



Βιοπληροφορική

Ενότητα 5^η: Φυλογενετική ανάλυση 2

Ηλίας Καππάς
Τμήμα Βιολογίας



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

- Φυλογενετικές προσεγγίσεις
- Αλγόριθμοι
- UPGMA – αποστάσεις
- Μοριακό ρολόι
- NJ – αποστάσεις
- MP – χαρακτήρες
- ML – χαρακτήρες



Μεθοδολογία

Βήματα φυλογενετικής ανάλυσης:

- Ανάκτηση ακολουθιών
- Στοίχιση
- Επιλογή μοντέλου εξέλιξης
- Κατασκευή φυλογενετικού δέντρου
- Αξιολόγηση φυλογενετικού δέντρου

Εξελικτικό μοντέλο ... μαθηματική περιγραφή (παραδοχές) –
evolutionary model αναπαραγωγή εξελικτικής διαδικασίας

Κριτήριο βελτιστοποίησης ... αριθμητική ποσότητα που
optimality criterion μεγιστοποιείται ή ελαχιστοποιείται

Στρατηγική αναζήτησης ... αλγόριθμος για την εξερεύνηση δέντρων
search strategy που βελτιστοποιούν το κριτήριο



Φυλογενετικές προσεγγίσεις

Κατασκευή φυλογενετικών δέντρων → προσεγγίσεις

(1) Φαινετική *phenetics* → Ομοιότητα *similarity* ?

Καμία αναφορά σε μοντέλο εξελικτικών σχέσεων

Μέτρηση διαφορών *mismatches* μεταξύ ειδών

Μετατροπή σε μήτρα γενετικών αποστάσεων

Κατασκευή δέντρου μέσω ιεραρχικής ομαδοποίησης

(2) Κλαδιστική *cladistics* → Γενεαλογική σχέση *genealogical relationship*

Εναλλακτικά μονοπάτια εξέλιξης

Κατασκευή βέλτιστου δέντρου *optimal tree* σύμφωνα με κάποιο μοντέλο εξελικτικών αλλαγών



Πληροφορία - Αλγόριθμοι

Κατασκευή φυλογενετικών δέντρων → τύποι πληροφορίας

→ αλγόριθμοι

		Αλγόριθμος	
		κριτήριο βελτιστοποίησης	ομαδοποίηση
Τύπος πληροφορίας	χαρακτήρες	<u>PARSIMONY</u> <u>MAXIMUM LIKELIHOOD</u>	
	αποστάσεις	MINIMUM EVOLUTION LEAST SQUARES	<u>UPGMA</u> <u>NEIGHBOR-JOINING</u>



Πληροφορία

Μέθοδοι χαρακτήρα → χρήση χαρακτήρων

Είδος A	A	T	G	G	C	T	A	T	T	C	T	T	A	T	A	G	T	A	C	G
Είδος B	A	T	C	G	C	T	A	G	T	C	T	T	A	T	A	T	T	A	C	A
Είδος C	T	T	C	A	C	T	A	G	A	C	C	T	G	T	G	G	T	C	C	A
Είδος D	T	T	G	A	C	C	A	G	A	C	C	T	G	T	G	G	T	C	C	G
Είδος E	T	T	G	A	C	C	A	G	T	T	C	T	C	T	A	G	T	T	C	G

Μέθοδοι απόστασης → μετατροπή διαφορών σε δια-ειδικές αποστάσεις → κατασκευή μήτρας γενετικών αποστάσεων

	A	B	C	D	E	
Είδος A	—	0.20	0.50	0.45	0.40	
Είδος B	0.23	—	0.40	0.55	0.50	
Είδος C	0.87	0.59	—	0.15	0.40	<i>p distance uncorrected</i>
Είδος D	0.73	1.12	0.17	—	0.25	
Είδος E	0.59	0.89	0.61	0.31	—	

Kimura 2-parameter distance → όρος εξισορρόπησης



Αλγόριθμοι

Ομαδοποίηση *clustering*

Δια-ειδικές αποστάσεις (σταθερός όρος)

Μεγάλη ταχύτητα

Μοναδικό δέντρο

Βελτιστοποίηση *optimality*

Δια-ειδικές αποστάσεις ή χαρακτήρες (μεταβλητός όρος)

Κριτήριο βελτιστοποίησης (ελάχιστα μήκη βραχιόνων,
ελάχιστα βήματα, μέγιστη πιθανοφάνεια)

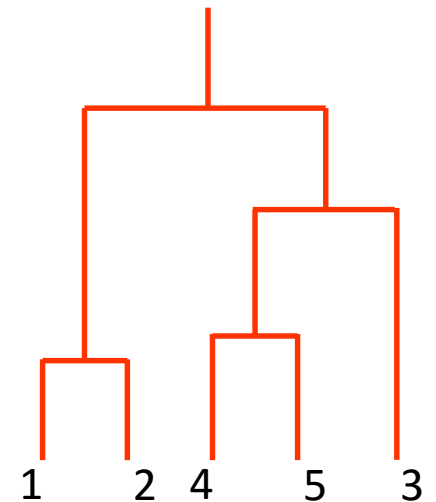
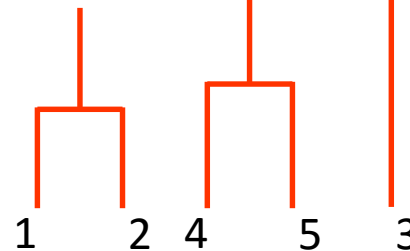
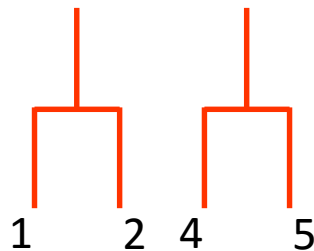
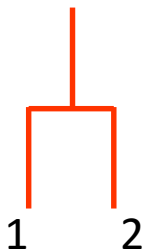
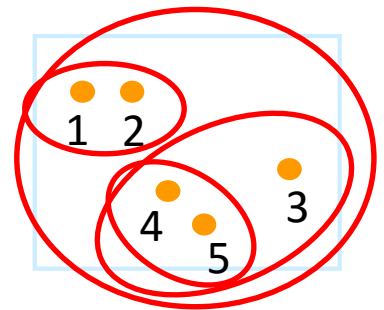
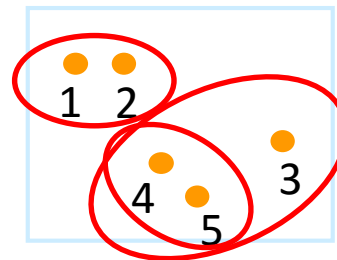
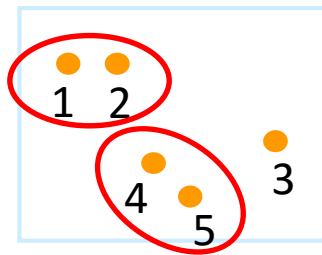
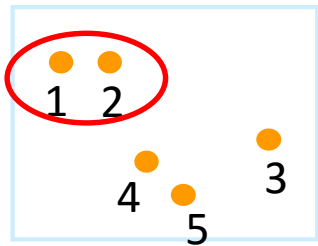
Ειδικός αλγόριθμος για max ή min συνάρτησης

Πολλά δέντρα



UPGMA – αποστάσεις (1/2)

UPGMA *Unweighted Pair Group Method using Arithmetic Averages* (1/4)



URGMA – αποστάσεις (2/2)

URGMA → χαρακτηριστικά - παραδοχές

- Δέντρο πάντοτε με ρίζα
- Σταθερός ρυθμός εξέλιξης μεταξύ των taxa
→ μοριακό ρολόι molecular clock hypothesis***

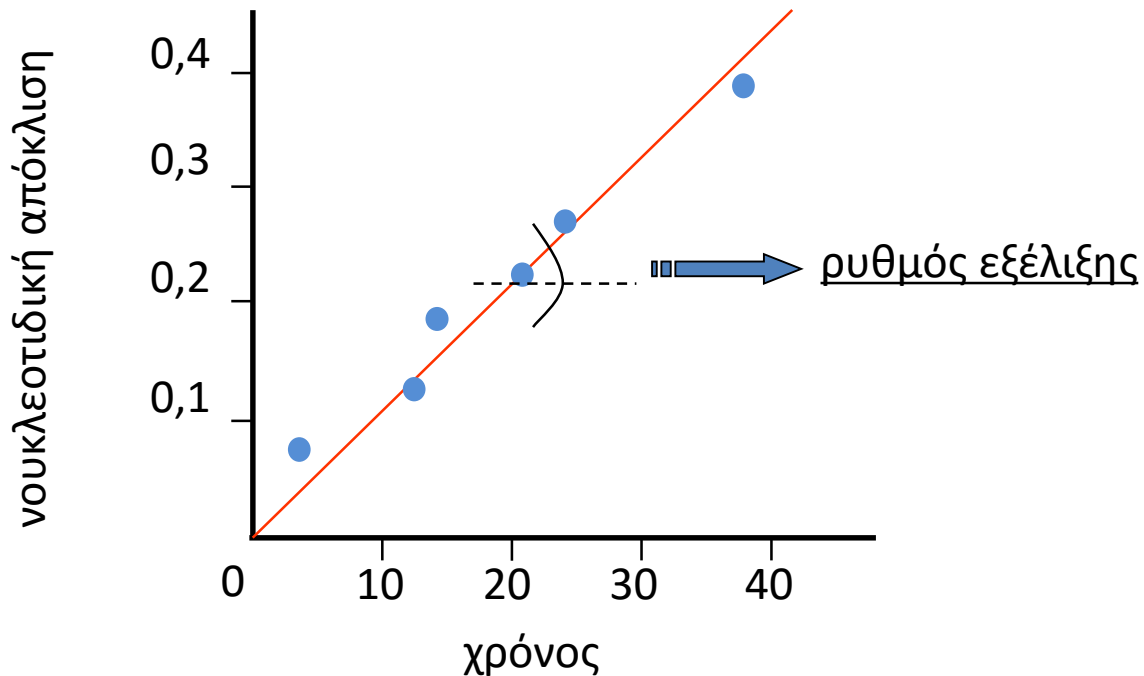
(Zuckerkandl & Pauling, 1962; Margoliash, 1963)

- Όχι ακριβής



Μοριακό ρολόι (1/6)

Υπόθεση μοριακού ρολογιού



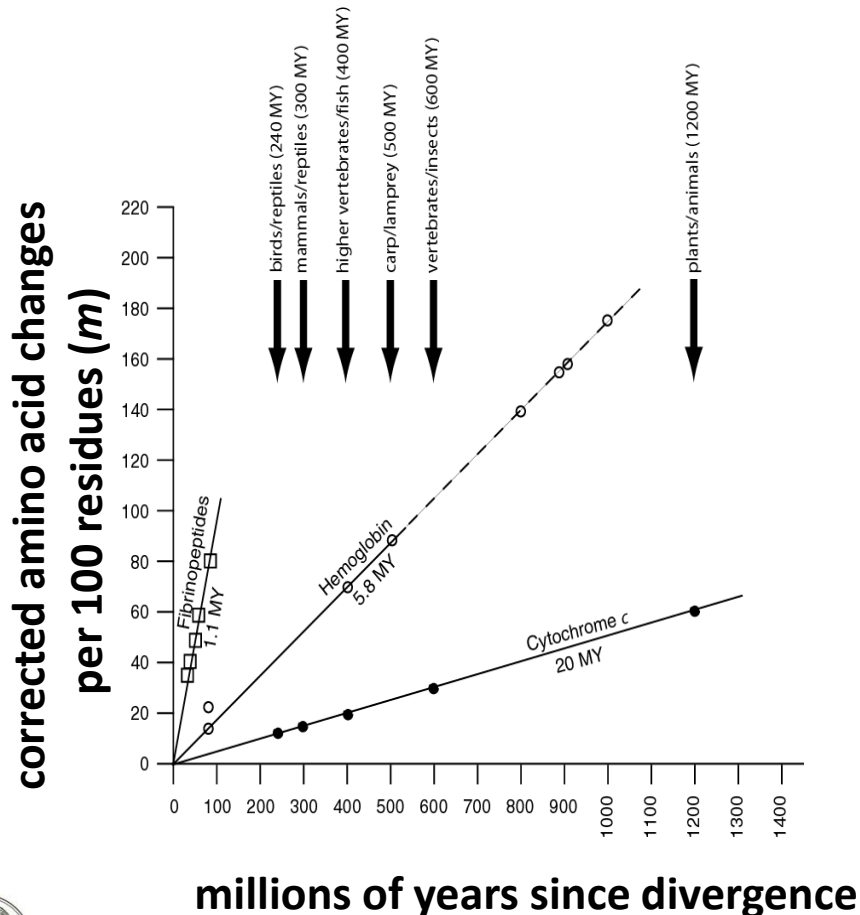
Χρόνοι απόκλισης

- Βιογεωγραφία
- Απολιθώματα
- Μοριακή απόκλιση



Μοριακό ρολόι (2/6)

Εικόνα 1. Ρυθμοί εξέλιξης διαφορετικών πρωτεϊνών.



Dickerson (1971)

- Για κάθε πρωτεΐνη, οι τιμές σχηματίζουν ευθεία γραμμή. Συνεπώς, ο ρυθμός αμινοξικών αντικαταστάσεων έχει παραμείνει σταθερός για κάθε πρωτεΐνη
- Ο μέσος ρυθμός αλλαγής διαφέρει για κάθε πρωτεΐνη. Ο χρόνος για 1% απόκλιση μεταξύ 2 εξελικτικών γραμμών είναι 20 MY (cytochrome c), 5.8 MY (hemoglobin) και 1.1 MY (fibrinopeptides)
- Οι παρατηρούμενες διαφορές στο ρυθμό αλλαγής αντικατοπτρίζουν λειτουργικούς περιορισμούς λόγω φυσικής επιλογής



Μοριακό ρολόι (3/6)

Υπόθεση μοριακού ρολογιού ... σημασία

Αν οι ακολουθίες εξελίσσονται με σταθερούς ρυθμούς, τότε ...

υπολογισμός χρόνου απόκλισης των ακολουθιών ... οργανισμών

(διαδικασία ανάλογη της ραδιοχρονολόγησης)

N = συνολικός αριθμός αντικαταστάσεων *total number of substitutions*

L = αριθμός συγκρινόμενων νουκλεοτιδίων μεταξύ 2 ακολουθιών

number of nucleotide sites compared between 2 sequences

$K = N/L$ = αριθμός αντικαταστάσεων ανά θέση

number of substitutions per nucleotide site



Μοριακό ρολόι (4/6)

Ρυθμός νουκλεοτιδικής αντικατάστασης r και χρόνος απόκλισης T

r = ρυθμός αντικατάστασης *rate of substitution* ... hemoglobin alpha
= 0.56×10^{-9} / θέση / χρόνο *per site per year*

$K = 0.093$ = αριθμός αντικαταστάσεων / θέση ... rat *versus* human
number of substitutions per nucleotide site

$$r = K / 2T$$

$$T = .093 / (2)(0.56 \times 10^{-9}) = 80 \text{ million years}$$

Εξαιρέσεις της υπόθεσης του μοριακού ρολογιού !!!



Μοριακό ρολόι (5/6)

Σύγκριση εξελικτικών ρυθμών

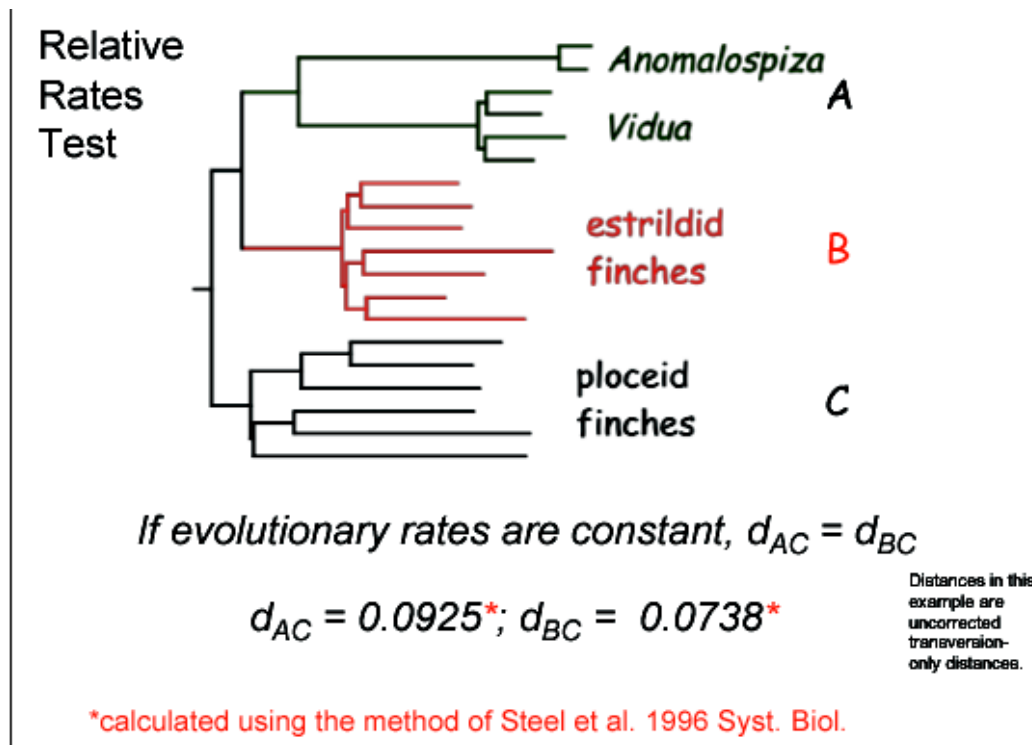
Οι ρυθμοί μπορούν να διαφέρουν μεταξύ εξελικτικών γραμμών:

- διαφορές στον τρόπο ζωής ... χρόνος γενεάς
- διαφορές στη φυσιολογία ... μεταβολικός ρυθμός
- διαφοροποιούσα φυσική επιλογή
- πληθυσμιακή ιστορία



Μοριακό ρολόι (6/6)

παράδειγμα σύγκρισης ...



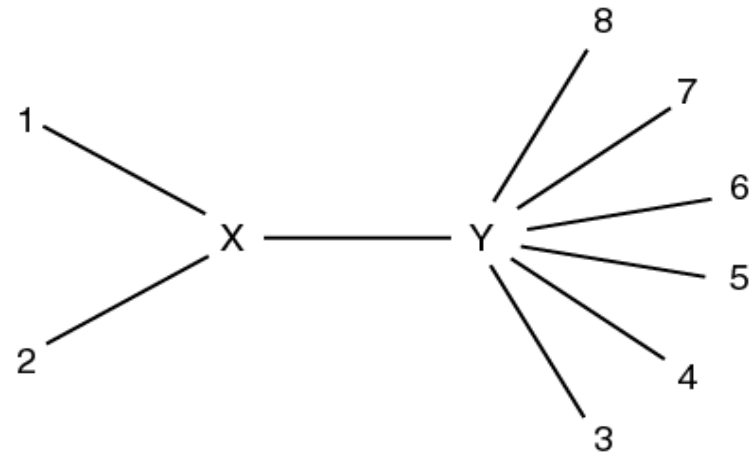
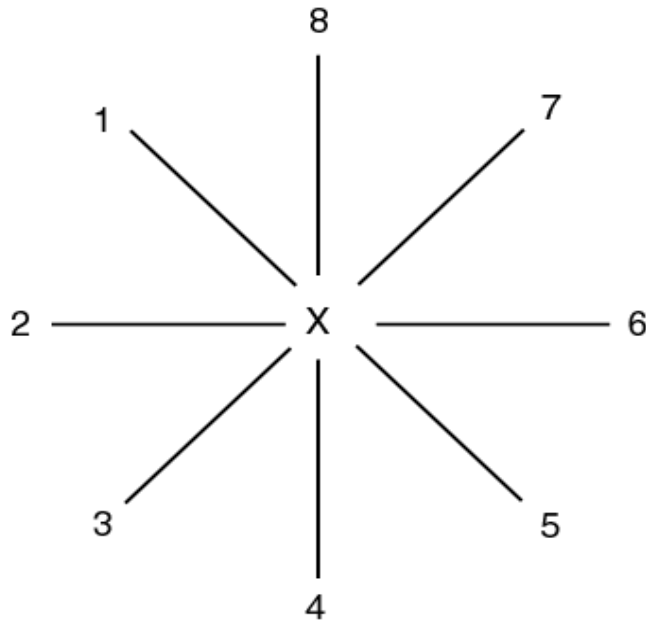
Εικόνα 2. Έλεγχος σχετικών ρυθμών εξέλιξης.



NJ – αποστάσεις (1/2)

NJ Neighbor-Joining (2/4)

Αρχή Ελάχιστης Εξέλιξης *Minimum Evolution (ME)*



$$d_{XY} = 1/2(d_{1Y} + d_{2Y} - d_{12})$$

Ελαχιστοποίηση μηκών βραχιόνων σε κάθε βήμα



NJ – αποστάσεις (2/2)

NJ → χαρακτηριστικά - παραδοχές

- Δέντρο χωρίς ρίζα
- Διαφορετικός ρυθμός εξέλιξης μεταξύ των taxa
- Εξαιρετικά γρήγορος αλγόριθμος → δημοφιλής
- Εξαιρετικά ακριβής



MP – χαρακτήρες (1/4)

Μέγιστη Φειδωλότητα *Maximum Parsimony (MP)* (3/4)

“Plurality should not be assumed without necessity”

...one should not increase, beyond what is necessary, the number of entities required to explain anything → Ockham’s razor

(*lex parsimoniae* = Law of Parsimony)

κριτήριο βελτιστοποίησης



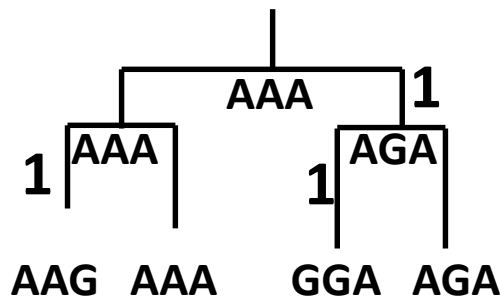
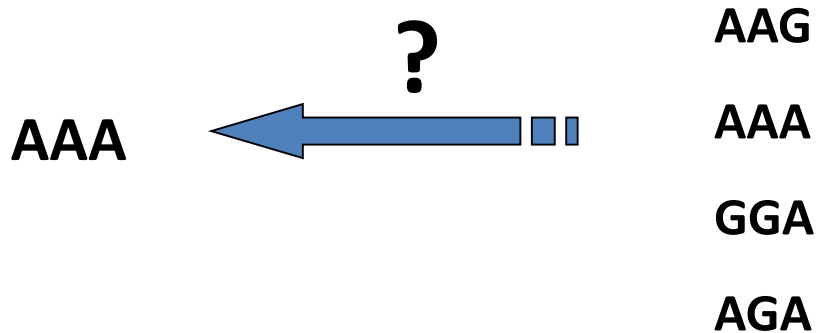
τα λιγότερα εξελικτικά βήματα

→ αλλαγές → αντικαταστάσεις νουκλεοτιδίων - αμινοξέων

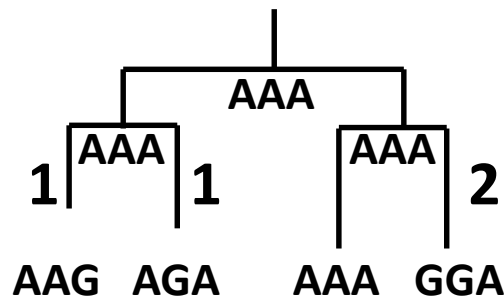


MP – χαρακτήρες (2/4)

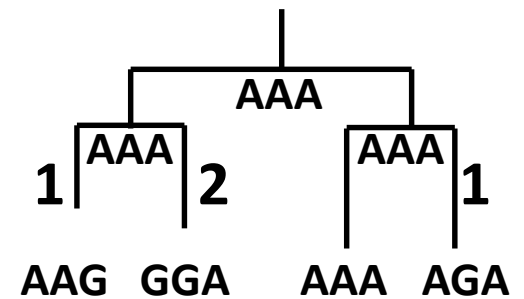
Μέγιστη Φειδωλότητα



βήματα = 3



βήματα = 4



βήματα = 4



MP – χαρακτήρες (3/4)

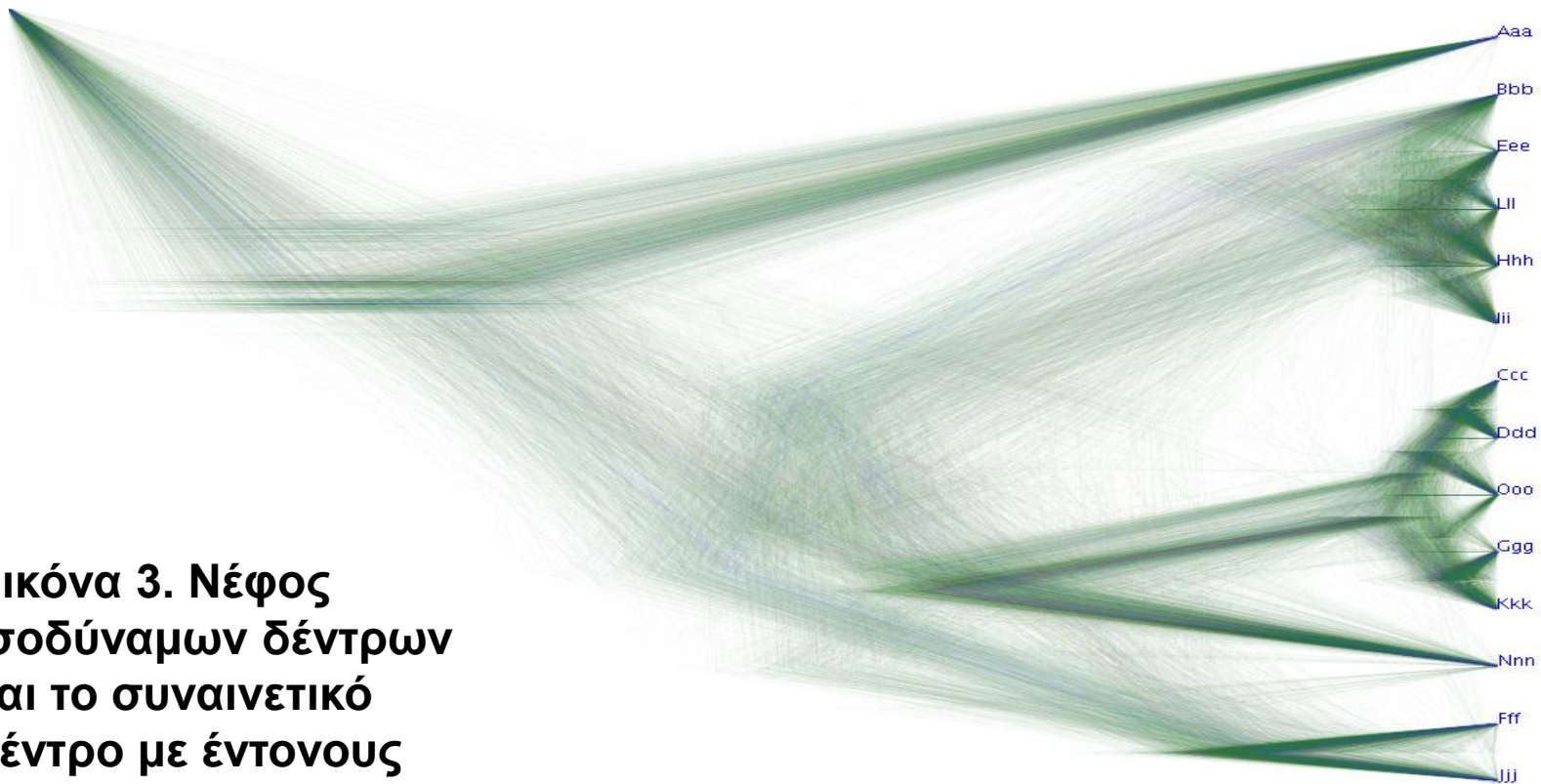
MP → χαρακτηριστικά - παραδοχές

- Μόνο πληροφοριακές θέσεις *parsimony informative sites*
- Δυνατότητα indels ως 5^{ος} χαρακτήρας
- Απλή standard ή σταθμισμένη weighted φειδωλότητα
- Δύο βασικοί αλγόριθμοι → *exhaustive search*
→ *heuristic search*
- Μικρή νουκλεοτιδική απόκλιση
- Σχετικά σταθερός ρυθμός εξέλιξης
- Διακριτά βήματα → μεγάλος αριθμός δέντρων
→ συναινετικό δέντρο *consensus tree* → *strict*
→ *majority rule*



MP – χαρακτήρες (4/4)

συναινετικό δέντρο *consensus tree*



**Εικόνα 3. Νέφος
ισοδύναμων δέντρων
και το συναινετικό
δέντρο με έντονους
κλάδους.**



ML – χαρακτήρες (1/2)

Μέγιστη Πιθανοφάνεια *Maximum Likelihood (ML)* (4/4)

Γέφυρα μεταξύ distance ↔ parsimony

Προσπάθεια εύρεσης του εξελικτικού μοντέλου και του δέντρου που περιγράφουν καλύτερα (μέγιστη πιθανοφάνεια) την παρατηρούμενη γενετική απόκλιση μεταξύ των taxa

κριτήριο βελτιστοποίησης



η μεγαλύτερη πιθανοφάνεια ...

ΟΧΙ του δέντρου ή του εξελικτικού μοντέλου

Πιθανότητα θέσης → άθροισμα εναλλακτικών

→ γινόμενο πιθανοτήτων όλων των θέσεων

→ συνολική πιθανοφάνεια δέντρου



ML – χαρακτήρες (2/2)

ML → χαρακτηριστικά - παραδοχές

- Μεγαλύτερο εύρος παραμετροποίησης

→ απλά ↔ εξειδικευμένα μοντέλα

- Συνήθως η πιο ακριβής μέθοδος

- Υπολογισμοί πιθανοτήτων → σπανίως
ισοδύναμα δέντρα

- Υπολογιστική ταχύτητα – χρόνος → ρύθμιση
αλγόριθμου



Σύγκριση μεθόδων

Neighbor-joining	Maximum parsimony	Maximum likelihood
Μόνο αποστάσεις	Μόνο κοινούς προκύπτοντες χαρακτήρες	Χρησιμοποίηση όλων των δεδομένων
Ελαχιστοποιεί αποστάσεις μεταξύ κοντινότερων γειτόνων	Ελαχιστοποιεί συνολική απόσταση	Μεγιστοποιεί πιθανοφάνεια δέντρου δοθέντων συγκεκριμένων παραμέτρων
Πολύ γρήγορος	Αργός	Πολύ αργός
Εύκολη παγίδευση σε τοπικά βέλτιστα	Παραδοχές δεν ισχύουν όταν η εξέλιξη είναι γρήγορη	Ισχυρά εξαρτώμενος από το μοντέλο εξέλιξης
Κατάλληλος για προκαταρκτικό δέντρο	Καλύτερη επιλογή (<30 taxa, ομοπλασίες σπάνιες)	Κατάλληλος για μικρά data sets και για σύγκριση δέντρων κατασκευασμένων με άλλες μεθόδους



Φυλογενετικά προγράμματα

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

**Λίστα >200 φυλογενετικών προγραμμάτων
ανάλυσης**



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)
- **MEGA:** <http://www.megasoftware.net/>
- **BioEdit:** <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>



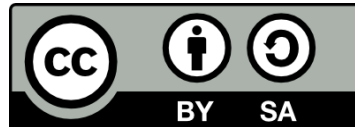
Σημείωμα Αναφοράς

Copyright Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης,
Ηλίας Καππάς, «Βιοπληροφορική, Φυλογενετική
ανάλυση 2». Έκδοση: 1.0. Θεσσαλονίκη 2014.
Διαθέσιμο από τη δικτυακή διεύθυνση:
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

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

