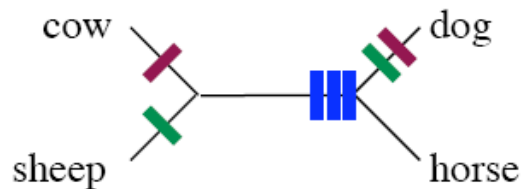
# Μέγιστη φειδωλότητα (Maximum Parsimony) 5/8





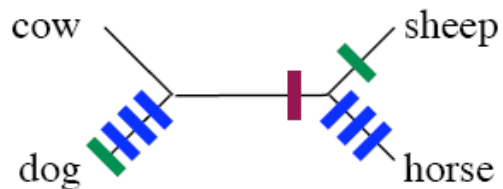
# Μέγιστη φειδωλότητα (Maximum Parsimony) 6/8

cow AAAAACCA  
 sheep ACTACCAA  
 dog AACCCACC  
 horse AAGCAAAC

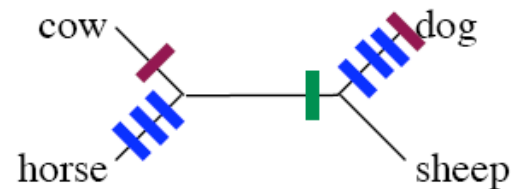


3 sites with 1 event  
 2 sites with 2 events  
 = 7 events needed

**Most parsimonious tree**



1 sites with 1 event  
 4 sites with 2 events  
 = 9 events needed



1 sites with 1 event  
 4 sites with 2 events  
 = 9 events needed



# Μέγιστη φειδωλότητα (Maximum Parsimony) 7/8

- Επιτρέπει την ανακατασκευή **προγονικών ακολουθιών**.
- Δημιουργία **δέντρου συναίνεσης**, όταν υπάρχουν περισσότερα από ένα βέλτιστα δέντρα.
- **Δεν** λαμβάνει υπόψη τις **πολλαπλές αντικαταστάσεις**.
  - Είναι προτιμότερη η χρήση της για κοντινές εξελικτικά ακολουθίες.





# Μέγιστη φειδωλότητα (Maximum Parsimony) 8/8

- **Υψηλό υπολογιστικό κόστος**
  - Παράγει όλα τα δέντρα μέχρι να καταλήξει στο βέλτιστο.
  - Δημιουργία **αλγορίθμων** που αναγνωρίζουν το καλύτερο δέντρο **χωρίς να βαθμολογούν** όλα τα εναλλακτικά.



# Μέγιστη πιθανοφάνεια (Maximum Likelihood) 1/2

- Με βάση κάποιο **εξελικτικό μοντέλο** αναζητά το δένδρο που έχει τη **μεγαλύτερη πιθανότητα** να περιγράψει τα δεδομένα.
- **Υψηλό υπολογιστικό κόστος**
  - Παράγει όλα τα δυνατά δέντρα μέχρι να καταλήξει στο βέλτιστο.



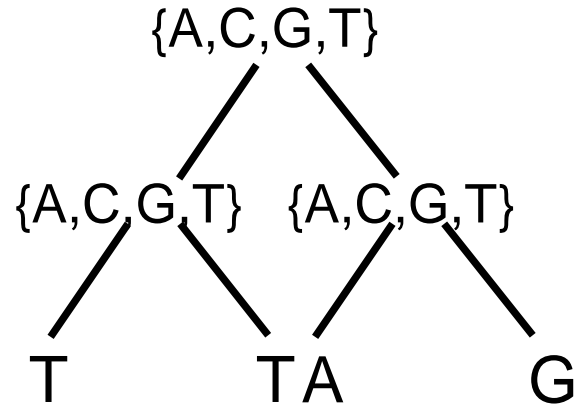
# Μέγιστη πιθανοφάνεια (Maximum Likelihood) 2/2

I. ACGCGTTGGG

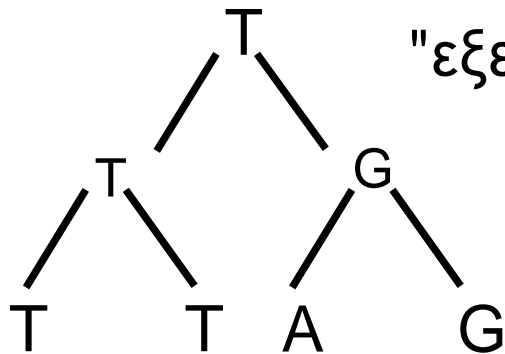
II. ACGCGTTGGG

III. ACGCAATGAA

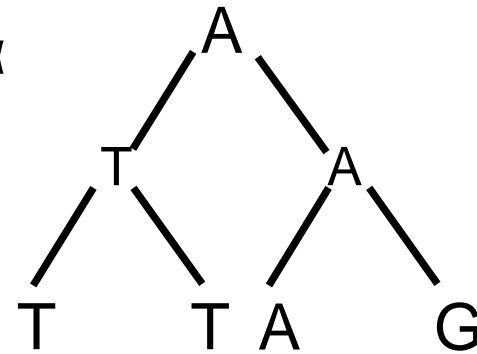
IV. ACACAGGGAA



$4 \times 4 \times 4 = 64$  δυνατοί συνδυασμοί



"εξελικτικά" δέντρα



$$P(T \rightarrow T)^3 P(T \rightarrow G) P(G \rightarrow A) P(G \rightarrow G)$$

$$P(A \rightarrow A)^2 P(A \rightarrow T) P(T \rightarrow T)^2 P(A \rightarrow G)$$



# Αξιολόγηση Φυλογενετικών Δέντρων 1/2

## A Bootstrapping

### Alignment of sequences

Species 1	ATGTTGGATGGTGAT
Species 2	ATGTTGGAAGGAGAA
Species 3	ATGTTAGGAGAAGAA
Species 4	ATGTCAGCAGCCGCC

### Bootstrapping alignment #1

Species 1	ATGTTGGAGGGTAAT
Species 2	ATGTTGGAGGGAAAA
Species 3	ATGTTAGGGGAAAAA
Species 4	ATGTCAGCGGCCCCC

### Bootstrapping alignment #2

Species 1	ATGGGGGATGGTGAT
Species 2	ATGGGGGAAGGAGAA
Species 3	ATGGGAGGAGAAGAA
Species 4	ATGGGAGCAGCCGCC

## B Jackknifing

### Alignment of sequences

Species 1	ATGTTGGATGGTGAT
Species 2	ATGTTGGAAGGAGAA
Species 3	ATGTTAGGAGAAGAA
Species 4	ATGTCAGCAGCCGCC

### Jackknifed alignment 1 (10 columns kept)

Species 1	AT-T-GGA-GG-GA-
Species 2	AT-T-GGA-GG-GA-
Species 3	AT-T-AGG-GA-GA-
Species 4	AT-T-AGC-GC-GC-

### Jackknifed alignment 2 (10 columns kept)

Species 1	--GTTGG--GGT-AT
Species 2	--GTTGG--GGA-AA
Species 3	--GTTAG--GAA-AA
Species 4	--GTCAG--GCC-CC



# Αξιολόγηση Φυλογενετικών Δέντρων 2/2

## ● Επανάληψη

- **Δειγματοληψία** στηλών της πολλαπλής στοίχισης
  - Bootstrap (με επανάθεση)
  - Jackknife (χωρίς επανάθεση)
- **Ανακατασκευή** φυλογενετικού δέντρου

## ● Προσδιορισμός του **ποσοστού εμφάνισης** ενός συγκεκριμένου κόμβου στα νέα φυλογενετικά δέντρα



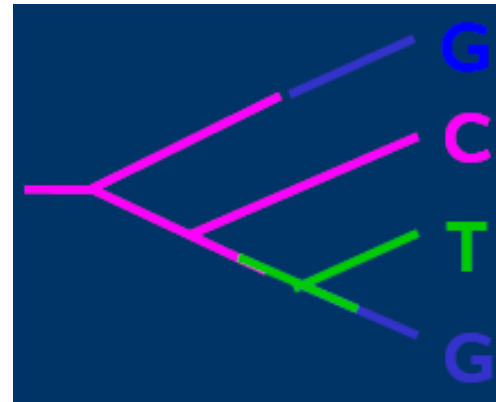
# Long Branch Attraction

## ● Ομοπλασία

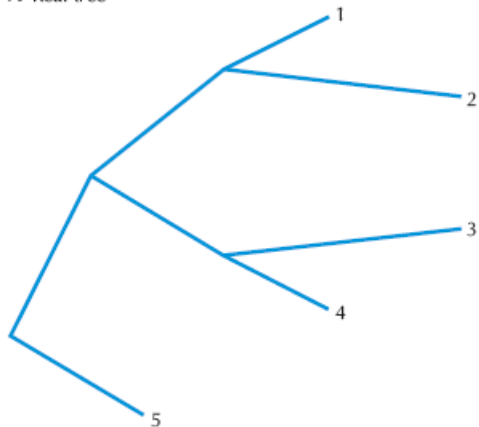
- Ομοιότητα που οφείλεται σε διαφορετικά εξελικτικά γεγονότα

## ● long branch attraction

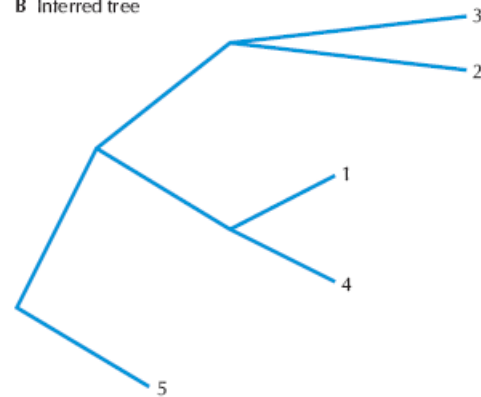
- Μακριά κλαδιά → περισσότερες γενετικές αλλαγές  
→ μεγαλύτερη πιθανότητα κάποιες θέσεις να έχουν ταυτόσημους χαρακτήρες



A Real tree



B Inferred tree





# Προγράμματα Φυλογενετικής Ανάλυσης

- PHYLIP

- <http://evolution.genetics.washington.edu/phylip.html>

- PhyML

- <http://www.atgc-montpellier.fr/phyml/>

- PAUP\* (εμπορικά διαθέσιμο)

- <http://paup.csit.fsu.edu/>



# Βιβλιογραφία

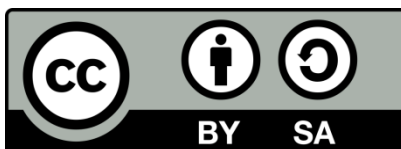
- David Mount, "Bioinformatics: Sequence and Genome Analysis", Cold Spring Harbor Laboratory Press; 2nd edition (March 12, 2013).
- Jonathan Pevsner, "Bioinformatics and Functional Genomics", Wiley-Blackwell; 2nd edition (May 4, 2009).
- Andreas D. Baxevanis, B. F. Francis Ouellette, "Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins", Wiley-Interscience; 3rd edition (October 29, 2004).





# Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδεια χρήσης, η άδεια χρήσης αναφέρεται ρητώς.





# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα Γεωπονικού Πανεπιστημίου Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ  
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ  
*επένδυση στην κοινωνία της γνώσης*  
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ  
2007-2013  
πρόγραμμα για την ανάπτυξη  
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



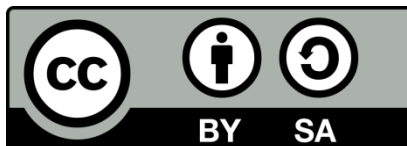
# Σημείωμα Αναφοράς

Copyright Γεωπονικό Πανεπιστήμιο Αθηνών 2015. Τμήμα Βιοτεχνολογίας, Θηραίου Τριάς. «Βιοπληροφορική». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:  
<https://mediasrv.aua.gr/eclass/courses/OCDB100/>



# Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων, π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Η άδεια αυτή ανήκει στις άδειες που ακολουθούν τις προδιαγραφές του Ορισμού Ανοικτής Γνώσης [2], είναι ανοικτό πολιτιστικό έργο [3] και για το λόγο αυτό αποτελεί ανοικτό περιεχόμενο [4].

[1] <http://creativecommons.org/licenses/by-sa/4.0/>

[2] <http://opendefinition.org/okd/ellinika/>

[3] <http://freedomdefined.org/Definition/EI>

[4] <http://opendefinition.org/buttons/>



# Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
  - το Σημείωμα Αδειοδότησης
  - τη δήλωση Διατήρησης Σημειωμάτων
  - το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)
- μαζί με τους συνοδευόμενους υπερσυνδέσμους.