

# Η εφαρμογή της πληροφορικής στην ανάγνωση της εξελικτικής ιστορίας των ειδών

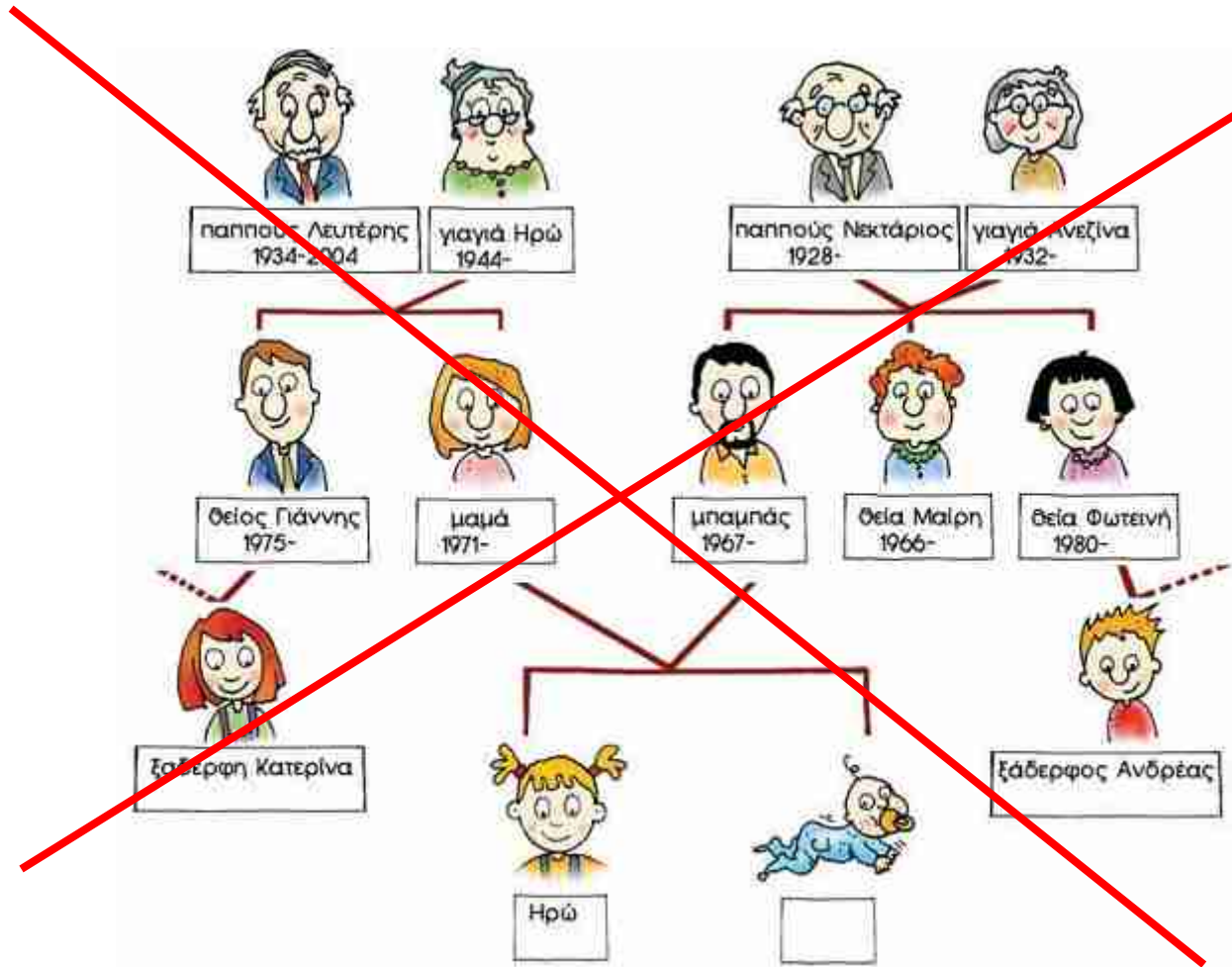
Αλέξανδρος Σταματάκης

ERA Chair, Institute of Computer Science, Foundation for Research and Technology - Hellas  
Research Group Leader, Heidelberg Institute for Theoretical Studies  
Full Professor, Dept. of Informatics, Karlsruhe Institute of Technology

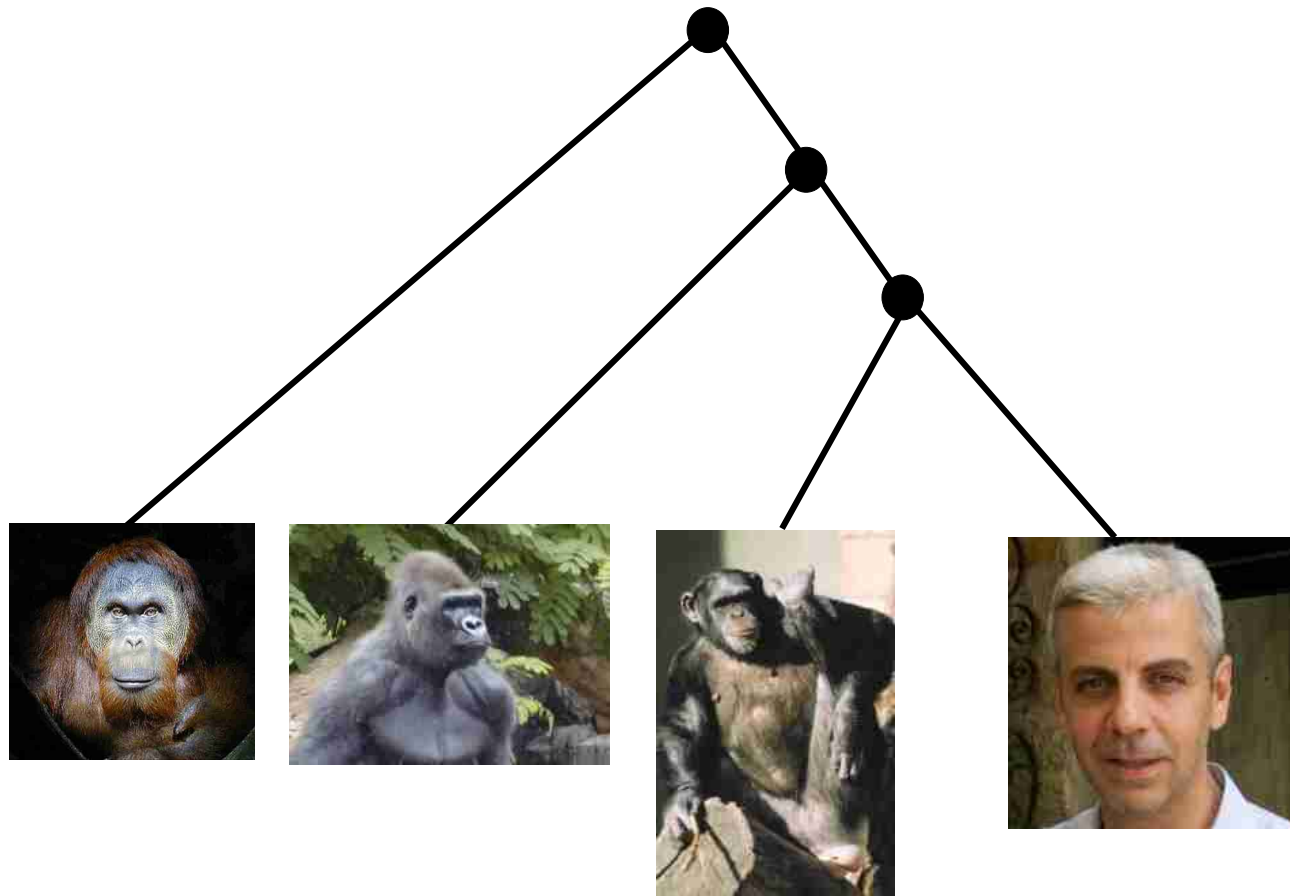
[www.biocomp.gr](http://www.biocomp.gr) (εργαστήριο Κρήτης)

[www.exelixis-lab.org](http://www.exelixis-lab.org) (εργαστήριο Χαϊδελβέργης)

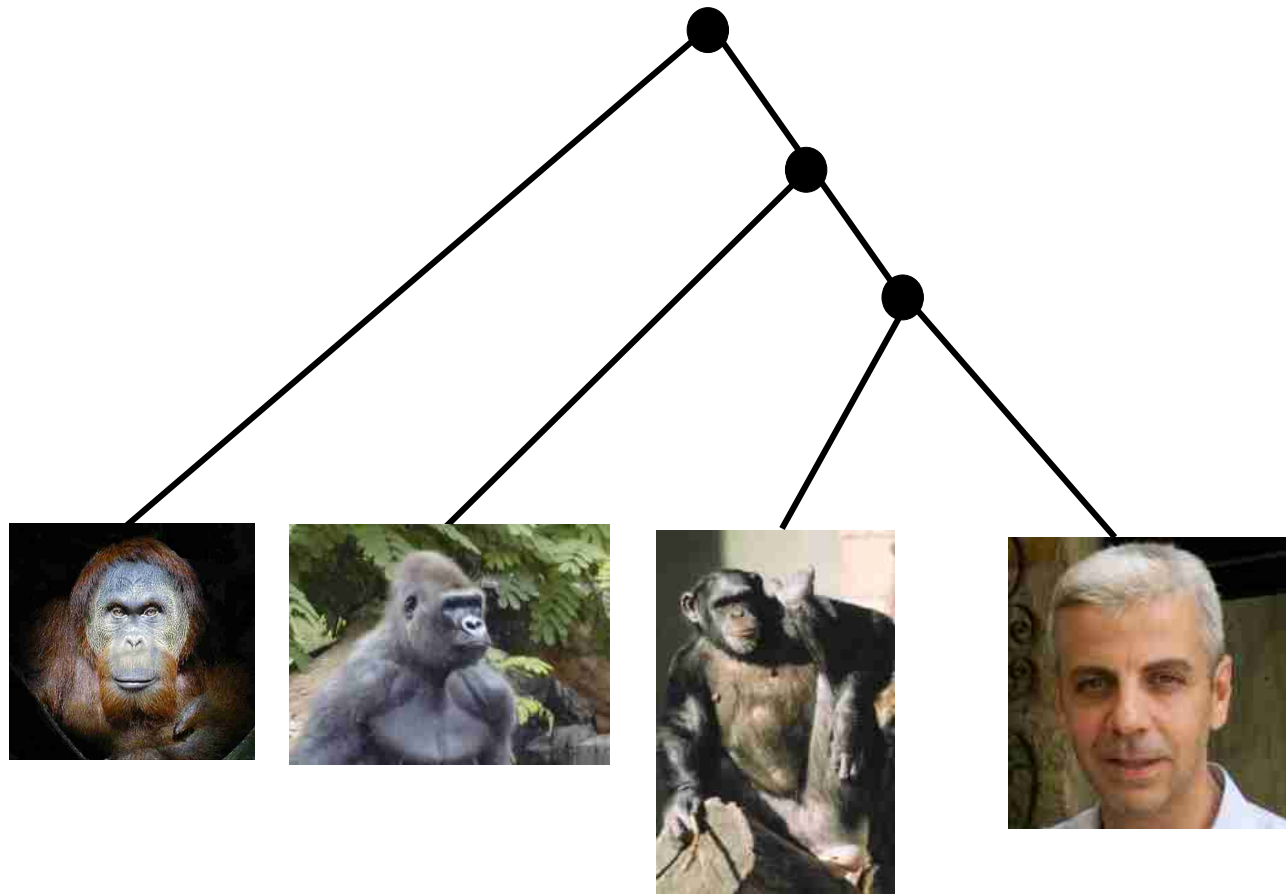
# Δεν θα μιλήσουμε για γενεαλογικά δέντρα



# Αλλά για φυλογενετικά δέντρα

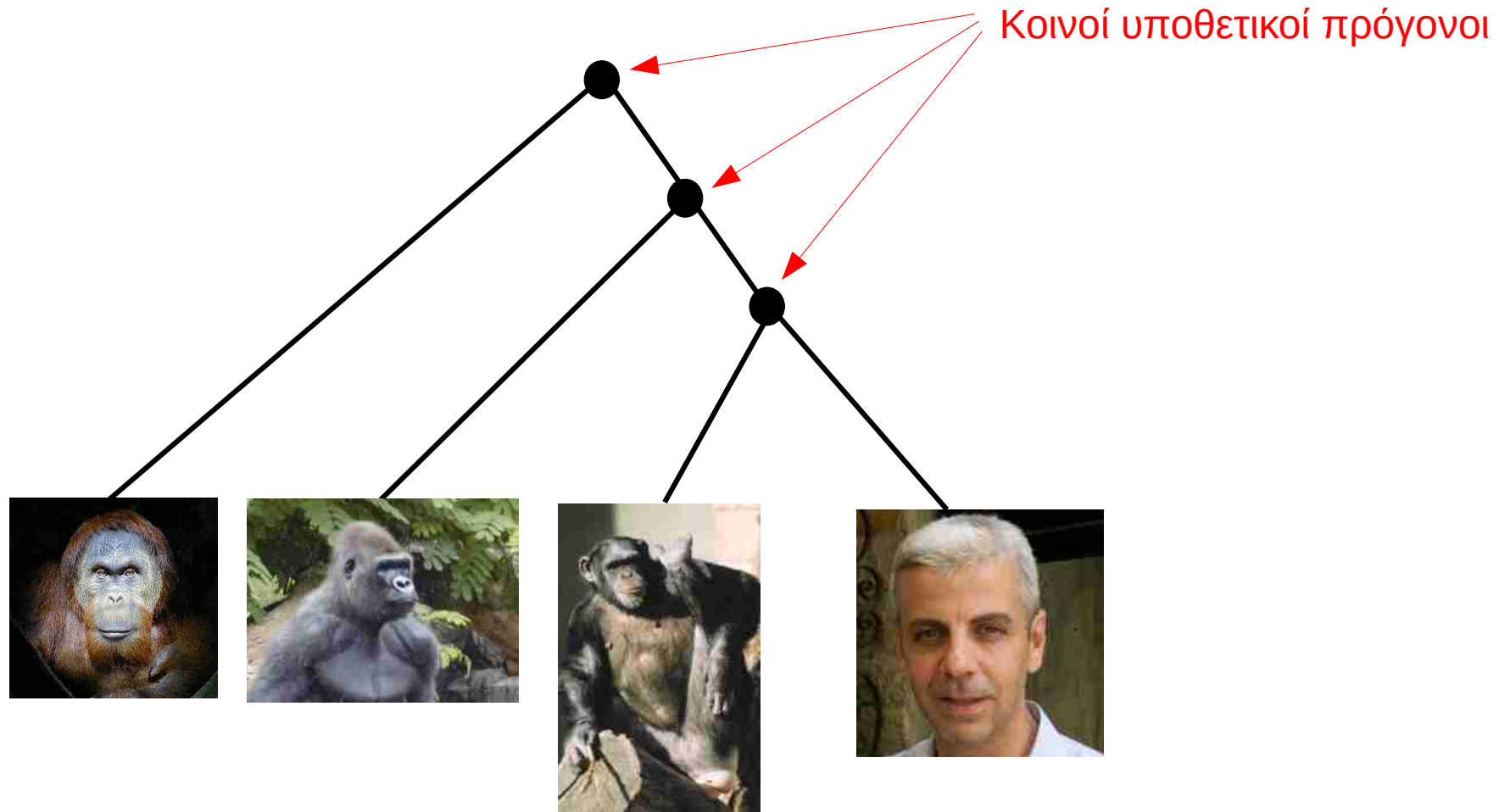


# Αλλά για φυλογενετικά δέντρα



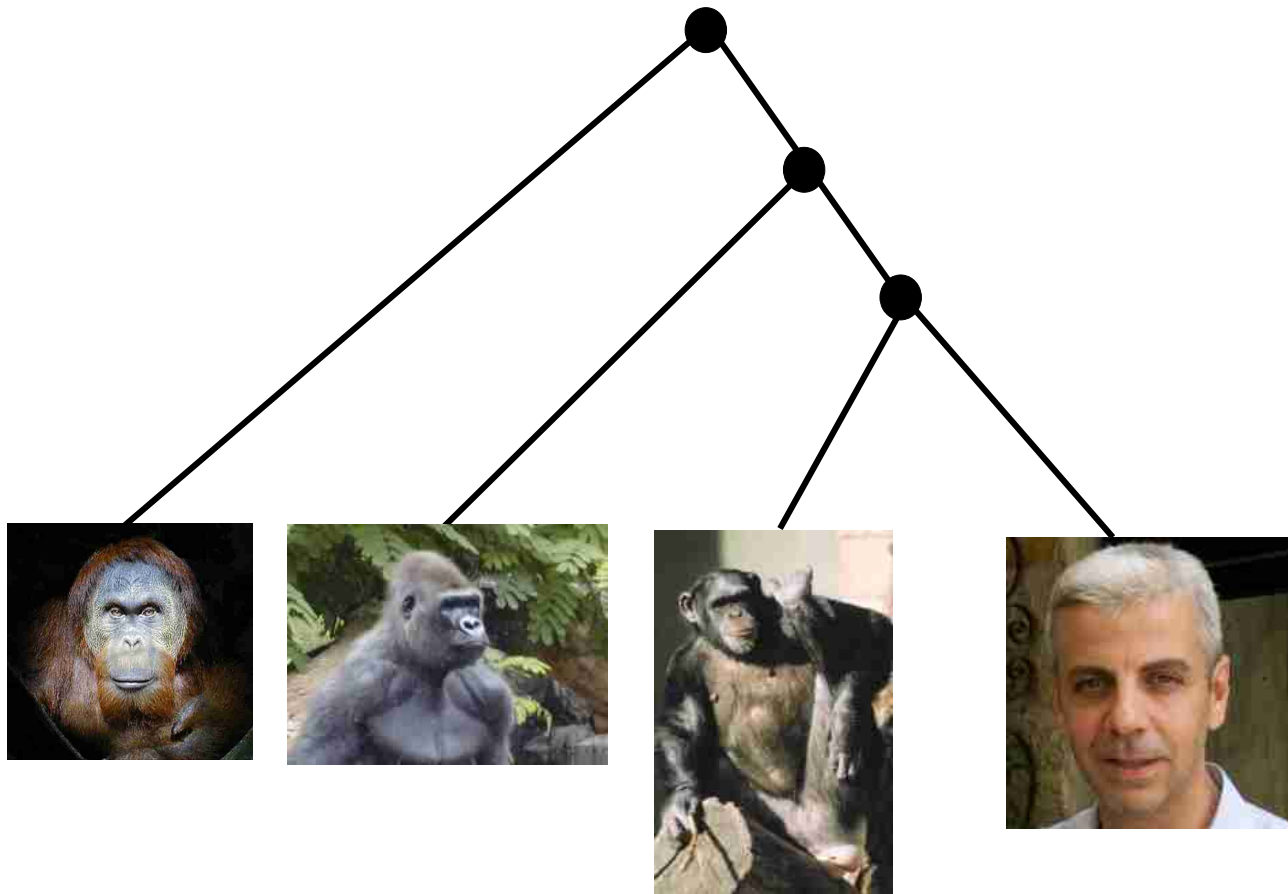
Μας ενδιαφέρει να ανατρέξουμε στο παρελθόν για να καταλάβουμε την εξελικτική ιστορία ανάμεσα **σε διαφορετικά είδη** !!!!

# Αλλά για φυλογενετικά δέντρα

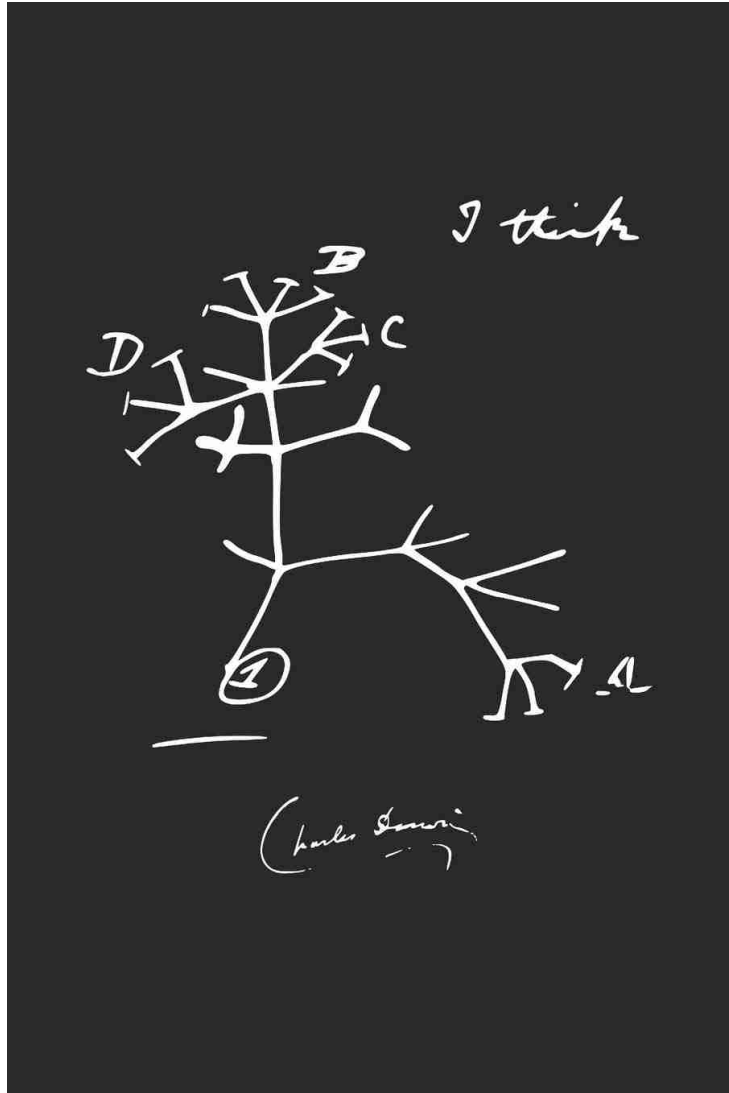


# Αλλά για φυλογενετικά δέντρα

Πώς κατασκευάζουμε ένα τέτοιο δέντρο;



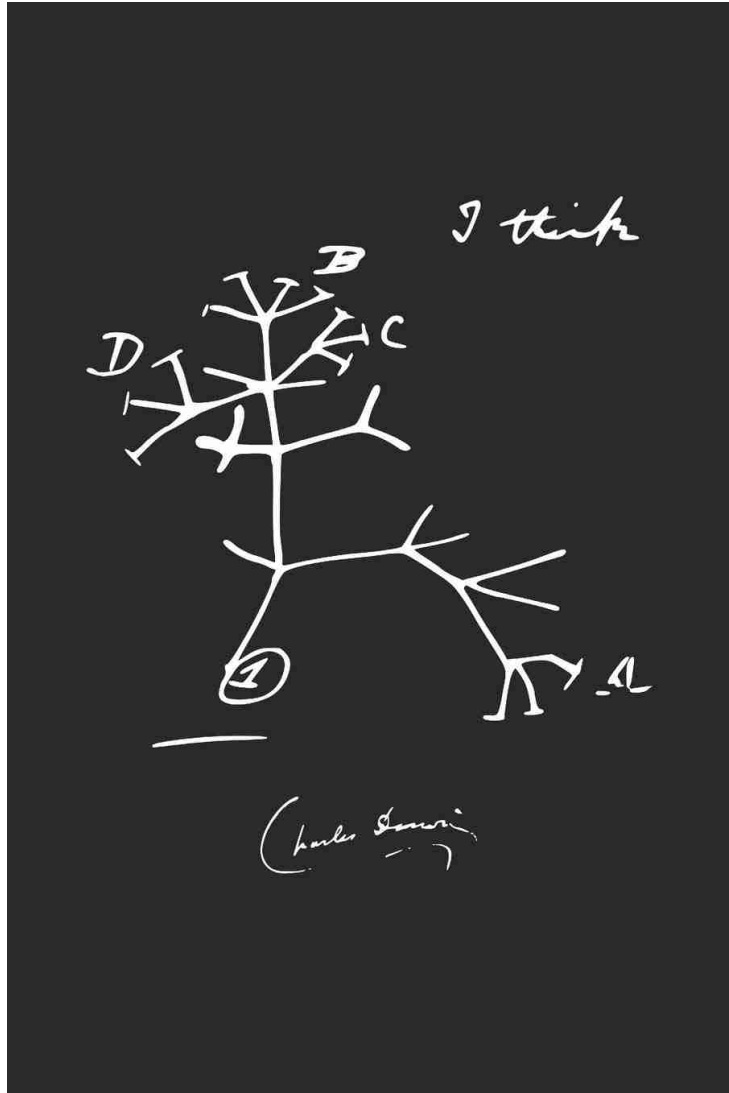
# Τον παλιό καιρό



Δαρβίνος 1837

**Συγκρίνοντας μορφολογικά χαρακτηριστικά**

# Τον παλιό καιρό



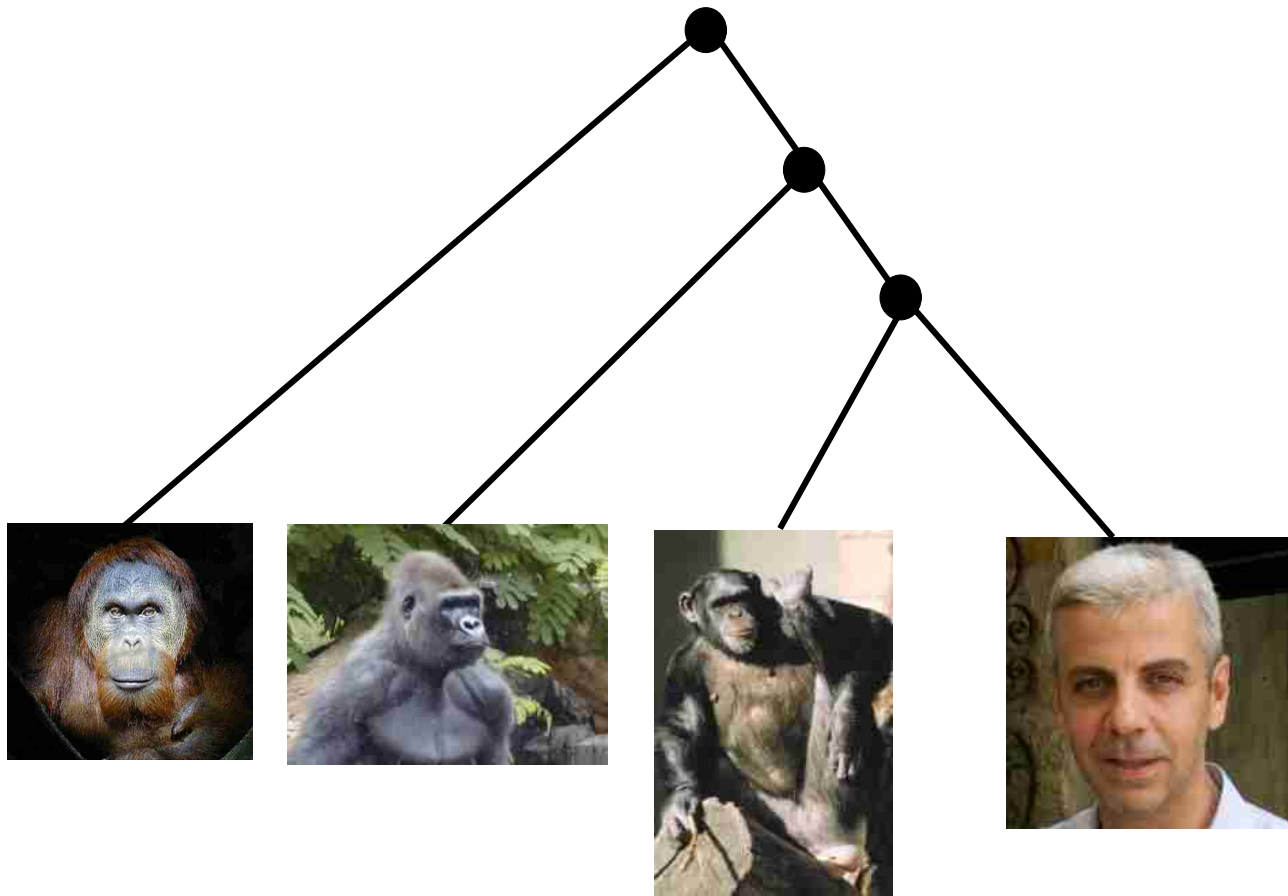
Δαρβίνος 1837

Συγκρίνοντας μορφολογικά χαρακτηριστικά

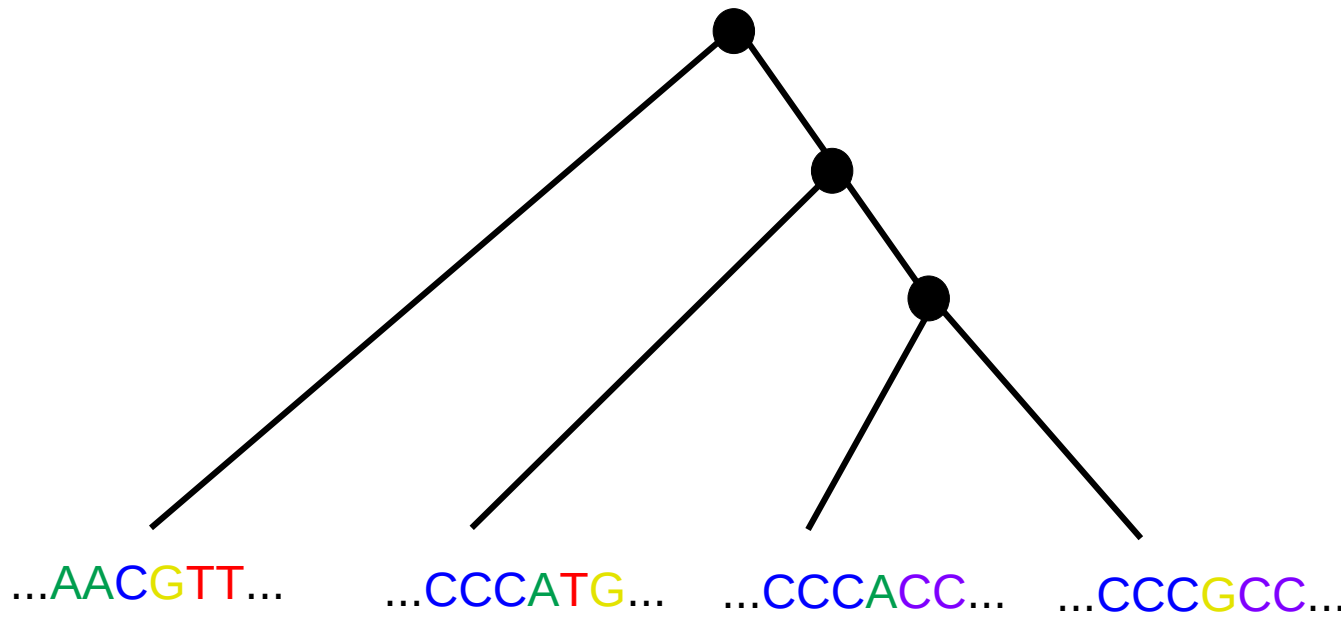
→ είναι κάπως υποκειμενικό



# Σήμερα

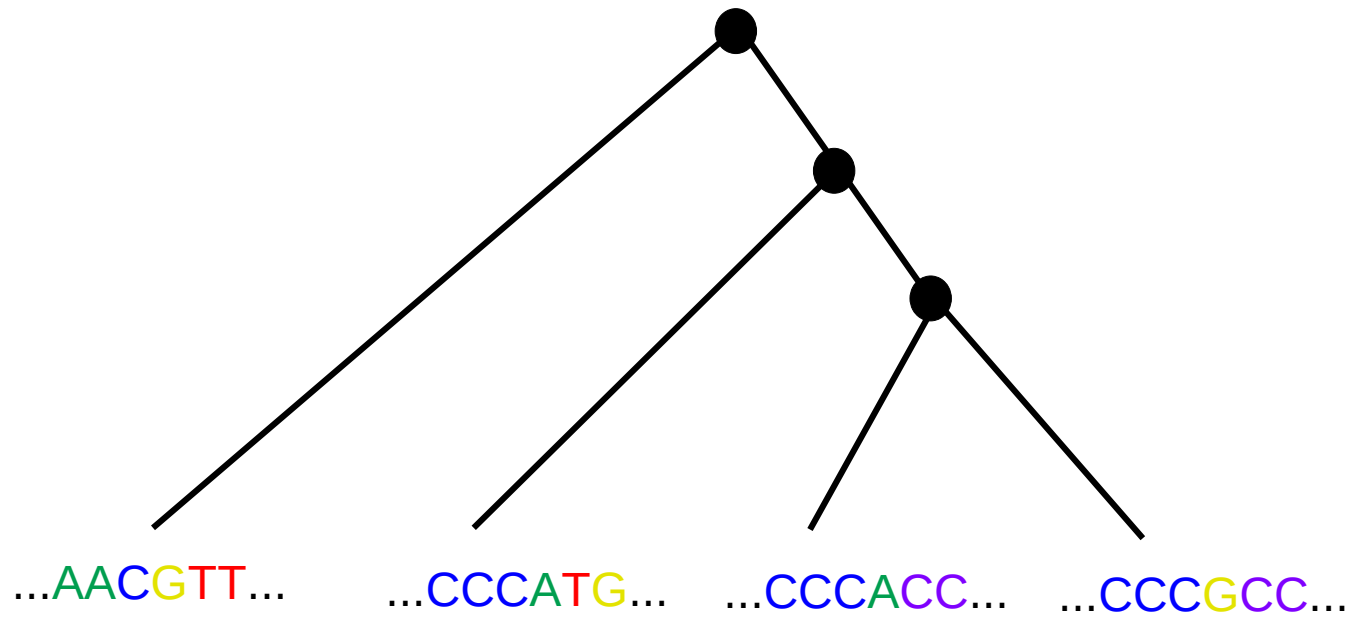


# Σήμερα



Εξαγωγή DNA

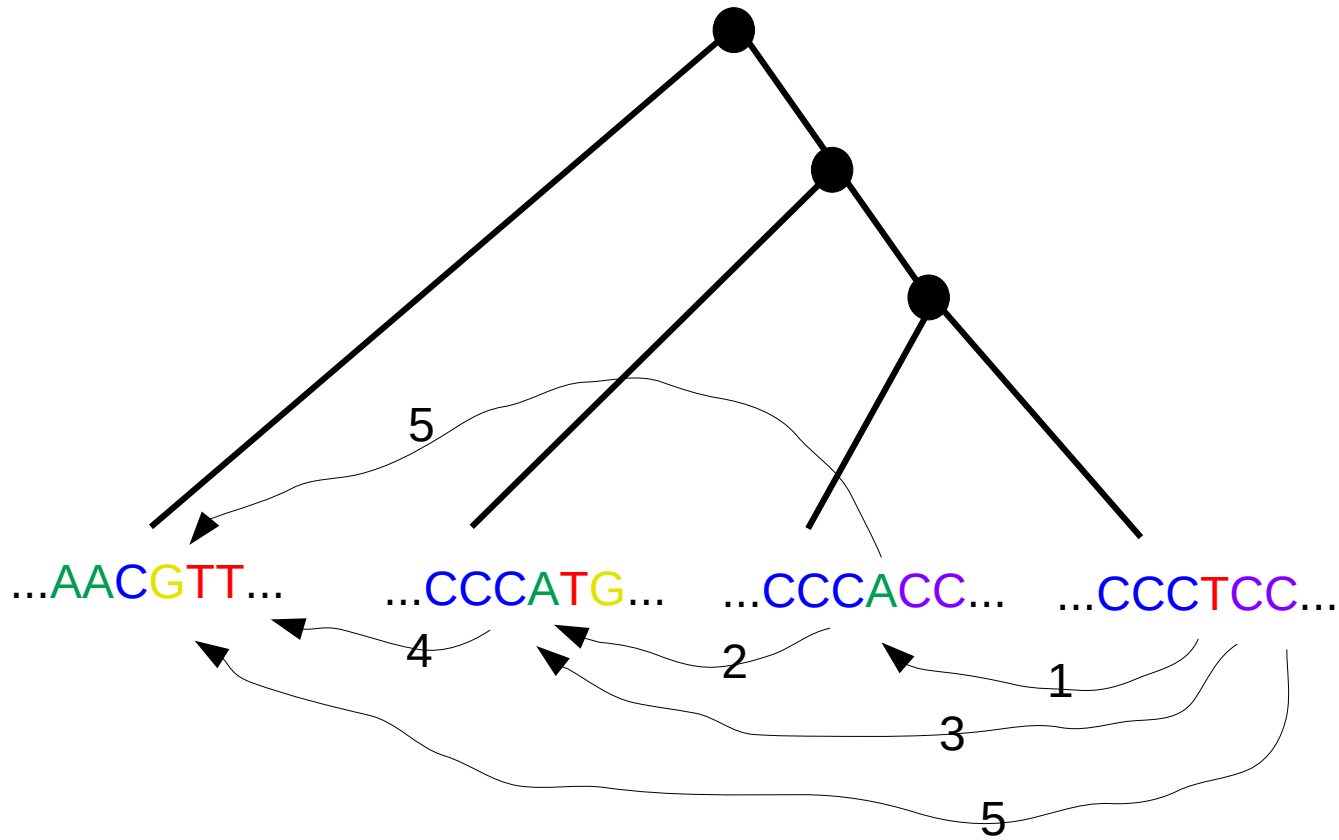
# Σήμερα



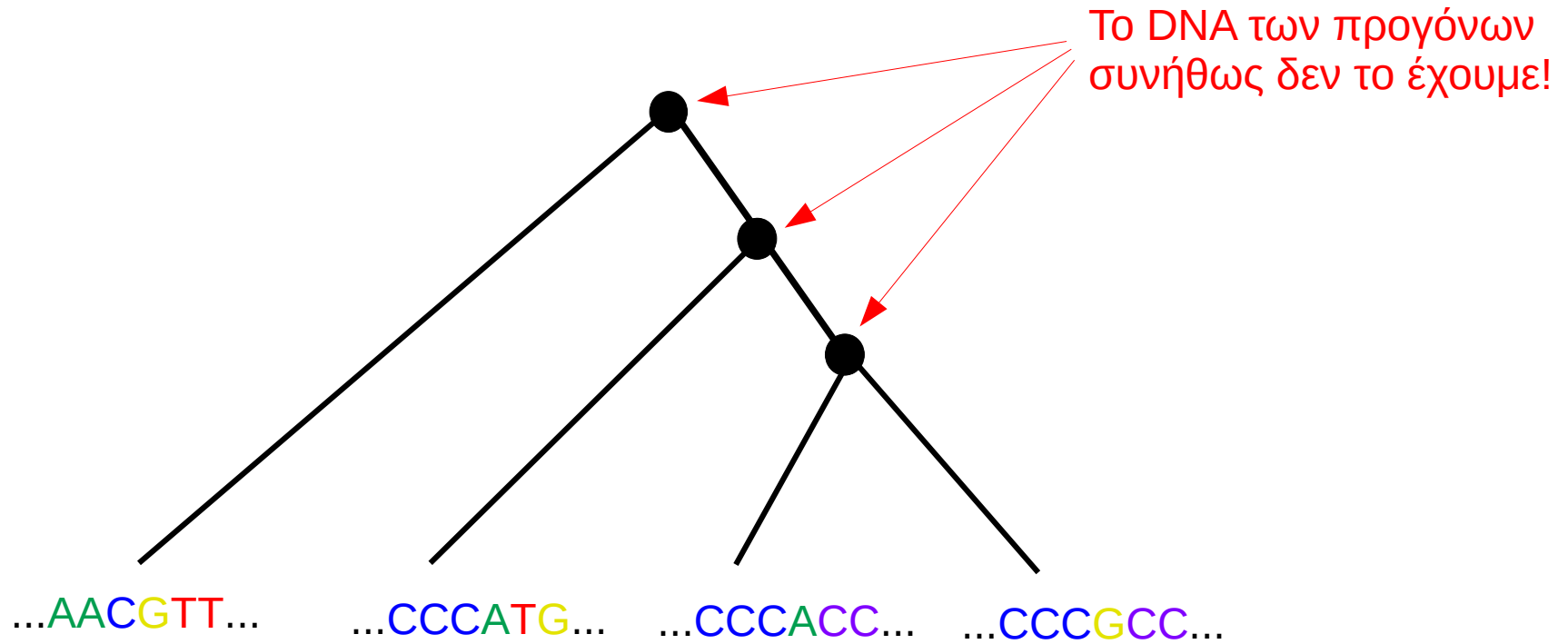
Τι παρατηρείτε;



# Σήμερα



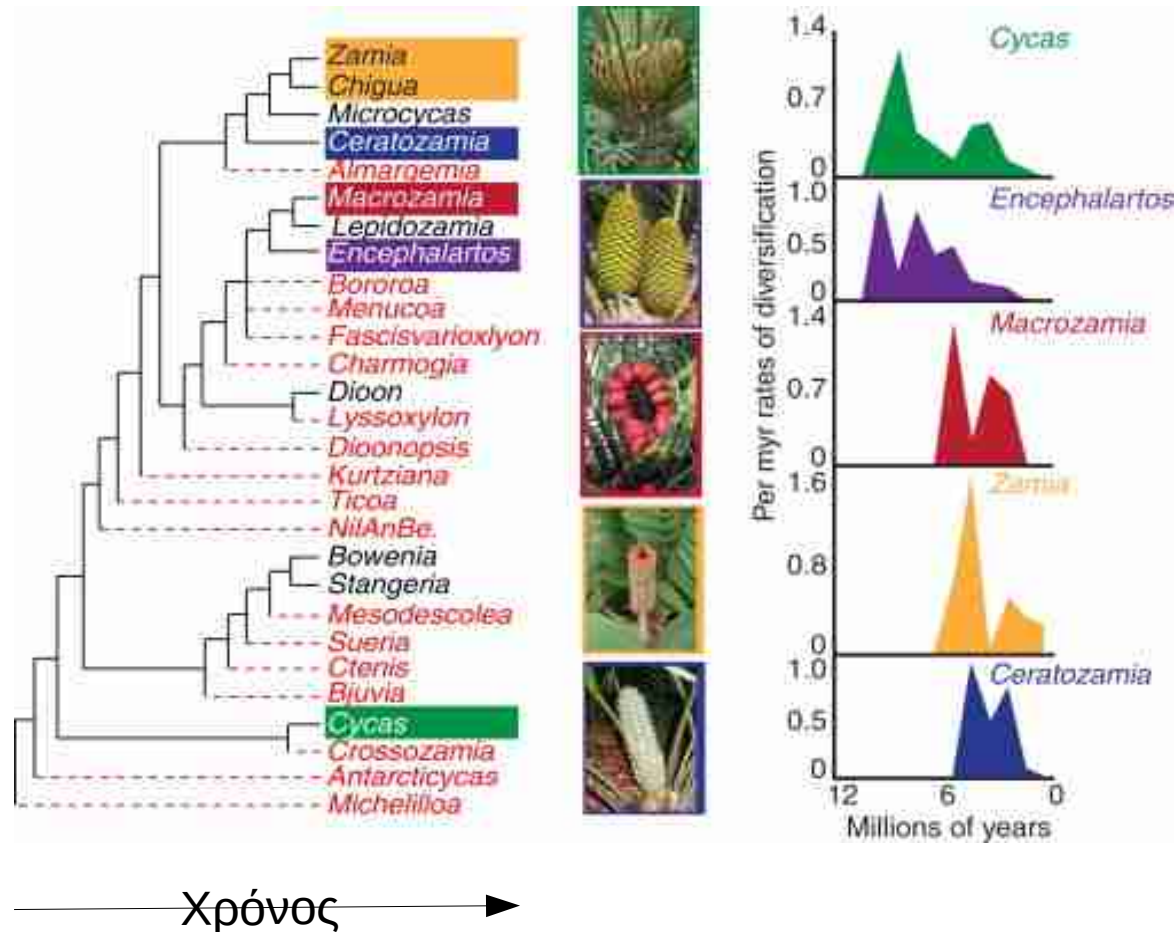
# Σήμερα



# Δομή Ομιλίας

- **Εφαρμογές Φυλογενετικών Δέντρων**
- Υπολογιστική Πολυπλοκότητα
- Άλλες Προκλήσεις

# Ρυθμοί διαφοροποίησης



**Από :** Charles C. Davis, Hanno Schaefer: "Plant Evolution: Pulses of Extinction and Speciation in Gymnosperm Diversity", *Current Biology*, 2011.

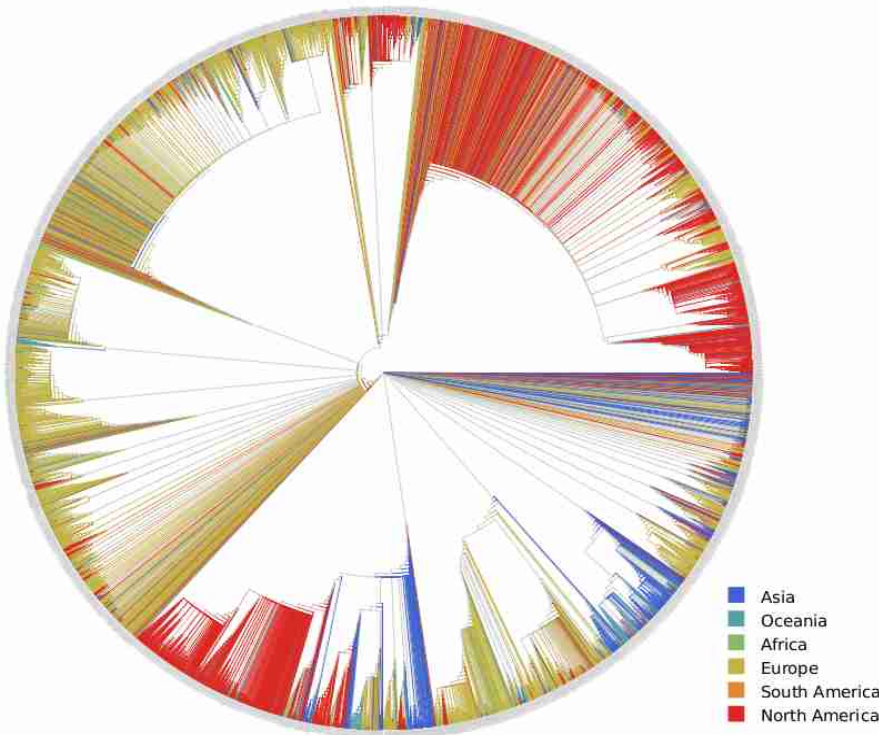


# Και βεβαίως και ο κορονοϊός

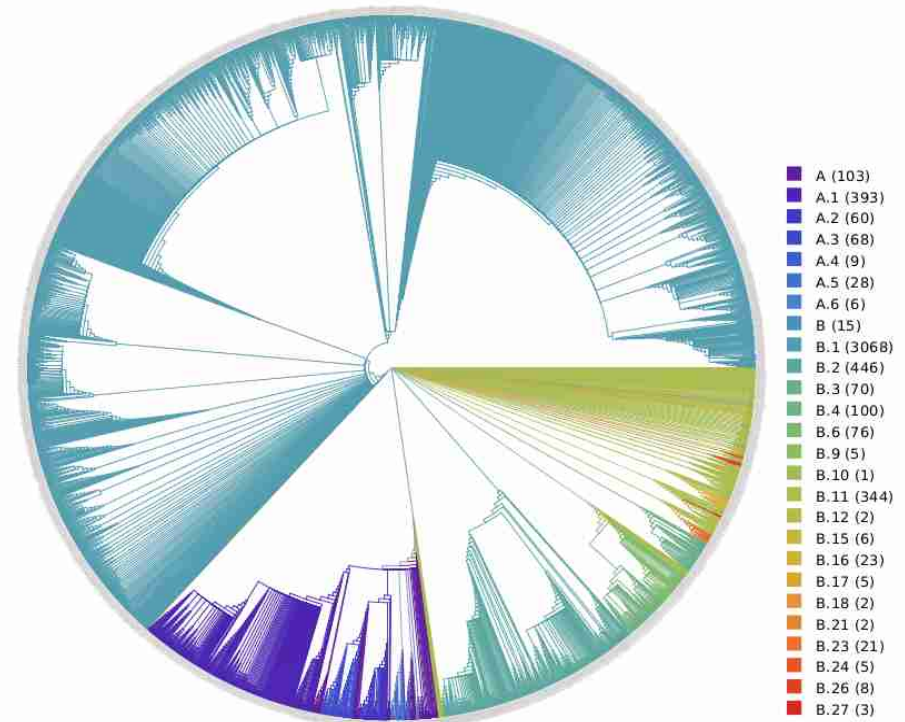
## Phylogenetic analysis of SARS-CoV-2 data is difficult

Benoit Morel<sup>1,3</sup>, Pierre Barbera<sup>4,5</sup>, Lucas Czech<sup>1</sup>, Ben Bettisworth<sup>1</sup>, Lukas Hübner<sup>1,2</sup>, Sarah Lutteropp<sup>1</sup>, Dora Serdari<sup>1</sup>, Evangelia-Georgia Kostaki<sup>5</sup>, Ioannis Mamais<sup>6</sup>, Alexey M Kozlov<sup>1</sup>, Pavlos Pavlidis<sup>2</sup>, Dimitrios Paraskevis<sup>3</sup>, and Alexandros Stamatakis<sup>1,2</sup>

Γεωγραφικές περιοχές



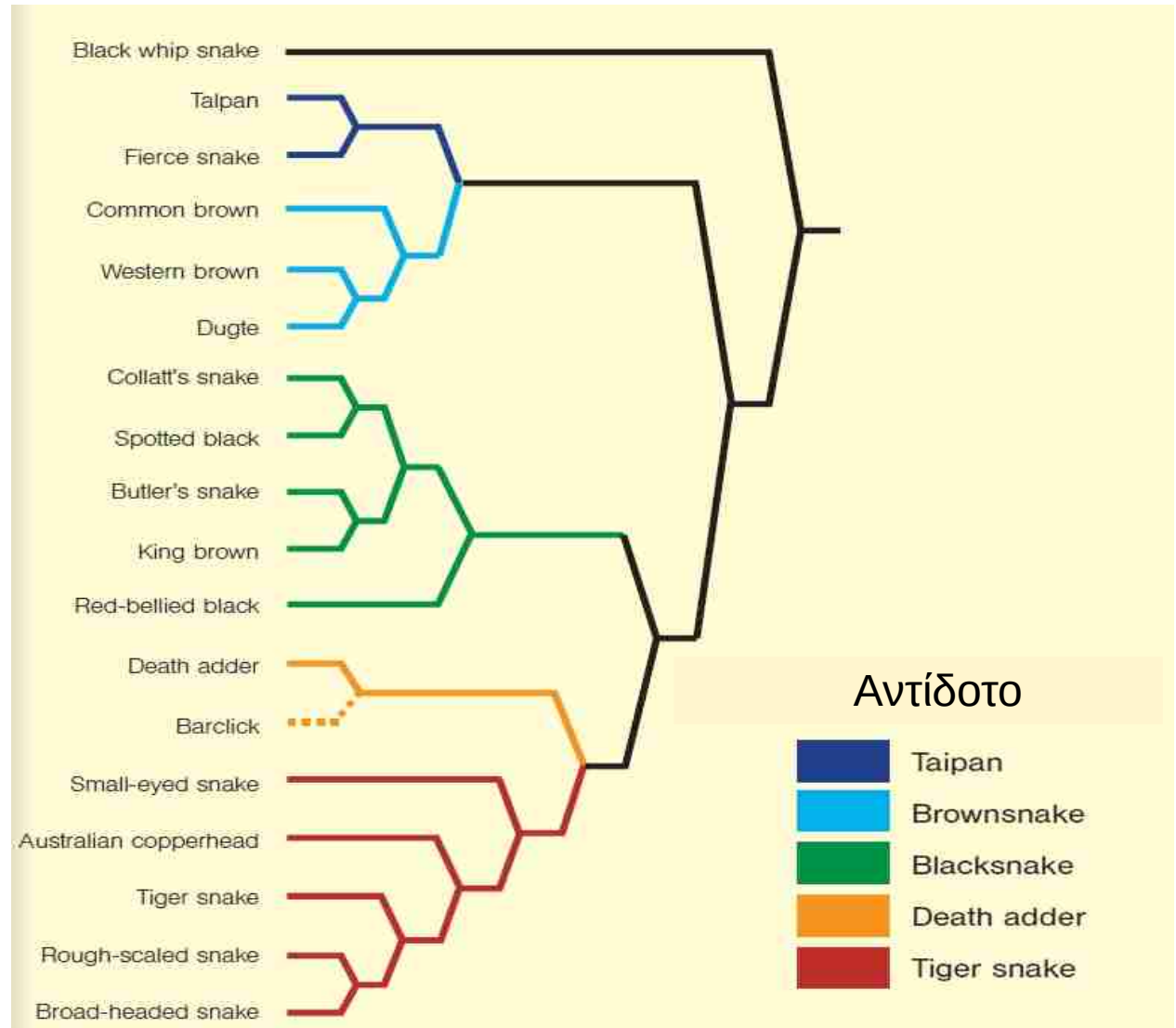
Στελέχη ιού





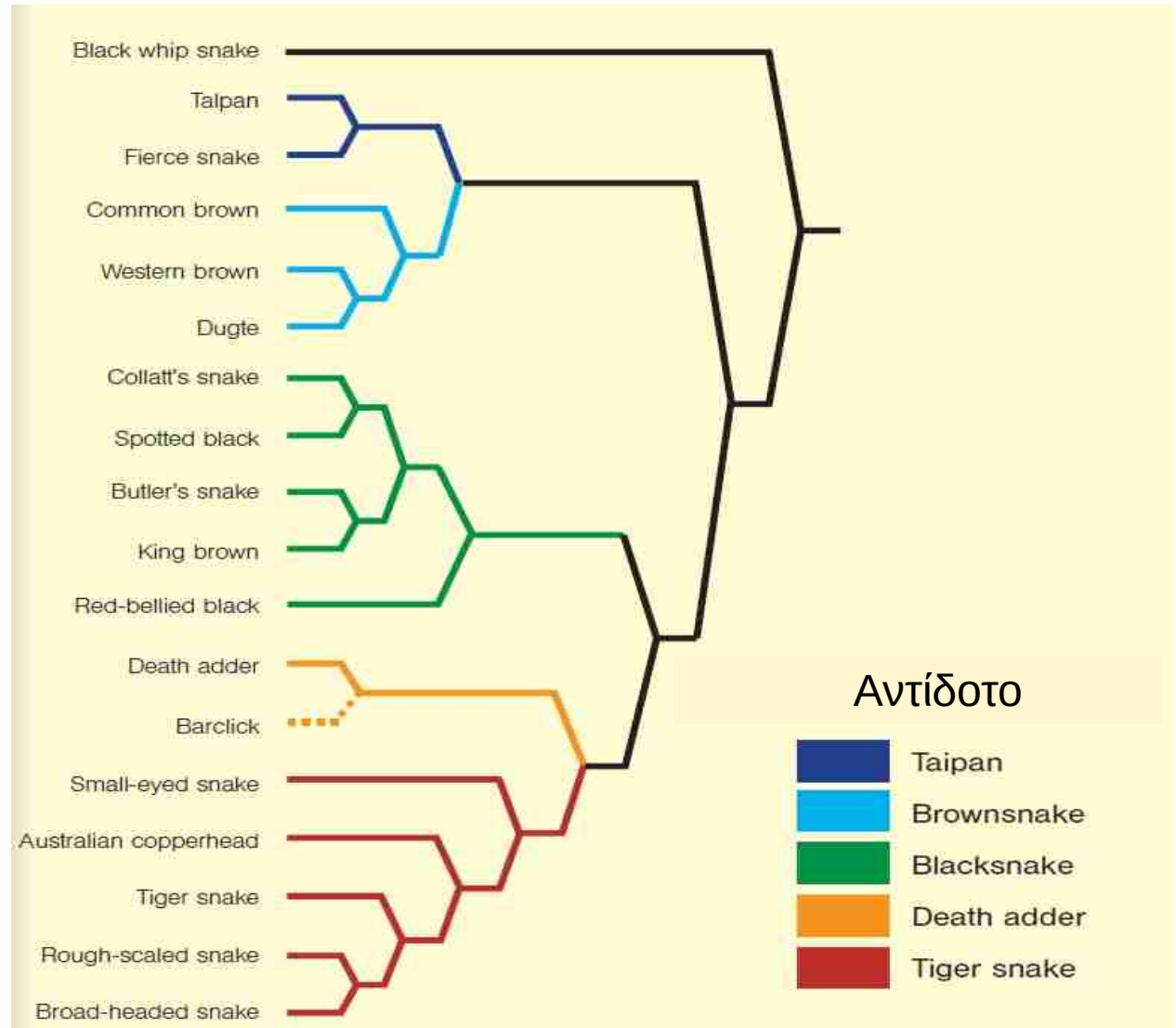
# Φίδια της Αυστραλίας

Η Αυστραλία έχει περισσότερα δηλητηριώδη φίδια από οποιαδήποτε άλλη ήπειρο και πολλοί άνθρωποι πεθαίνουν από δαγκώματα φιδιών κάθε χρόνο.



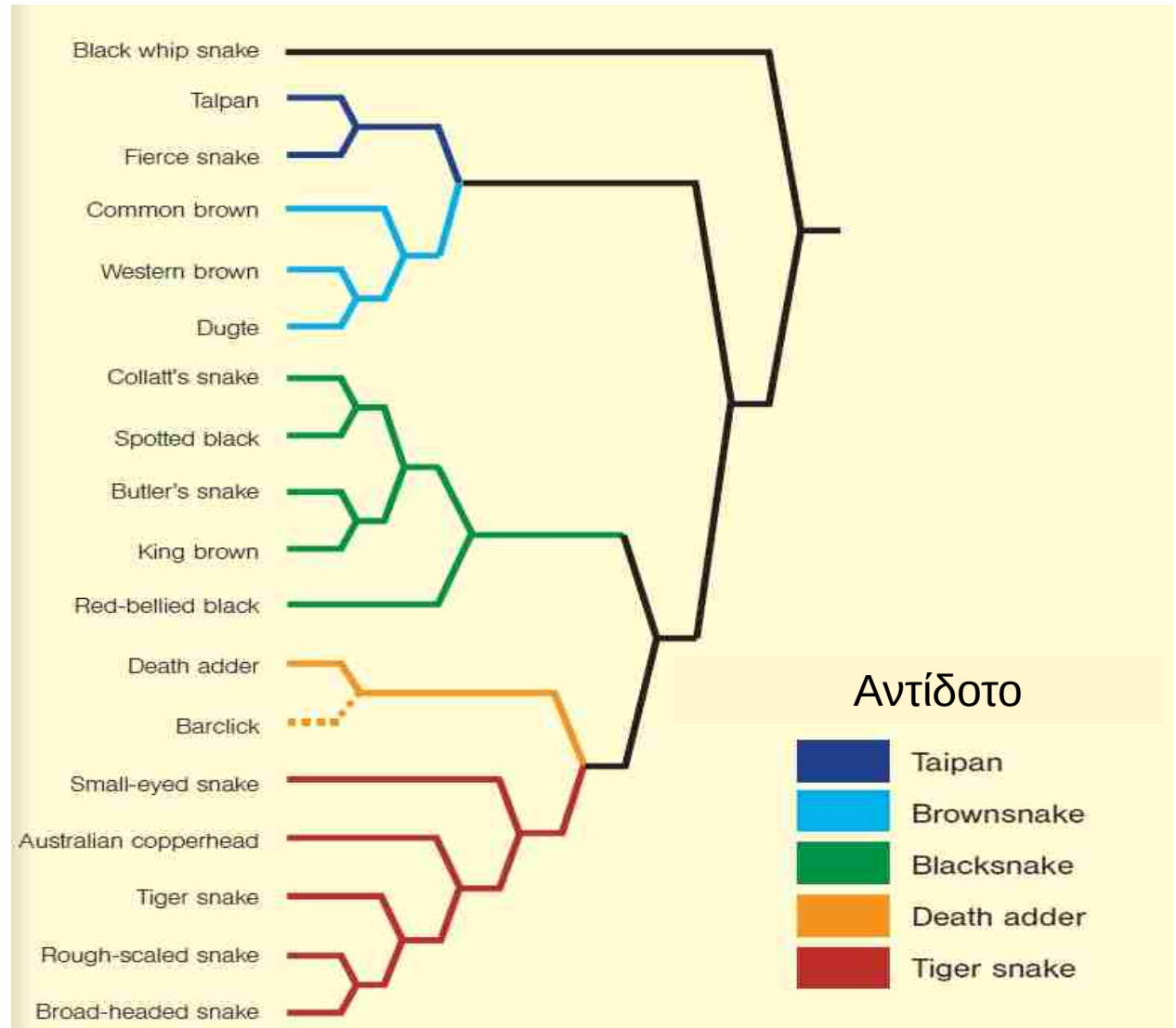
# Φίδια της Αυστραλίας

Η ανάπτυξη αποτελεσματικών αντιδωτών αποτελεί επομένως υψηλή προτεραιότητα, αλλά λίγα είναι γνωστά για το δηλητήριο των περισσότερων ειδών.



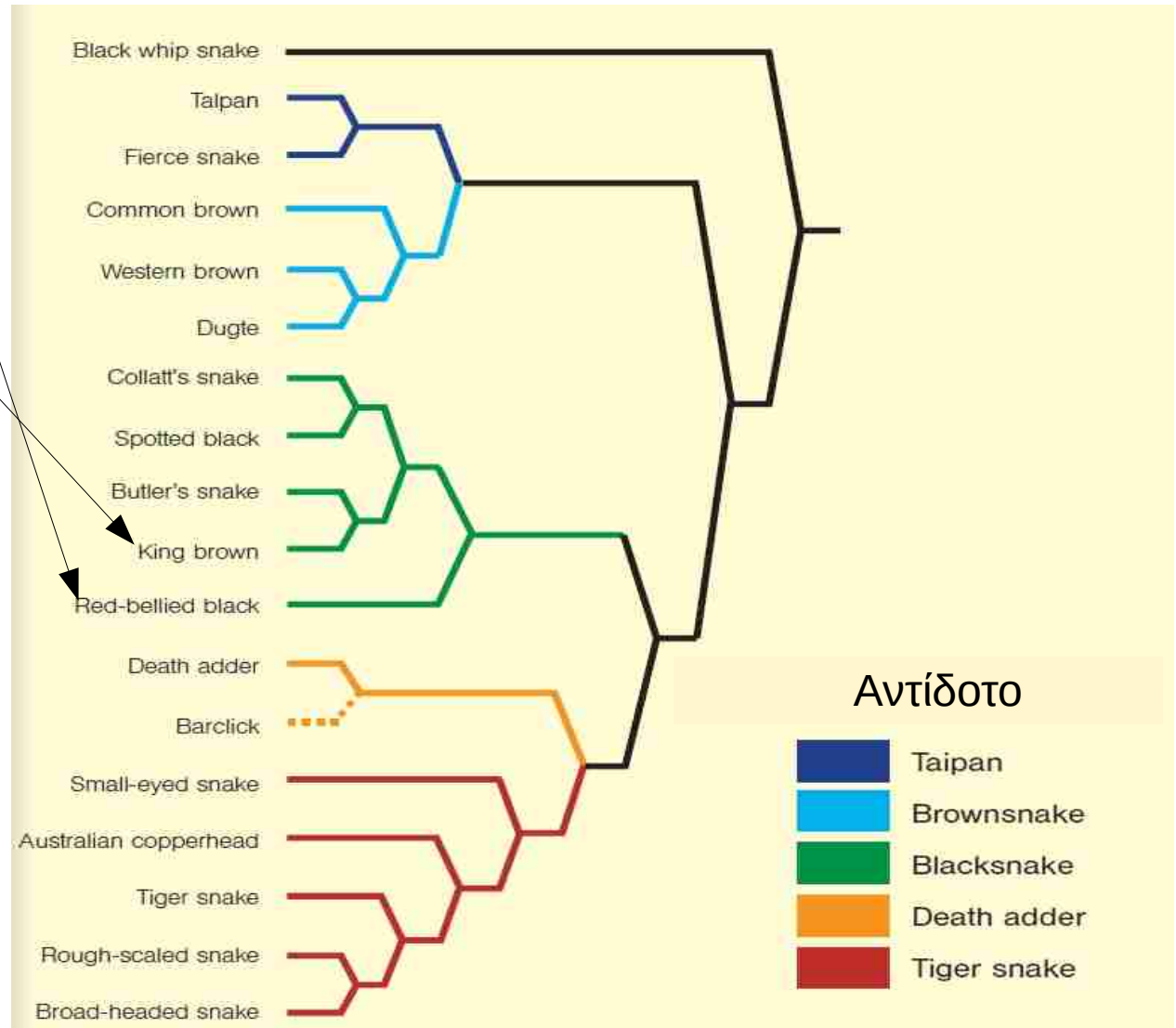
# Φίδια της Αυστραλίας

Η φυλογενετική ανάλυση βοηθά σε αυτό το έργο επειδή οι ιδιότητες του δηλητηρίου συσχετίζονται ισχυρά με τις εξελικτικές σχέσεις.



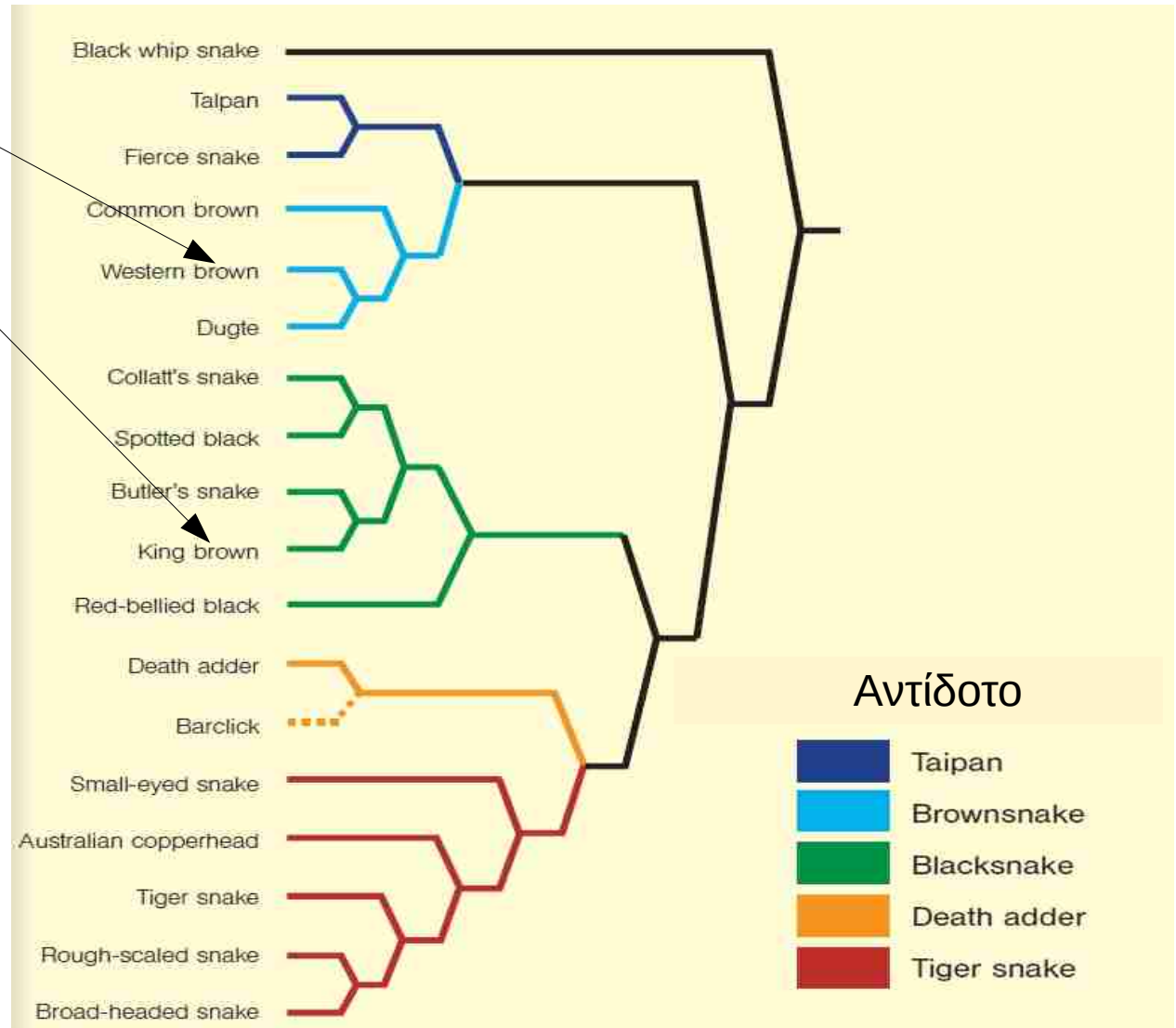
# Φίδια της Αυστραλίας

Αν και το μαύρο φίδι με την κόκκινη κοιλιά φαίνεται μορφολογικά πολύ διαφορετικό από το king brown, είναι στην πραγματικότητα στενά συνδεδεμένο και μπορεί να αντιμετωπιστεί με το ίδιο αντίδοτο.



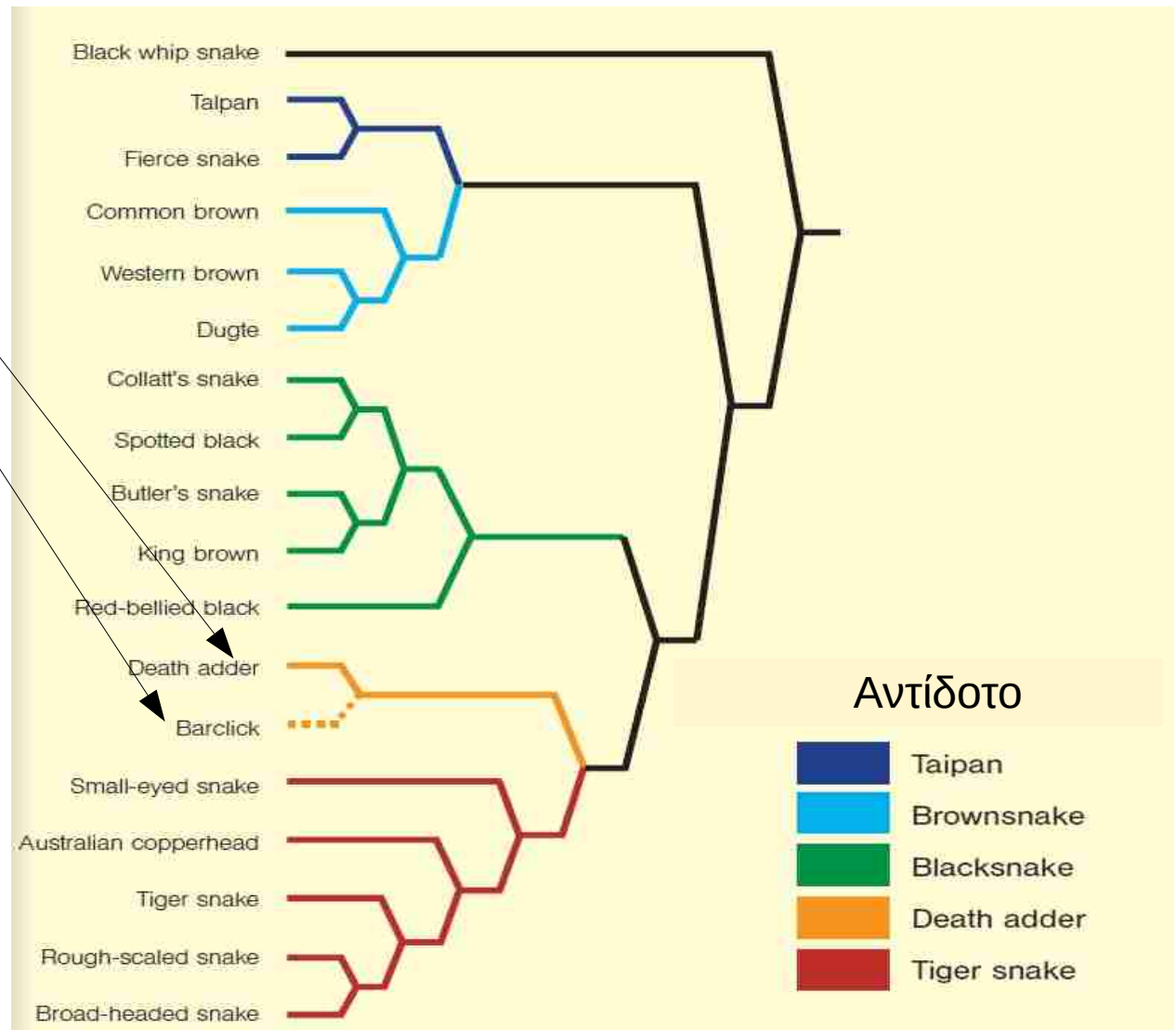
# Φίδια της Αυστραλίας

Αντίστροφα, το **western brown** μοιάζει αρκετά με το **king brown**, αλλά η εξελικτική τους σχέση είναι μακρινή και επομένως ανταποκρίνεται καλύτερα σε διαφορετικά αντίδοτα.



# Φίδια της Αυστραλίας

Η φυλογένεση είναι επίσης προγνωστική: η πρόσφατη απόδειξη ότι το ελάχιστα γνωστό **barclick** σχετίζεται στενά με τον **αθροιστή θανάτου** (Death adder) προβλέπει ότι το πρώτο είναι επίσης εξαιρετικά επικίνδυνο και μπορεί να ανταποκριθεί στο ευρέως διαθέσιμο αντίδοτο του αθροιστή θανάτου.





# Φίδια της Αυστραλίας

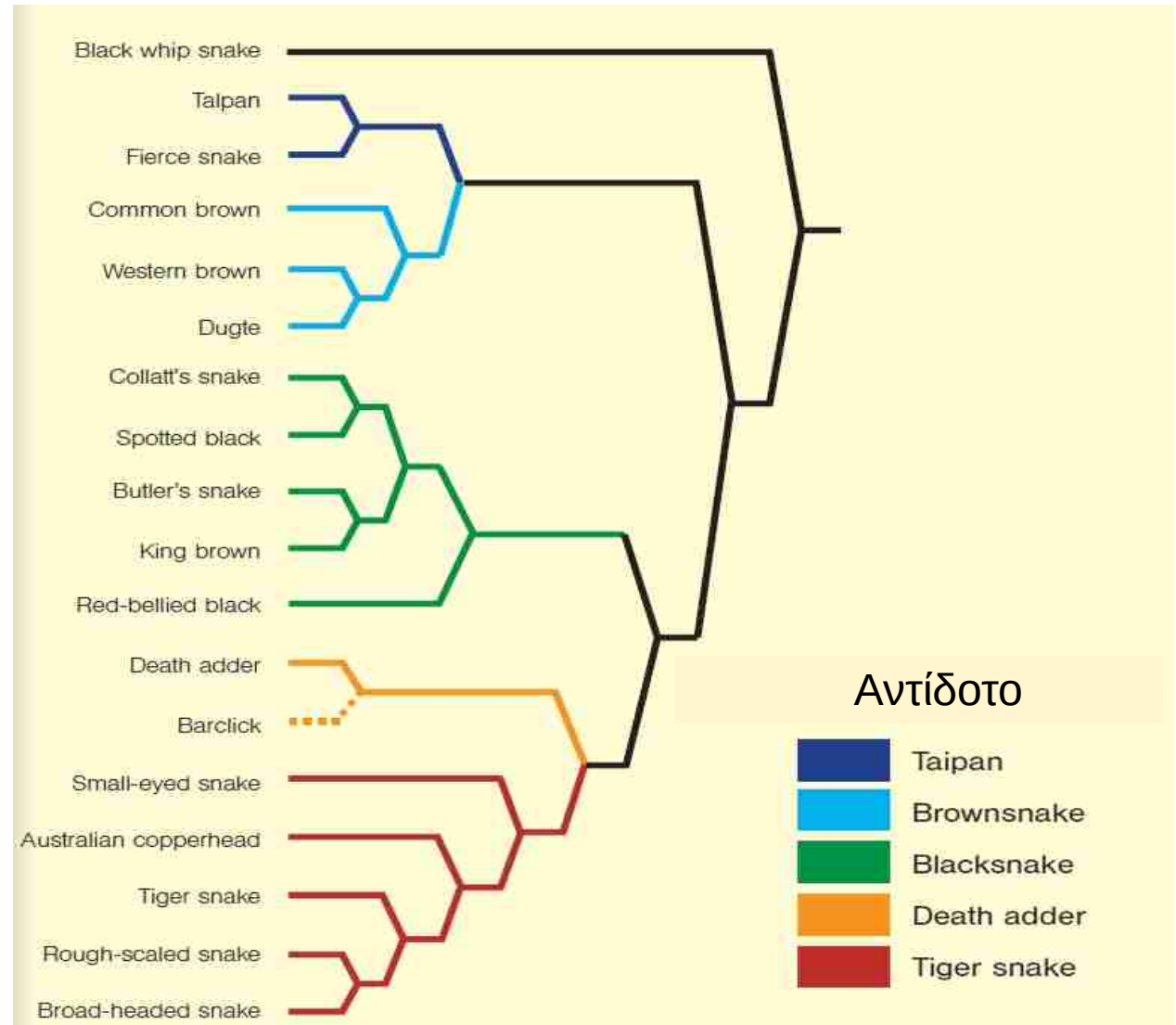
Australia has more poisonous snakes than any other continent, and many people die from **snakebites** each year. Developing **effective antivenins** is thus a **high priority**, but little is known about the venins of most species.

Phylogenetic analysis is helping with this task because **venin properties correlate strongly with evolutionary relationships**.

Although the **red-bellied black snake** looks **very different** from the **king brown**, it is actually **closely related** and can be treated with the same antivenin.

Conversely, the **western brown** looks **very similar** to the **king brown**, but it is only **distantly related** and thus responds best to **different antivenin**.

The **phylogeny is also predictive**: the recent demonstration that the poorly-known **barclick** is closely related to the **death adder** (orange lineage) **predicts** that the former is also **highly dangerous** and might respond to **widely-available death adder antivenin**.



# Ανώνυμες Αλληλουχίες

...AACGTT...

...CCCACC...

...CCCATG...

...CCCGCC...

Ανώνυμες αλληλουχίες από είδη που δεν γνωρίζουμε



# Ανώνυμες Αλληλουχίες

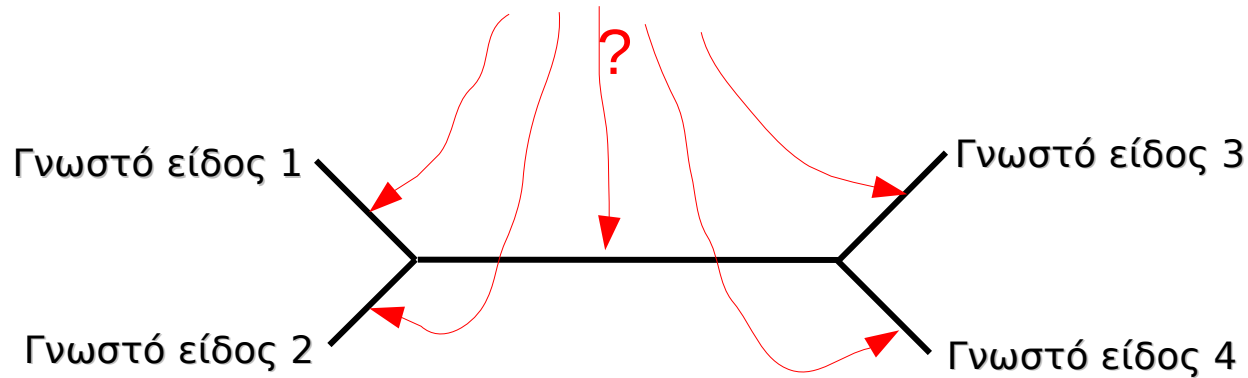
...AACGTT...

...CCCACC...

...CCCATG...

...CCCGCC...

Ανώνυμες αλληλουχίες από είδη που δεν γνωρίζουμε



Δέντρο αναφοράς

# Ανώνυμες Αλληλουχίες

...AACGTT...

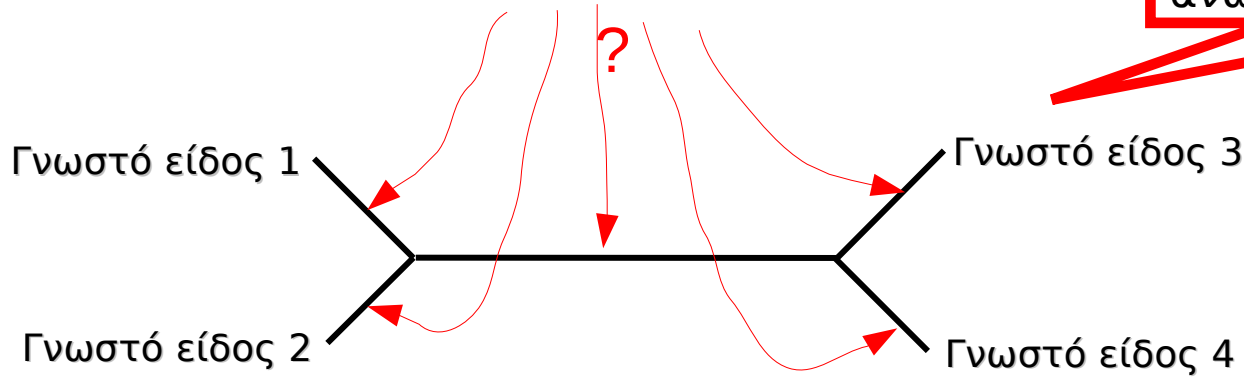
...CCCACC...

...CCCATG...

...CCCGCC...

Ανώνυμες αλληλουχίες από είδη που δεν γνωρίζο

Φυλογενετική τοποθέτηση  
ανώνυμων αλληλουχιών



Δέντρο αναφοράς

# Ανώνυμες Αλληλουχίες

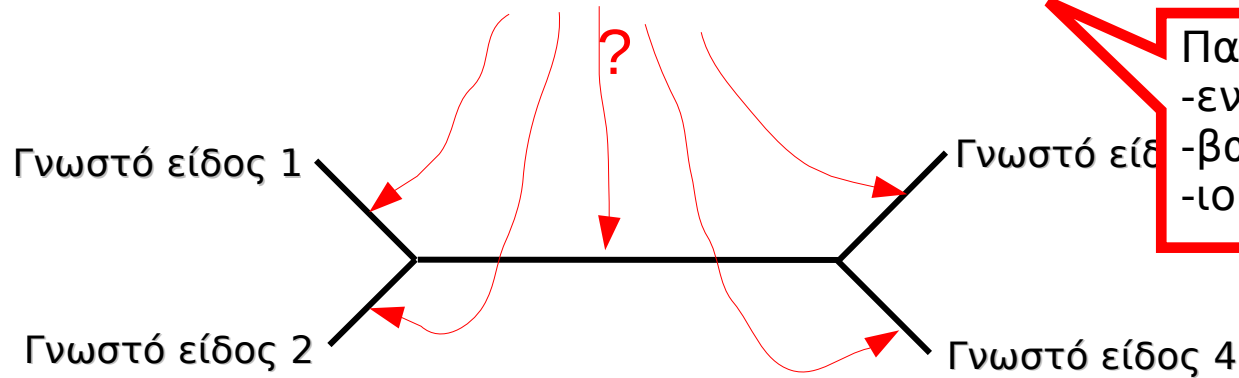
...AACGTT...

...CCCACC...

...CCCATG...

...CCCGCC...

Ανώνυμες αλληλουχίες από είδη που δεν γνωρίζουμε



Παραδείγματα  
-εναέριες συγκρούσεις  
-βακτηρίδια  
-ιοί (κορονοϊός)

Δέντρο αναφοράς

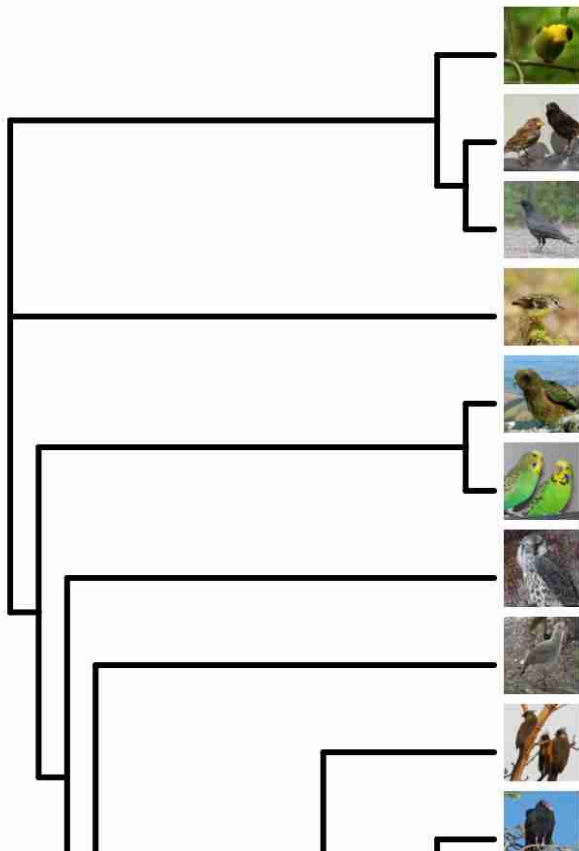


# Παιχνίδι φυλογενετικής τοποθέτησης για μαθητές – εναέριες συγκρούσεις

- <https://cme.h-its.org/exelixis/eseb/public/en/core/title.html>

## Aerial Collisions

*A research team found out, how birds around the world are related to each other.*



Do you see, which birds are closely related to each other? That is fascinating, right? Click on the images to find out more about the bird on the image.

[What is a phylogenetic tree?](#)

The molecular laboratory received a bird sample from a plane that should be identified. Can you help them?

[Become a real bird researcher](#)

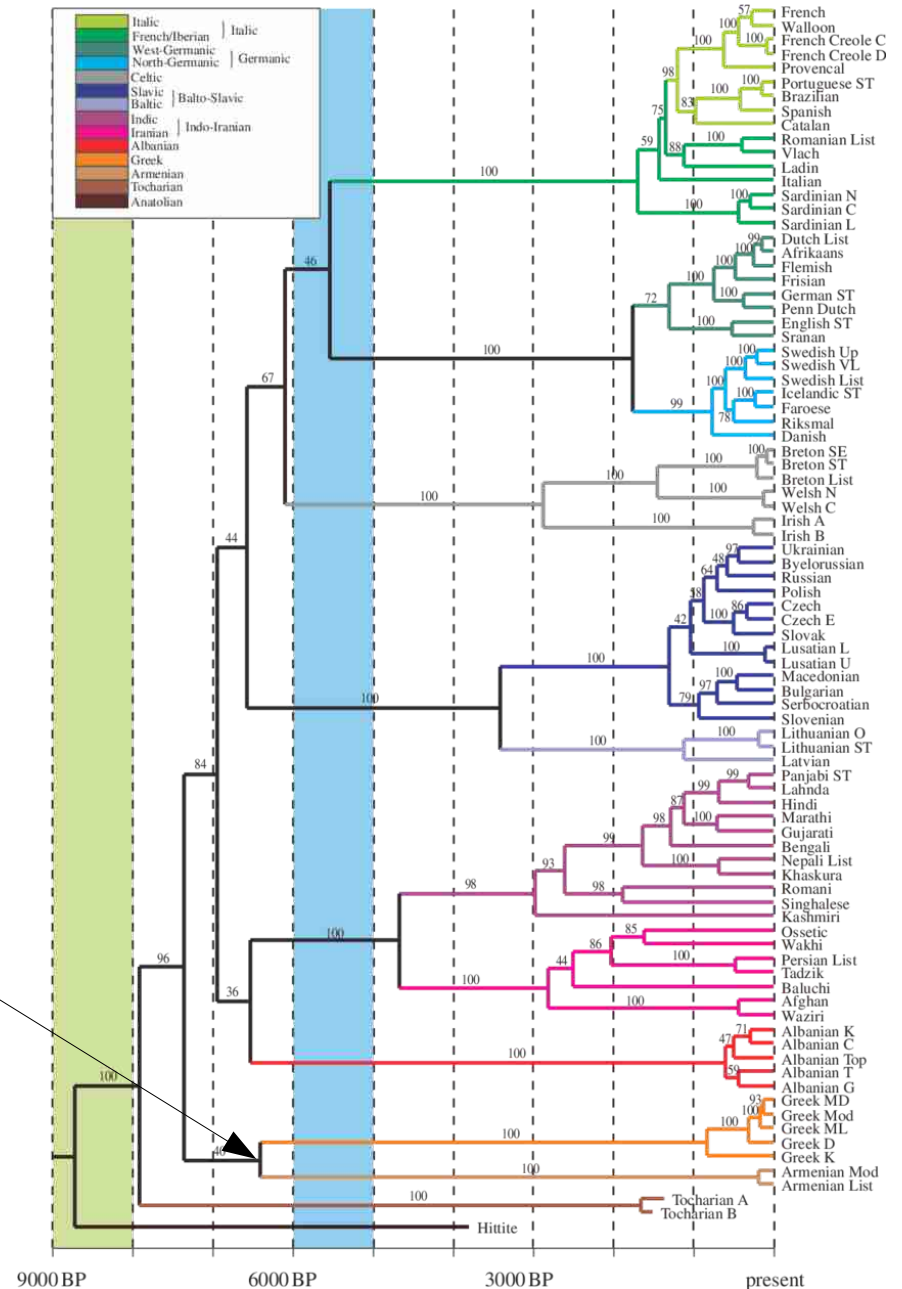
Where do the DNA samples come from? Watch this short movie!



# Εξέλιξη γλωσσών

- Μπορούμε, με τα ίδια μοντέλα, λογισμικά και κριτήρια να κατασκευάσουμε και εξελικτικά δέντρα για τις γλώσσες

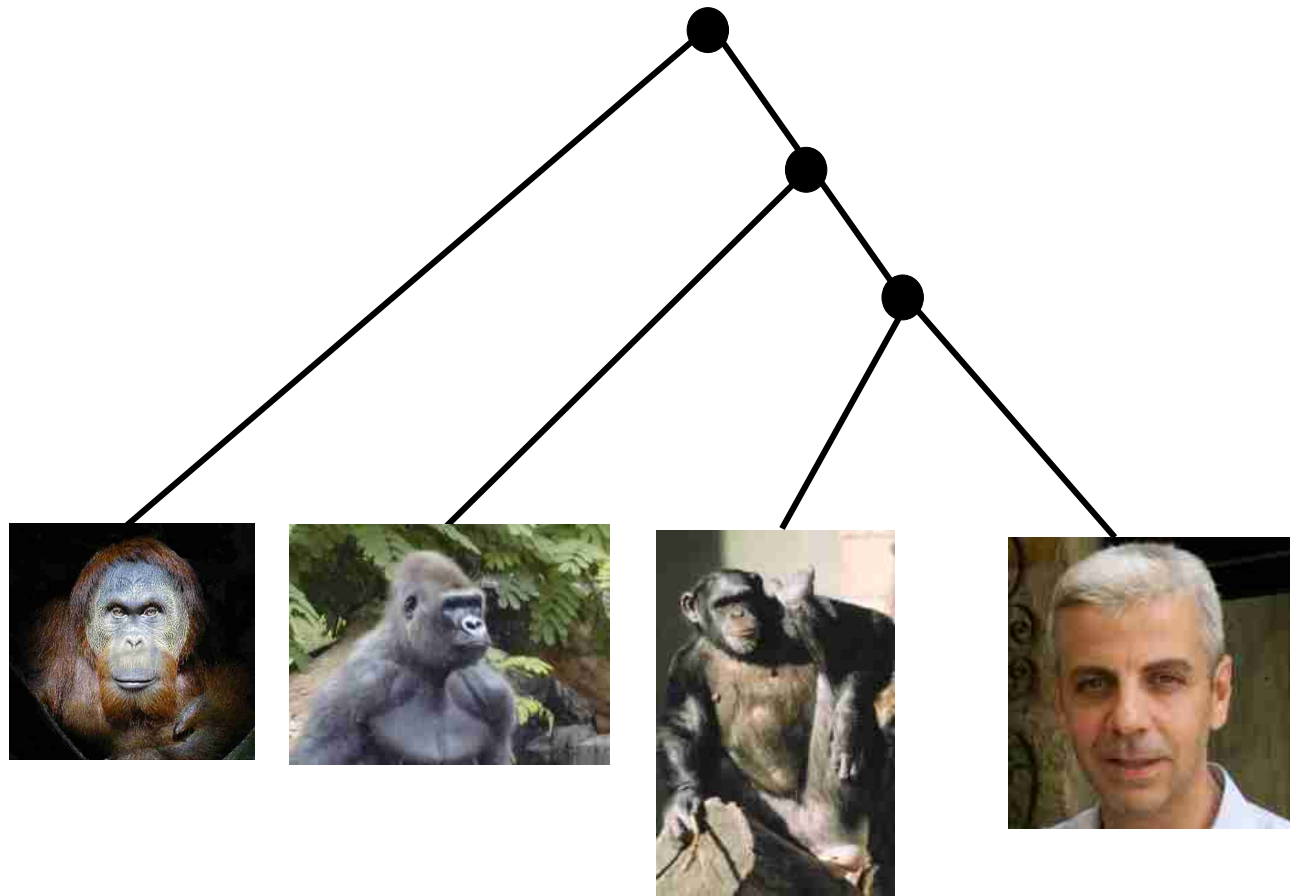
Ελληνικά και Αρμενικά



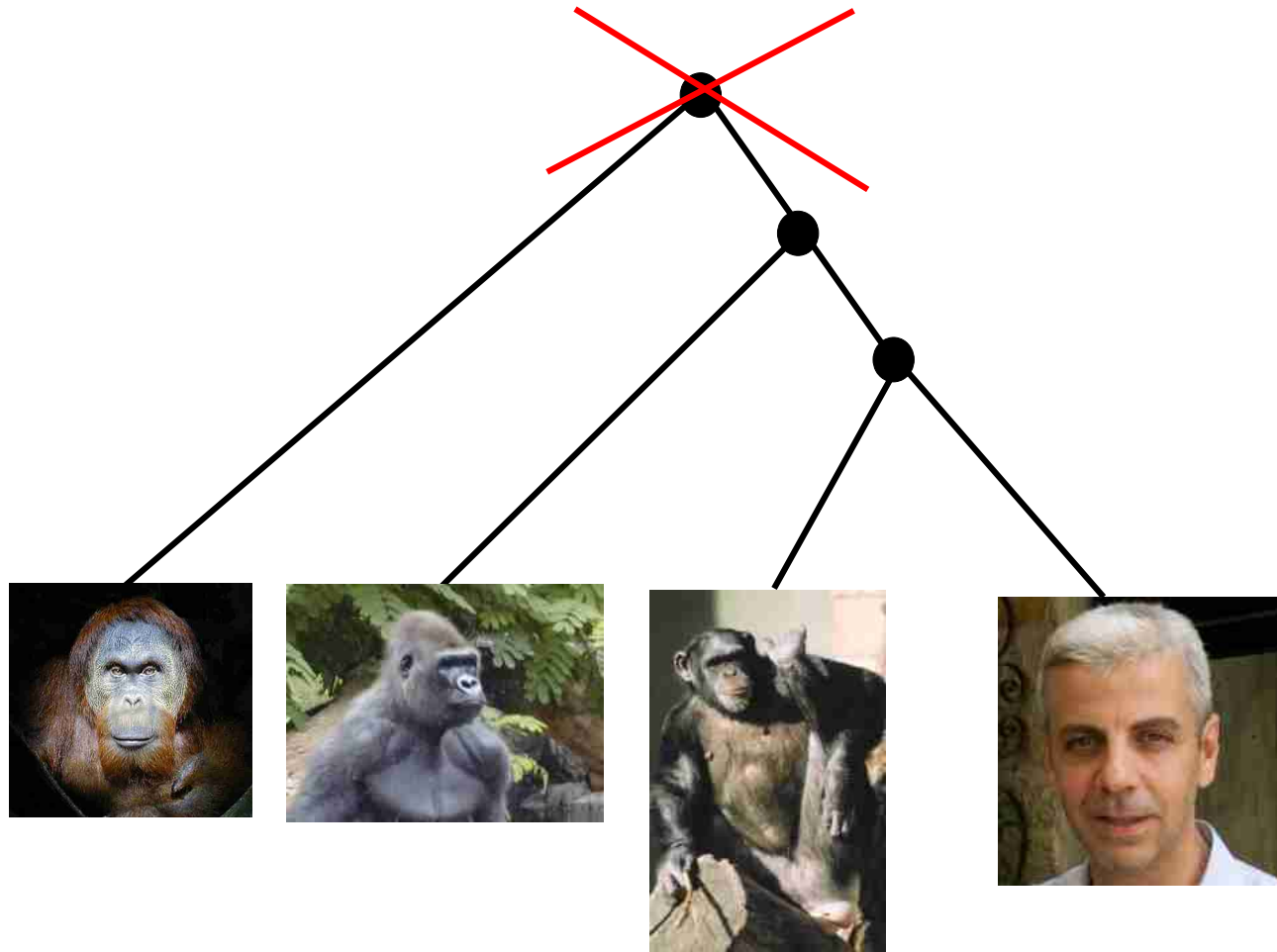
# Δομή Ομιλίας

- Εφαρμογές Φυλογενετικών Δέντρων
- **Υπολογιστική Πολυπλοκότητα**
- Άλλες Προκλήσεις

# Υπολογίζοντας Δέντρα

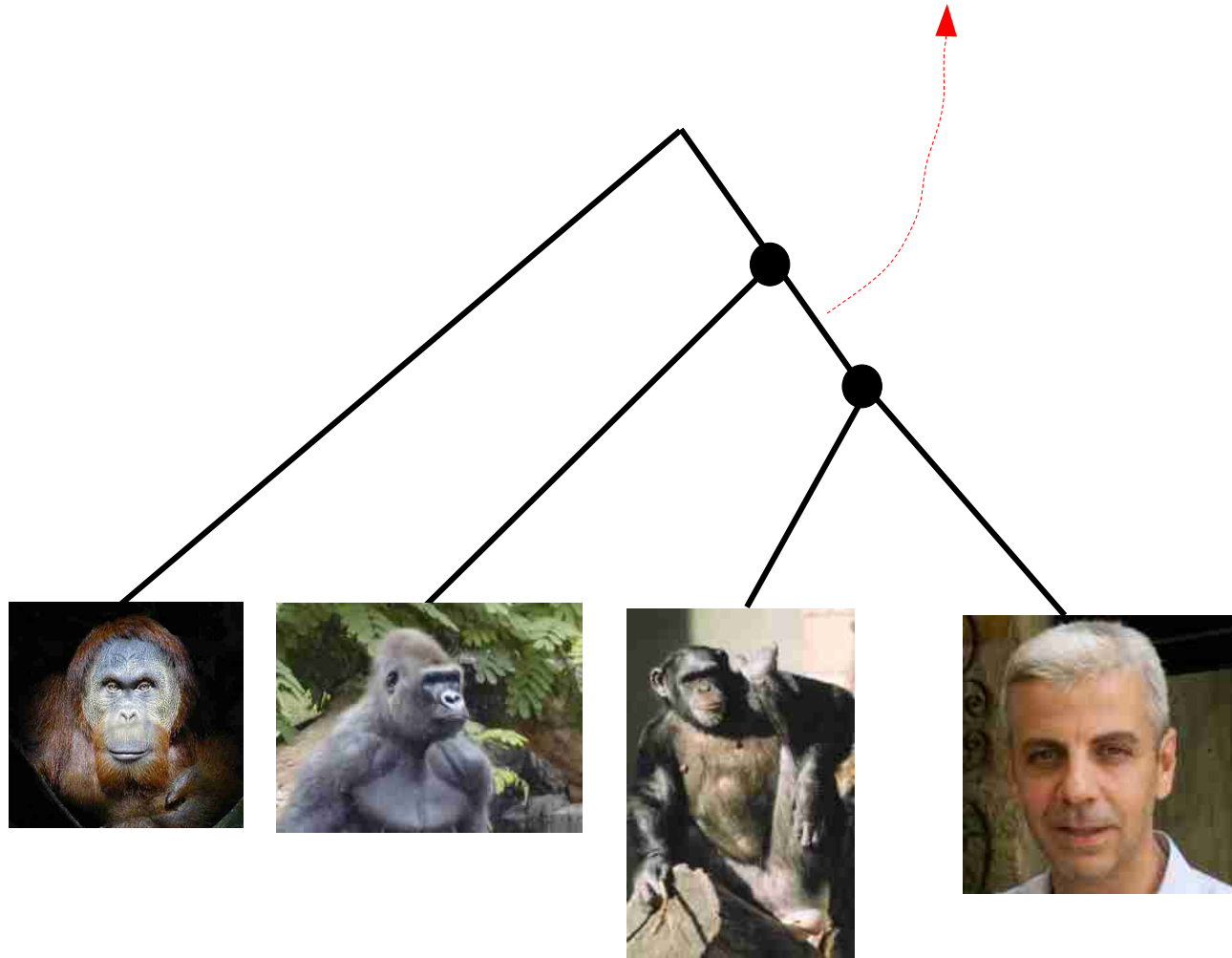


# Υπολογίζοντας Δέντρα

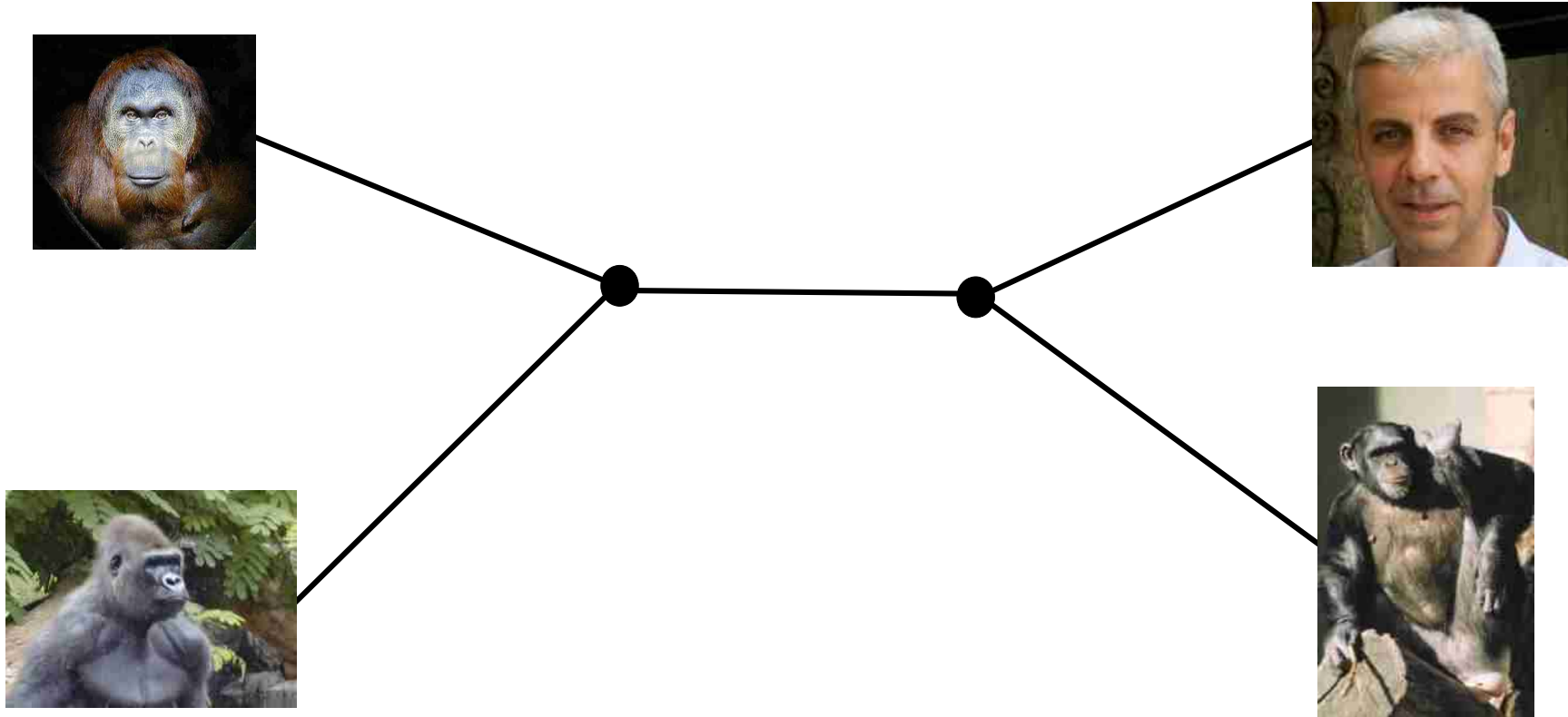




# Υπολογίζοντας Δέντρα

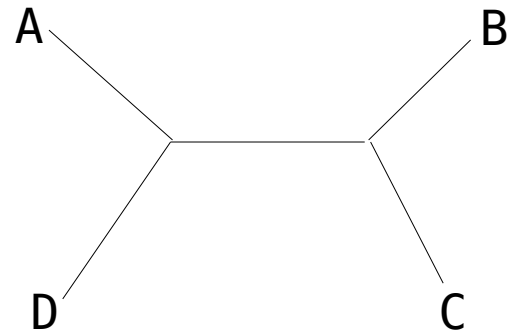
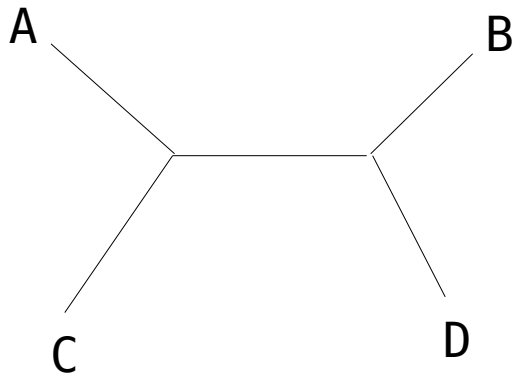
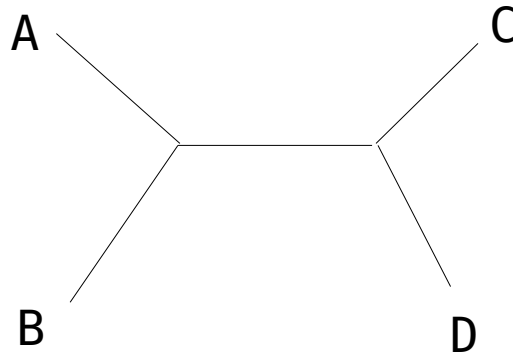


# Υπολογίζοντας Δέντρα

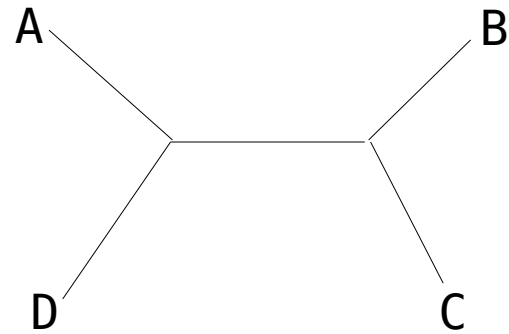
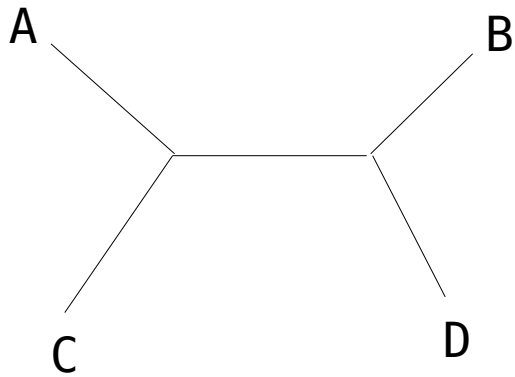
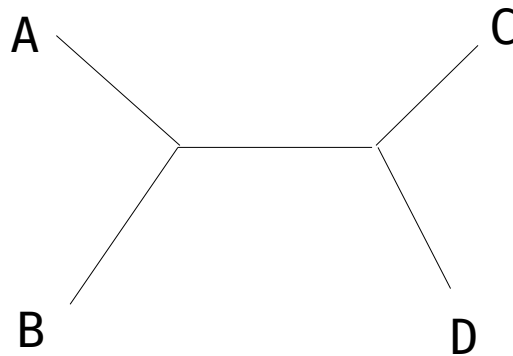


Πόσα φυλογενετικά δέντρα με 4 είδη  
*A, B, C, D* υπάρχουν;

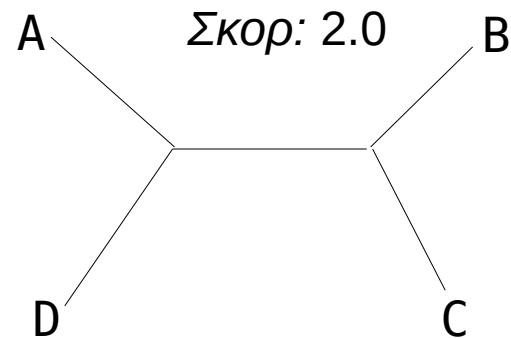
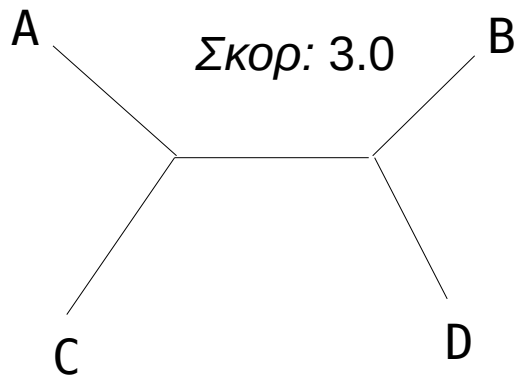
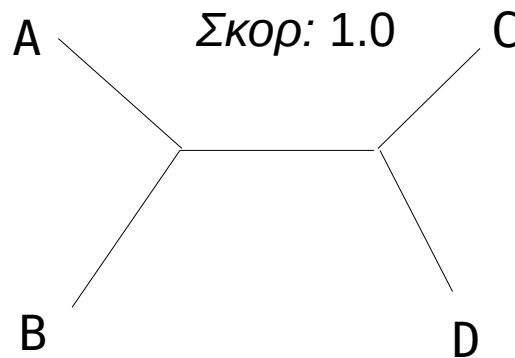
Πόσα φυλογενετικά δέντρα με 4 είδη  
*A, B, C, D* υπάρχουν;



Πώς διαλέγουμε το πιο πιθανό;

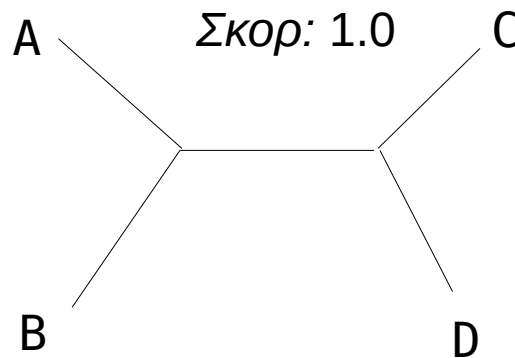


# Πώς διαλέγουμε το πιο πιθανό;

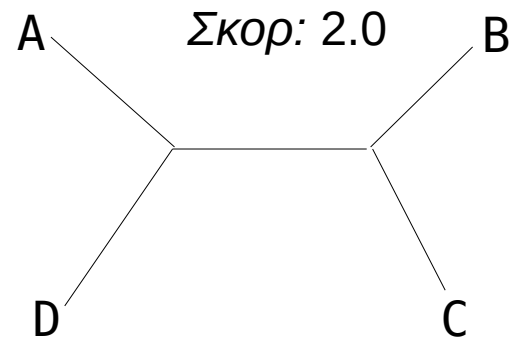
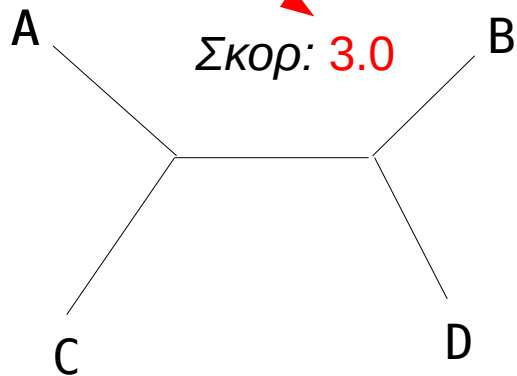


Χρειαζόμαστε κριτήρια (μαθηματικά **μοντέλα**) για να επιλέξουμε το πιο πιθανό δέντρο – δηλαδή το δέντρο που ταιριάζει/εξηγεί πιο καλά τα δεδομένα DNA που έχουμε

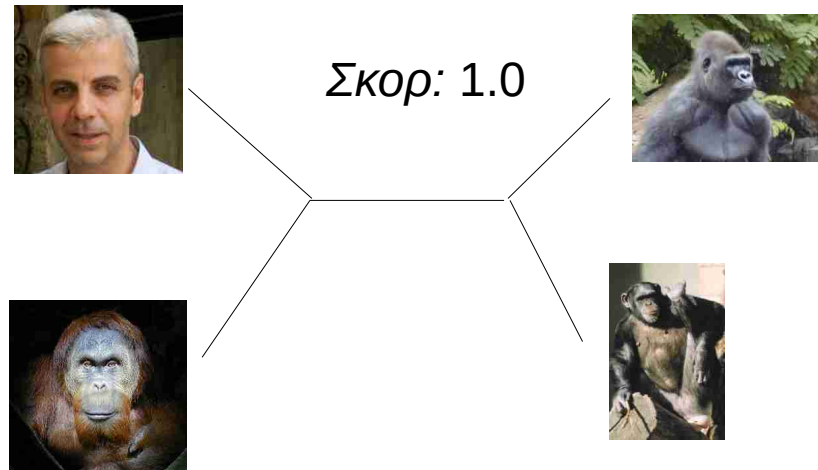
# Άρα, πρέπει να βρούμε το δέντρο με το βέλτιστο σκορ



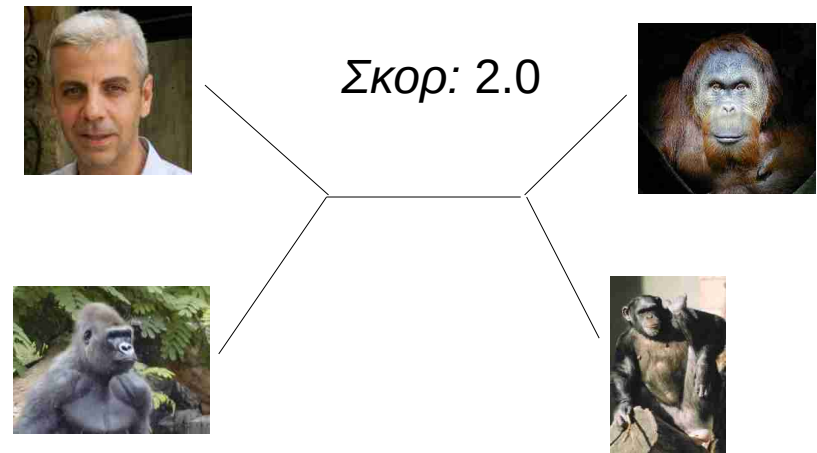
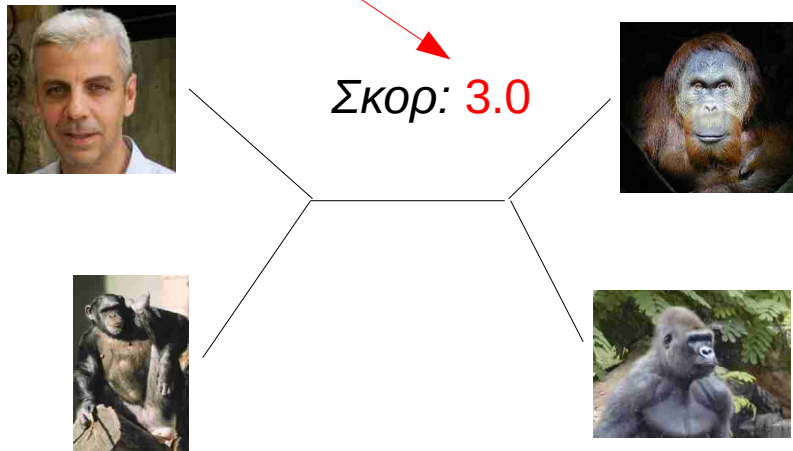
Βέλτιστο δέντρο



# Άρα, πρέπει να βρούμε το δέντρο με το βέλτιστο σκορ

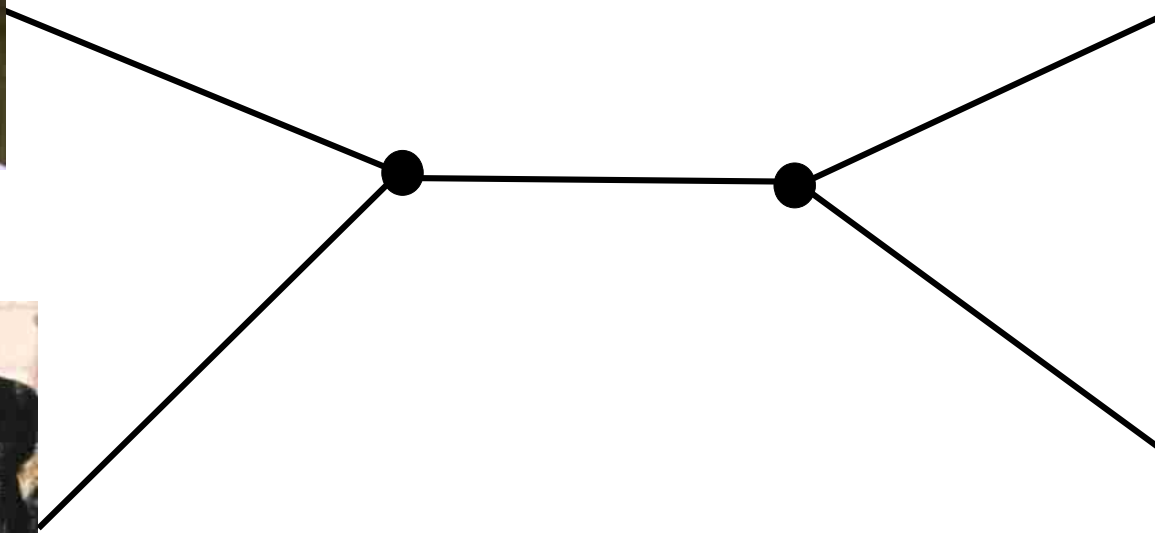
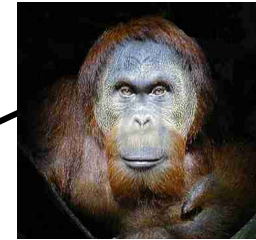


Βέλτιστο δέντρο





# Βέλτιστο Δέντρο - Πού είναι το DNA;

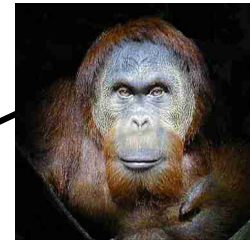


# Βέλτιστο Δέντρο - Πού είναι το DNA;

...CCCGCC...



...AACGTT...



...CCCAAC...

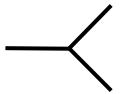


...CCCATG...

# Συμπέρασμα

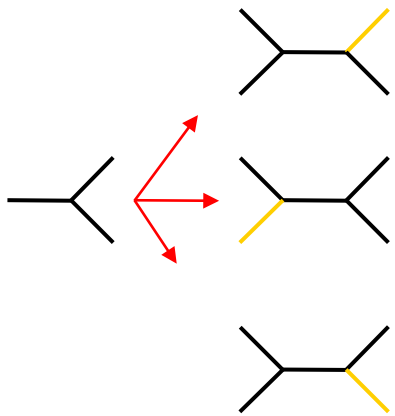
- Για 4 είδη έχουμε 3 διαφορετικά δέντρα
- Άρα, πρέπει απλά να υπολογίσουμε το σκορ μας (**στον υπολογιστή**) για αυτά τα τρία δέντρα για να βρούμε το δέντρο με το βέλτιστο σκορ
- Για 5, 6, 7, ... είδη πόσα δέντρα πρέπει να σκοράρουμε όμως – πόσα δέντρα υπάρχουν;

# Ας κατασκευάσουμε δέντρα



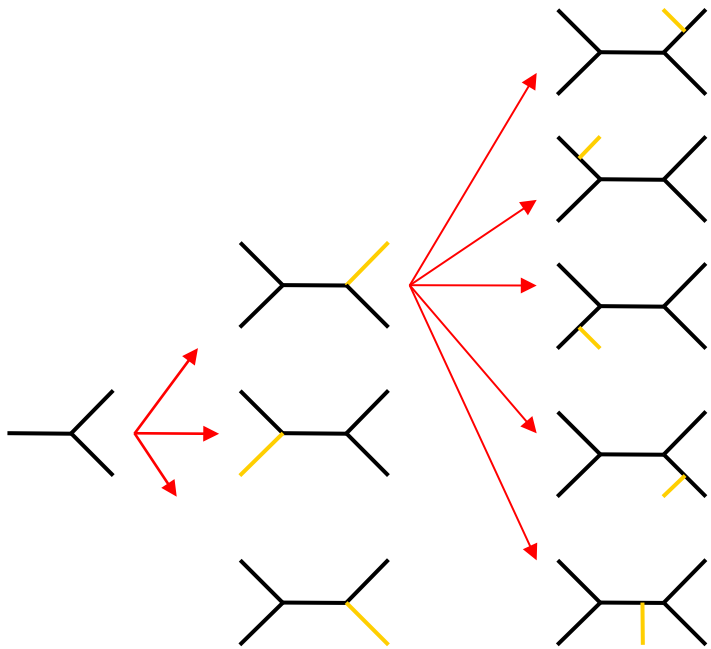
3 είδη  $\rightarrow$  1 δέντρο

# Ας κατασκευάσουμε δέντρα



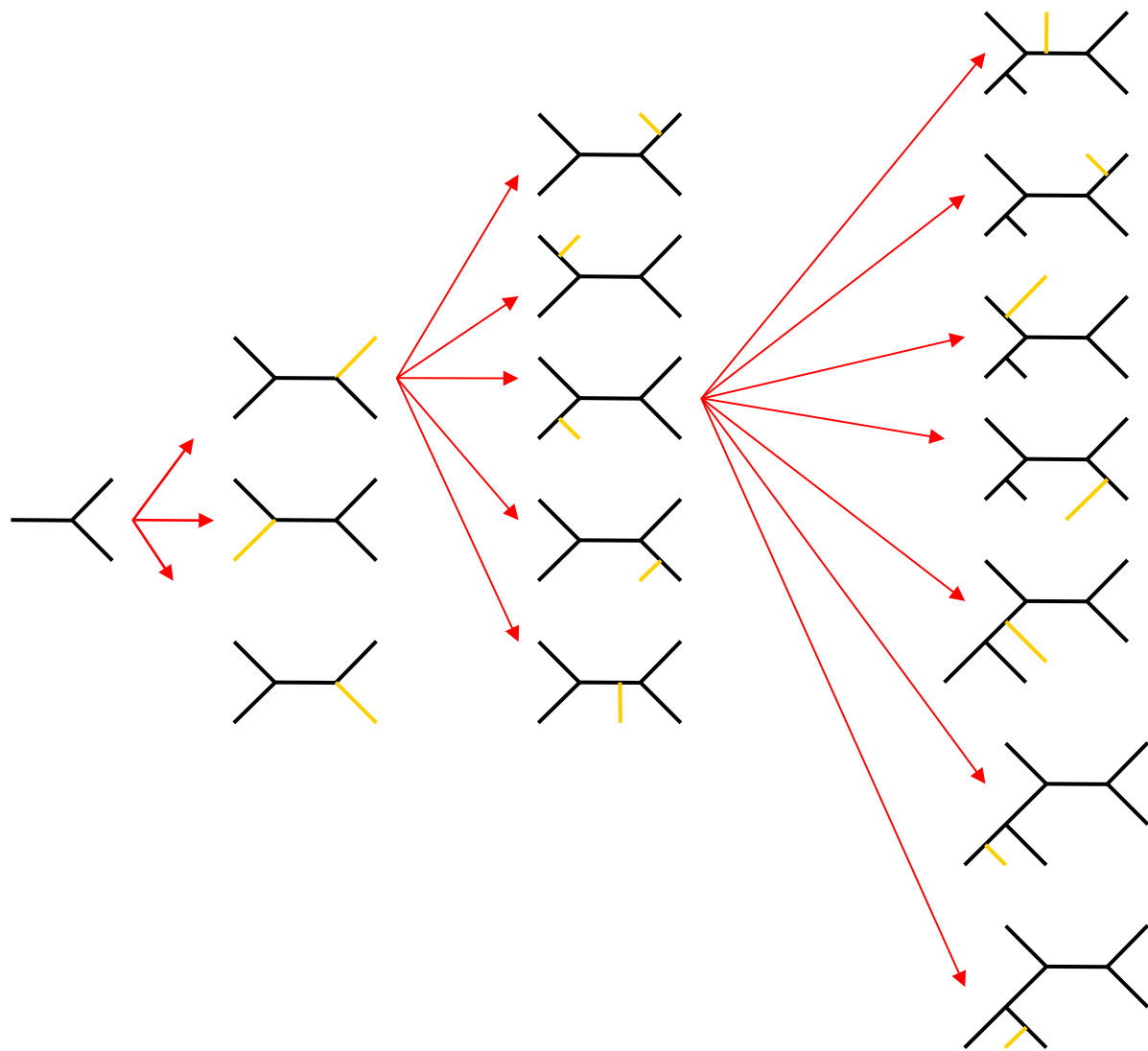
4 είδη  $\rightarrow$  3 δέντρα

# Ας κατασκευάσουμε δέντρα



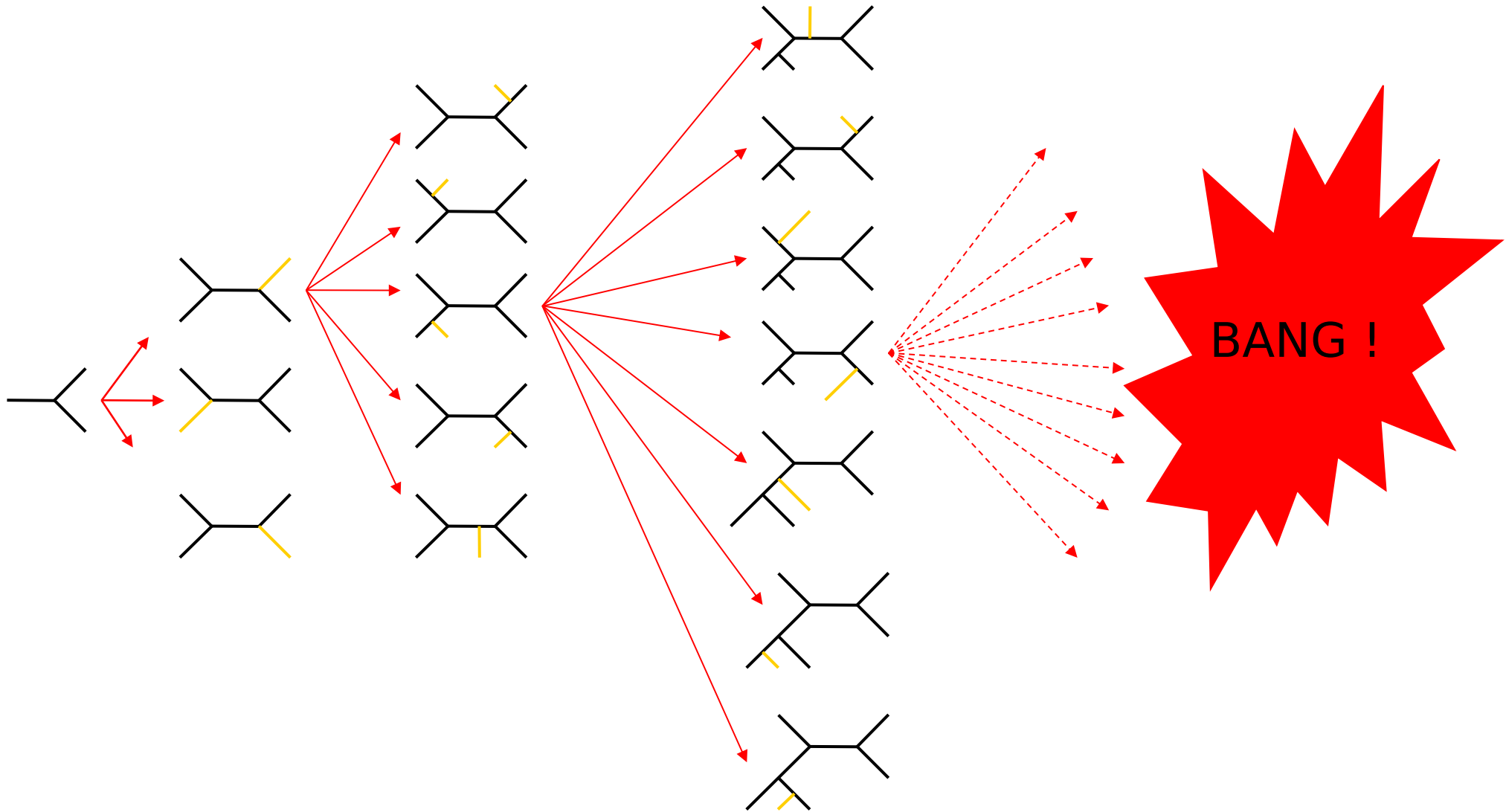
5 είδη  $\rightarrow$  15 δέντρα

# Ας κατασκευάσουμε δέντρα



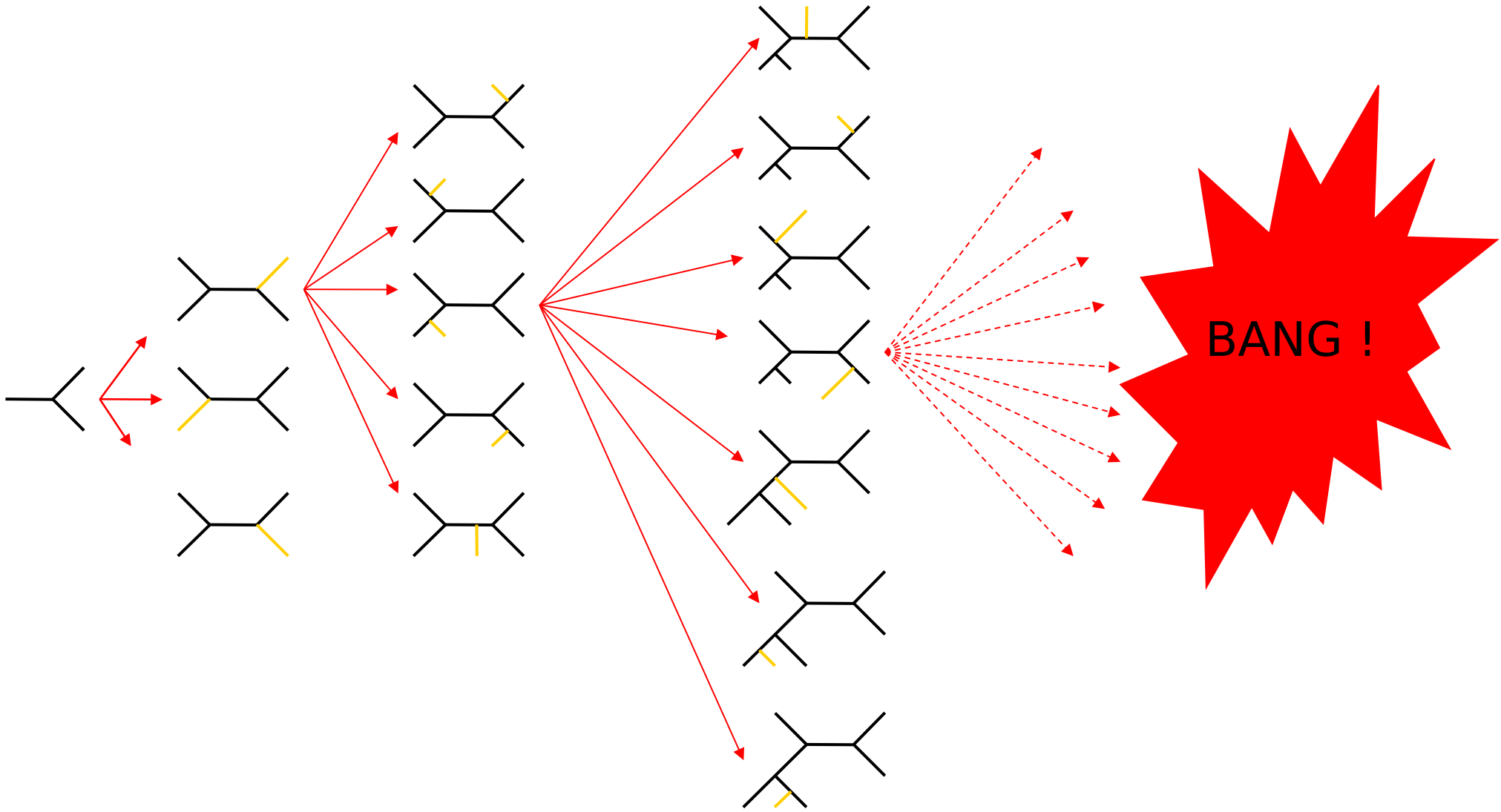
6 είδη  $\rightarrow$  105 δέντρα

Ο αριθμός των δέντρων αυξάνεται  
υπερεκθετικά με τον αριθμό των ειδών

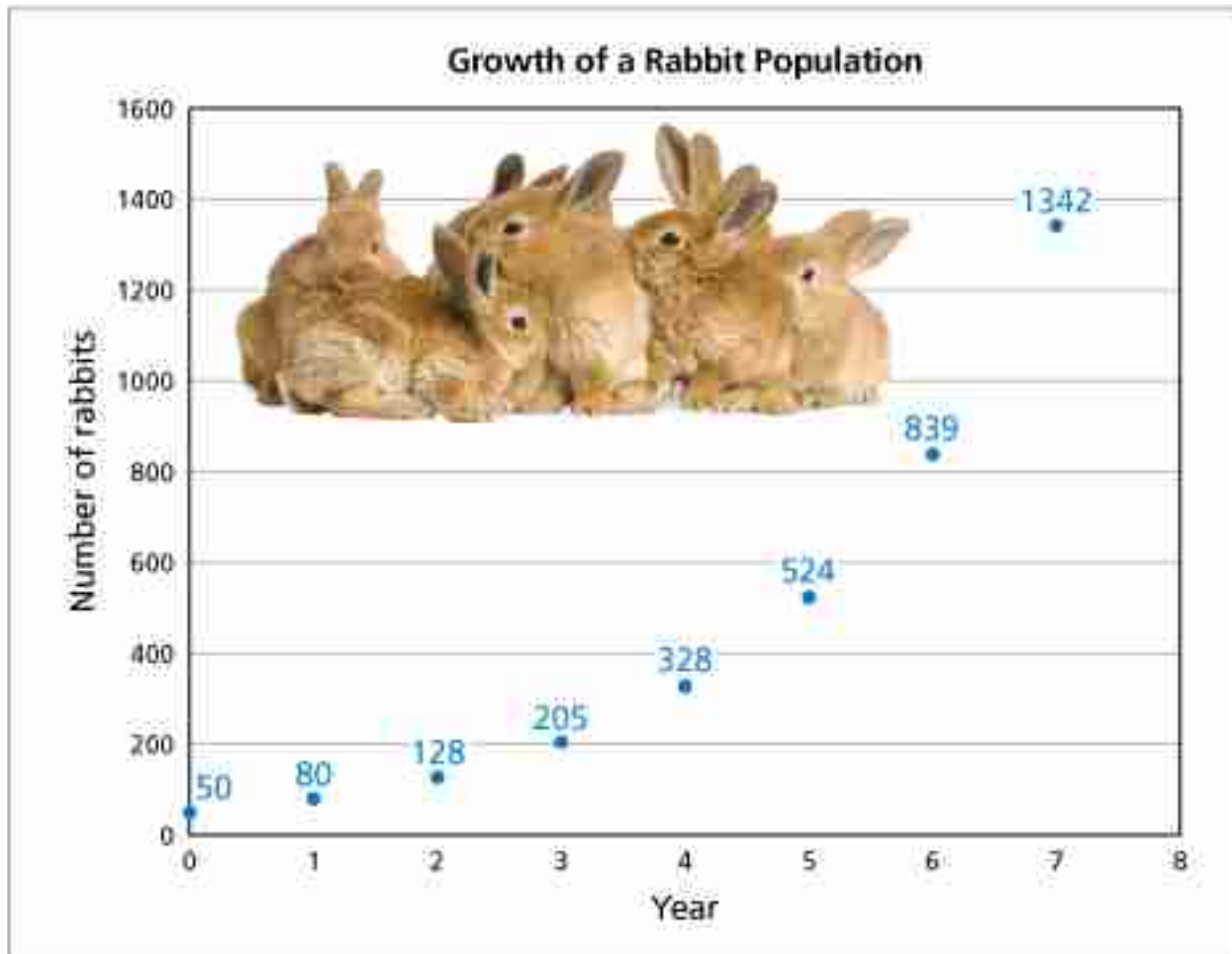




Την έχουμε βάψει ...



# Εκθετική Αύξηση



# Γιατί;

- Για να βρούμε το βέλτιστο δέντρο πρέπει να υπολογίσουμε το σκορ για **ΌΛΑ** τα δέντρα
  - δεν υπάρχει δυνατότητα να “κόψουμε δρόμο”
    - 10 είδη –  $2 \times 10^6$  δέντρα
    - 11 είδη –  $3 \times 10^7$  δέντρα
    - 15 είδη –  $8 \times 10^{12}$  δέντρα
    - 20 είδη –  $2 \times 10^{20}$  δέντρα
    - 50 είδη –  $3 \times 10^{74}$  δέντρα
- Αν υποθέσουμε, αρκετά ρεαλιστικά, ότι ο υπολογισμός του σκορ για ένα δέντρο παίρνει 1 δευτερόλεπτο

# Χρόνος Υπολογισμού

- 10 είδη → 555 ώρες
- 11 είδη → 8333 ώρες η 347 μέρες
- 15 είδη → 92592592 μέρες η 260091 χρόνια
- 20 είδη → περίπου  $2 \times 10^{15}$  μέρες η  $5.5 \times 10^{12}$  χρόνια
- 50 είδη → περίπου  $3 \times 10^{69}$  μέρες η  $9.5 \times 10^{66}$  χρόνια
  
- Και ένα δέντρο με 50 είδη σήμερα θεωρείται μικρό
- Σαν να βαράς το κεφάλι σου στον τοίχο ... ένα πράμα

# Ο αριθμός δέντρων για 2000 είδη

```
stamatak@exelixis:~/Desktop/GIT/TreeCounter$ ./treeCounter -n 2000
```

GNU GPL tree number calculator released June 2011 by Alexandros Stamatakis

```
Number of unrooted binary trees for 2000 taxa: 30049638174211656151632910065681814981377232074237013089504954043012636525258308210827685996688247000464352735214265634288295
8915023446000631493969130632970436056184861877465482277991223536809233455563199910834597693126756525012899867433187752811401960991631522367030609121735709762379847705467667
7795324797182614385273338226727784250737252849916669687584403510579587020686505817687044666318123742901021438506432471360934491667021135969756940300666252646479269124551031
4942366195542824118277625114848758254581227914289801132648902674033761294712745767036267579086843169660718609847941818865957214557044744572288661729053583520744253688123124
0106613156948861960941195646736200342575241335277575085829161096422575727699767991408283343210161327401652830993803904592327690690035972919709940739349563486203899010742687
2822975974655377102257672676842858011877224950106218117340523208265397342962227352536590515865631383272031119841987467599738646318290320383252308597997992216101227215780805
2481458312068440167606239306009711616729715504728487799634337531348994230372437347879131989085953764070134849446113877572576952408702461720107874297380462275052545706689372
3194182064407068918840038705902897721975164544959758216621306205064617761099485663734168183584989329076993382067801052437284614924034229611551826097782286191926720712951895
8936009959130974233072316382518428110330571017441156884305131865877544376308500311451110723837039707465182232040406154708273078629957549331031275208616700660791298014262230
0565123522718063819509335872651728623589020520016144361756075654286471422126613004434807084067501589247673166341539540575074474994909831496473031080411401891849735912811228
3787740498848340562102420566424463860093899650857429619472690543015281237526510965815284699797036792171129035568098180791695879516141592810495281798558472925344478644244359
9808531537204796814969465991768614533701051985928577157482455943377242369582576242663016946320482495182255939287403177623433881048604630975191556923871167513095213415098816
7154643078623526062377864068386804246902527491139319276802611515990582603886733172930713673903403618637463980605764836474670274446727880885337074254421922726677747003329403
320103828803511268902625518309679194835867892937016376817530482063389438714979311523536982296251116307148294599211620803302684762013335690441089668145436150905155877581167
9770012563912151116237444170497371704604029481104114822286466131918821997571383368352072526055202769823974613218495249264897050790398360256255606289852288839561357874156576
6488999260873286612630642543260248979229113560071640573984516375245243376943755857384725545564397599604255914640112221144755235573176239973057747183956531217416532295986675
9012941161239240722093250369673124884491553759210650656015416720774159236240868667675348286512964888739059707578802473393463470848159011639772797747480417316268700916728735
6121642268468160683198959801260376485615312781611689587215123123308760063473381097253118423339640390937378395066835578735307886358646400563299499490631187424029092779272693
30032244537759579722487345689151114585570783850541681667667425811301958063621907500790295031088209097271748136436989473971079932777700676301730617566538739726037777173008441
343940512366905544932486165082539957795036326704947844293498853172797348177797146567175151178876396434069332458076346110734214328195049909680874027397688914704517472055543
8969396668742601477241894693212902453317334188286773194653544113302100866575081713240342647580489218623663461607955943720516395154096949806486242309796947211169665961580041
7883992232264628498942352276926391124360767508997671789683459337890742346354557193530665615379900277916265636361861974048590938234062235459769733137213659343717585590664439
6461328300113672601934068706442339489199215304385281541659630118549423634863524585746642834609062916279564926584723003608555598989761916293248140094592489899468468862322586
0170551468905649837276003927487069550463788819741699429049102853180410776516007263384721638903782270013843599599730265727198804324664312359758555217196960513921010226596324
7887830977405333131551762978152880718652603176327264828094499370456258099938053058497699570080289379808014902901005293898472279947167804821689424159118284257696464786505731
5325178302336307298251692210346584265894447464916123854689718507968172909139032182834111184821384767728316548653212317382004131990510518967022201887049585687180509590730360
6930402937216038968917605587676955382318093705826257083898387409098468656634271397500013291835105943321729879825243707508272087959859437157667660155782699660343197752623308
8989962587800628009560944416932377944955441033696586261556256010669390303203878970983673786087056641433585106111658314520424513208508589994932364831689671194951671619567622
7070906973889588855579562466641536561723549301807394004760529801721771391686788000277851966173070061284517307582503735643102065112443730825229625040453160590741343881872563
4779138306605909318802522310085340176840261401539616989192075147108033757708849740141834599753972059878682064879116064969858177601153972058498222698907181349432691801821173
3188063653910893689811714891357456680542807485170175858266639633570189354449832669762835092657922201746372190273119641751489944010079636876017826747107019945473218887832742
6088966724371574713420600093704251309893630537459784279980403132989417266492290425730958368534416215640557290282066224003863237526380910233269897838860423759625601567975262
6950798639868104294832333160267216555178120899264677804935741326387137408423885546538336158643451305439624281397279559725995110706314305992615495622958320232708057681156690
4895866105220300573725298472118747827136713666058669271094875563974858489475910819727033878284439864486743456200958161930314727345961900499318424337975243662489363321244850
5971992523668529249305346252764137853413208943128901523738092556045987090912766662329678703328882059134949580074074473143388800724532321747309659741967114441453127132790205
1010047670104350638857953478447255389801541923317027519896180635152682543173193832925891931530164130548972311128666465492971930479296432829556719092881692091042334122007454
2420499008725850462080511048758830594959903111887366685094148821725734576355233964038481318213167408359006916400053262258184783765067804451177717328658189899215358309447765
350341796875
```

Approximately 3.00 times 10<sup>6</sup>6328

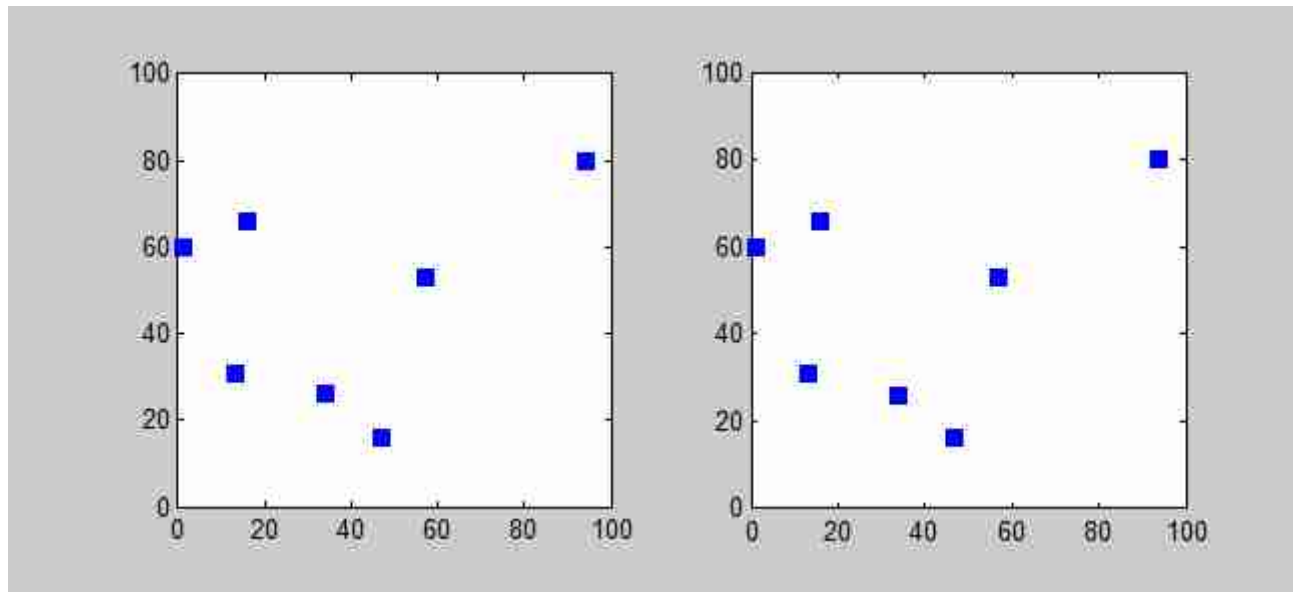
# NP-σκληρά Προβλήματα

- Στην πληροφορική (και στην καθημερινότητα) υπάρχουν πολλά τέτοια προβλήματα, όπως η εύρεση του βέλτιστου φυλογενετικού δέντρου – οπού χτυπάς το κεφάλι σου στον τοίχο → λόγω του (υπερ)εκθετικά αυξανόμενου αριθμού λύσεων δεν μπορούμε να βρούμε την βέλτιστη λύση ... γιατί θα περιμένουμε για αιώνες

# Ένα άλλο παράδειγμα προσαρμοσμένο στην Κρήτη

- Ο περιπλανώμενος τουρίστας
  - **μαθηματικό πρόβλημα του πλανόδιου πωλητή**
- Ένας τουρίστας θέλει να δει πχ 7 αξιοθέατα του νησιού,  
Σε ποιο αξιοθέατο πρέπει να ξεκινήσει και με ποια σειρά πρέπει να τα επισκεφτεί ώστε να βελτιστοποιήσει την κατανάλωση βενζίνης του (να ελαχιστοποιήσει τα χιλιόμετρα) και να επιστρέψει στο σημείο εκκίνησης (ένα από το 7 αξιοθέατα);
- Ούτε για αυτό δεν μπορούμε να βρούμε την βέλτιστη λύση για μεγάλο αριθμό αξιοθέατων
- Για  $n$  αξιοθέατα υπάρχουν  $n!$  διαφορετικές διαδρομές
  - 7 αξιοθέατα – επιλογές:  $7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$
  - $n=7$ ,  $7! = 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 = 5040$
  - $n=8$ ,  $8! = 40320$και πάει λέγοντας ...

# Ένα Παράδειγμα

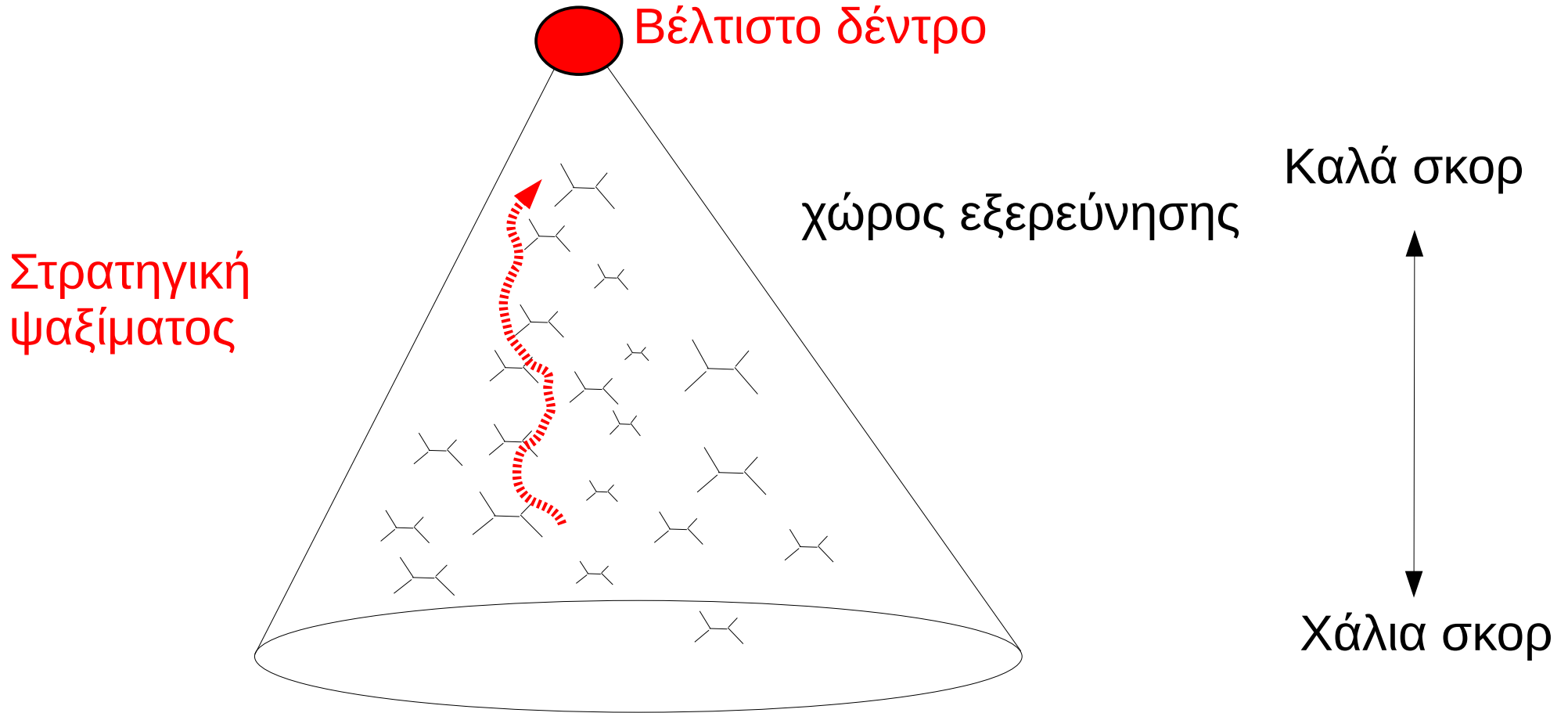




# Άλλα Παραδείγματα

- Βέλτιστη κατανομή θέσεων σε τραπέζια ενός γάμου ώστε φίλοι να κάθονται μαζί και εχθροί σε ξεχωριστά τραπέζια
- Βέλτιστη κατανομή αιθουσών στον πανεπιστήμιο ώστε οι περισσότεροι καθηγητές να κάνουν μάθημα την ώρα που θέλουν

# Πίσω στα δέντρα – λύσεις στην πράξη



# Πίσω στα δέντρα – λύσεις στην πράξη



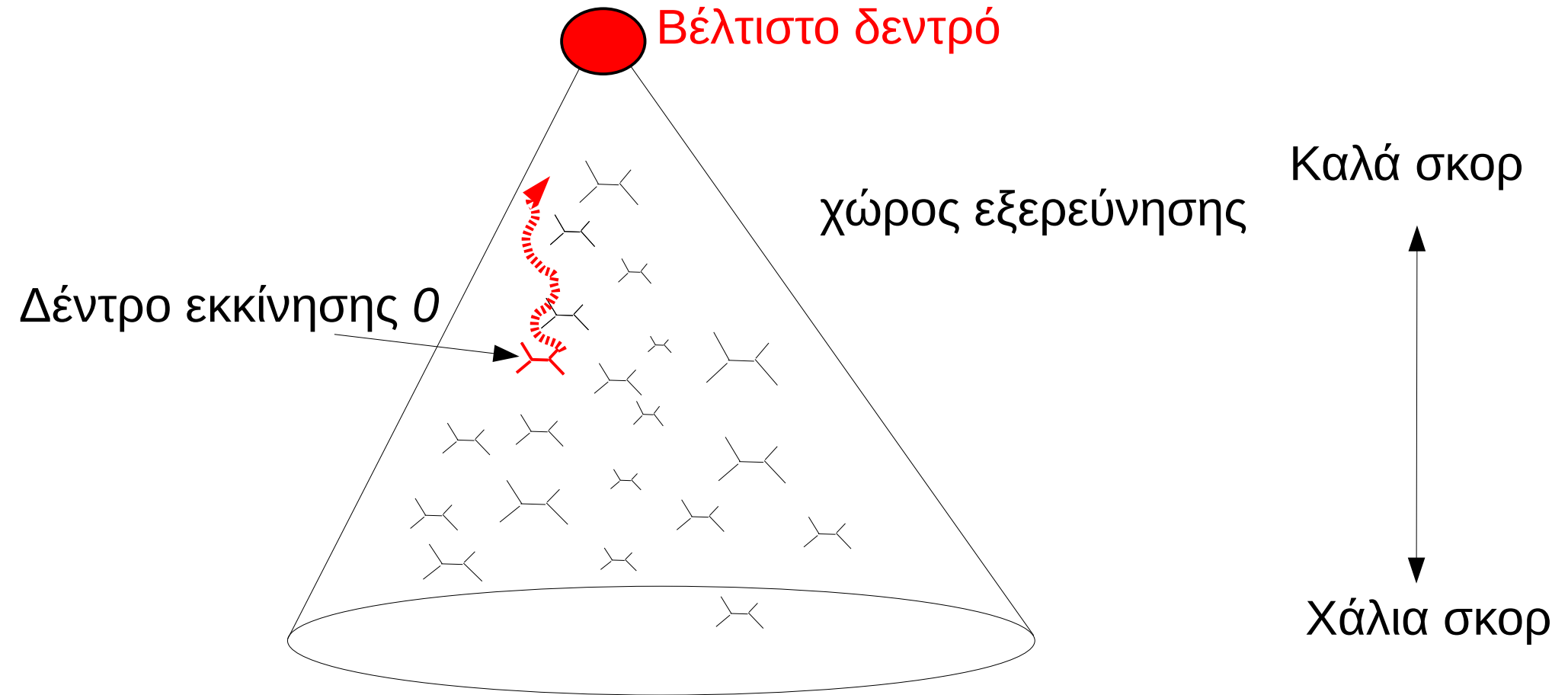
**Το πρόβλημα να βρούμε το βέλτιστο δέντρο είναι NP-σκληρό για όλα τα γνωστά κριτήρια**

# Πίσω στα δέντρα – λύσεις στην πράξη

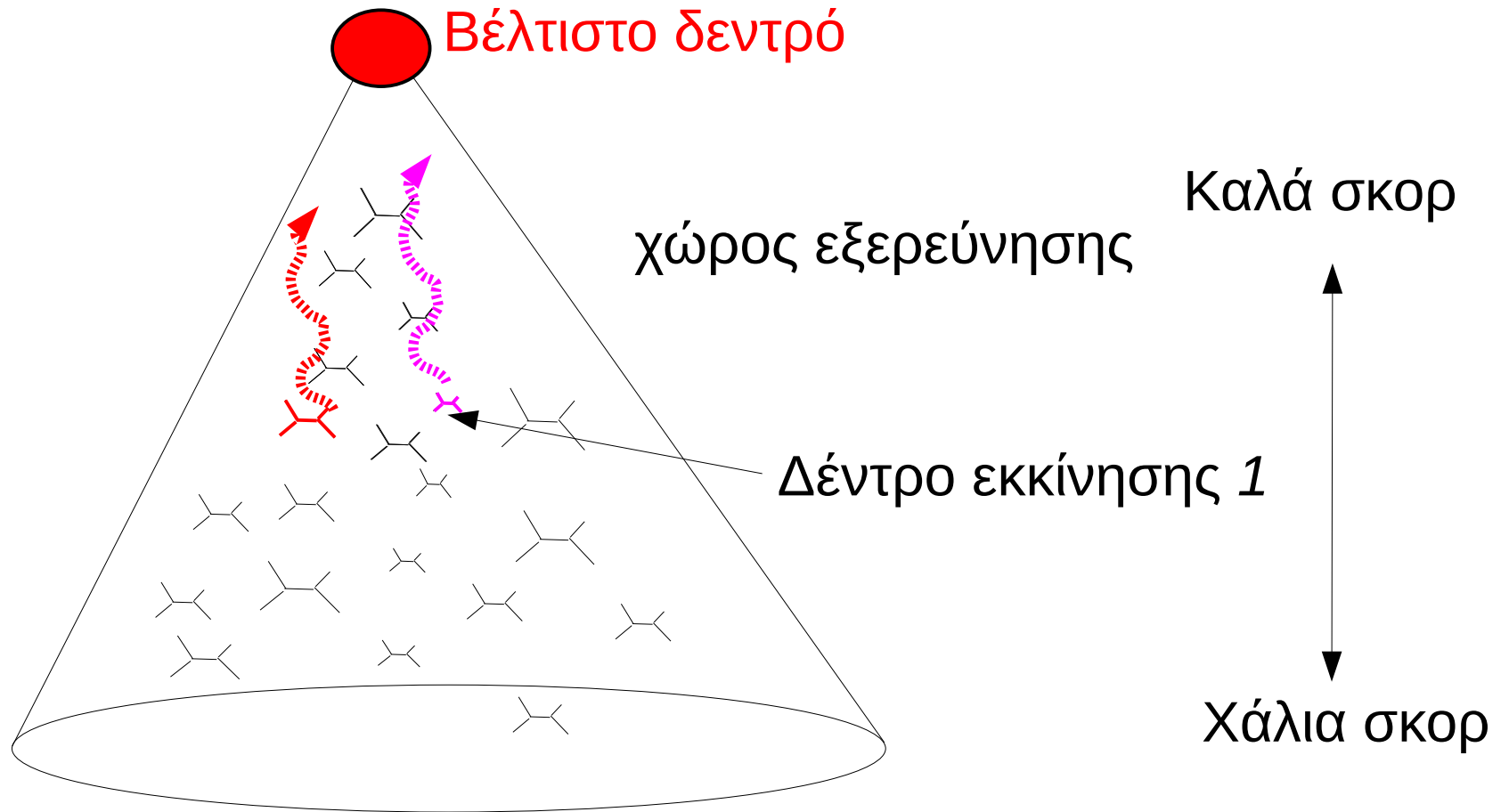


Το πρόβλημα να βρούμε το βέλτιστο δέντρο είναι NP-σκληρό για όλα τα γνωστά κριτήρια

# Δέντρα εκκίνησης



# Δέντρα εκκίνησης

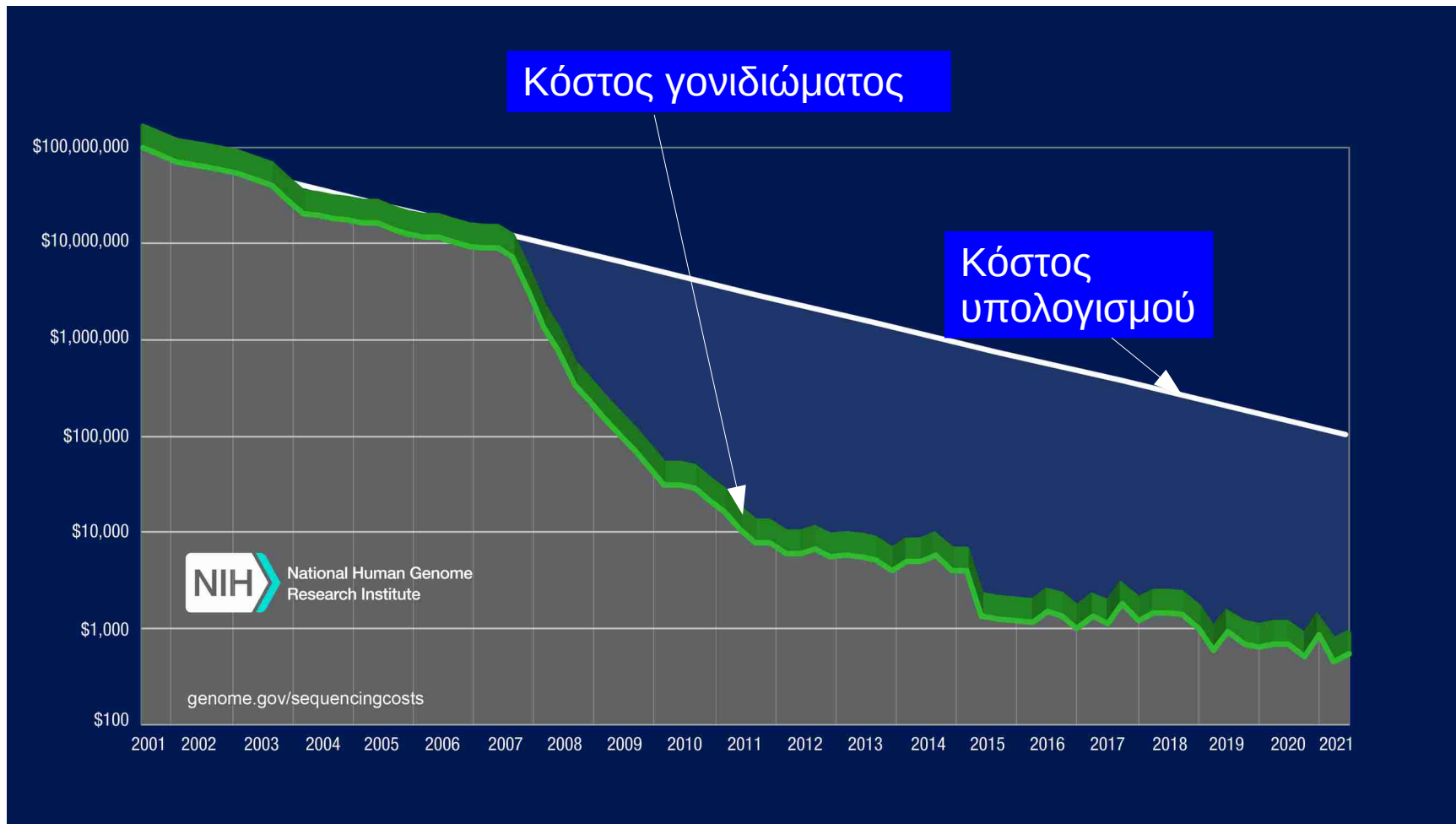


# Δομή Ομιλίας

- Εφαρμογές Φυλογενετικών Δέντρων
- Υπολογιστική Πολυπλοκότητα
- Άλλες Προκλήσεις

# Πνιγόμεστε στα Δεδομένα

Κόστος



Χρόνος



# Δεδομένα στο πέρας του χρόνου

- Δεκαετία του 90
  - 10-20 είδη & ένα γονίδιο
  - Ένα γονίδιο: περίπου 1000 χαρακτήρες DNA ανά είδος
- 2001 (όταν άρχισα το διδακτορικό)
  - 150 είδη & ένα γονίδιο (βακτηρίδια)
- 2011
  - 55,000 είδη & 10 γονίδια - περίπου 10,000 χαρακτήρες DNA ανά είδος (φυτά)
- 2014
  - 140 είδη & ολόκληρο το γονιδίωμα - πάνω από 1,000,000 χαρακτήρες ανά είδος (πουλιά)  
"Whole-genome analyses resolve early branches in the tree of life of modern birds". *Science*, 46(6215):1320-1331, 2014
- 2019
  - 350 είδη & ολόκληρο το γονιδίωμα - πάνω από 1,000,000 χαρακτήρες ανά είδος (πουλιά)  
→ χρειάστηκαν **500,000** ώρες υπολογισμού για να υπολογίσουμε ΈΝΑ δέντρο

# 500,000 ώρες υπολογισμού;

- Χρησιμοποιούμε παράλληλους επεξεργαστές



Παράλληλος επεξεργαστής για την έρευνα στο Μόναχο – πάνω από 60,000 πυρήνες

# Προκλήσεις

- Η παραλληλοποίηση απαιτεί πολλή δουλειά του προγραμματιστή, ώστε όταν τρέχουμε το πρόγραμμα σε 10 πυρήνες να τρέχει και 10 φορές πιο γρήγορα
- Καταναλώνουμε τεράστιες ποσότητες ενέργειας με 500,000 ώρες υπολογισμού

# Καινοτομίες

New Results

Comment on this paper

## A Fast and Memory-Efficient Implementation of the Transfer Bootstrap

Sarah Lutteropp, Alexey M. Kozlov, Alexandros Stamatakis

doi: <https://doi.org/10.1093/bioinformatics/btx205>

480x επιτάχυνση

## EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences

Pierre Barbera, Alexey M Kozlov, Lucas Czech, Benoit Morel, Diego Darriba, Tomáš Flouri, Alexandros Stamatakis

Systematic Biology, Volume 67, Issue 3, 353–369, <https://doi.org/10.1093/sysbio/yty033>

30x επιτάχυνση

CORRECTED PROOF

## RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference

Alexey M Kozlov, Diego Darriba, Tomáš Flouri, Benoit Morel, Alexandros Stamatakis

Bioinformatics, btz305, <https://doi.org/10.1093/bioinformatics/btz305>

Published: 09 June 2017, 4x επιτάχυνση

## Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo

P Kapli, S Lutteropp, J Zhang, K Kobert, P Pavlidis, A Stamatakis, T Flouri

Bioinformatics, Volume 33, Issue 11, 1 June 2017, Pages 1629–1638, <https://doi.org/10.1093/bioinformatics/btx025>

Published: 20 January 2017, 1000x επιτάχυνση

Θα εξοικονομήσουμε πολλή ενέργεια έτσι;

# Το παράδοξο του Jevons

New Results

Comment on this paper

## A Fast and Memory-Efficient Implementation of the Transfer Bootstrap

Sarah Lutteropp, Alexey M. Kozlov, Alexandros Stamatakis

doi: <https://doi.org/10.1093/bioinformatics/btz305>

480x επιτάχυνση

## EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences

Pierre Barbera, Alexey M Kozlov, Lucas Czech, Benoit Morel, Diego Darriba, Tomáš Flouri, Alexandros Stamatakis

Systematic Biology, Volume 68, Issue 3, 1 June 2019, Pages 363–369, <https://doi.org/10.1093/sysbio/sby033>

30x επιτάχυνση

CORRECTED PROOF

## RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference

Alexey M Kozlov, Diego Darriba, Tomáš Flouri, Benoit Morel, Alexandros Stamatakis

Bioinformatics, btz305

Published: 20 June 2017

4x επιτάχυνση

## Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo

P Kapli, S Lutteropp, J Zhang, K Kobert, P Pavlidis, A Stamatakis, T Flouri

Bioinformatics, Volume 33, Issue 11, 1 June 2017, Pages 1627–1638, <https://doi.org/10.1093/bioinformatics/btx025>

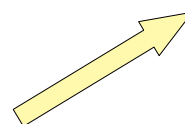
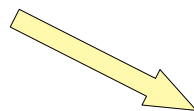
Published: 20 June 2017

1000x επιτάχυνση

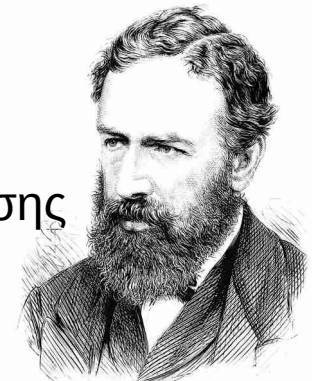
W. S. Jevons „The Coal Question“ (1865)

Βελτιωμένη αποδοτικότητα

Αυξημένος ρυθμός κατανάλωσης



Χαμηλότερο κόστος



# Φιλοσοφικά Ζητήματα

- Στις περισσότερες περιπτώσεις δεν γνωρίζουμε το πραγματικό φυλογενετικό δέντρο
- Πρέπει να υπολογίσουμε όχι μόνο ένα δέντρο αλλά και την αβεβαιότητά του
- Ανάλογα με τα δεδομένα που έχουμε το δέντρο υπολογίζεται εύκολα ή δύσκολα



# Μορφές Δεδομένων

Αυτά;

Ποια δεδομένα είναι πιο δύσκολα?

*E1*

*E2*

·

·

·

·

·

·

·

·

·

·

·

·

·

·

·

*E10000*

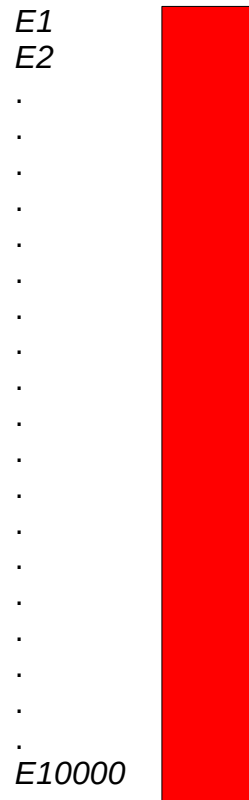






# Δυσκολία Δεδομένων

Λίγο μπακαλίστικο είναι αυτό → μπορούμε να ορίσουμε τον βαθμό δυσκολίας μαθηματικά;



Δύσκολο



Εύκολο

# Και όχι μόνο, μπορούμε να προβλέψουμε τον βαθμό δυσκολίας

JOURNAL ARTICLE

## From Easy to Hopeless—Predicting the Difficulty of Phylogenetic Analyses

Julia Haag , Dimitri Höhler, Ben Bettisworth, Alexandros Stamatakis

*Molecular Biology and Evolution*, Volume 39, Issue 12, December 2022, msac254,

<https://doi.org/10.1093/molbev/msac254>

**Published:** 17 November 2022

Μέθοδος τεχνητής νοημοσύνης

Παίρνει τα δεδομένα και κάνει πρόγνωση του βαθμού δυσκολίας για την κατασκευή δέντρου μεταξύ 0.0 (εύκολο) και 1.0 (δύσκολο)

Η πρόγνωση του βαθμού δυσκολίας μας επιτρέπει να προσαρμόσουμε ανάλογα την στρατηγική ανάλυσης των δεδομένων μας

Σας ευχαριστώ πολύ

