



ΤΕΧΝΟΛΟΓΙΚΟ
ΕΚΠΑΙΔΕΥΤΙΚΟ
ΙΔΡΥΜΑ ΑΘΗΝΑΣ

ATHENS
UNIVERSITY
OF APPLIED
SCIENCES



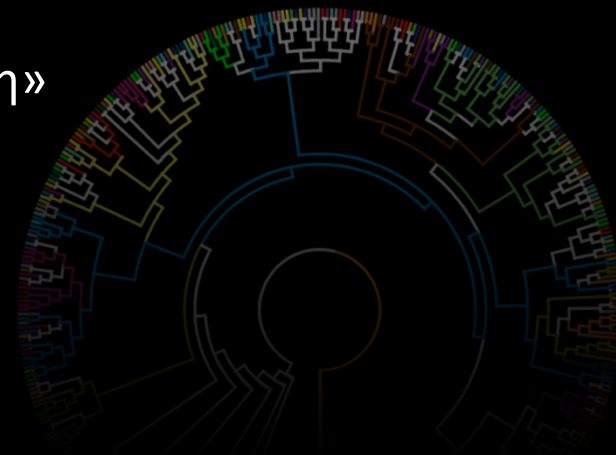
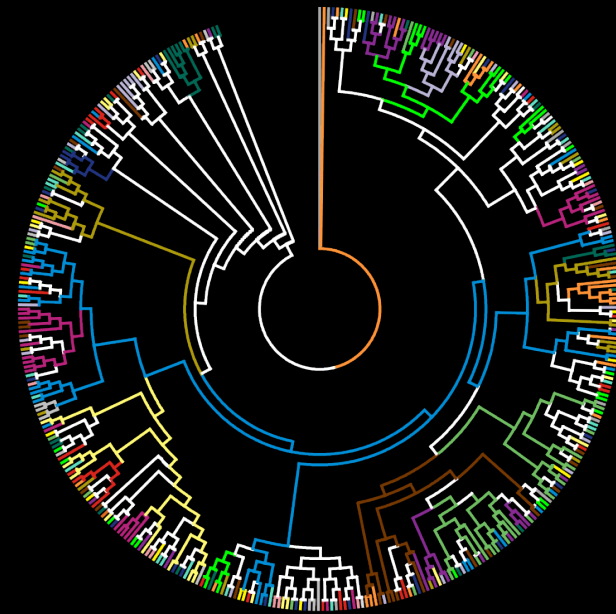
UNIVERSITY OF
LIVERPOOL

Εισαγωγή στη Μοριακή Επιδημιολογία και Επιδημιολογική Επιτήρηση

Πρόγραμμα Μεταπτυχιακών Σπουδών
«Βιοϊατρικές Μέθοδοι και Τεχνολογία στη Διάγνωση»

Apostolos Beloukas

abeloukas@teiath.gr; beloukas@liverpool.ac.uk



Δευτέρα, 22 Μαΐου 2017

Εκπαιδευτικοί στόχοι

- ❖ Τι είναι η Μοριακή Επιδημιολογία και ποια/ες οι διαφορές με την κλασσική επιδημιολογία
- ❖ Τι είναι η φυλογένεια ή φυλογενετικό δένδρο
 - ❖ Ορολογία (πχ extant, ancestral, branch point, branch length)
- ❖ Γιατί και πως φτιάχνουμε φυλογενετικό δένδρο
 - Algorithms to build phylogenetic trees
 - Distance-based methods
 - Parsimony methods
 - Minimize the number of changes
 - Probabilistic methods
 - Tree that best explains the data using probabilistic models
- ❖ Ιογενείς λοιμώξεις και Επιδημιολογική επιτήρηση
 - origin, expansion, geographic distribution, evolution

Today's Random Medical News

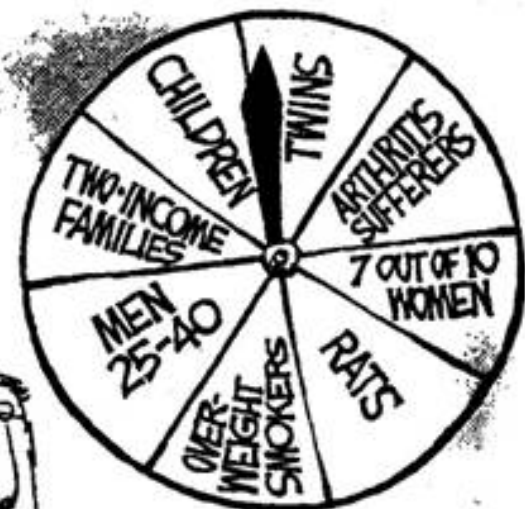
from the New England
Journal of
Panic-Inducing
Gobbledygook



CAN CAUSE



IN



ACCORDING TO A
REPORT RELEASED
TODAY...



Jim Borgman
The Cincinnati Enquirer
King Features Syndicate

Εκπαιδευτικοί στόχοι

- ❖ Τι είναι η Μοριακή Επιδημιολογία και ποια/ες οι διαφορές με την κλασσική επιδημιολογία
- ❖ Τι είναι η φυλογένεια ή φυλογενετικό δένδρο
 - ❖ Ορολογία (πχ extant, ancestral, branch point, branch length)
- ❖ Γιατί και πως φτιάχνουμε φυλογενετικό δένδρο
 - Algorithms to build phylogenetic trees
 - Distance-based methods
 - Parsimony methods
 - Minimize the number of changes
 - Probabilistic methods
 - Tree that best explains the data using probabilistic models
- ❖ Ιογενείς λοιμώξεις και Επιδημιολογική επιτήρηση
 - origin, expansion, geographic distribution, evolution

Επιδημιολογία

Epi ... demi... Ology (Greek origin: επι... δημο.. ολογία...)

- ❖ Epidemiology = the science which deals with what befalls upon people.....
- ❖ Bridge between biomedical, social and behavioral sciences

Most popular old definitions

Oxford English Dictionary

The branch of medical science which treats epidemics

MacMahon and Pugh, 1970

The study of the distribution and determinants of disease frequency in human populations

John Last, 1988

The study of the distribution and determinants of health-related states or events in specified populations and the application of this study to the control of health problems

Kuller LH: American J of Epidemiology 1991;134:1051

Epidemiology is the study of "EPIDEMICS" and their prevention

Σύγχρονος Ορισμός

Study of the **occurrence and distribution** of health-related **diseases or events** in specified **populations**, including the study of the **determinants** influencing such states, and the application of this knowledge to **control** the public health problem

Επιδημιολογία είναι η μελέτη της **κατανομής** των **παραγόντων** που επηρεάζουν τη **συχνότητα** των **νοσημάτων*** στον **ανθρώπινο πληθυσμό** και η εφαρμογή της στον **έλεγχο** των υγειονομικών προβλημάτων.

*** Νόσημα: νόσος, κακώσεις, αναπηρία, θάνατος**

(Porta M, Last J, Greenland S. A Dictionary of Epidemiology, 2008)

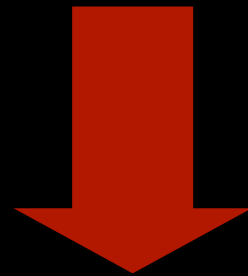
Σκοποί – Στόχοι Επιδημιολογίας

- ❖ To investigate nature / extent of **health-related phenomena** in the community / identify **priorities**
- ❖ To study **natural history** and prognosis of health-related problems
- ❖ To identify causes and **risk factors**
- ❖ To recommend / assist in application of / evaluate best **interventions** (preventive and therapeutic measures)
- ❖ To provide foundation for **public policy**

Classical vs molecular
epidemiology

Revolution in Molecular Biology has Impacted:

- Science
- Medicine
- Society
- **Epidemiology**



Molecular Epidemiology
Challenge for the 21st Century

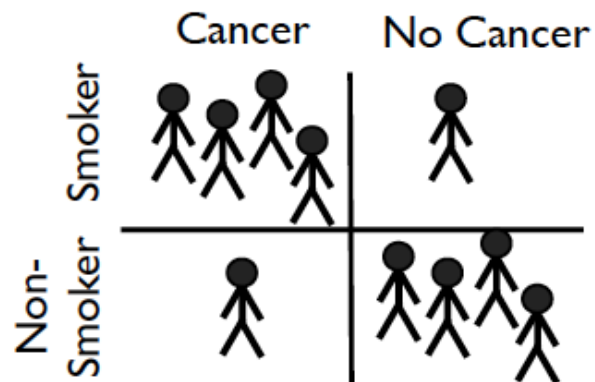
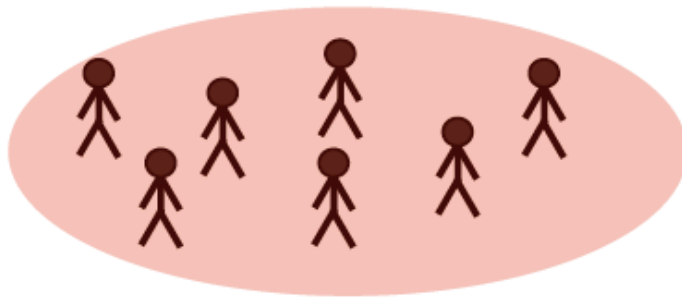
Μοριακή Επιδημιολογία

Use and measurement of **biomarkers** in epidemiology:

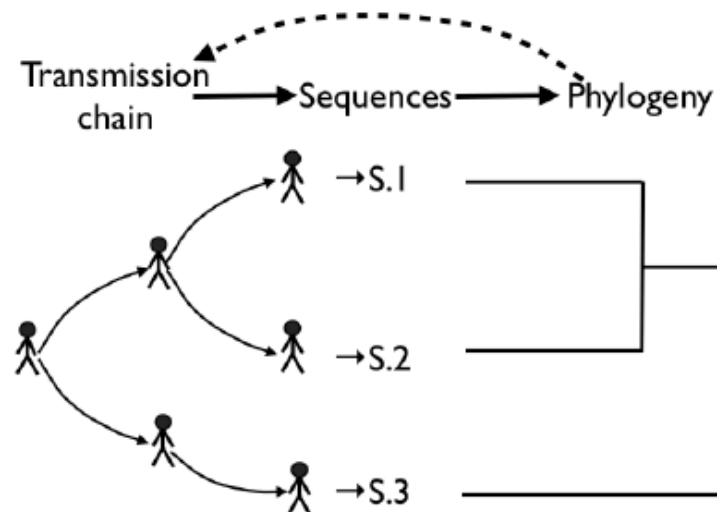
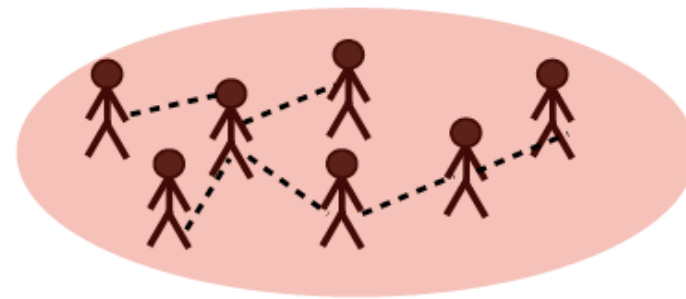
- ✓ Identification of **genetic** disease-related **markers**
- ✓ Study **molecular phylogeny** organisms /species
- ✓ Study **epidemics** of rapidly evolving pathogens
- ✓ Study **evolution** of pathogen epidemics and pathogens evolution

Κλασική vs Μοριακή Επιδημιολογία

Classical Epidemiology



Network/Molecular Epidemiology



Γενετική Επιδημιολογία

Use and measurement of **biomarkers** in epidemiology:

- ✓ Identification of **genetic** disease-related **markers**
- ✓ Study molecular phylogeny organisms /species
- ✓ Study epidemics of rapidly evolving pathogens
- ✓ Study evolution of pathogen epidemics and pathogens evolution

Παραδείγματα μοριακής επιδημιολογίας

Ταυτοποίηση γενετικών παραγόντων που σχετίζονται με νοσήματα:

- ✓ **HIV/AIDS**
- ✓ Ψωρίαση
- ✓ Ερυθηματώδης λύκος
- ✓ Νόσος Alzheimer
- ✓ Νόσος Parkinson
- ✓ Μεταβολισμός φαρμάκων

Συσχέτιση CCR5Δ32 με ανάπτυξη AIDS

- CCR5 chemoreceptor acts as co-receptor for HIV virion entry into the host cells
- 10% of Caucasian people in USA has a deletion of 32 nucleotides (CCR5Δ32)

Does CCR5Δ32 affects HIV infectivity and/or AIDS progression

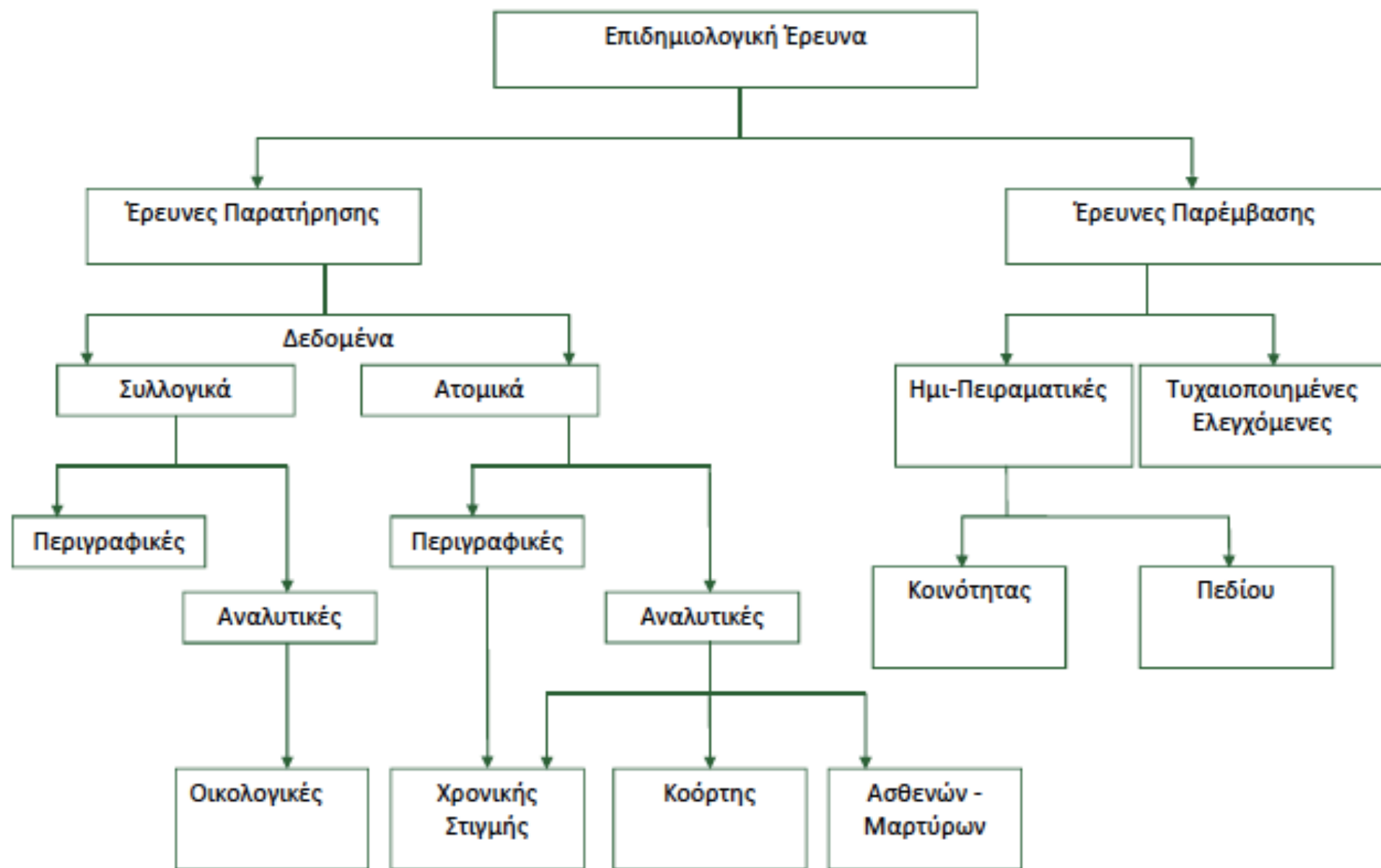
- Το 1996 βρέθηκε ότι ο υποδοχέας της χημειοκίνης CCR5 παίζει το ρόλο συνυποδοχέα για τη εισαγωγή (μόλυνση) των κυττάρων στόχων από τον HIV-1
- Βρέθηκε ότι το γονίδιο παρουσιάζει μια έλλειψη 32 νουκλεοτιδίων (ccR5Δ32) με συχνότητα 10% στον Καυκάσιο πληθυσμό στις ΗΠΑ

Συσχέτιση CCR5Δ32 με ανάπτυξη AIDS

Ο συγκεκριμένος δείκτης (ccR5Δ32) συσχετίζεται με διαφορετική μολυσματικότητα του ιού ή εξέλιξη της νόσου;

Πραγματοποίηση 2 είδους μελετών:

- Μελέτης επιπολασμού ccR5Δ32 σε HIV(-) και HIV(+) ασθενείς.
- Αναδρομική μελέτη (retro prospective) σε ομάδες (cohort) ασθενών με διαφορετικό ρυθμό εξέλιξης νόσου (AIDS)

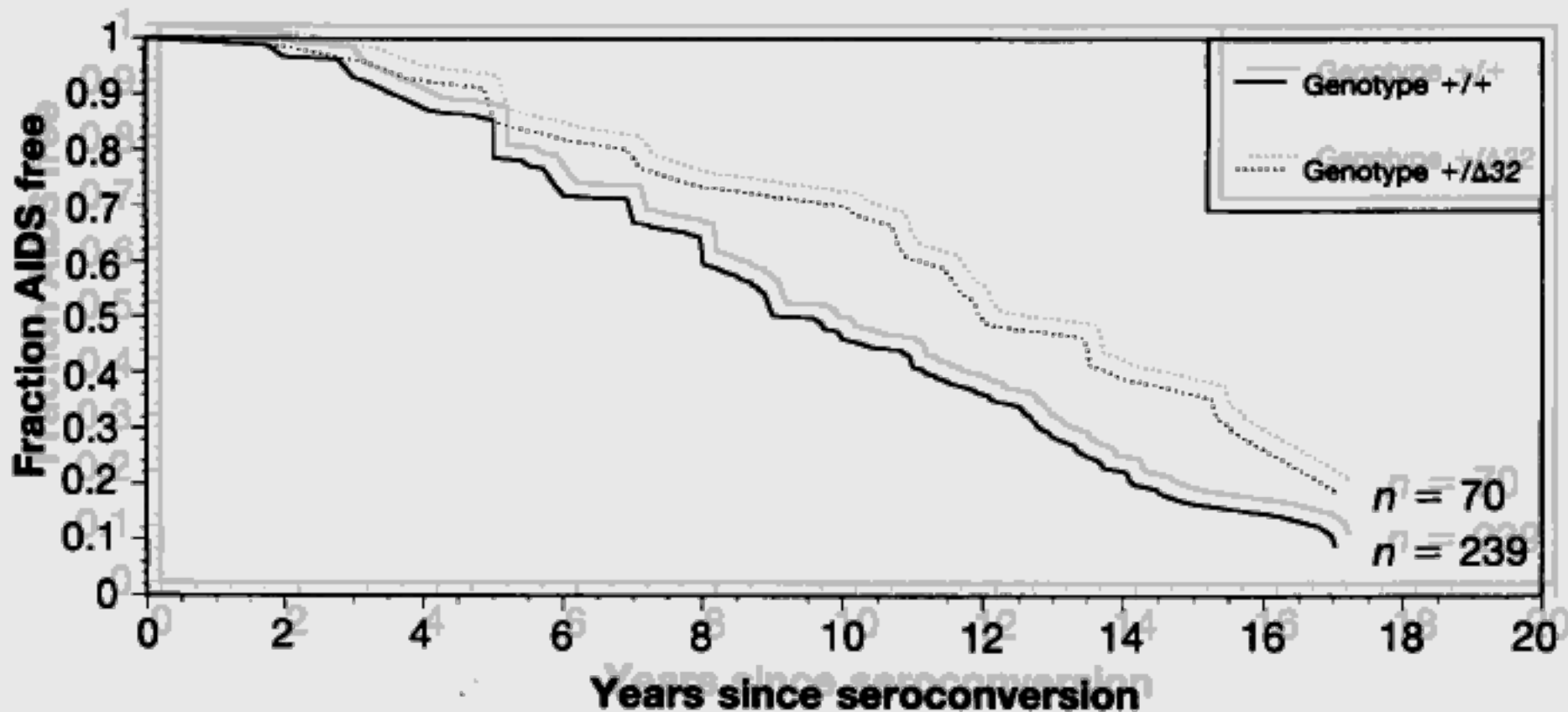


Cohort	Risk group	HIV-1 antibody status	Total	Number of patients (% of total) with <i>CKR5</i> genotype			Statistics	
				+/+	+/ Δ 32	Δ 32/ Δ 32	G	P
DCG	Homosexual men	Positive	137	118 (86)	19 (14)	0 (0)	2.0 (1.6)	0.38 (0.44)
		Negative	212	175 (83)	33 (16)	4 (2)		
		Total	349	293	52	4		
MAC	Homosexual men	Positive	265	201 (76)	64 (24)	0 (0)	9.9 (9.6)	0.007 (0.008)
		Negative	24	14 (58)	6 (25)	3 (17)		
		Total	289	215	70	3		
SFCC	Homosexual men	Positive	150	110 (73)	40 (27)	0 (0)	9.2 (7.9)	0.01 (0.02)
		Negative	42	34 (81)	5 (12)	3 (7)		
		Total	193	145	45	3		
HGDS	Hemophiliacs	Positive	133	106 (80)	27 (20)	0 (0)	0.62 (0.1)	0.73 (0.85)
		Negative	104	87 (84)	17 (16)	0 (0)		
		Total	237	193	44	0		
MHCS	Hemophiliacs	Positive	192	156 (81)	36 (19)	0 (0)	8.0 (8.9)	0.02 (0.01)
		Negative	191	158 (82)	26 (14)	7 (4)		
		Total	383	314	62	7		
ALIVE	Intravenous drug users	Positive	466	457 (98)	9 (2)	0 (0)	0.1 (NC)	0.77 (NC)
		Negative	39	39 (100)	0 (0)	0 (0)		
		Total	505	496	9	0		
All cohorts	Homosexual men and hemophiliacs	Positive	1343	1148 (85)	195 (15)	0 (0)	35.0 (33.3)	2.5×10^{-8} (5.8×10^{-6})
		Negative	612	508 (83)	87 (14)	17 (3)		
		Total	1955	1656	282	17		

Συσχέτιση CCR5Δ32 με ανάπτυξη AIDS

- Ο επιπολασμός των +/+ και +/-ccR5Δ32 δεν διαφέρει μεταξύ HIV(-) και HIV(+) ανθρώπων
- Υπήρχε όμως σημαντική διαφορά για τους ομόζυγους με ccR5Δ32 που δεν βρέθηκε κανείς να έχει μολυνθεί με HIV
- Ο γενετικός δείκτης αυτός συνδέεται με αυξημένη αντοχή στη μόλυνση με HIV

Συσχέτιση CCR5Δ32 με ανάπτυξη AIDS



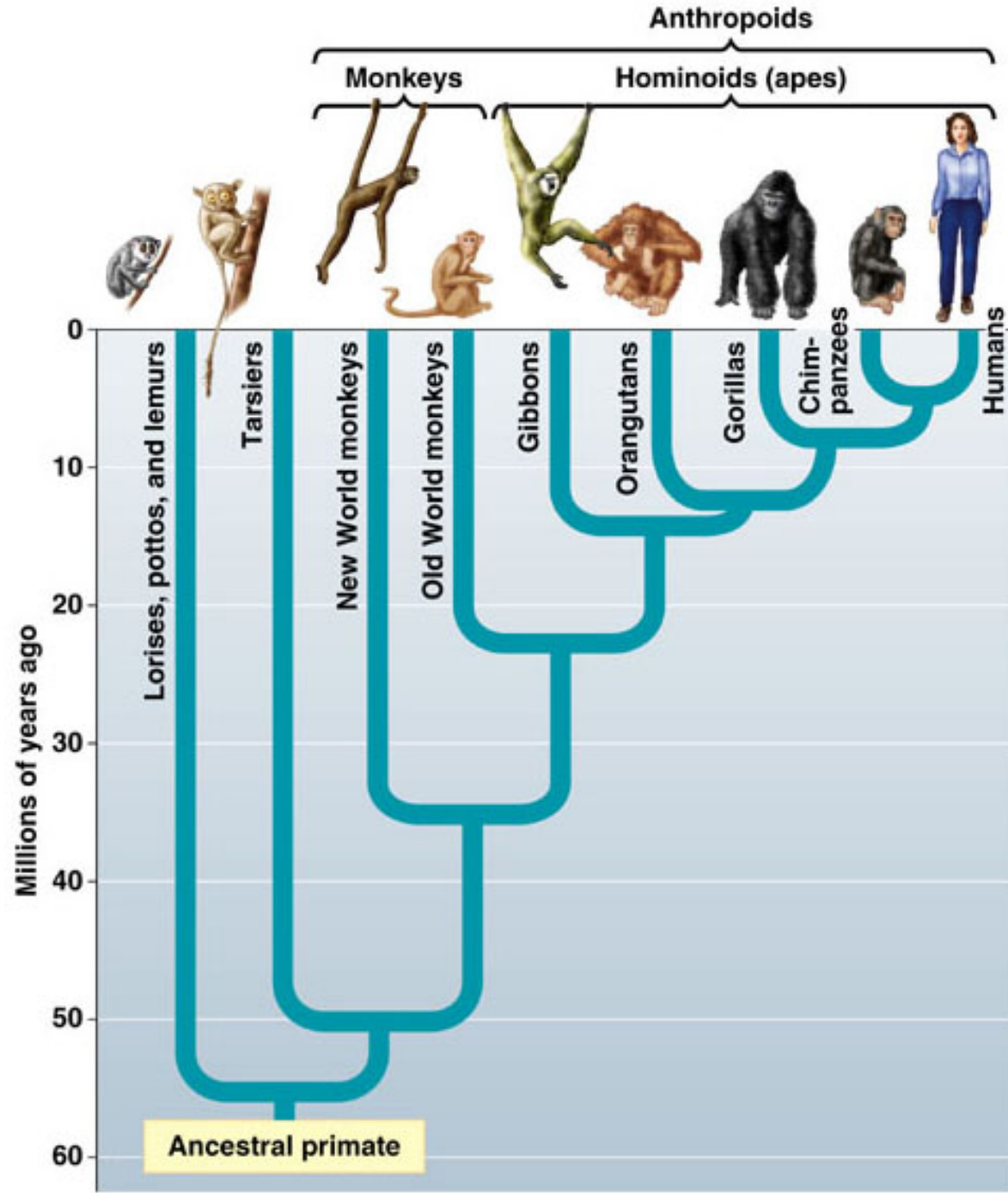
($P=0.005$)

Μελέτη μοριακής φυλογένειας οργανισμών-ειδών

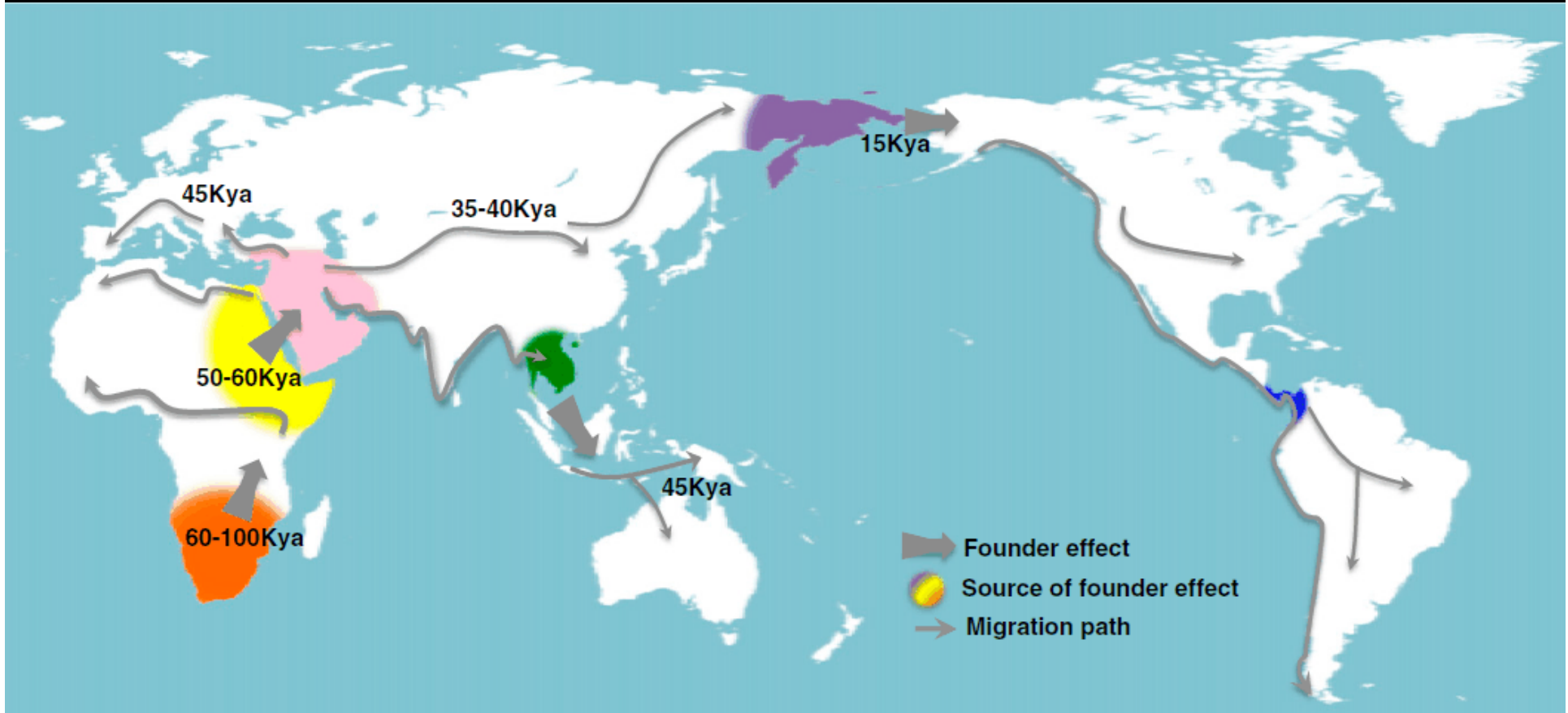
Molecular Epidemiology

- ✓ Identification of genetic disease-related markers
- ✓ Study molecular **phylogeny** organisms /species
- ✓ Study epidemics of rapidly evolving pathogens
- ✓ Study evolution of pathogen epidemics and pathogens evolution

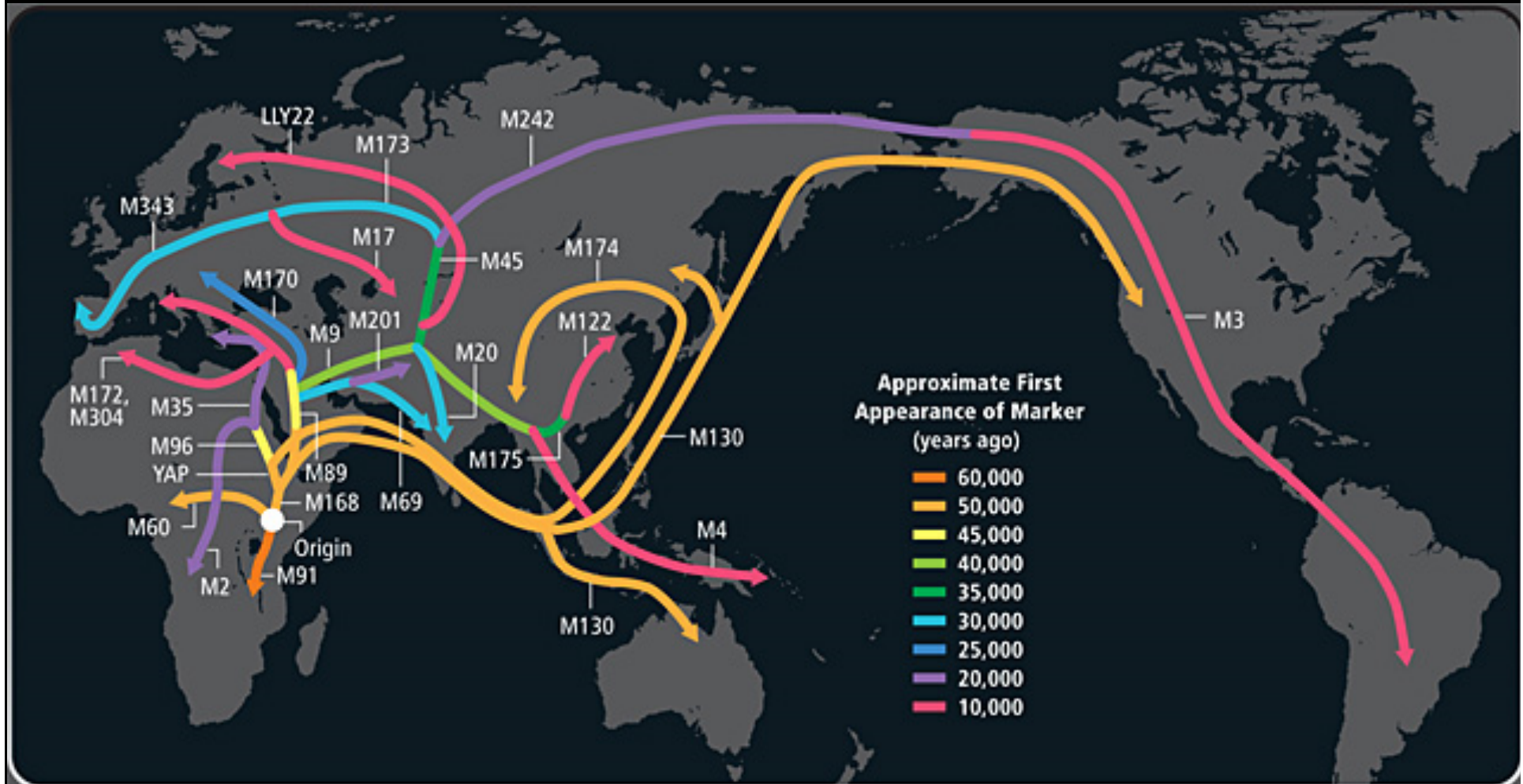
Προέλευση του Ανθρώπου



Μετανάστευση του Ανθρώπου



Μετανάστευση του σύγχρονου Ανθρώπου



Εκπαιδευτικοί στόχοι

- ❖ Τι είναι η Μοριακή Επιδημιολογία και ποια/ες οι διαφορές με την κλασσική επιδημιολογία
- ❖ **Τι είναι η φυλογένεια ή φυλογενετικό δένδρο**
 - ❖ **Ορολογία (πχ extant, ancestral, branch point, branch length)**
- ❖ Γιατί και πως φτιάχνουμε φυλογενετικό δένδρο
 - Algorithms to build phylogenetic trees
 - Distance-based methods
 - Parsimony methods
 - Minimize the number of changes
 - Probabilistic methods
 - Tree that best explains the data using probabilistic models
- ❖ Ιογενείς λοιμώξεις και Επιδημιολογική επιτήρηση
 - origin, expansion, geographic distribution, evolution

What is phylogenetics ?

Phylogenetics is the study of evolutionary relationships.

Relationships *among* species:

rodents

birds

snakes

crocodiles

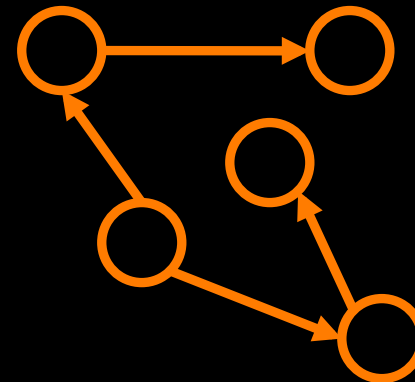
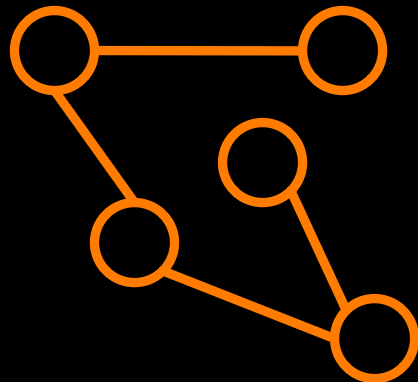
primates

marsupials

lizards

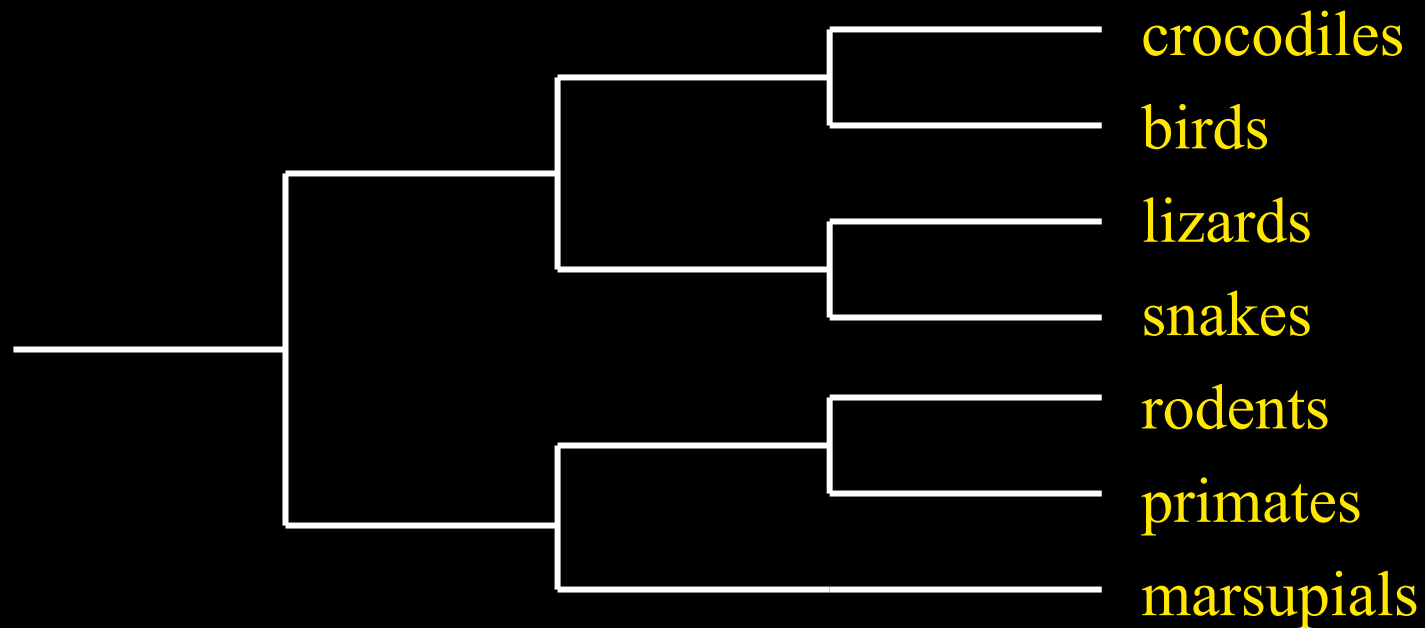
What is a tree ?

- Graph theoretically:
 - Undirected case: graph without cycles
 - Directed case: underlying undirected graph is a tree
 - Often it is required that $\text{in degree}(v) \leq 1$ for all v



What is phylogenetics ?

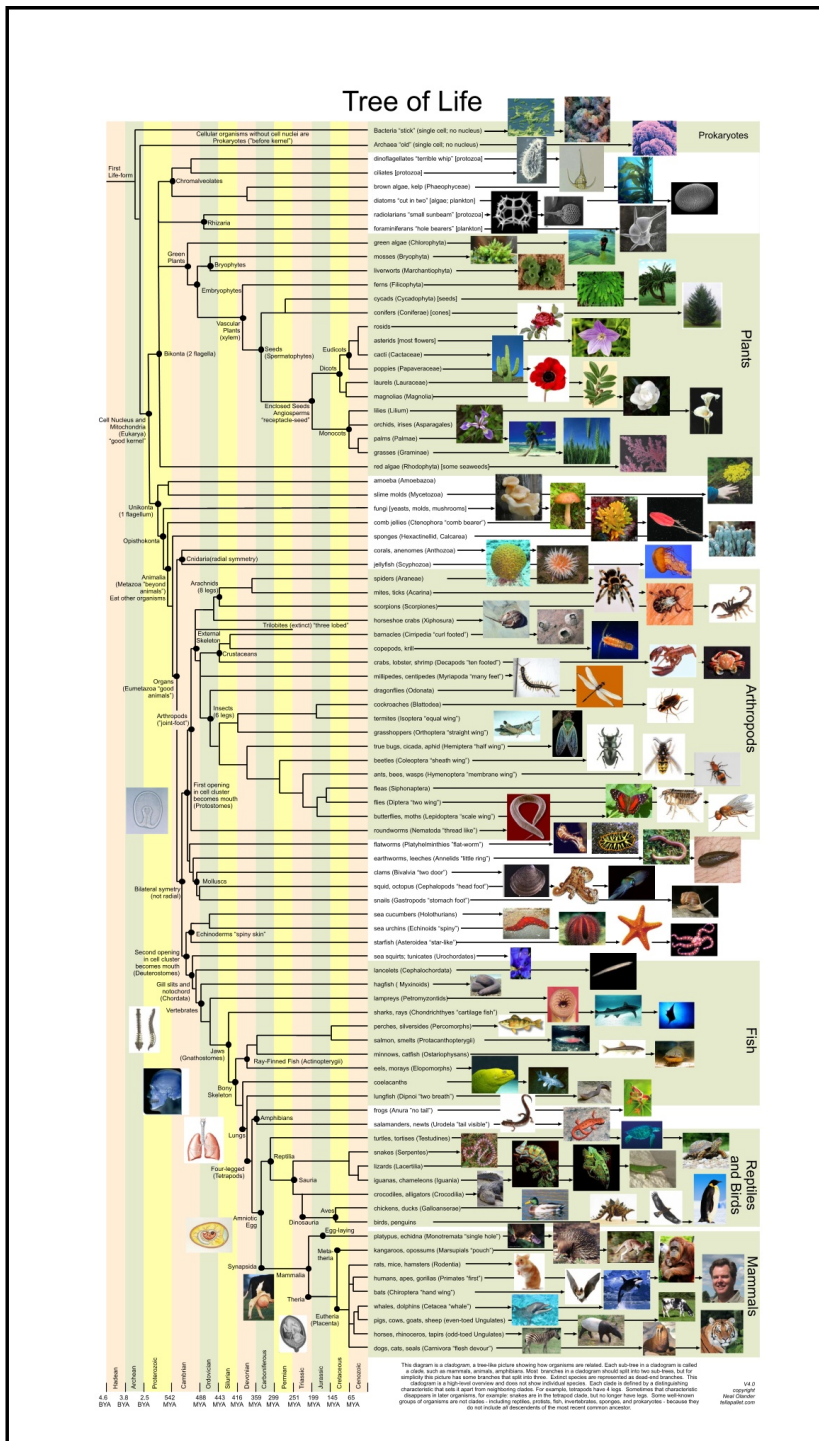
Relationships *among* species:



This is an example of a **phylogenetic tree**.

Φυλογενετικά Δένδρα

- A tree that describes evolutionary relationships among entities
 - Species, genes, strains
- This relationship is called “phylogeny”
- Leaves represent extant (current day) species
- Internal nodes represent ancestral species
- Phylogenetics:
 - The task for inferring the phylogenetic tree from observations in existing organisms



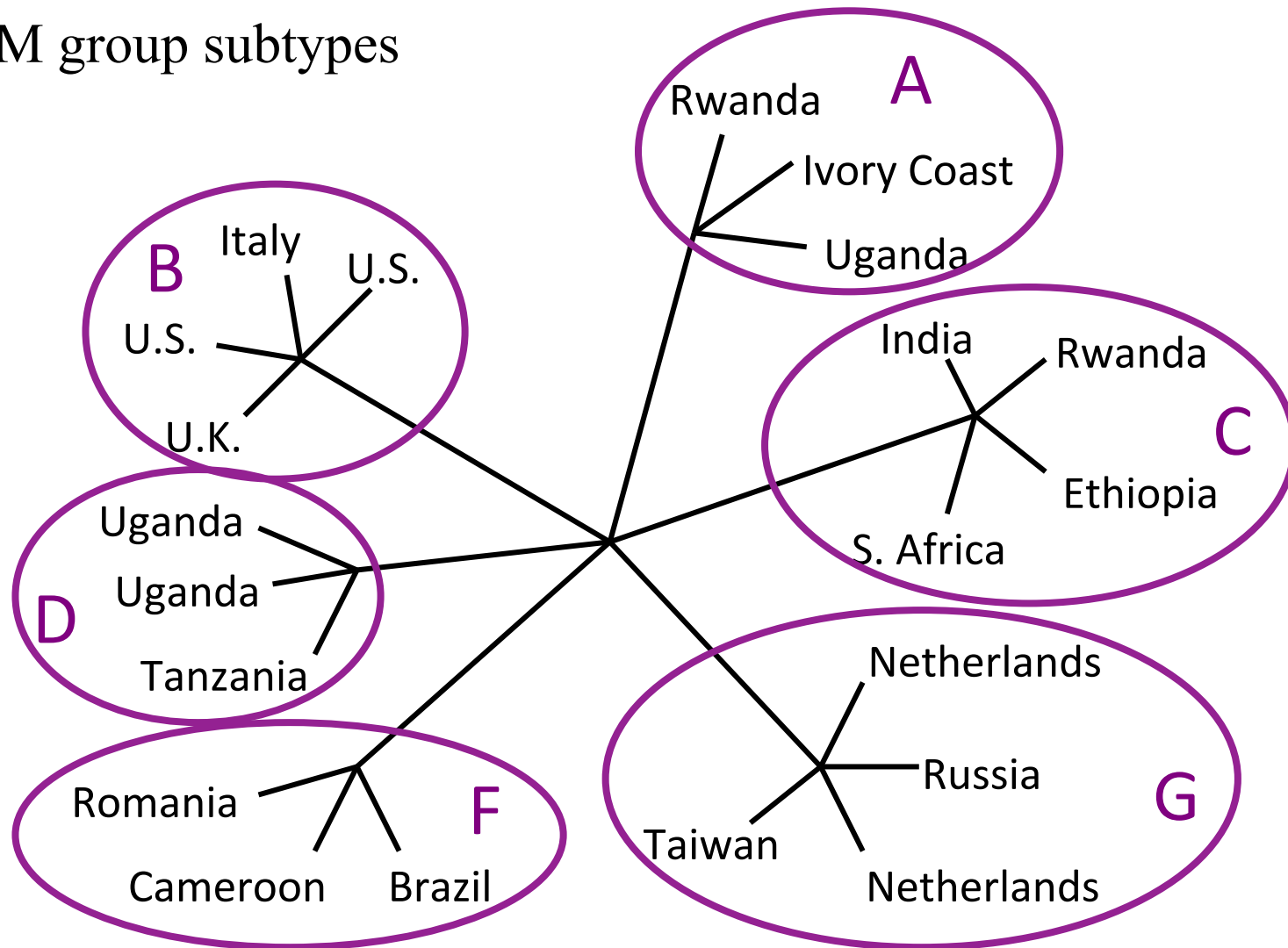
Relationships among species

Tree of life aims to represent the phylogeny of all species on earth

<https://tree.opentreeoflife.org/about/synthesis-release/v9.1>

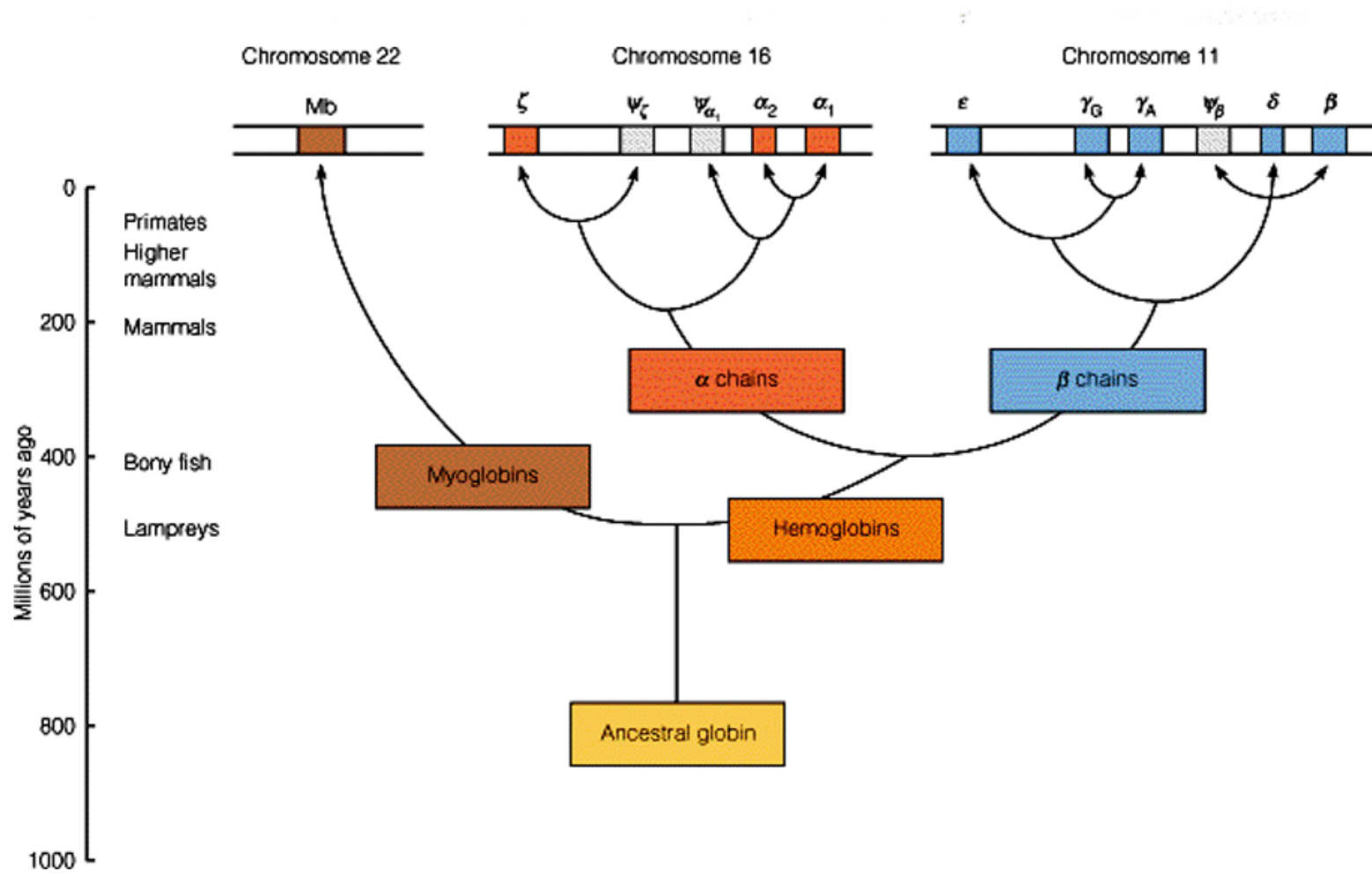
Relationships within species

HIV-1 M group subtypes

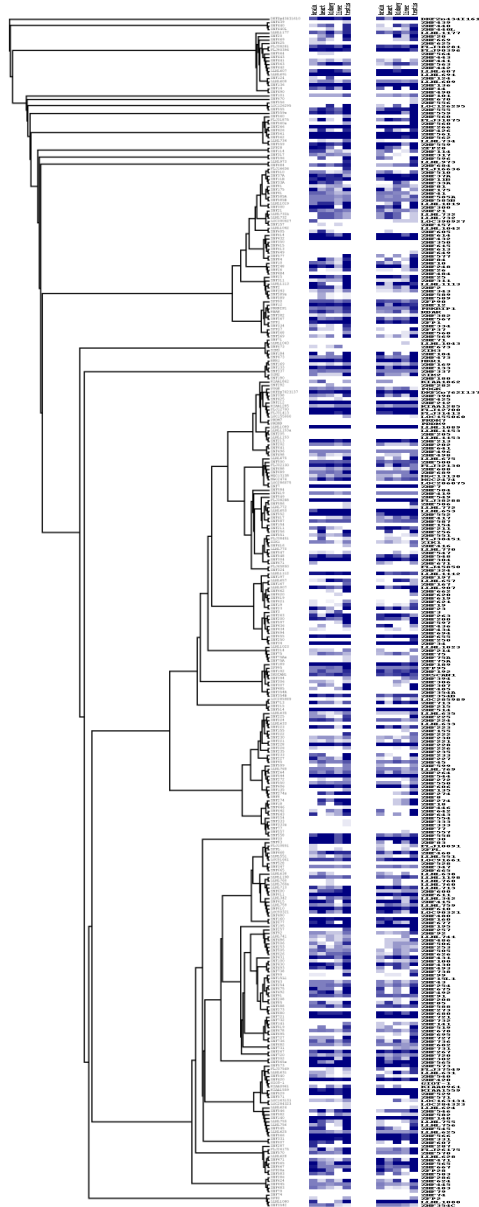


Relationships among genes

Example Gene Tree: Globins



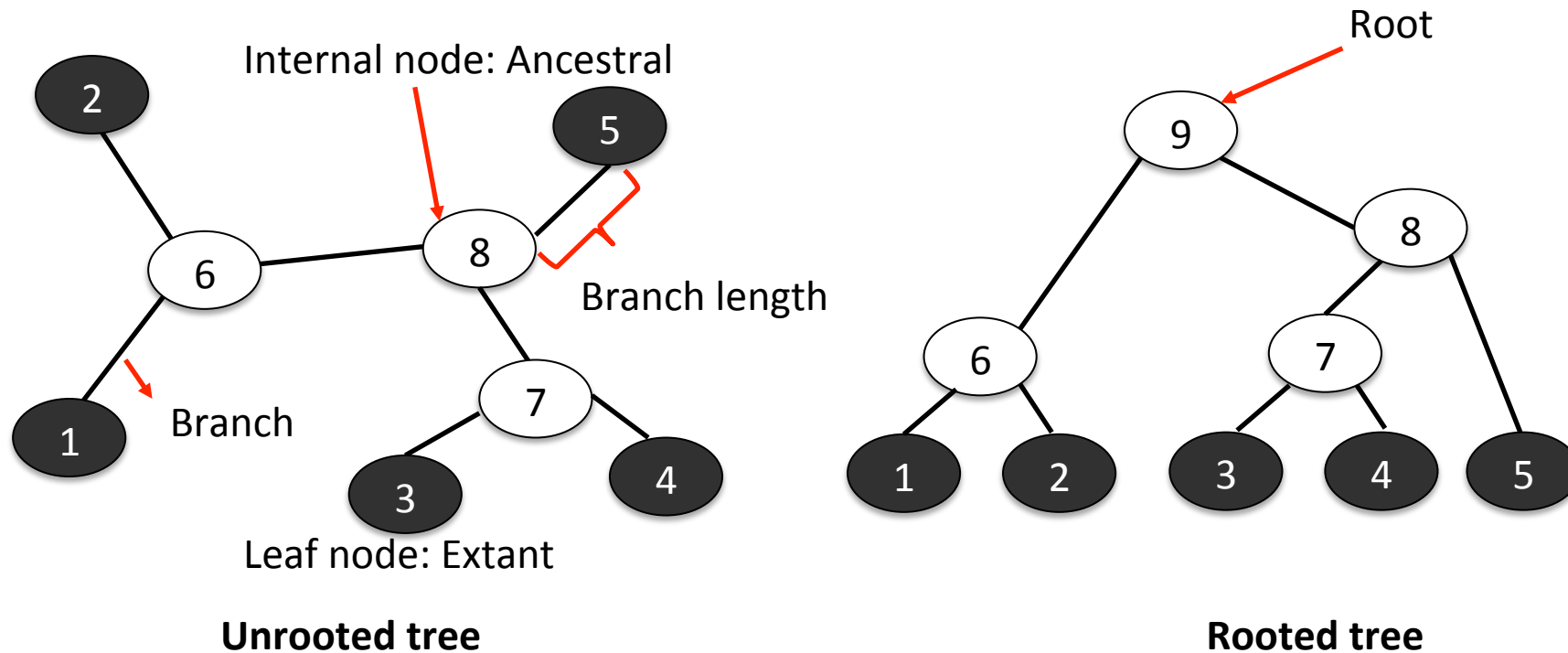
A tree of gene expression patterns in multiple tissues



Phylogenetic tree basics I

- Leaves represent entities (genes, species, individuals/strains) being compared
 - the term *taxon* (*taxa* plural) is used to refer to these when they represent species and broader classifications of organisms
 - For example if taxa are species, the tree is a species tree
- Internal nodes are ancestral units
- Phylogenetic trees can be rooted or unrooted
 - the root represents the common ancestor
- In a *rooted* tree, path from root to a node represents an evolutionary path
 - Gives directionality to evolutionary time
- An *unrooted* tree specifies relationships among taxa, but lacks directionality information

Phylogenetic tree basics II



Each tree topology represents a different evolutionary history

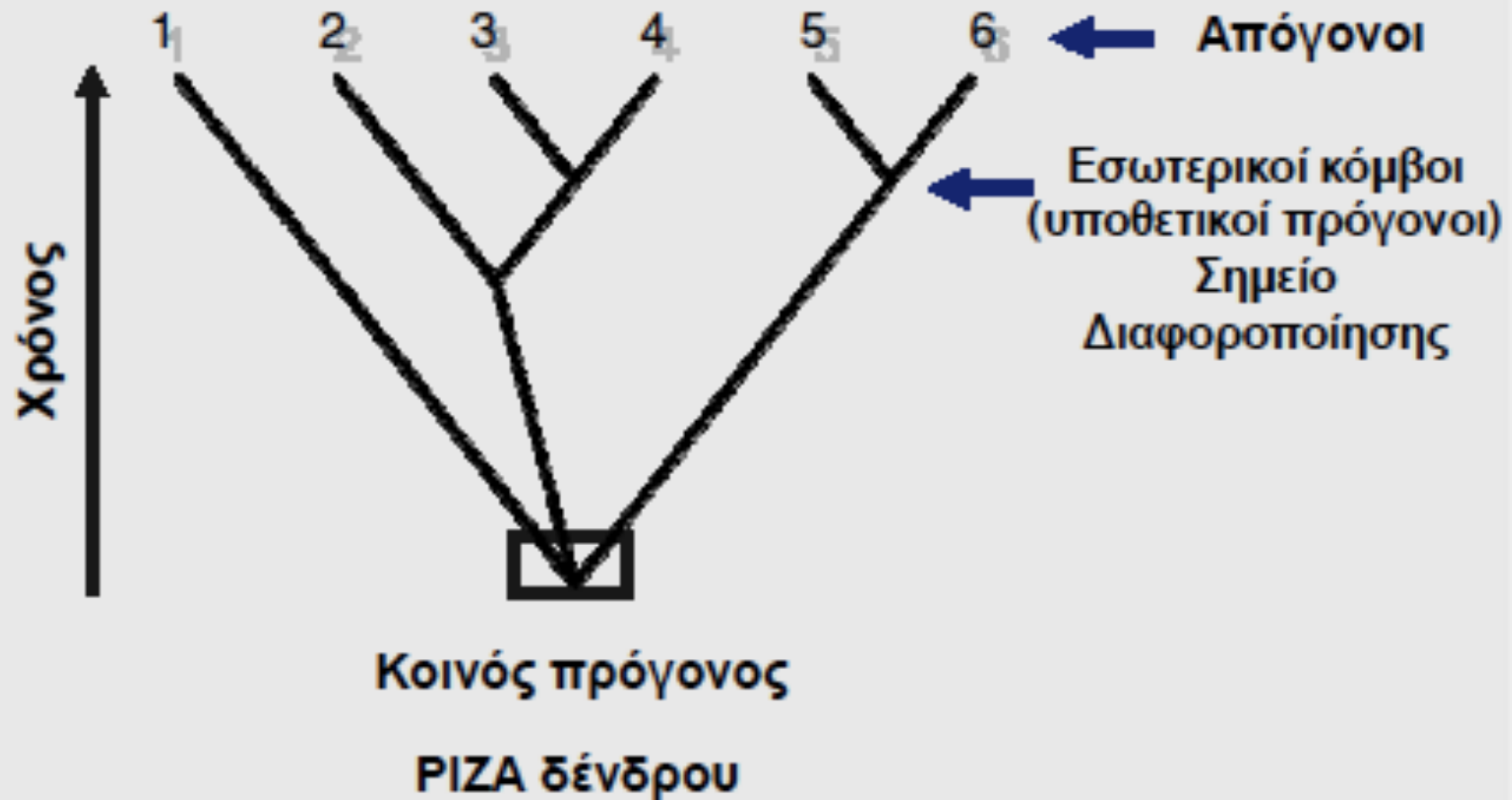
For a species tree, internal nodes represent speciation events

Branch length describes the evolutionary divergence between two nodes

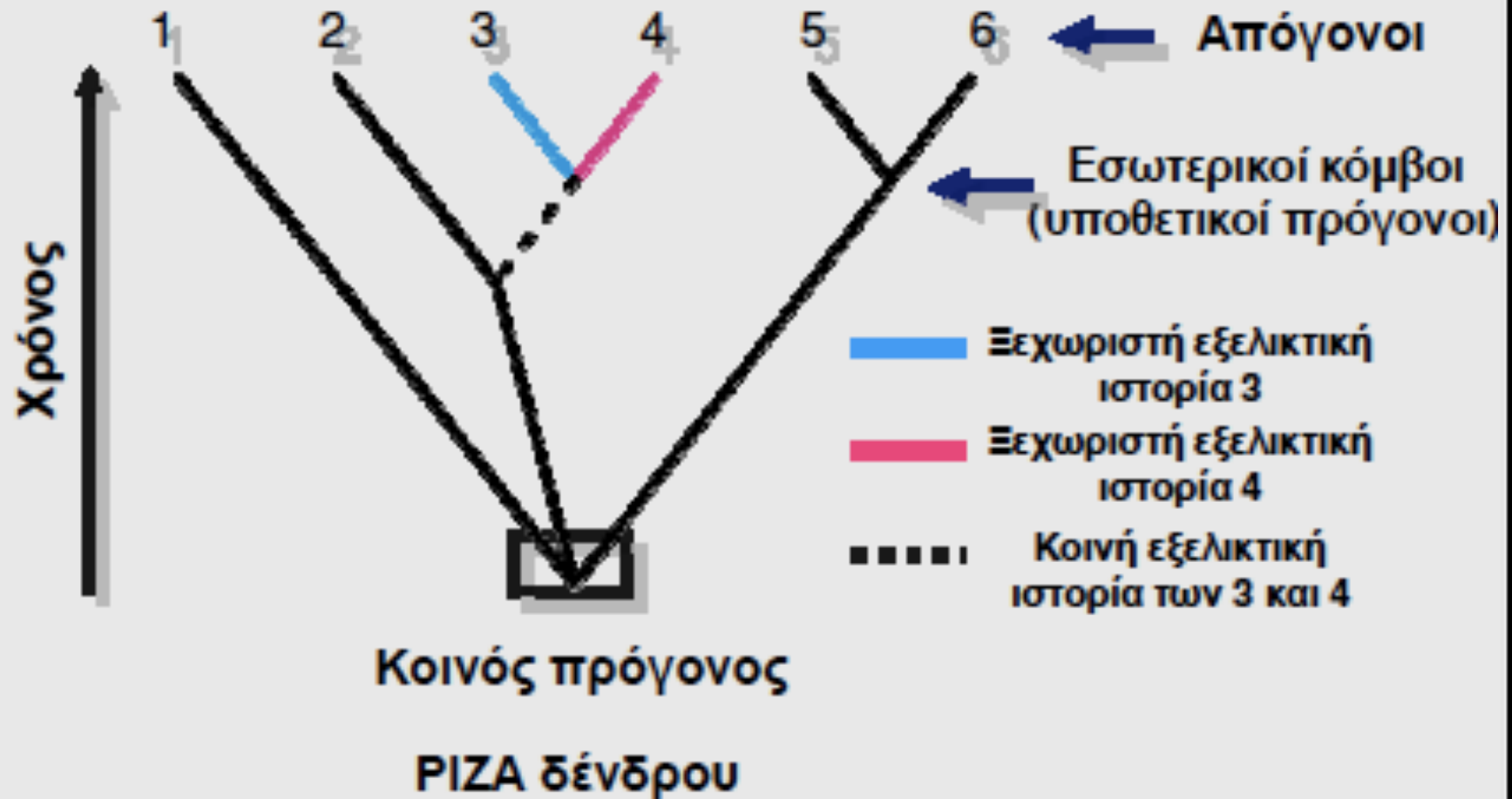
Στοιχεία Φυλογενετικού Δένδρου

- Η ερμηνεία ενός φυλογενετικού δένδρου μοιάζει πολύ με αυτή ενός γενεαλογικού δένδρου.
- Η ρίζα του δένδρου αναπαριστά τον κοινό (??) πρόγονο και οι άκρες τους απογόνους τους.
- Όσο μετακινούμαστε από τη ρίζα προς τις άκρες του δένδρου ακολουθούμε την εξελικτική πορεία του χρόνου.
- Στη φυλογένεια όταν ένας πληθυσμός διαχωρίζεται, αναπαρίσταται ως διακλάδωση.

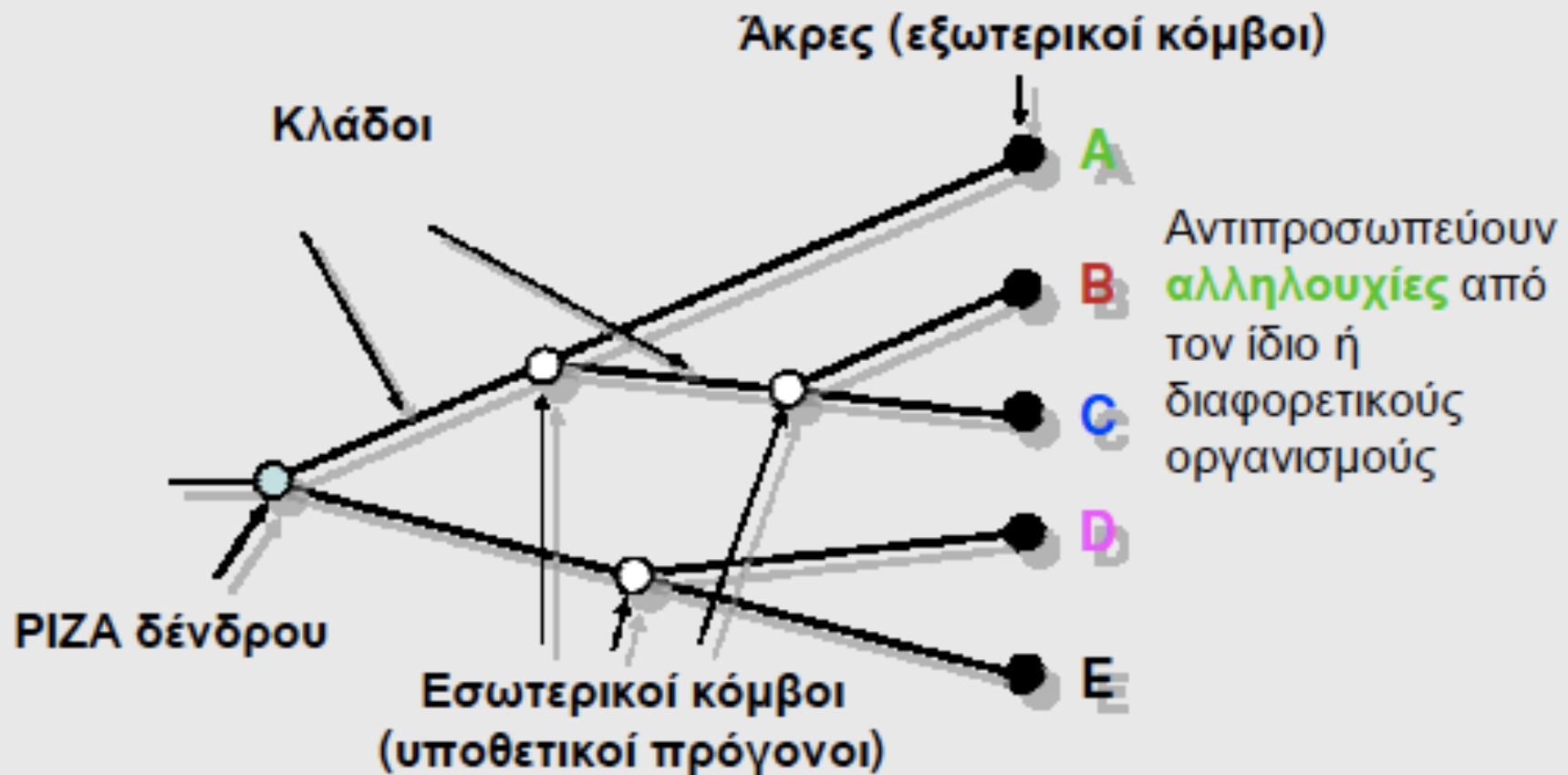
Στοιχεία Φυλογενετικού Δένδρου



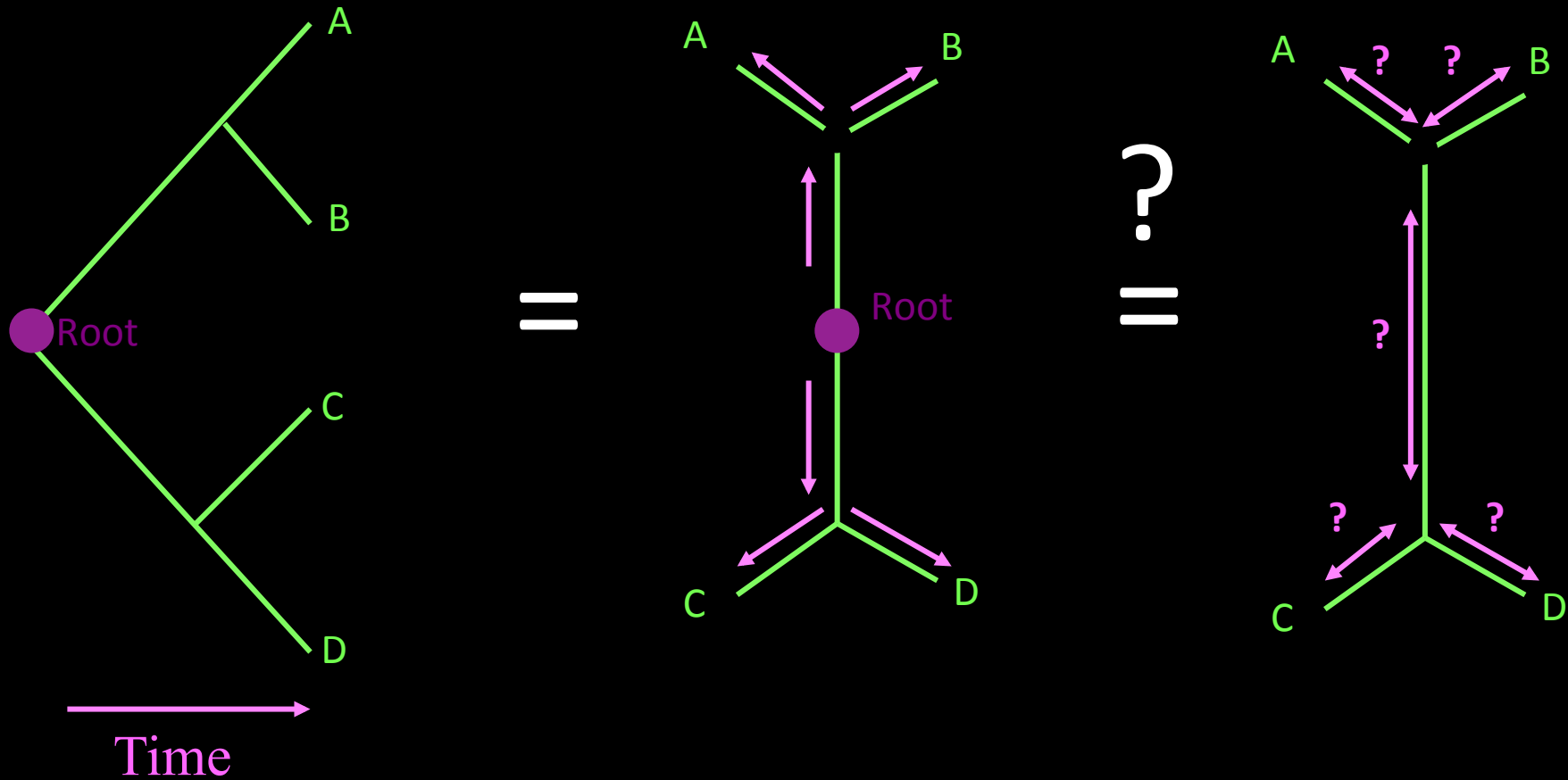
Στοιχεία Φυλογενετικού Δένδρου



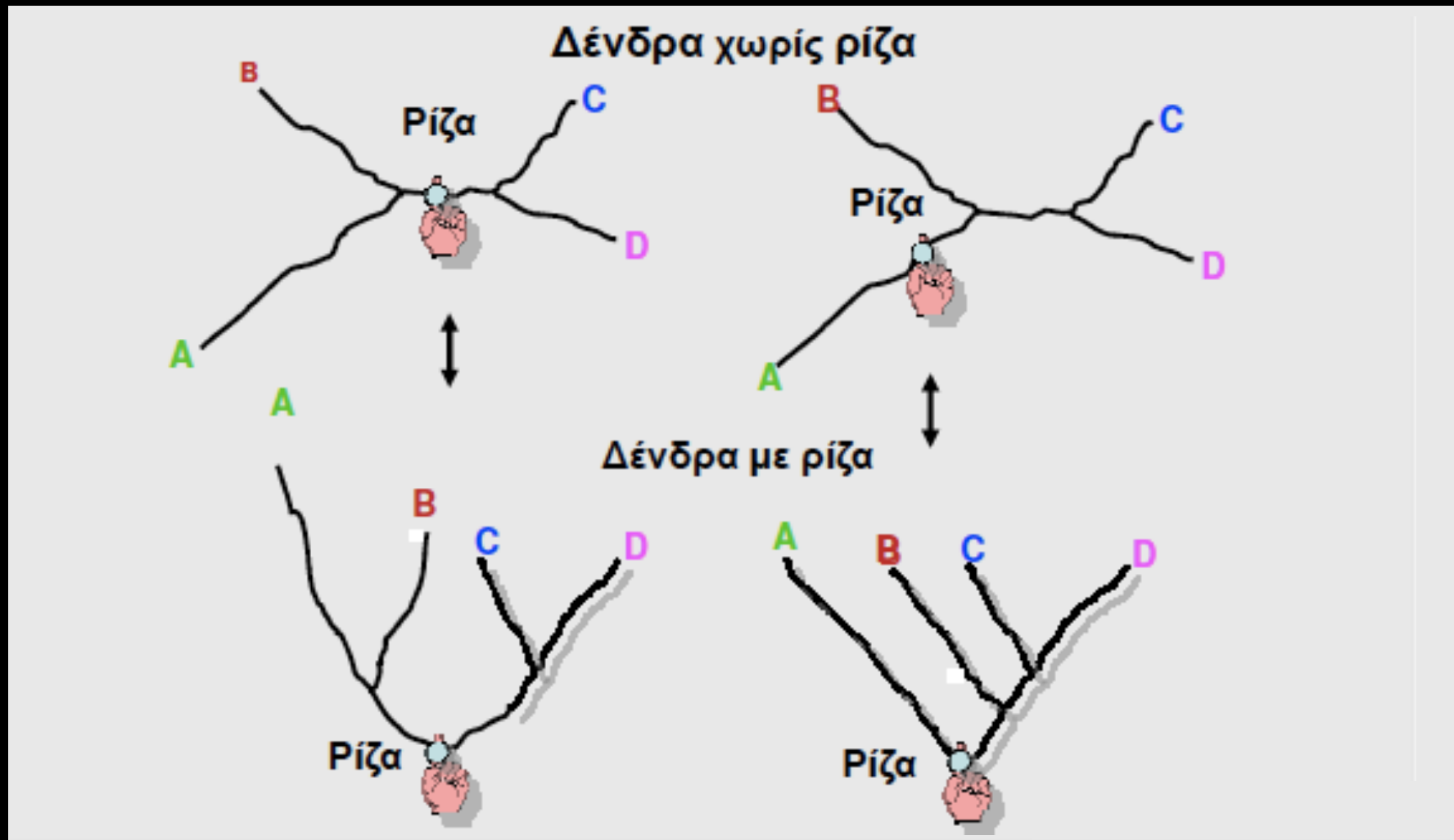
Στοιχεία Φυλογενετικού Δένδρου



Rooted vs Unrooted Trees



Φυλογενετικά δένδρα με ρίζα και χωρίς ρίζα



Η έννοια της κλαδοποίησης

- Ένας «κλάδος» (clade) αποτελεί μια «ομαδοποίηση» ΟΛΩΝ των αλληλουχιών που έχουν προκύψει από κοινό πρόγονο.
- Αν ψαλιδίσουμε τη ρίζα ενός «κλάδου», όλες οι αλληλουχίες που θα αποχωριστούν από το δένδρο αποτελούν το συγκεκριμένο κλαδί.
- Συνεπώς, τα μέλη ενός κλάδου αποτελούν τα μέλη που θα αποχωριστούν αν ψαλιδίσουμε τη ρίζα του (κατ' αναλογία με ένα τσαμπί σταφύλι, ή ένα κλαδί ενός φυσικού δένδρου)

Μοριακή Φυλογένεια – φυλογενετικά δένδρα

- Ο αριθμός των δυνατών δένδρων (χωρίς ρίζα) που είναι δυνατό να κατασκευαστούν από έναν αριθμό αλληλουχιών είναι:

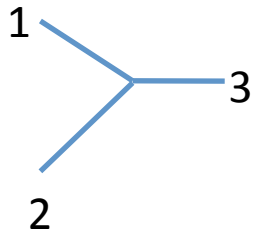
$$N = \frac{(2n - 5)!}{2^{n-3} (n-3)!}$$

- Ενδεικτικά αναφέρεται ότι για 5 αλληλουχίες ο αριθμός N είναι 15, ενώ για 10 ο αριθμός N αυξάνεται σε 2.027.025

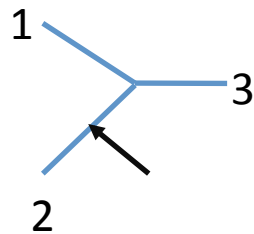
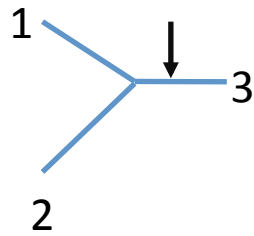
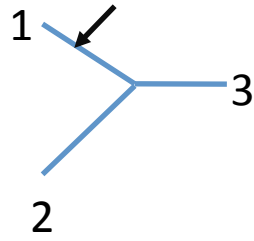
Phylogenetic tree counting I

- A rooted binary tree with n leaf nodes has
 - $n-1$ internal nodes
 - $2n-2$ edges/branches
- An unrooted binary tree with n leaf nodes has
 - $n-2$ internal nodes
 - $2n-3$ edges/branches
 - A root can be added to any of these branches to give $2n-3$ rooted trees for any unrooted tree
- E.g. for $n=3$ there is *one* unrooted tree and *three* rooted trees

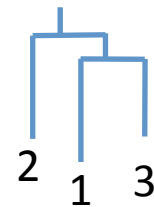
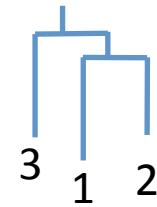
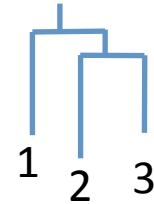
Phylogenetic tree counting II



An unrooted tree



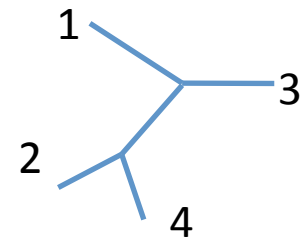
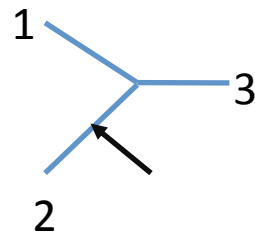
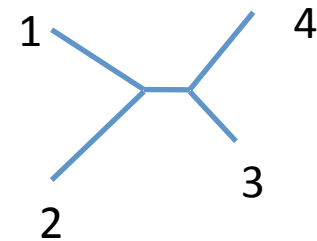
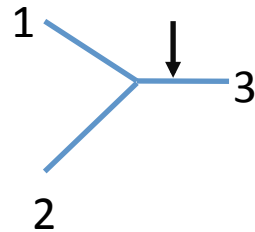
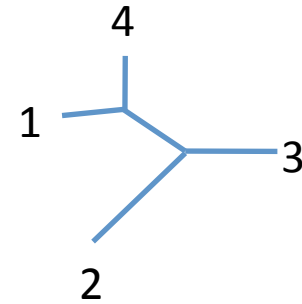
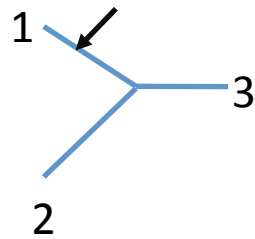
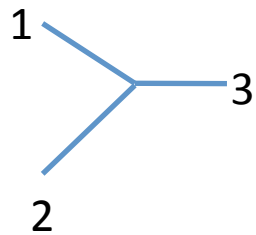
Possible positions for root



Rooted trees

Phylogenetic tree counting III

- Instead of adding a root we could add a branch for the $n+1^{\text{th}}$ taxon



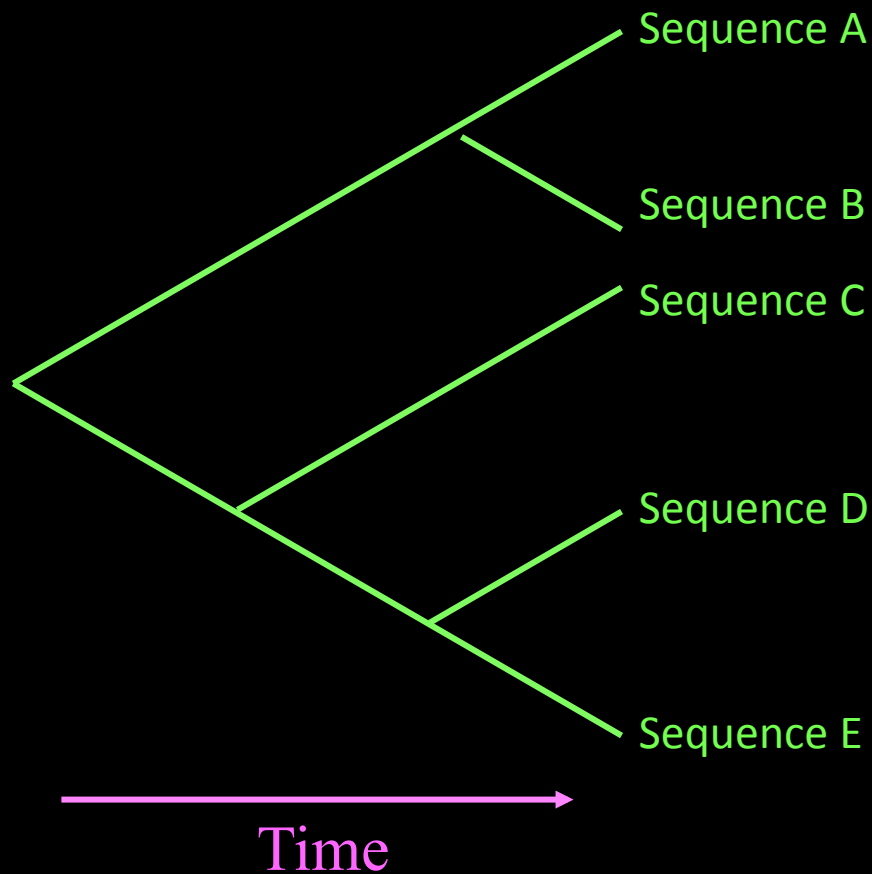
Number of Possible Trees

- given n sequences, there are $\prod_{i=3}^n (2i - 5)$ possible unrooted trees
- and $(2n - 3) \prod_{i=3}^n (2i - 5)$ possible rooted trees
- This grows very fast
 - For $n=10$, we have 2 million unrooted trees
 - For $n=20$, we have $2.2 * 10^{20}$

Αριθμός πιθανών φυλογενετικών δένδρων

Αλληλουχίες	Δένδρα χωρίς ρίζα	Δένδρα με ρίζα
3	1	3
4	3	15
5	15	105
6	105	945
10	2,027,025	34,459,425
30	8.69×10^{36}	4.95×10^{38}
N	$(2N - 5)! / 2^{N-3} (N - 3)!$	$(2N - 3)! / 2^{N-2} (N - 2)!$

Interpreting a Phylogeny

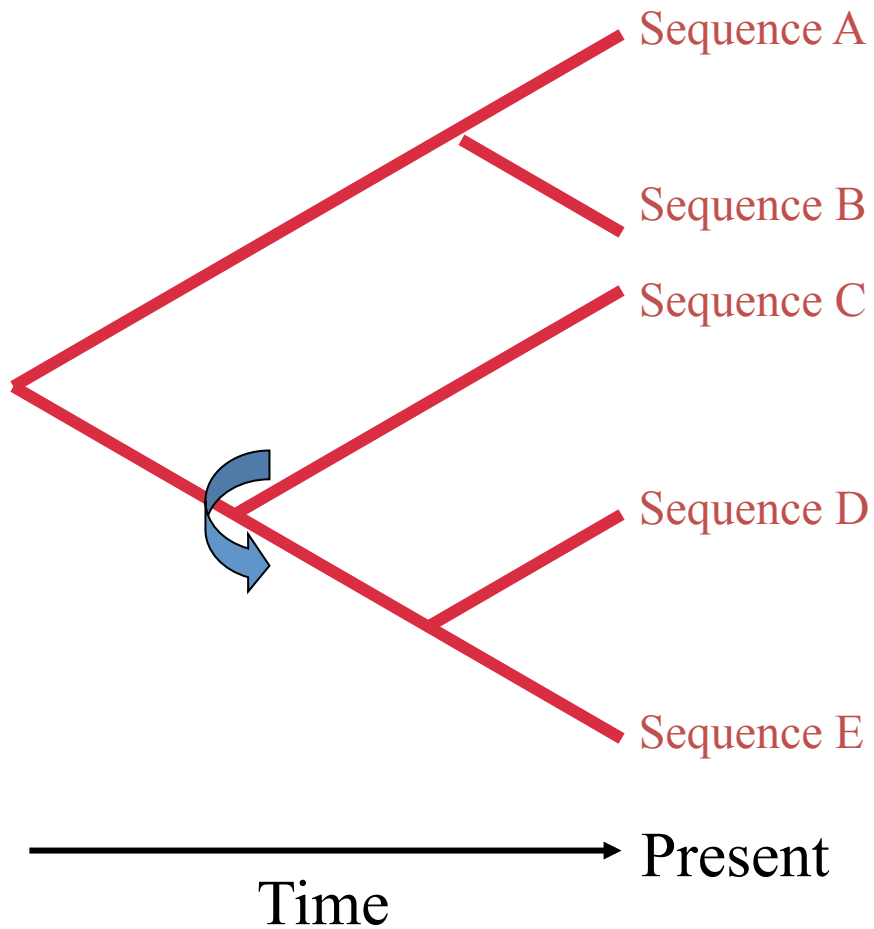


Which sequence is most closely related to B?

A, because B diverged from A more recently than from any other sequence.

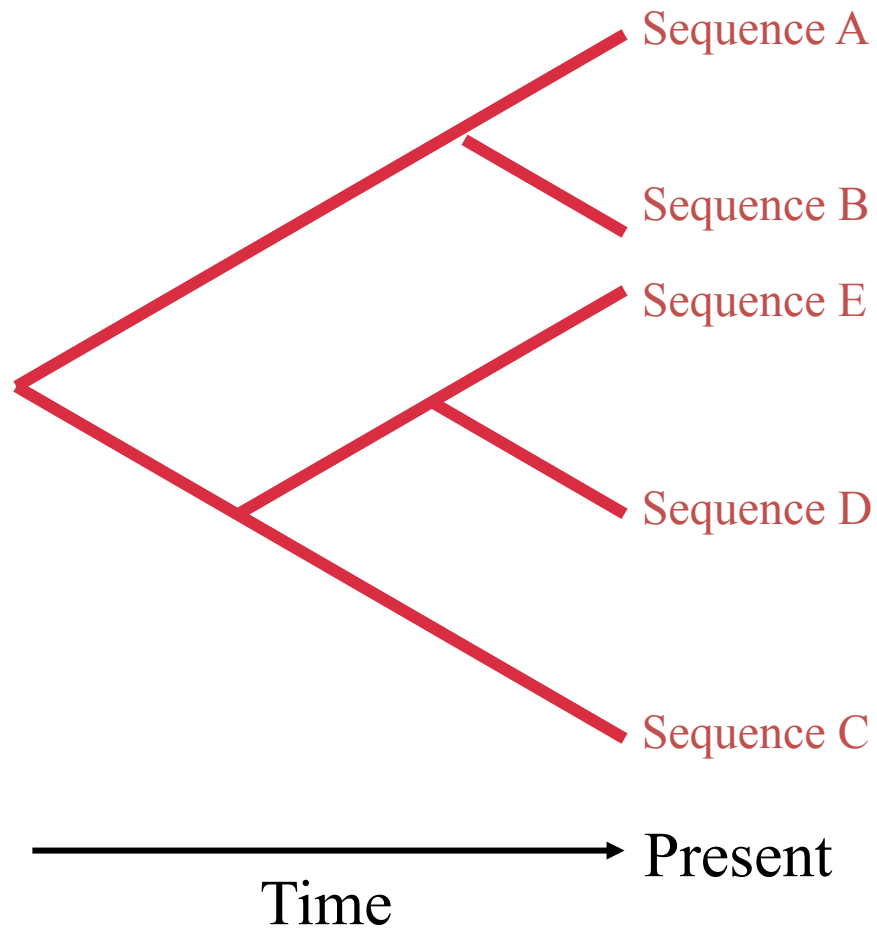
Physical position in tree is **not meaningful!** Only tree structure matters.

Phylogenetic concepts: Interpreting a Phylogeny



- *Physical* position in tree is not meaningful
- Swiveling can only be done at the nodes
- Only tree structure matters

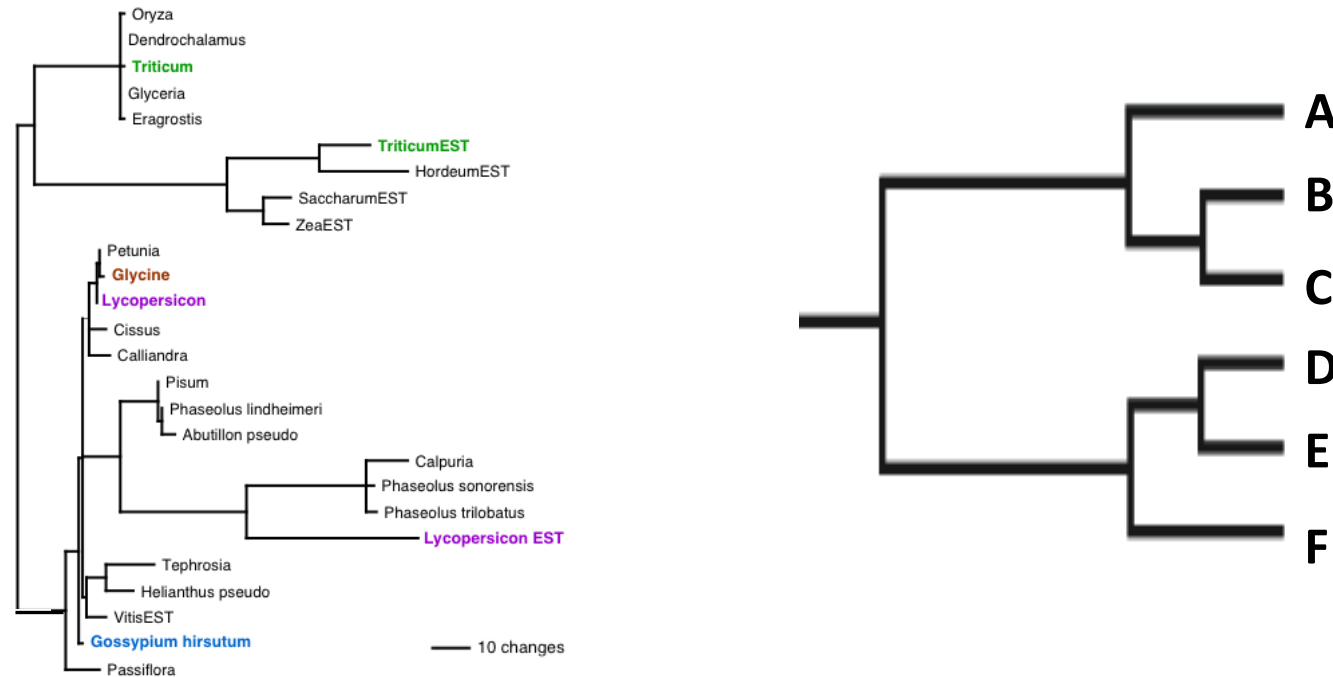
Phylogenetic concepts: Interpreting a Phylogeny



- *Physical* position in tree is not meaningful
- Swiveling can only be done at the nodes
- Only tree structure matters

Tree Terminology

Scaled vs. unscaled trees

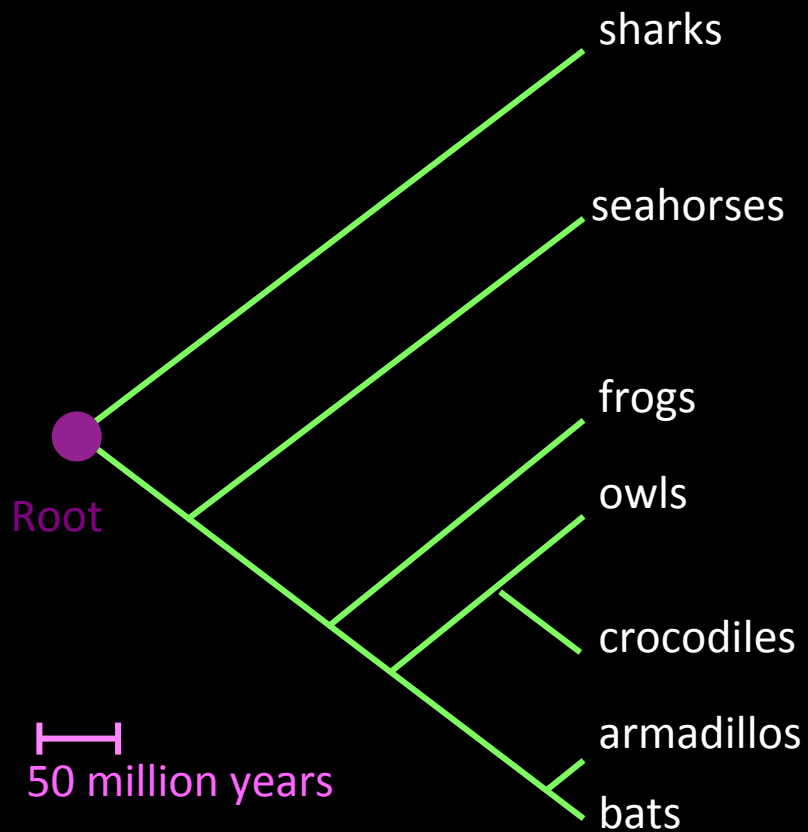


Scaled trees: Branch lengths are proportional to the number of nucleotide/ amino acid changes that occurred on that branch (usually a scale is included).

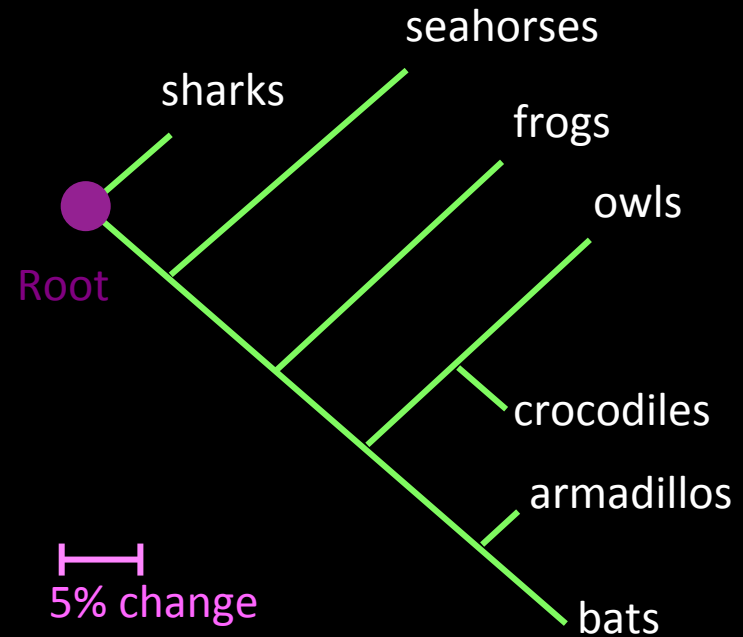
Unscaled trees: Branch lengths are not proportional to the number of nucleotide/ amino acid changes (usually used to illustrate evolutionary relationships only).

Tree Types

Evolutionary trees
measure *time*.



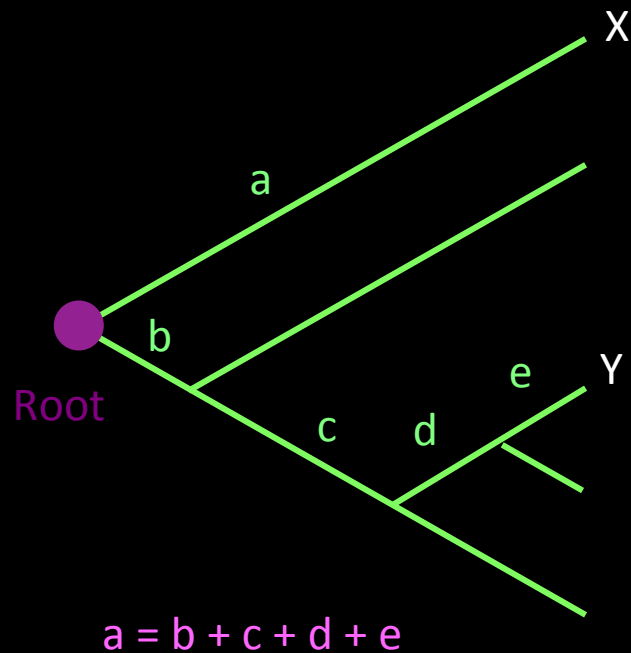
Phylograms
measure *change*.



Tree Properties

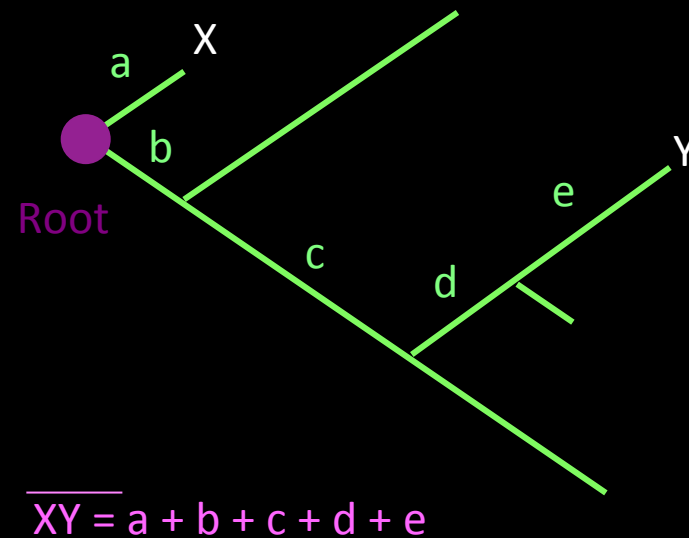
Ultrametricity

All tips are an equal distance from the root



Additivity

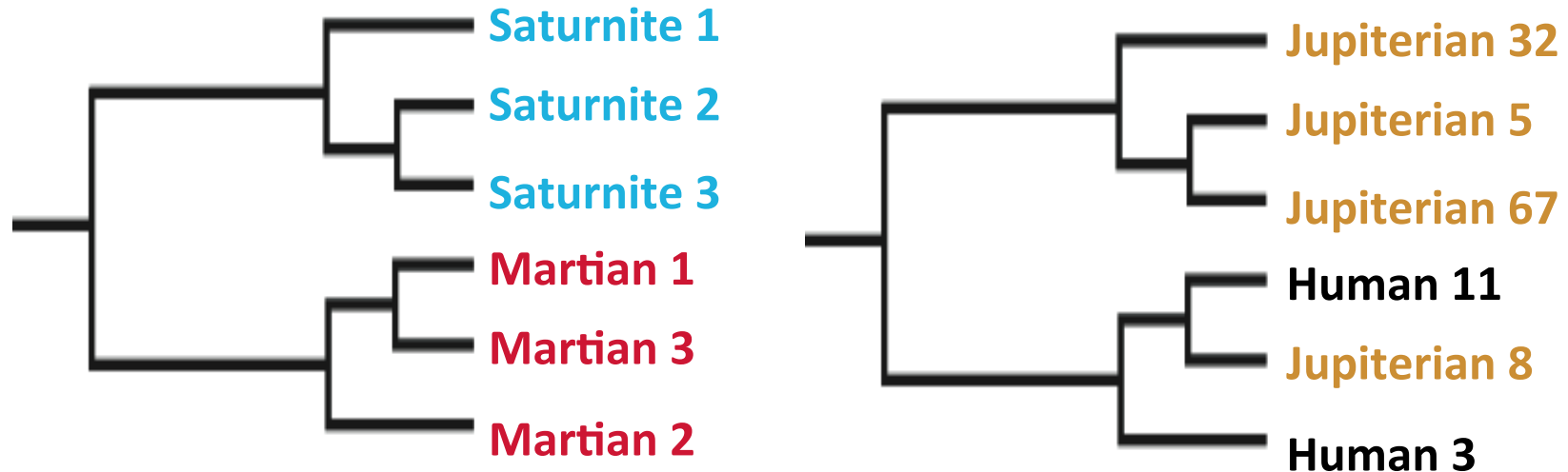
Distance between any two tips equals the total branch length between them.



In simple scenarios, evolutionary trees are ultrametric and phylograms are additive.

Tree Terminology

Monophyletic vs. paraphyletic



Monophyletic groups: All taxa within the group are derived from a single common ancestor and members form a natural clade.

Paraphyletic groups: The common ancestor is shared by other taxon in the group and members do not form a natural clade.