



# Βιοπληροφορική

## Ενότητα 4<sup>η</sup>: Φυλογενετική ανάλυση 1

Ηλίας Καππάς  
Τμήμα Βιολογίας



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

# Άδειες Χρήσης

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



# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη 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.



# Περιεχόμενα ενότητας

- Φυλογένεση
- Ομολογία – Ομοπλασία
- Ορισμοί – Ορολογία
- Προσθήκη ρίζας
- Μοντέλα εξέλιξης



# Φυλογένεση *phylogeny, phylogenetics*

Η εύρεση των γενεαλογικών σχέσεων ταξινομικών μονάδων

Η επιστήμη που διατυπώνει και ελέγχει υποθέσεις για τα ιστορικά πρότυπα προέλευσης λειτουργικών ταξινομικών μονάδων χρησιμοποιώντας εξελικτικά δέντρα

(*Operational Taxonomic Units* → είδη, πληθυσμοί, άτομα)

επιλεγμένο επίπεδο ταξινομικής δειγματοληψίας

ΧΑΡΑΚΤΗΡΕΣ *characters* → ποσοτικοί (συνεχείς, μεριστικοί)  
→ ποιοτικοί

Κλασική φυλογένεση → μορφολογία

Σύγχρονη φυλογενετική → DNA, πρωτεΐνες



Θέσεις *sites*



Φυλογενετικό δέντρο

*phylogenetic tree*



# Ομολογία *homology*

Ομοιότητα λόγω κοινής καταγωγής *common ancestry*

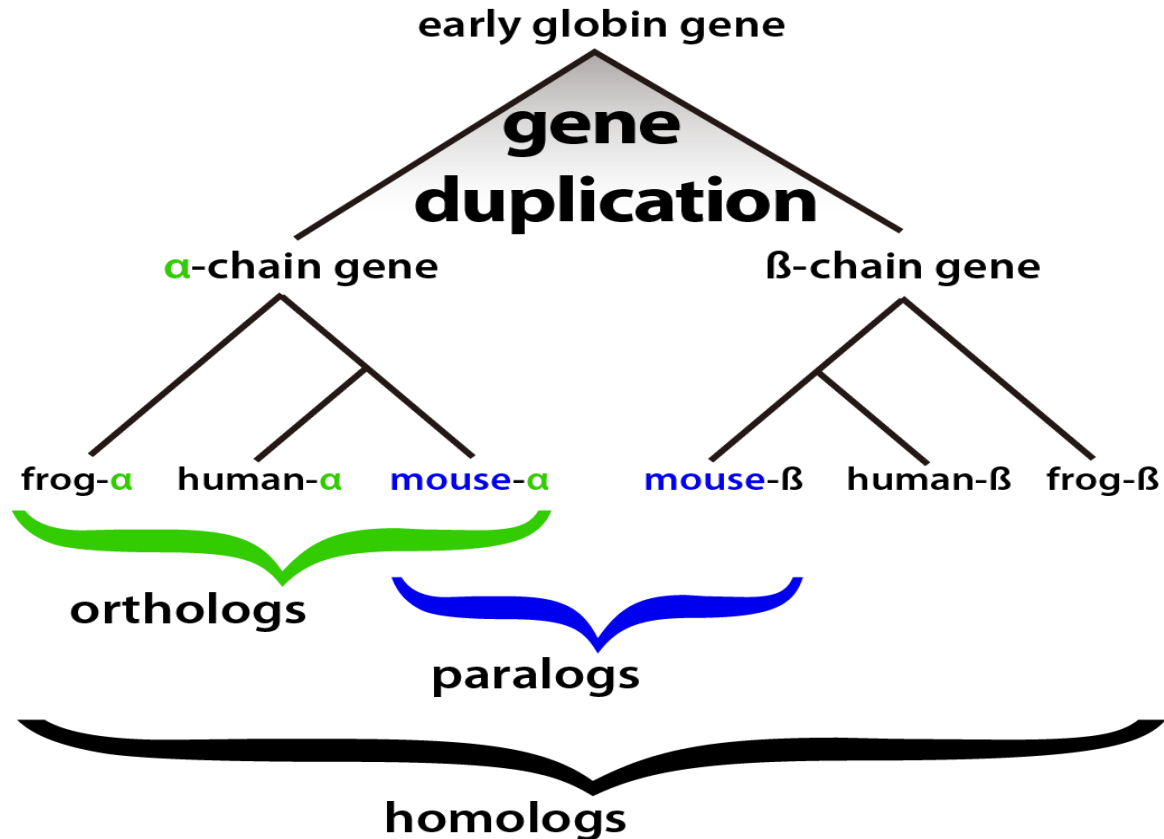
- Ορθόλογα *orthologues* → ειδογένεση
- Παράλογα *paralogues* → γονιδιακός διπλασιασμός
- Ξενόλογα *xenologues* → οριζόντια γονιδιακή μεταφορά

... το ίδιο γονίδιο σε διαφορετικούς συγγενικούς οργανισμούς

... διαφορετικά συγγενικά γονίδια στον ίδιο οργανισμό



# Ομολογία



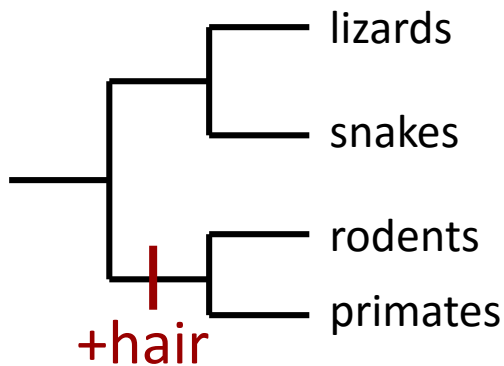
**Εικόνα 1:** Το ίδιο γονίδιο σε διαφορετικούς συγγενικούς οργανισμούς.



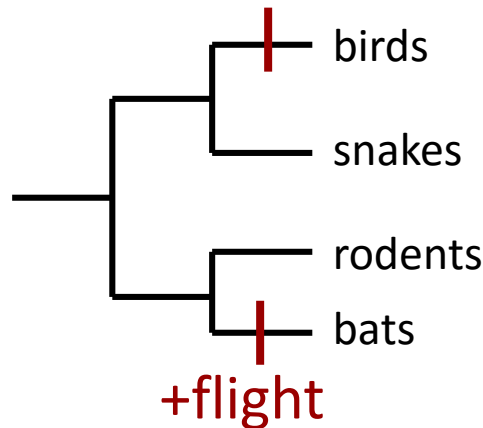
# Ομολογία και Ομοπλασία

**Ομολογία *homology*:** ταυτόσημος χαρακτήρας λόγω κοινής καταγωγής (εξελικτικό σήμα)

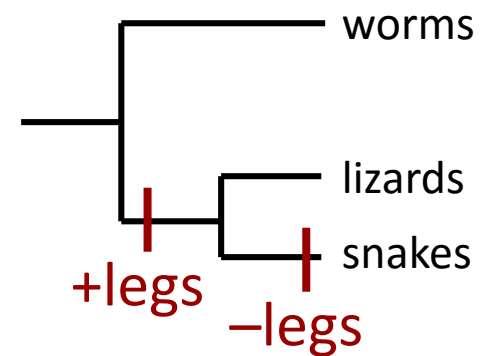
**Ομοπλασία *homoplasy*:** ταυτόσημος χαρακτήρας λόγω εξελικτικής σύγκλισης ή αναστροφής (εξελικτικός θόρυβος)



Homology



Homoplasy  
(convergence)



Homoplasy  
(reversal)





# Ορισμοί – Ορολογία (1/6)

- **Ομαδοποίηση *clustering*** → δημιουργία κλάσεων ομοίων στοιχείων
- **Δέντρο *tree*** → δισδιάστατο γράφημα που περιέχει κόμβους και βραχίονες  
(μεταξύ δύο κόμβων μόνο μία διαδρομή)
- **Τοπολογία *topology*** ή **πρότυπο διακλάδωσης *branching pattern*** → απεικόνιση των σχέσεων των ταξινομικών μονάδων
- **Μήκη βραχιόνων *branch lengths*** → ποσοτική απεικόνιση διαφορών ή χρονικής απόκλισης



# Ορισμοί – Ορολογία (2/6)

- **Ρίζα root** → μια ακολουθία ή ένα σημείο που θεωρούμε ως κοινό πρόγονο όλων των ακολουθιών του δείγματός μας

Unrooted tree → τοπολογία

Rooted tree → τοπολογία + κατεύθυνση εξέλιξης

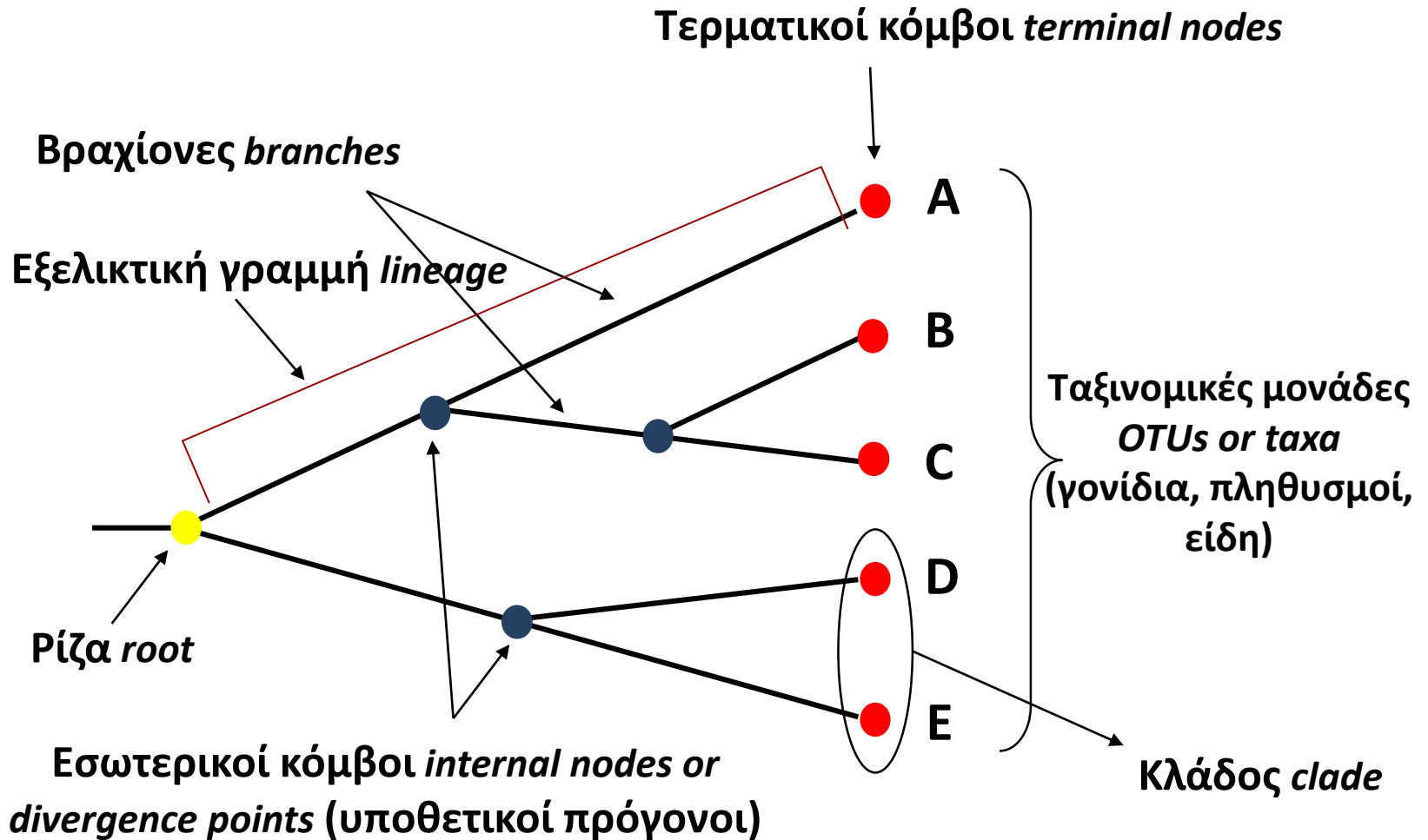
- **Εξωομάδα outgroup** → μία ακολουθία που απέκλινε εξελικτικά από την εσωομάδα πριν από την απόκλιση των μελών της εσωομάδας

ενδιάμεσος βαθμός συγγένειας με την **εσωομάδα ingroup**

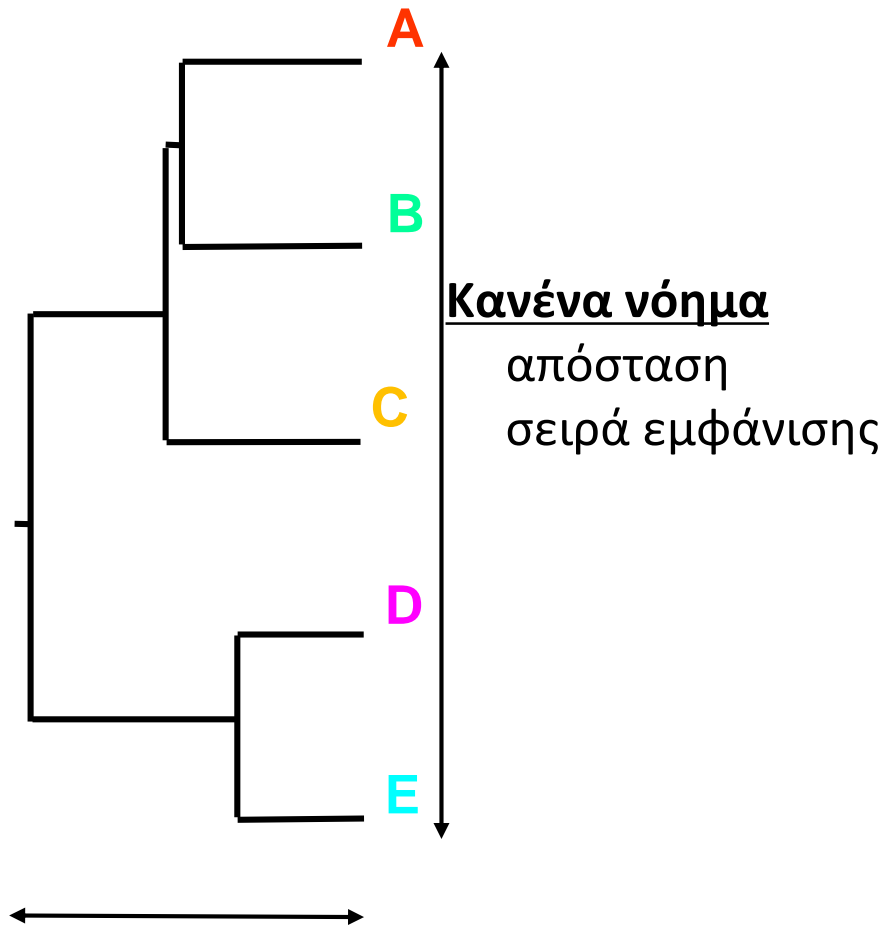
- **Σημείο του μέσου midpoint** → το μέσο του βραχίονα με το μεγαλύτερο μήκος



# Ορισμοί – Ορολογία (3/6)



# Ορισμοί – Ορολογία (4/6)



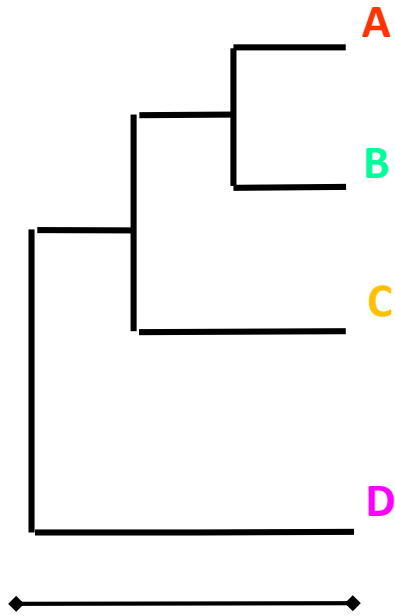
- Απουσία κλίμακας → κλαδόγραμμα *cladogram*
- Ανάλογη γενετικής απόστασης ή ποσού αλλαγής → φυλόγραμμα *phylogram*
- Ανάλογη του χρόνου → χρονόγραμμα *chronogram*

$((C,(A,B)),(D,E))$  → μορφή ένθετων παρενθέσεων *nested parentheses*



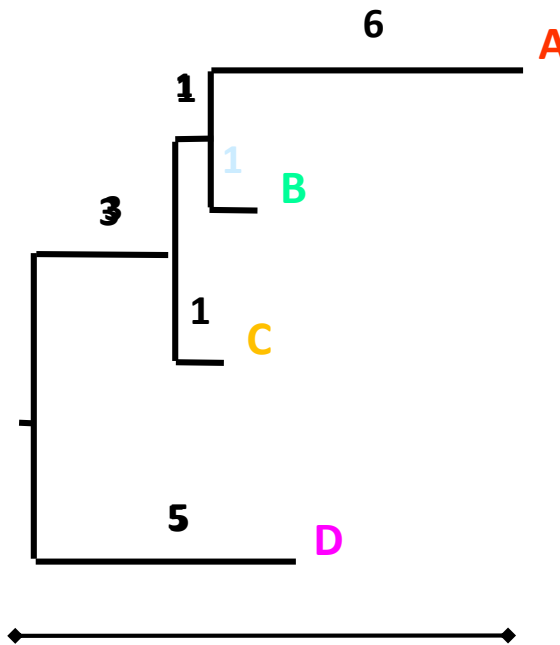
# Ορισμοί – Ορολογία (5/6)

Cladogram



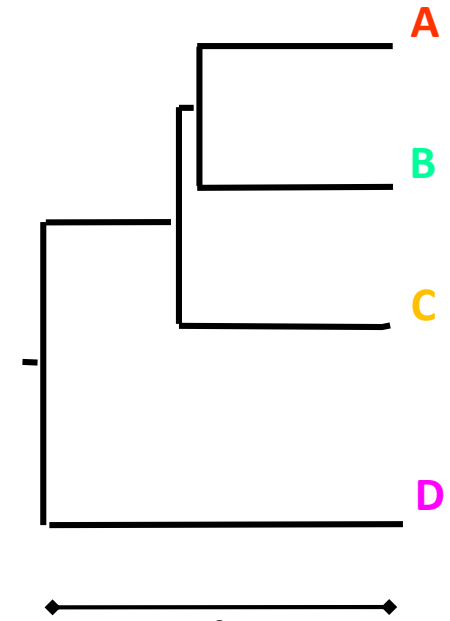
απουσία  
κλίμακας

Phylogram



ποσό αλλαγής

Ultrametric tree

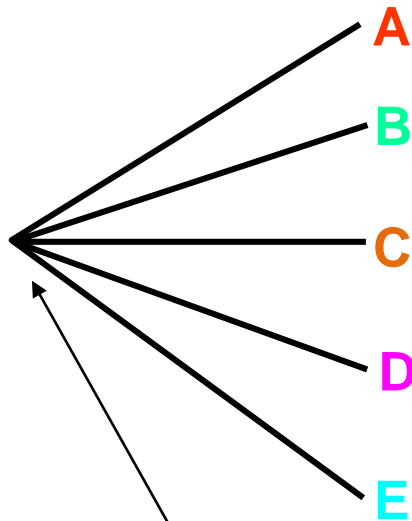


χρόνος



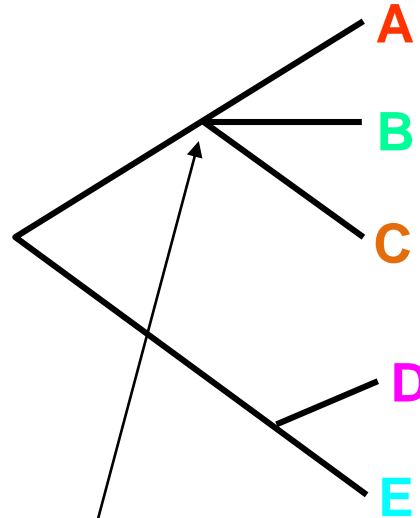
# Ορισμοί – Ορολογία (6/6)

Μη επιλυθείσα φυλογένεση  
*star phylogeny*

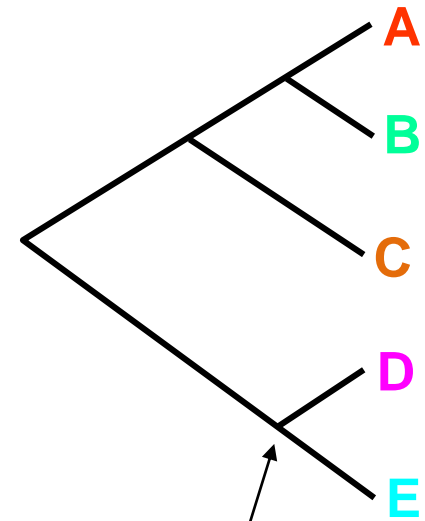


Πολυτομία *polytomy or multifurcation*

Μερικώς  
επιλυθείσα  
*partially resolved  
phylogeny*



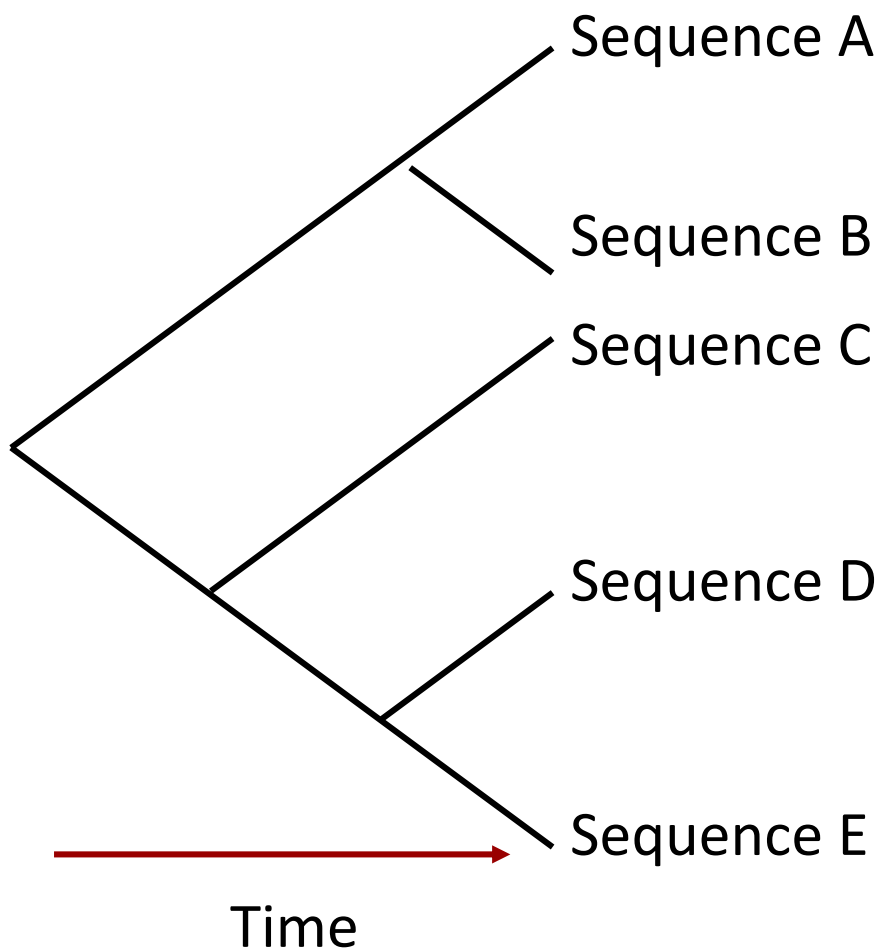
Πλήρως επιλυθείσα  
*fully resolved,  
bifurcating phylogeny*



Διακλάδωση *bifurcation*



# Ερμηνεία φυλογένεσης



Ποια ακολουθία είναι περισσότερο συγγενική προς την ακολουθία B?

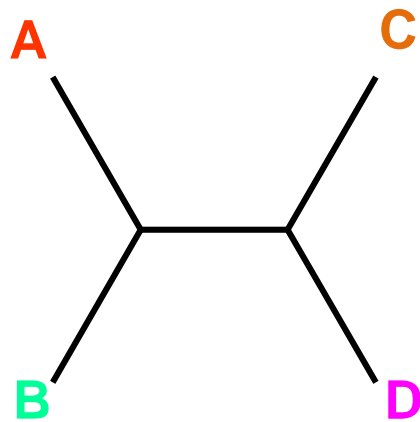
**A**, διότι η B απέκλινε από την A πιο πρόσφατα σε σχέση με οποιαδήποτε άλλη ακολουθία ... A και B κοντινότεροι συγγενείς



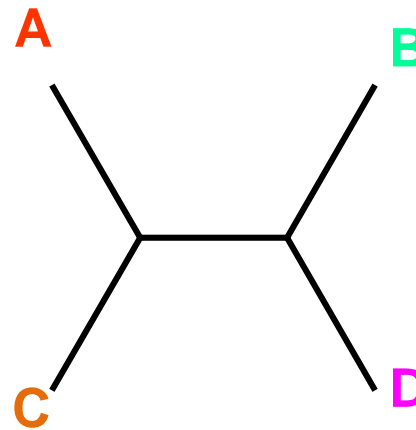
# Προσθήκη ρίζας (1/5)

Τρία πιθανά δέντρα χωρίς ρίζα για 4 taxa (A, B, C, D)

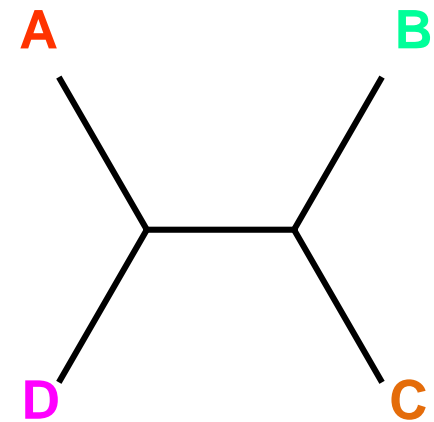
Tree 1



Tree 2



Tree 3





# Προσθήκη ρίζας (2/5)

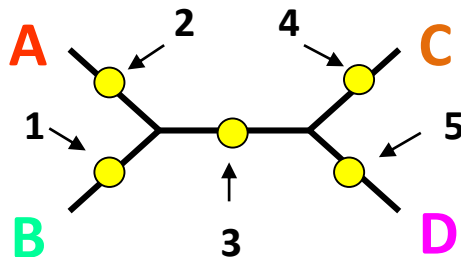
Taxa	Unrooted trees	x	Roots	=	Rooted trees
3	1		3		3
4	3		5		15
5	15		7		105
6	105		9		945
7	945		11		10.395
8	10.395		13		135.135
9	135.135		15		2.027.025
10	2.027.025		17		34.489.707
.	.		.		.
.	.		.		.
.	.		.		.
.	.		.		.
50	$3 \times 10^{74}$		97		$2,8 \times 10^{76}$

$$\frac{(2n-5)!}{n-3} \quad (n \geq 3) \quad \underline{\text{άτομα στο σύμπαν} \approx 10^{80}}$$

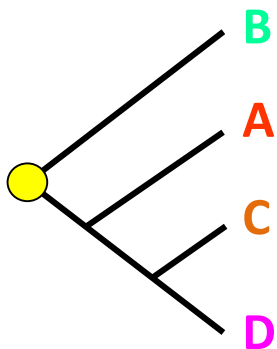


# Προσθήκη ρίζας (3/5)

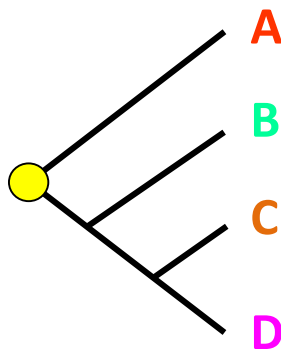
Unrooted tree



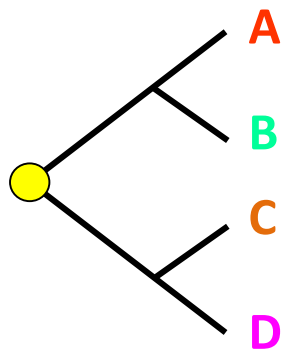
Rooted tree 1



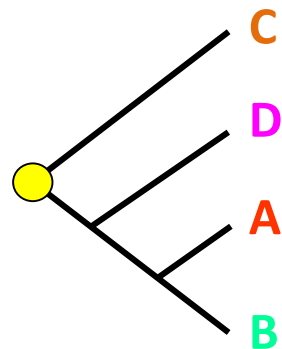
Rooted tree 2



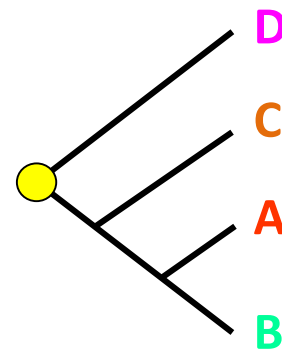
Rooted tree 3



Rooted tree 4



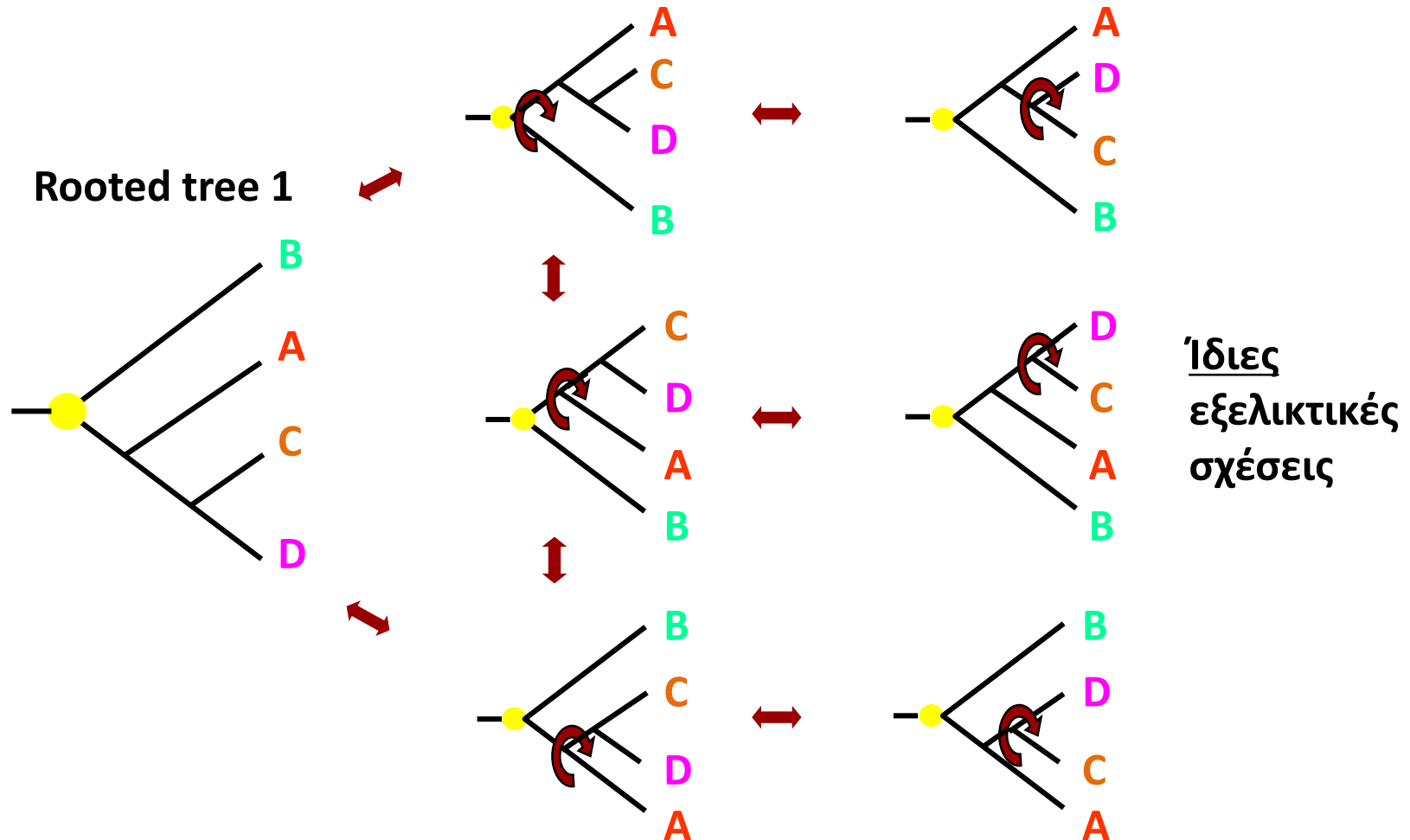
Rooted tree 5



Διαφορετικές εξελικτικές σχέσεις



# Προσθήκη ρίζας (4/5)



# Προσθήκη ρίζας (5/5)

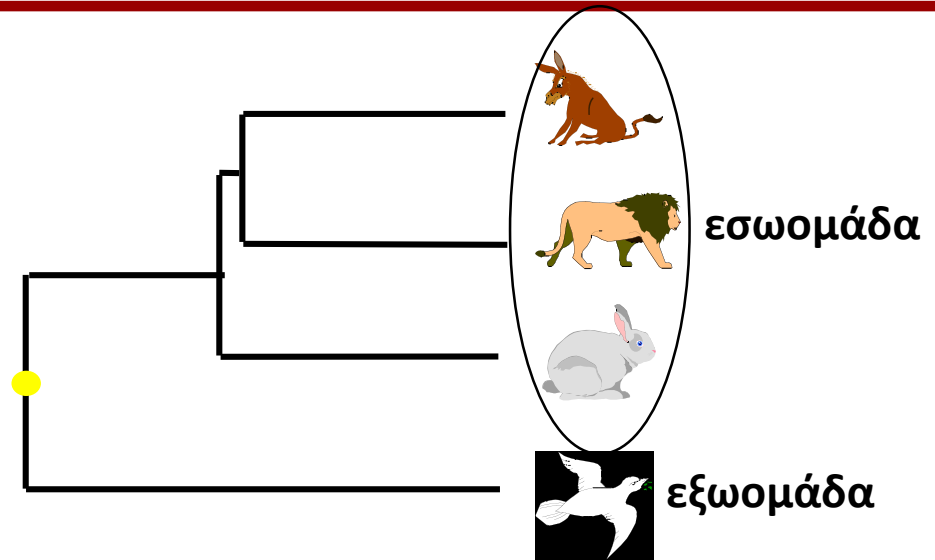
## (1) Εξωμάδα *outgroup*

Ταχον που βρίσκεται εξωτερικά της εσωμάδας *ingroup*

Απαραίτητη κάποια γνώση των εξελικτικών σχέσεων μεταξύ των μελετούμενων taxa

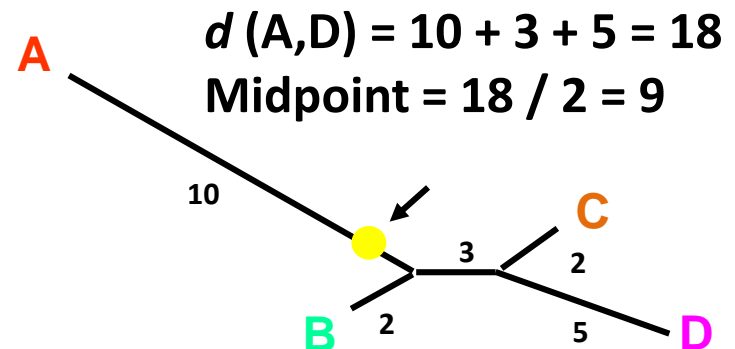
π.χ. πτηνό → θηλαστικά

$\alpha$ -globin →  $\beta$ -globins

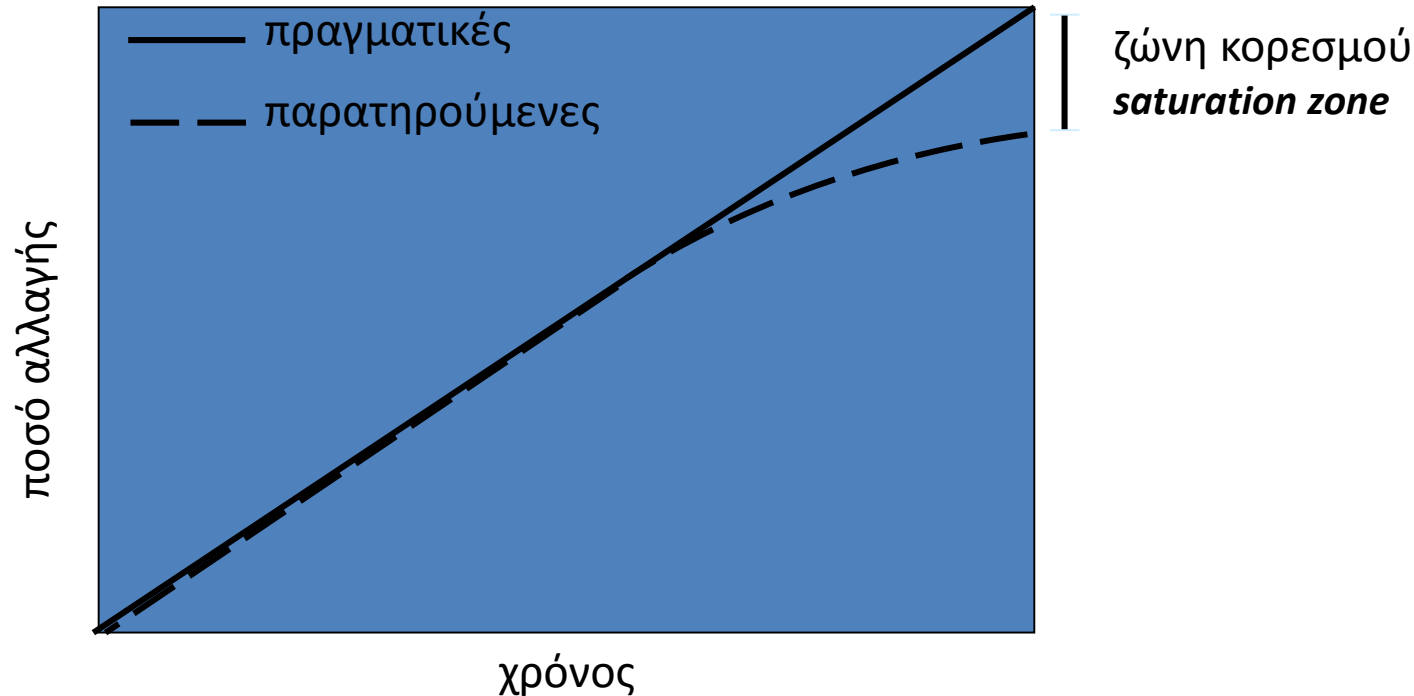


## (2) Σημείο του μέσου *midpoint*

Στο μέσο της απόστασης μεταξύ των δύο πιο απομακρυσμένων taxa



# Μοντέλα εξέλιξης (1/10)



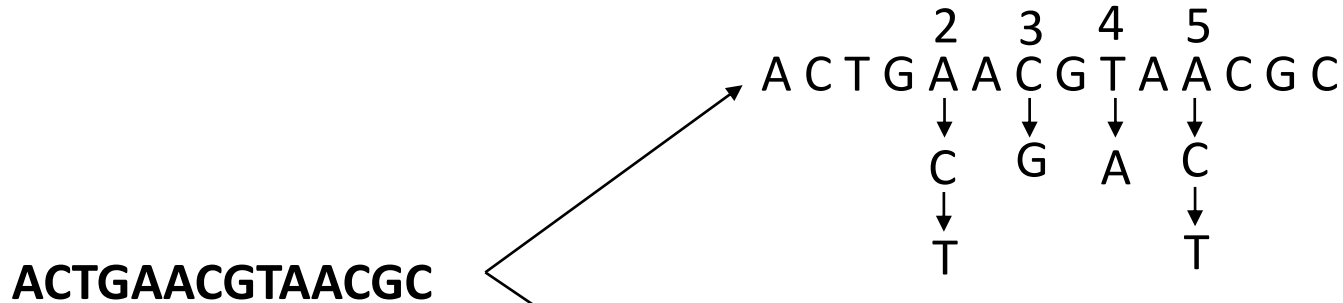
- (1) Υπολογισμός πραγματικών αλλαγών από τις παρατηρούμενες → μοντέλο εξέλιξης
- (2) Απαραίτητη η προσαρμογή των τιμών σε ένα φυλογενετικό δέντρο όσο το δυνατό καλύτερα

(1) → παραδοχές **assumptions** (2) → αλγοριθμικό **algorithmic**



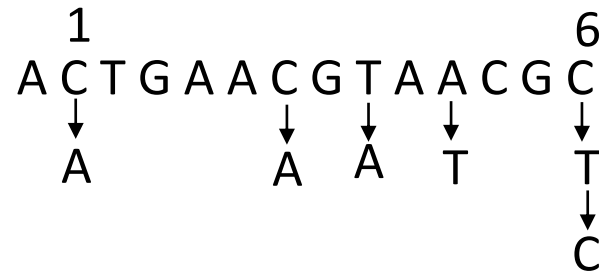
# Μοντέλα εξέλιξης (2/10)

## Προγονική ακολουθία *ancestral sequence*



## Αντικαταστάσεις *substitutions*

- 1 απλή *single*
- 2 διαδοχική *sequential*
- 3 συμπτωματική *coincidental*
- 4 παράλληλη *parallel*
- 5 συγκλίνουσα *convergent*
- 6 ανάστροφη *back*



# Μοντέλα εξέλιξης (3/10)

## Παραδοχές *assumptions*

- Σωστή ακολουθία
- Ομολογία
- Σωστή στοίχιση
- Κοινή ιστορική σχέση κάθε ακολουθίας με τις υπόλοιπες (όχι μείγματα ακολουθιών)
- Η αλλαγή σε μια θέση είναι ανεξάρτητη από τις αλλαγές σε άλλες θέσεις
- Όλες οι ακολουθίες έχουν εξελιχθεί με στοχαστικές διαδικασίες
- Επαρκής δειγματοληψία
- Πολυμορφισμός αντιπροσωπευτικός του ευρύτερου πληθυσμού
- Επαρκές φυλογενετικό σήμα



# Μοντέλα εξέλιξης (4/10)

## Κρίσιμα σημεία

Εισαγωγές - Απαλείψεις βάσεων *insertions-deletions (indels or gaps)*

χωρίς ικανοποιητική απάντηση

Συχνότητες βάσεων *base frequencies*

Ομοιογενές μοτίβο αλλαγών *homogeneity assumption*

Μεταθέσεις - Μεταστροφές *transitions-transversions* →  $t_s > t_v$

βιοχημική δομή βάσεων, χημικές ιδιότητες  
συμπληρωματικότητας

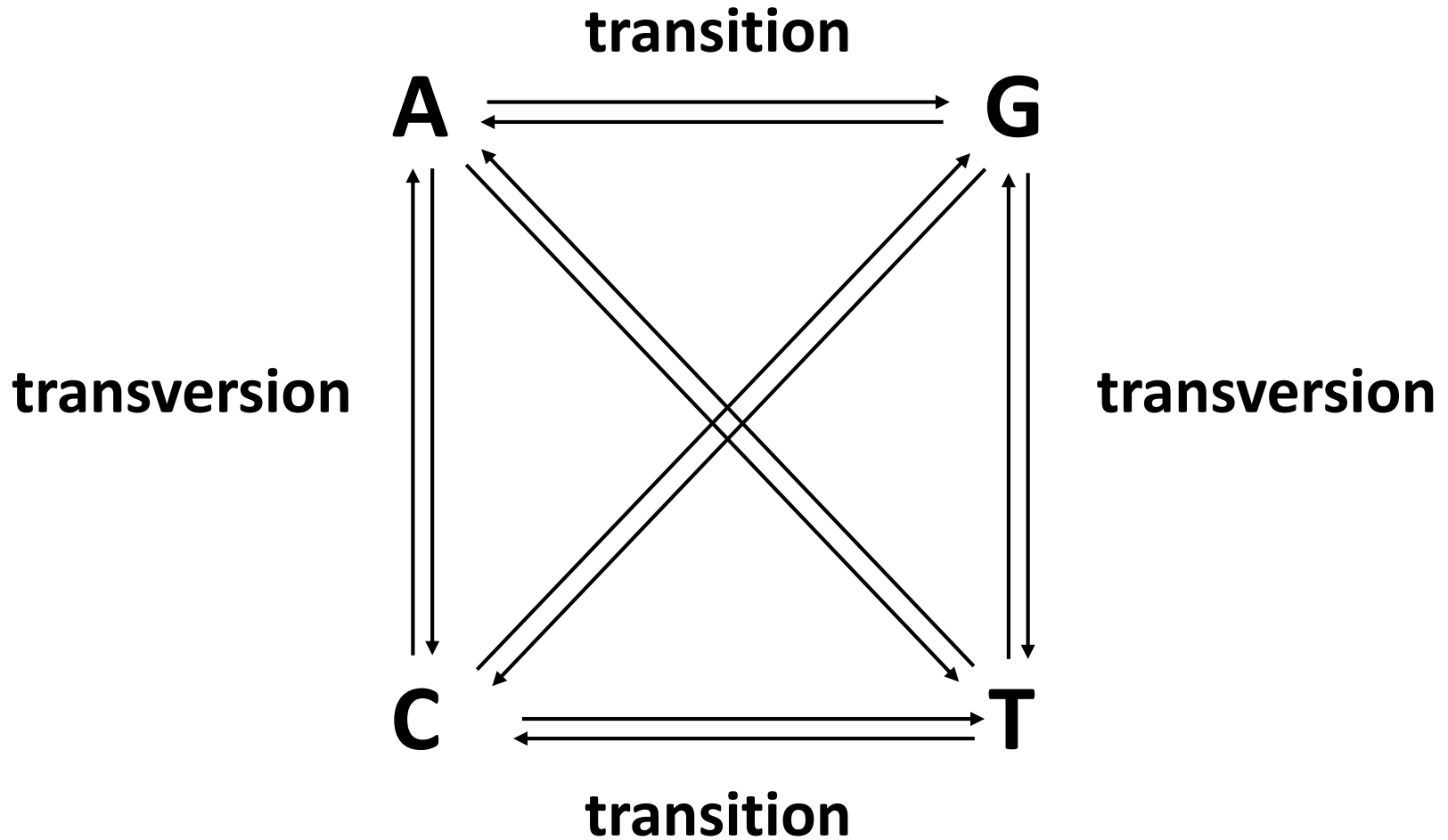
Ενιαίος ρυθμός εξέλιξης σε κάθε θέση *uniform rate*

## Παραμετροποίηση μοντέλων





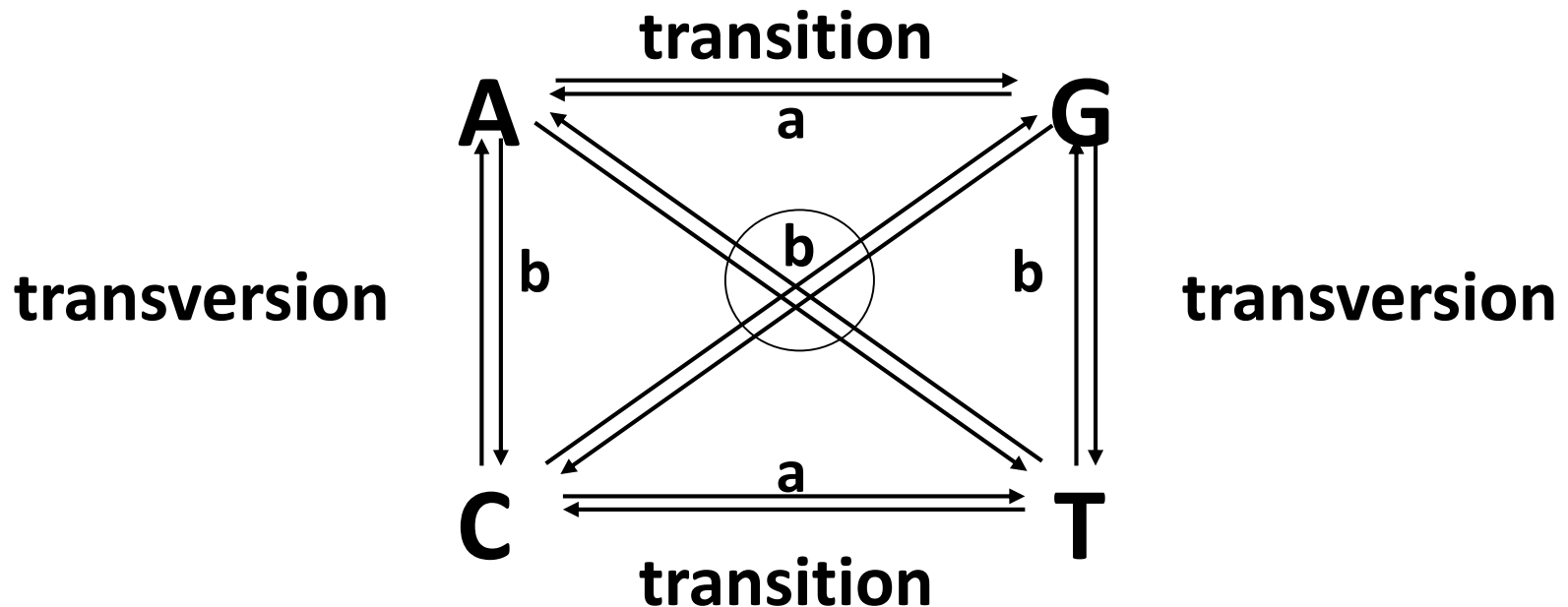
# Μοντέλα εξέλιξης (5/10)



# Μοντέλα εξέλιξης (6/10)

Μοντέλο Jukes & Cantor (1969)

$$D = \left(-\frac{3}{4}\right) \ln \left(1 - \frac{4}{3} p\right)$$



# Μοντέλα εξέλιξης (7/10)

## Πλεονεκτήματα μοντέλων

$$p = 3/60 = \underline{0.05} \quad \Rightarrow \quad D = \left(-\frac{3}{4}\right) \ln \left(1 - \frac{4}{3} 0.05\right) = \mathbf{0.052}$$

Jukes & Cantor

$$p = 30/60 = \underline{0.5} \quad \Rightarrow \quad D = \left(-\frac{3}{4}\right) \ln \left(1 - \frac{4}{3} 0.5\right) = \mathbf{0.82}$$

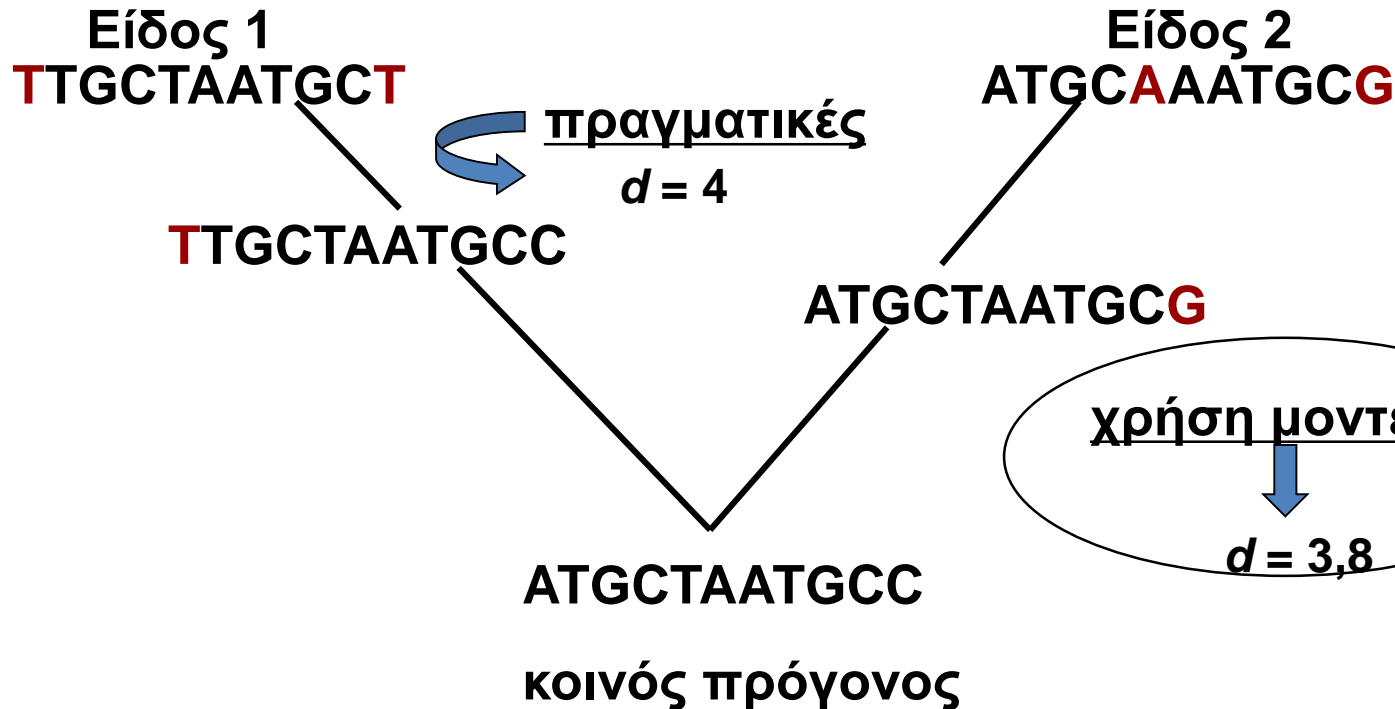


# Μοντέλα εξέλιξης (8/10)

## Πλεονεκτήματα μοντέλων

Είδος 1 **TTGCTAATGCT**  $\Rightarrow$   $d = 3$  παρατηρούμενες

Είδος 2 **ATGCAAATGCG**



# Μοντέλα εξέλιξης (9/10)

## Generalized Model of Nucleotide Substitution - the Q matrix

$$Q = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{pmatrix} -\mu(a\pi_C + b\pi_G + c\pi_T) & \mu a\pi_C & \mu b\pi_G & \mu c\pi_T \\ \mu g\pi_A & -\mu(g\pi_A + d\pi_G + e\pi_T) & \mu d\pi_G & \mu e\pi_T \\ \mu h\pi_A & \mu j\pi_C & -\mu(h\pi_A + j\pi_C + f\pi_T) & \mu f\pi_T \\ \mu i\pi_A & \mu k\pi_C & \mu l\pi_G & -\mu(i\pi_A + k\pi_C + l\pi_G) \end{pmatrix} \end{matrix}$$

Where:

$\mu$  = mean instantaneous substitution rate

$\pi$  = frequency of nucleotide denoted by the subscript

$a, b, c, \dots, l$  = relative rates of transformation from one base to another

Nearly all models of DNA substitution are special cases of this matrix (derived by implementing various simplifying assumptions).

**Εικόνα 2:** Γενικευμένο μοντέλο νουκλεοτιδικών αντικαταστάσεων.



# Μοντέλα εξέλιξης (10/10)

## Jukes-Cantor Model

- Assumes that all substitutions occur at the same rate ( $a = b = c = d = e = f$ ) and bases are found in equal frequency ( $\pi_A = \pi_C = \pi_G = \pi_T$ ). Then:

$$Q = \begin{pmatrix} -\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\ \frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\ \frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu \\ \frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu \end{pmatrix}$$

**Εικόνα 3:** Μήτρα νουκλεοτιδικών αντικαταστάσεων για το μοντέλο Jukes-Cantor.



# Προτεινόμενη Βιβλιογραφία

- Hall BG (2004): Phylogenetic Trees Made Easy (Sinauer, Sunderland)
- Nei M, Kumar S (2000): Molecular Evolution and Phylogenetics (Oxford University Press, New York)



# Σημείωμα Χρήσης Έργων Τρίτων

---

- Εικόνα 1: <http://commons.wikimedia.org/wiki/File:Homology.png>, by Popo H. Liao, CC BY-SA 3.0, <http://creativecommons.org/licenses/by-sa/3.0/deed.en>





# Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης,  
Ηλίας Καππάς, «Βιοπληροφορική, Φυλογενετική  
ανάλυση 1». Έκδοση: 1.0. Θεσσαλονίκη 2014.  
Διαθέσιμο από τη δικτυακή διεύθυνση:  
[http://opencourses.auth.gr/eclass\\_courses](http://opencourses.auth.gr/eclass_courses).



# Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά - Παρόμοια Διανομή [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».



Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

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





# Σας ευχαριστώ!

Ηλίας Καππάς, Λέκτορας Τμήματος  
Βιολογίας Α.Π.Θ.



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



# Τέλος ενότητας

Επεξεργασία: Στυλιανή Μηνούδη  
Θεσσαλονίκη, Εαρινό εξάμηνο 2014



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

# 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.

