

RAxML : 20 χρόνια ανάπτυξης λογισμικού ανοιχτού κώδικα για την εξελικτική βιολογία

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

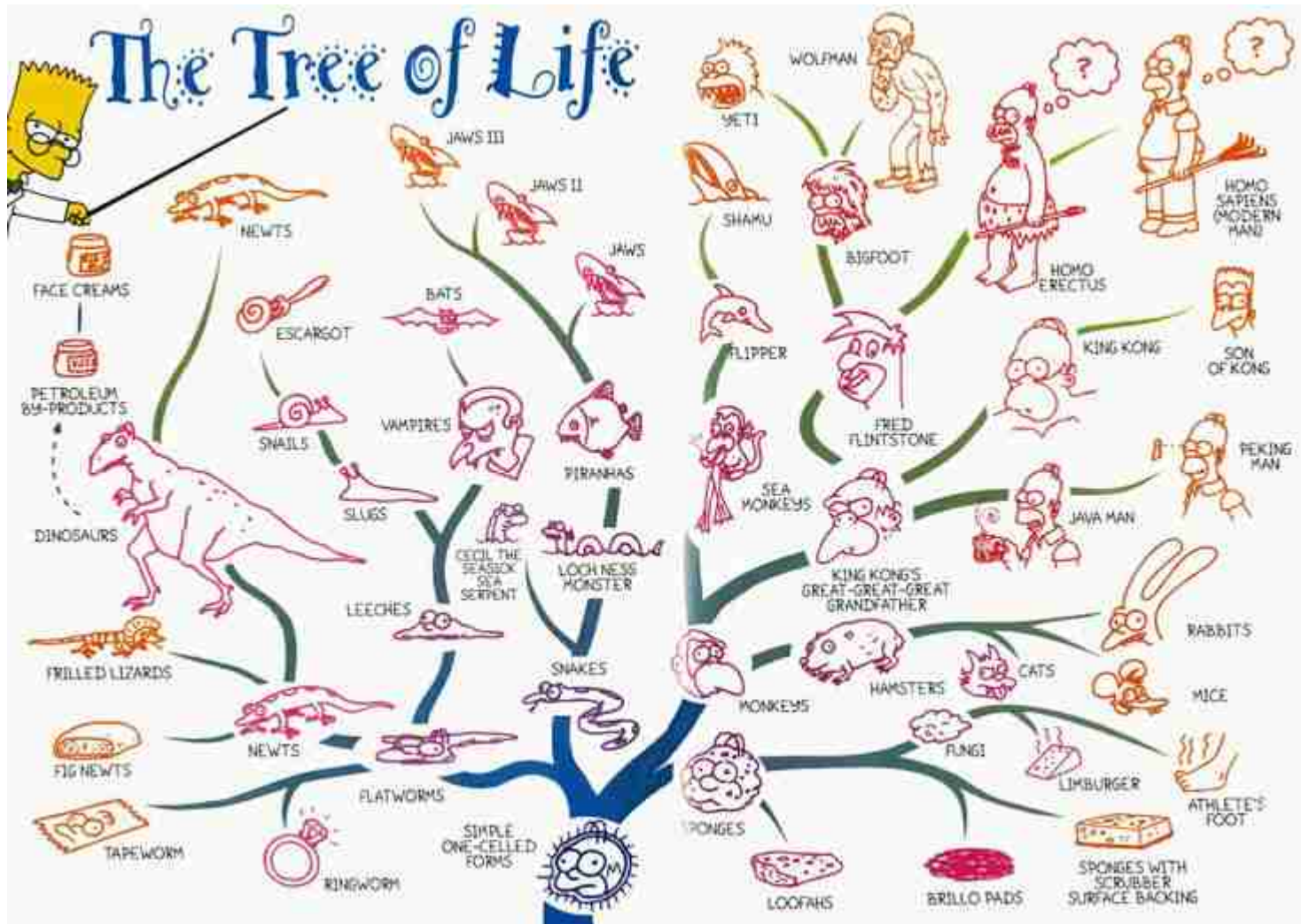
www.biocomp.gr (εργαστήριο Κρήτης)

www.exelixis-lab.org (εργαστήριο Χαϊδελβέργης)

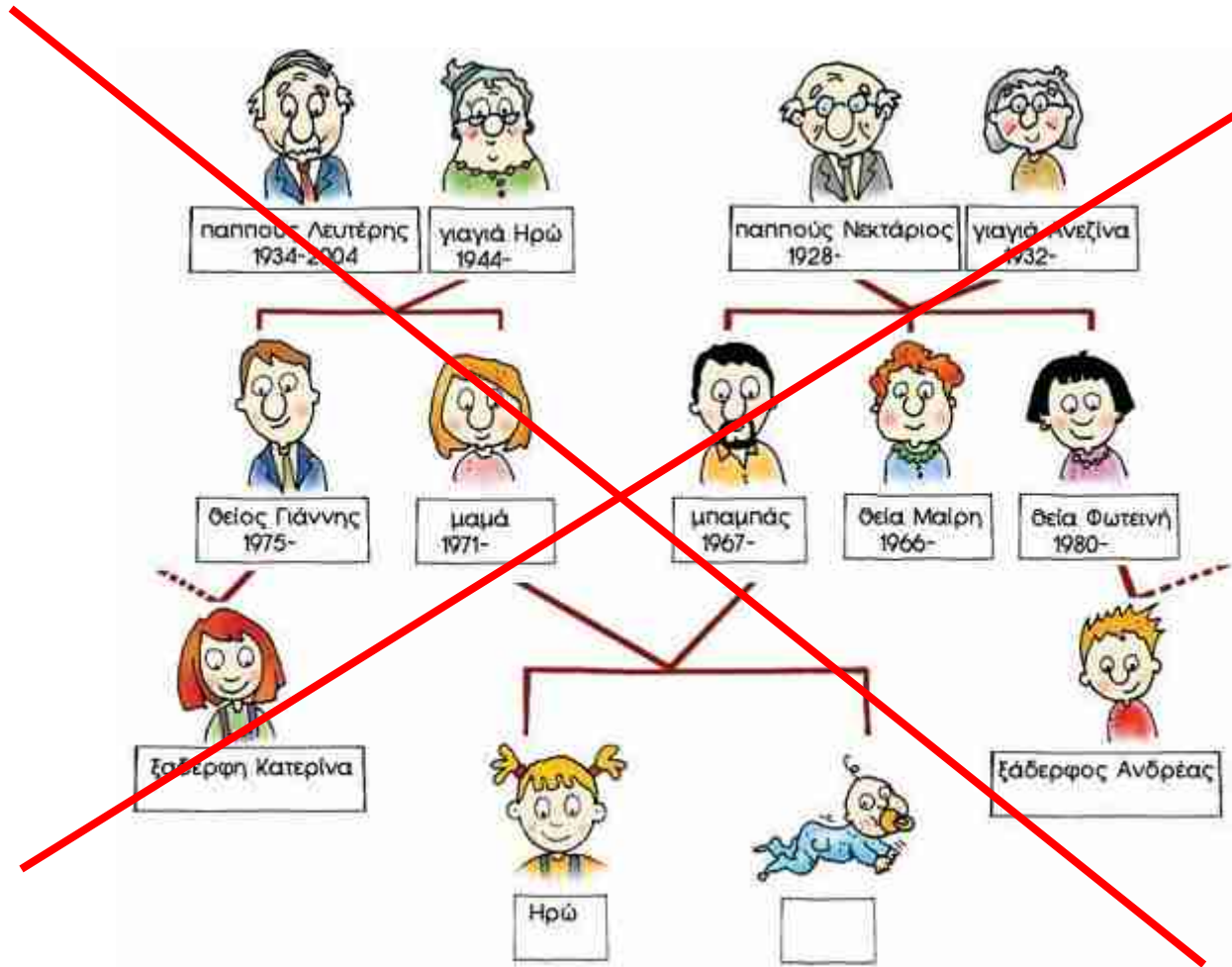
Δομή Ομιλίας

- **Τι υπολογίζει το RAxML;**
- Εφαρμογές
- Υπολογιστική Πολυπλοκότητα
- Γιατί ανοιχτός κώδικας;
- Προβλήματα και Προκλήσεις

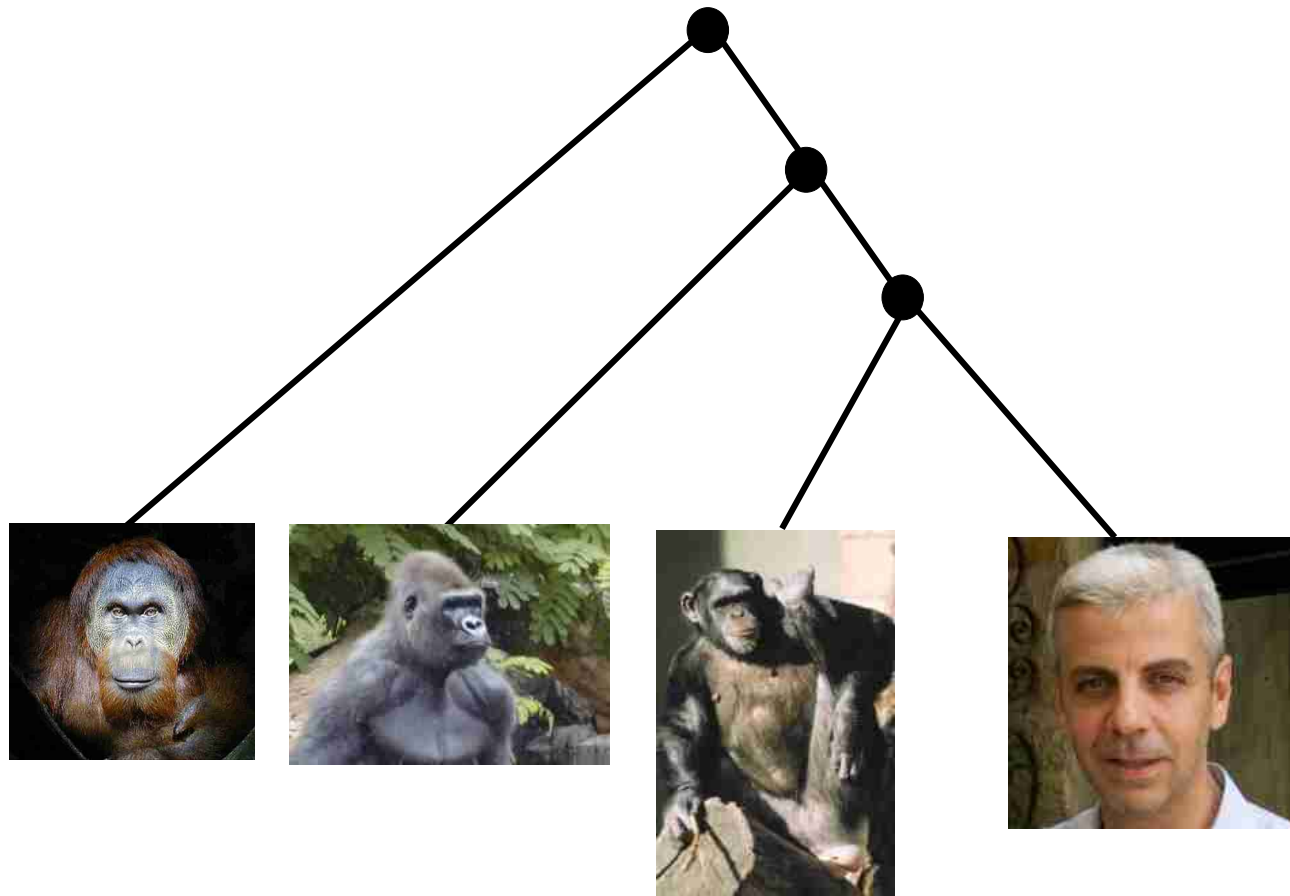
Φυλογενετικά Δέντρα



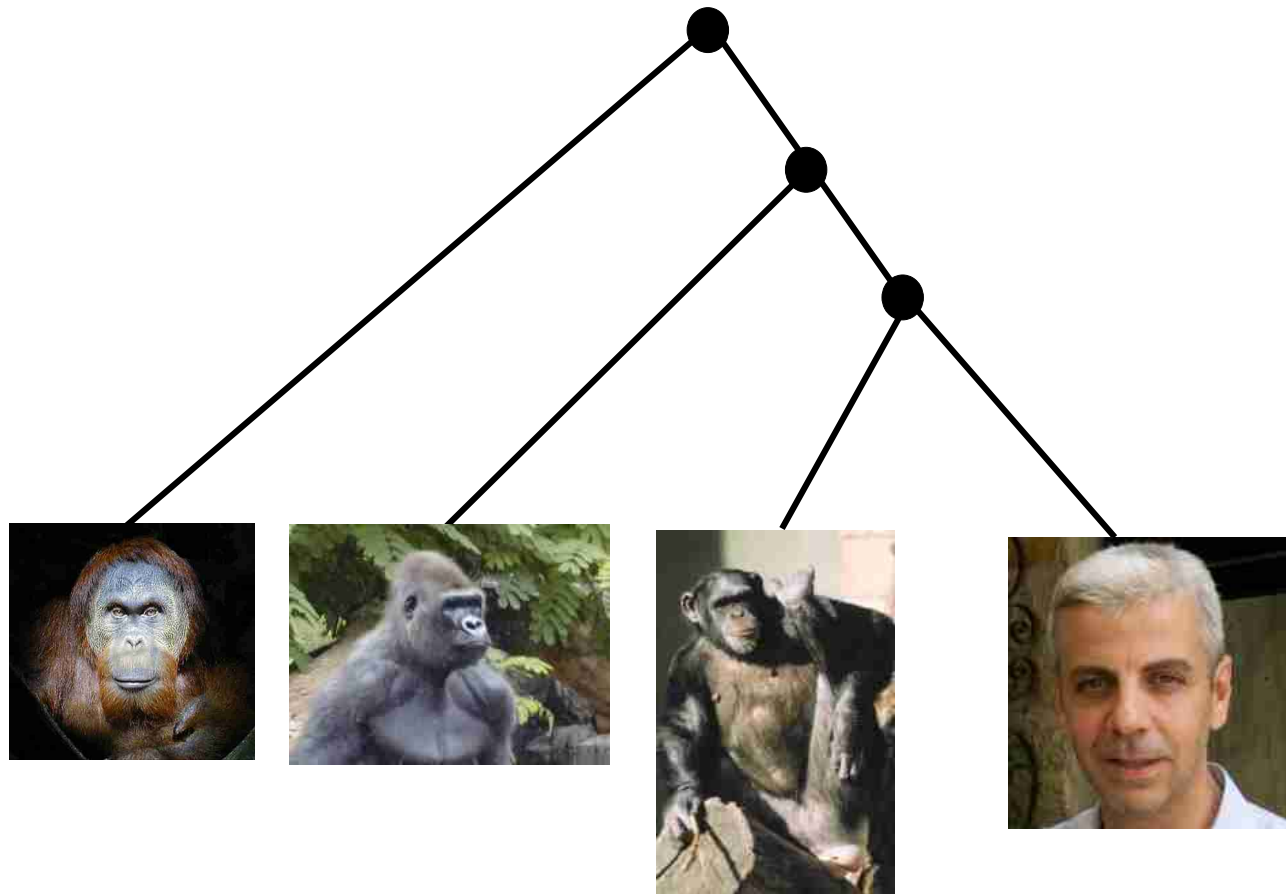
Δεν θα μιλήσουμε για γενεαλογικά δέντρα



Αλλά για φυλογενετικά δέντρα

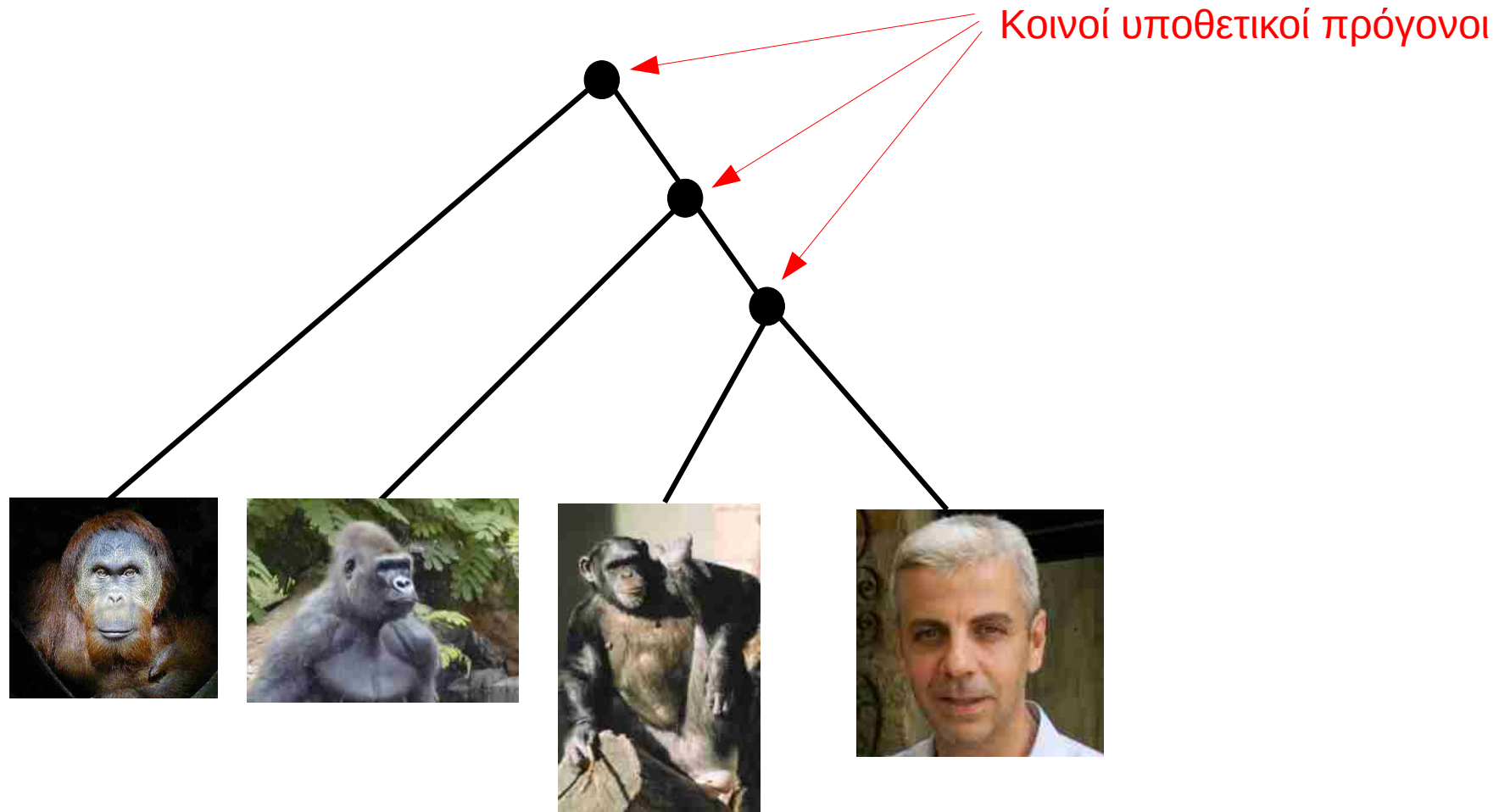


Αλλά για φυλογενετικά δέντρα



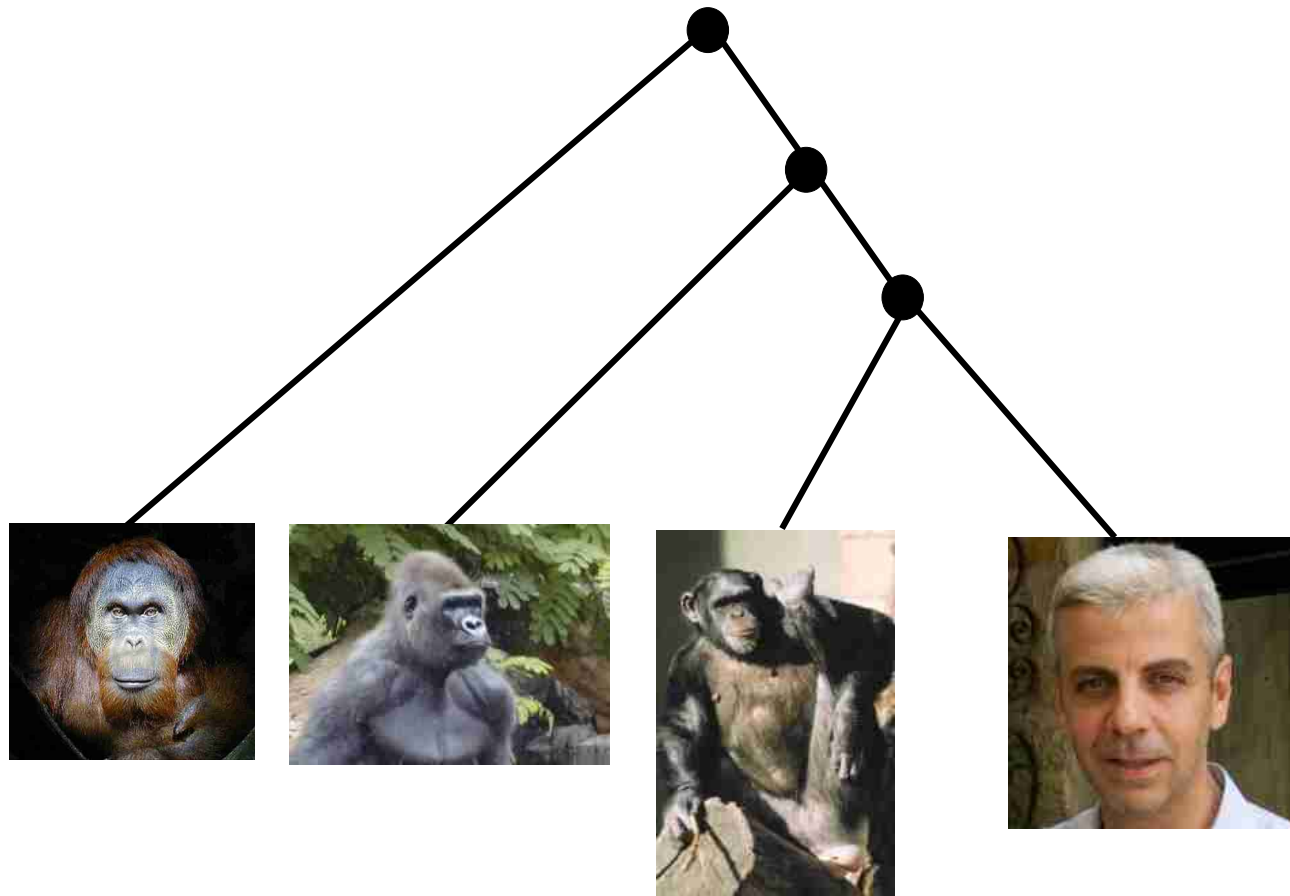
Μας ενδιαφέρει να ανατρέξουμε στο παρελθόν για να καταλάβουμε την εξελικτική ιστορία ανάμεσα **σε διαφορετικά είδη** !!!!

Αλλά για φυλογενετικά δέντρα

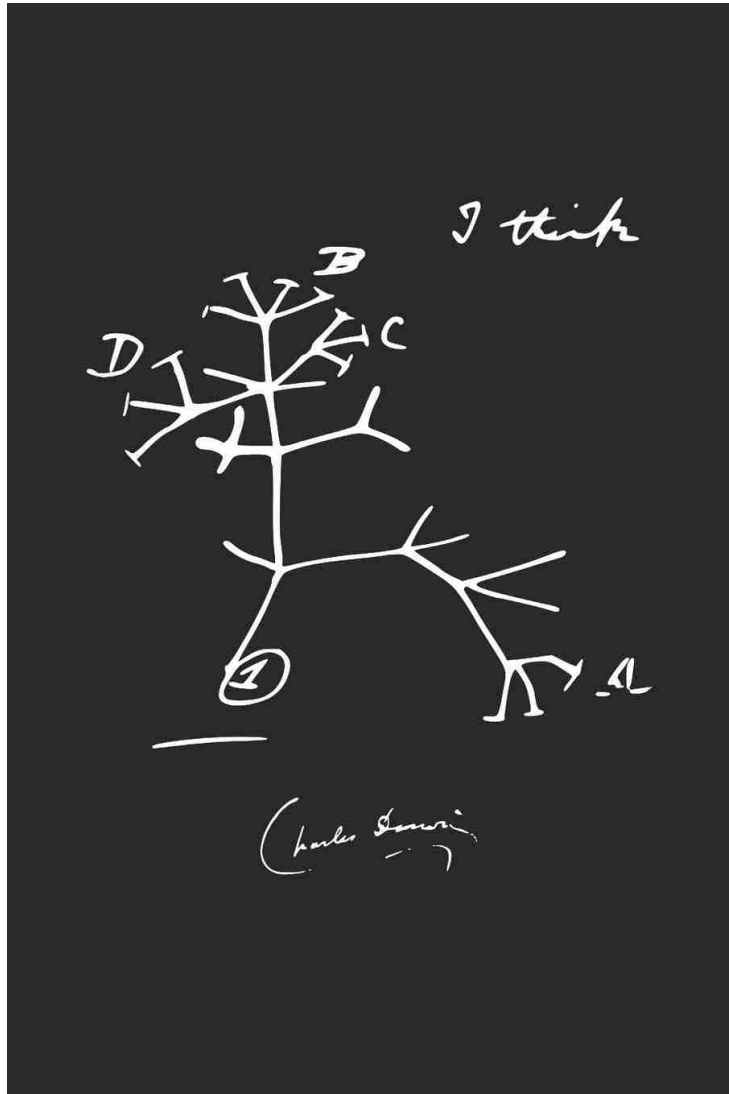


Αλλά για φυλογενετικά δέντρα

Πώς κατασκευάζουμε ένα τέτοιο δέντρο;



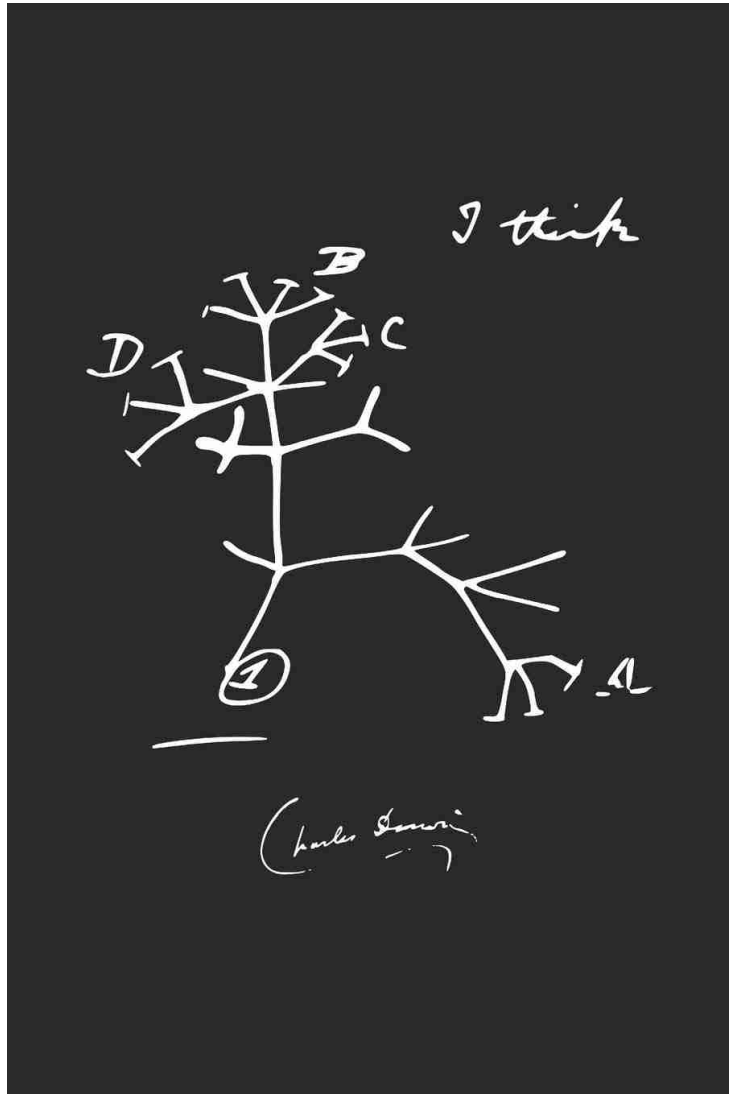
Τον παλιό καιρό



Δαρβίνος 1837

Συγκρίνοντας μορφολογικά χαρακτηριστικά

Τον παλιό καιρό

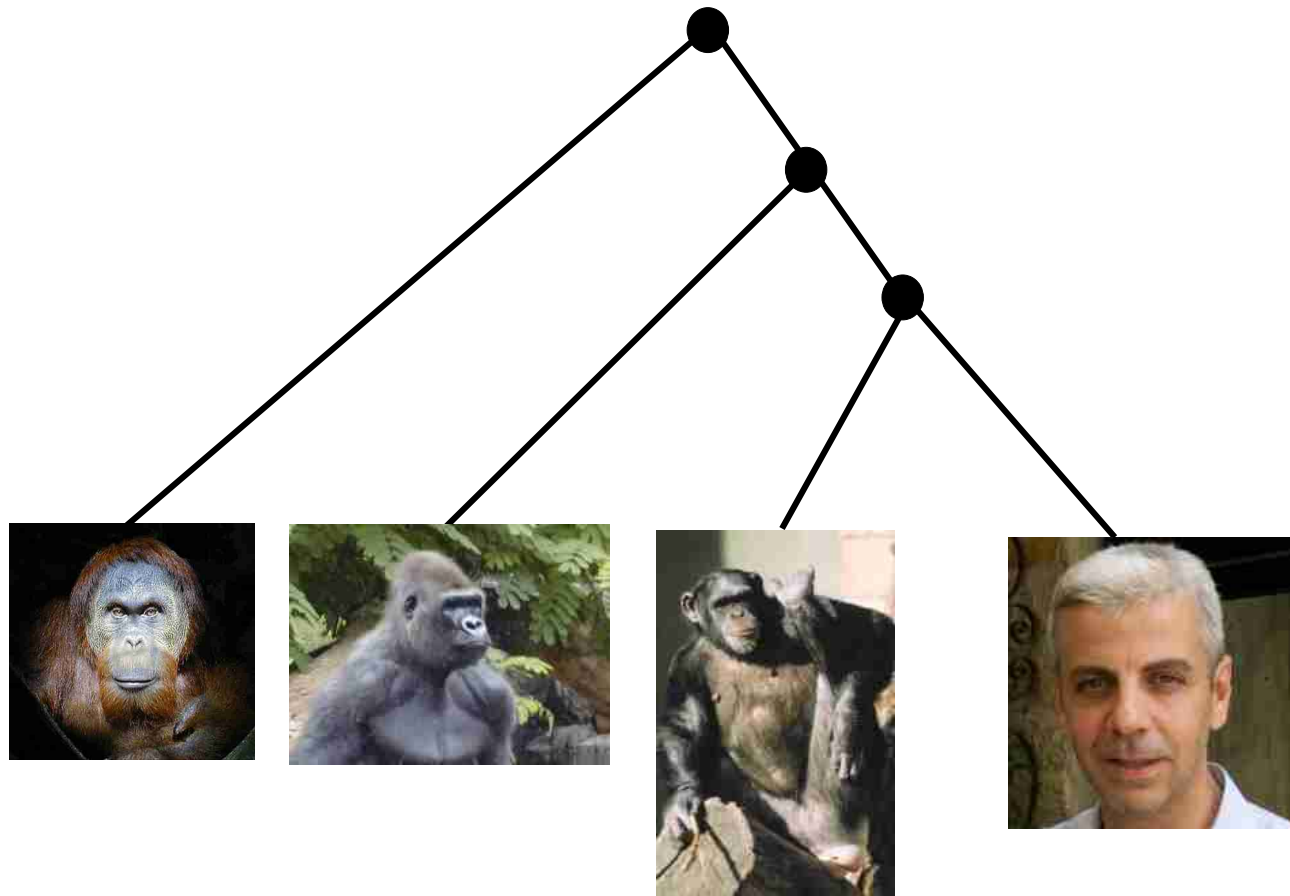


Δαρβίνος 1837

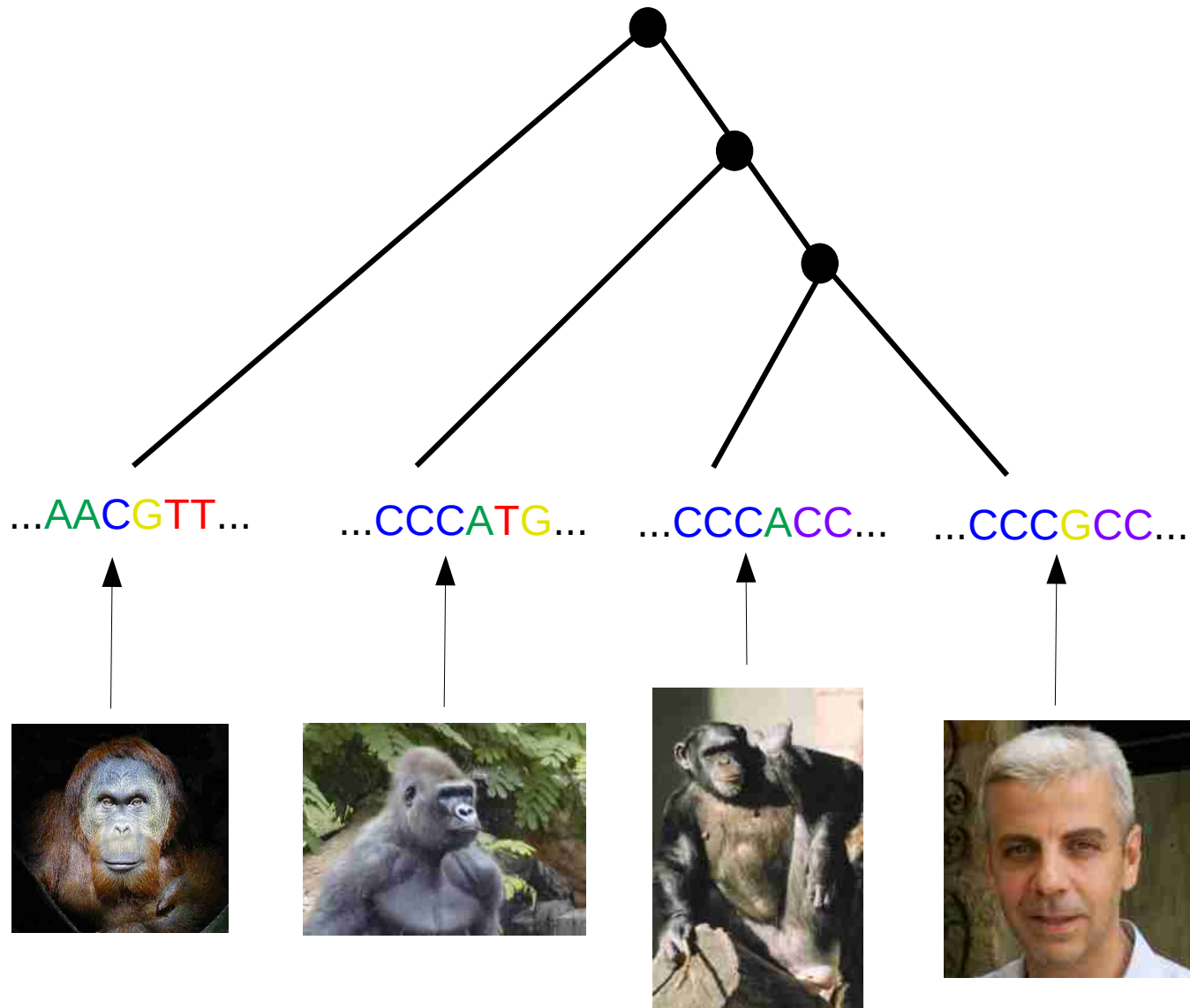
Συγκρίνοντας μορφολογικά χαρακτηριστικά

→ είναι κάπως υποκειμενικό

Σήμερα

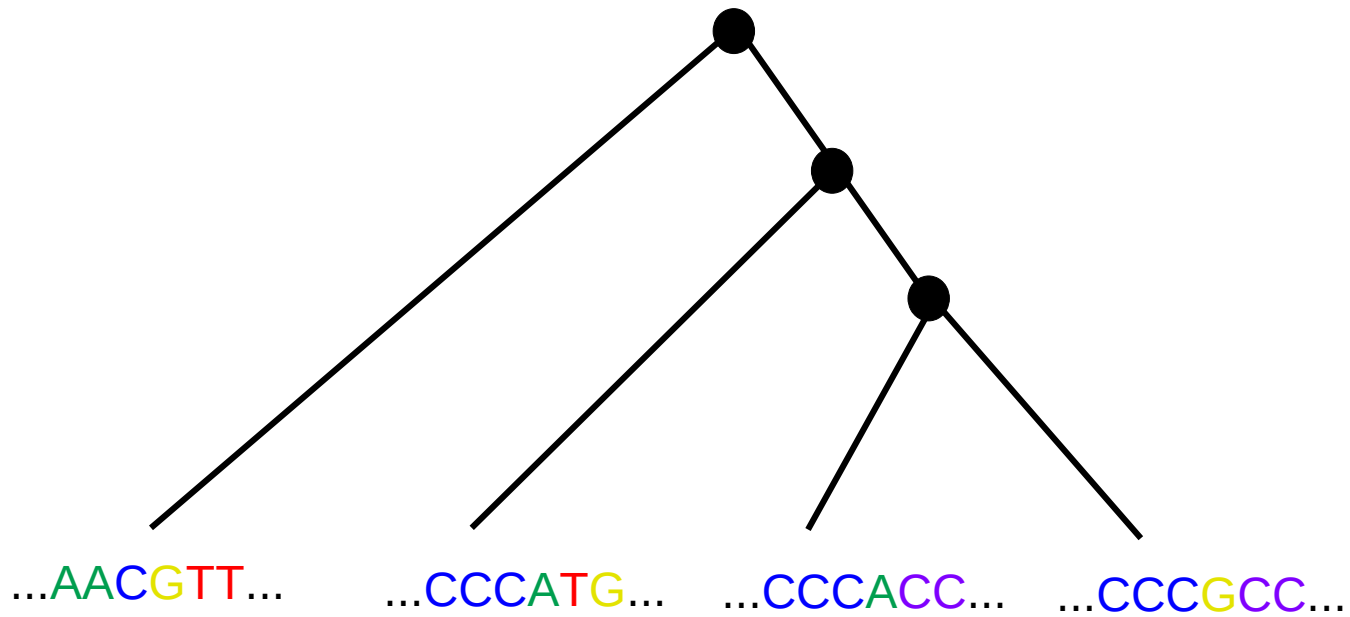


Σήμερα



Εξαγωγή DNA

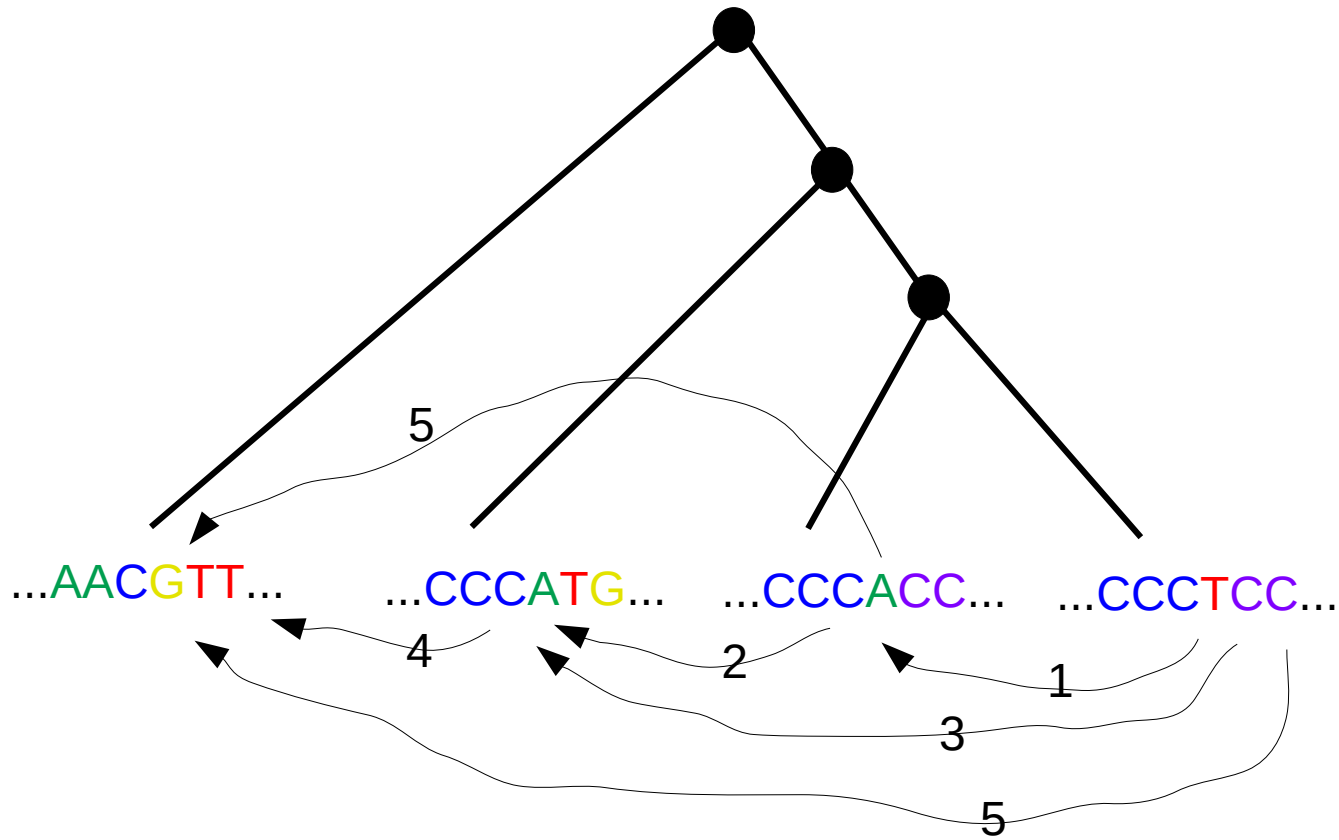
Σήμερα



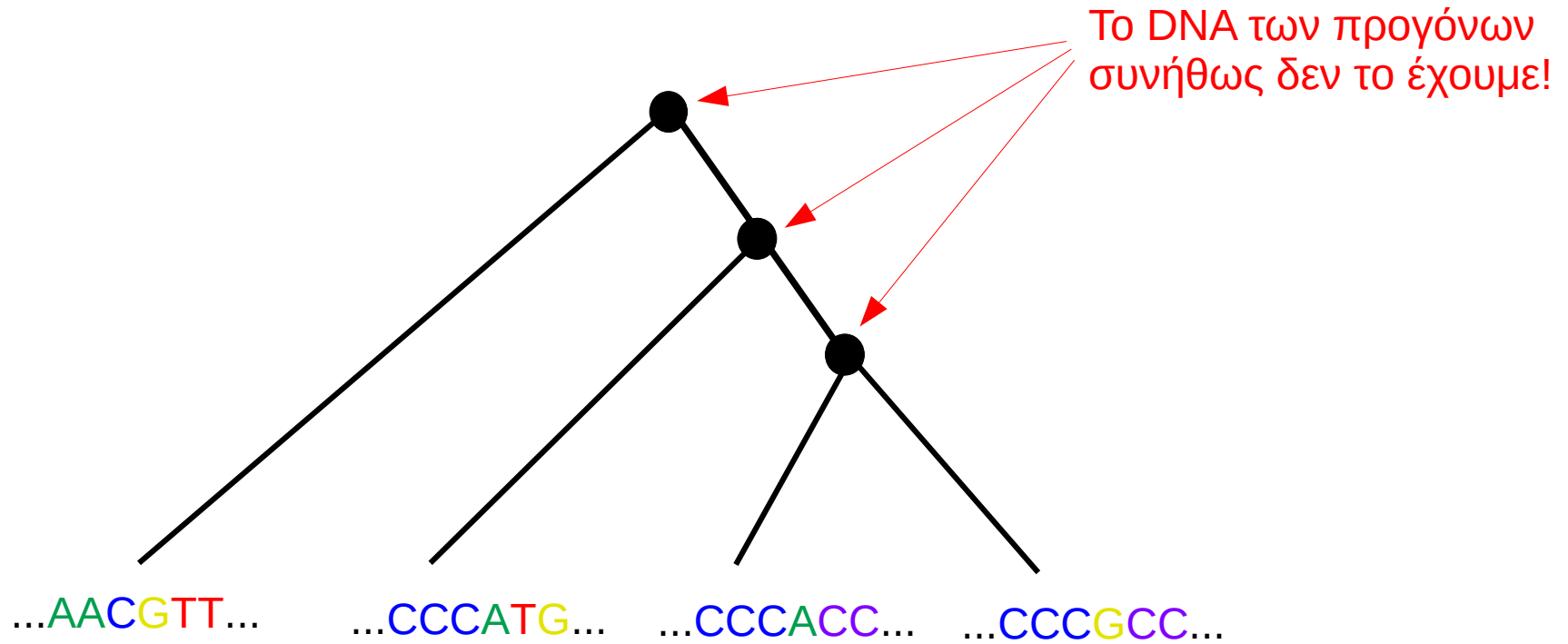
Τι παρατηρείτε;



Σήμερα



Σήμερα



Δομή Ομιλίας

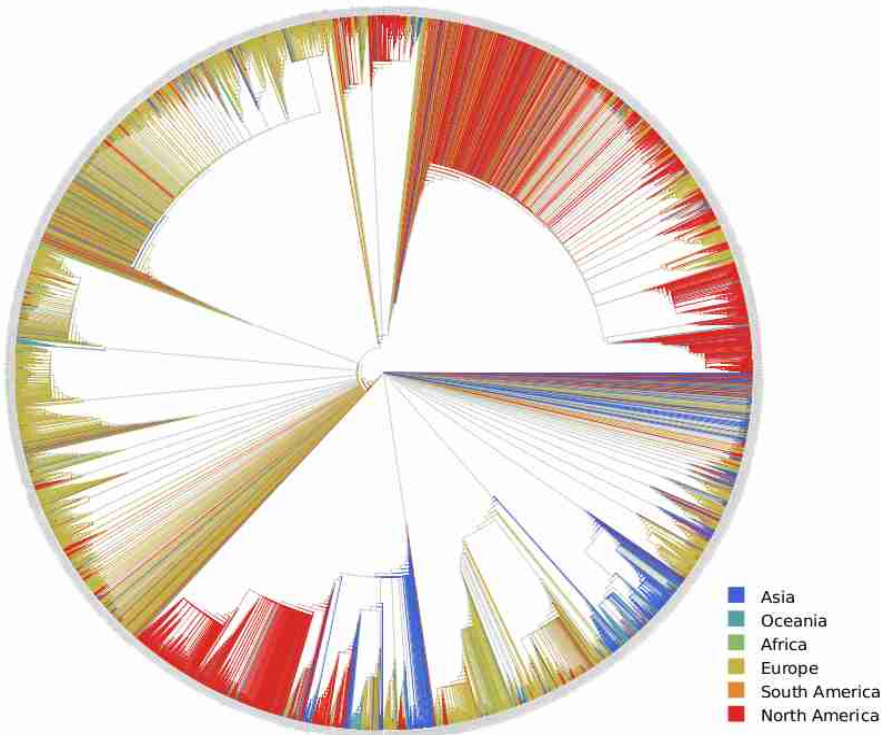
- Τι υπολογίζει το RAxML;
- **Εφαρμογές**
- Υπολογιστική Πολυπλοκότητα
- Γιατί ανοιχτός κώδικας;
- Προβλήματα και Προκλήσεις

Ο κορονοϊός

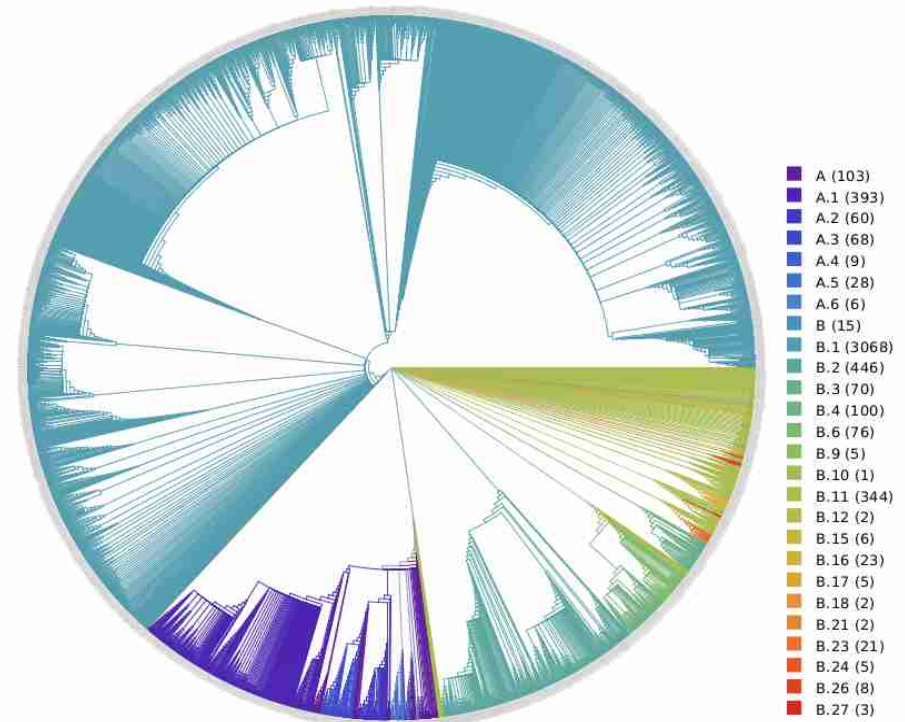
Phylogenetic analysis of SARS-CoV-2 data is difficult

Benoit Morel^{1,3}, Pierre Barbera^{4,5}, Lucas Czech³, Ben Bettisworth¹, Lukas Hübner^{1,2}, Sarah Lutteropp¹, Dora Serdari¹, Evangelia-Georgia Kostaki⁵, Ioannis Mamais⁶, Alexey M Kozlov¹, Pavlos Pavlidis², Dimitrios Paraskevis³, and Alexandros Stamatakis^{1,2}

Γεωγραφικές περιοχές



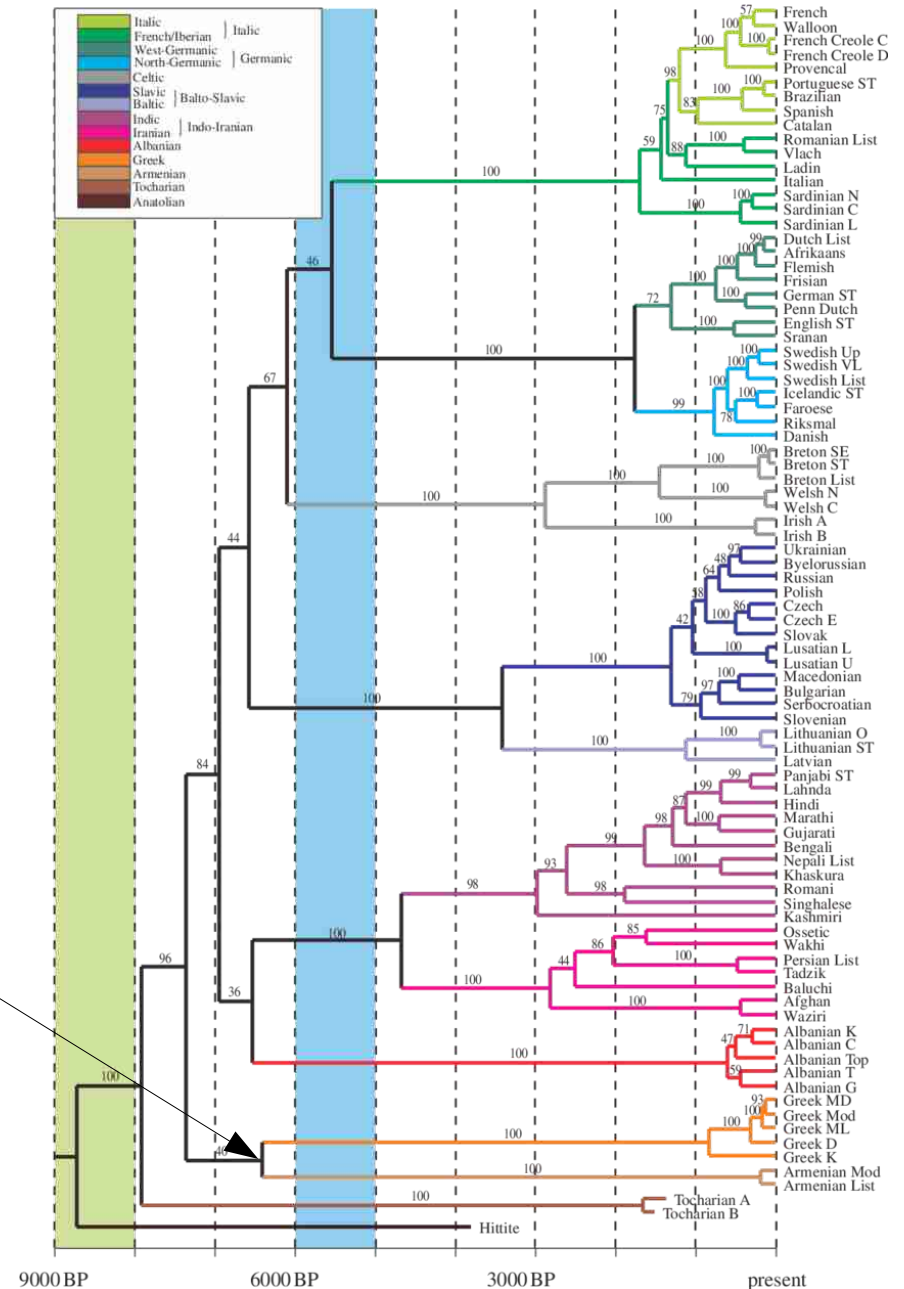
Στελέχη ιού



Εξέλιξη γλωσσών

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

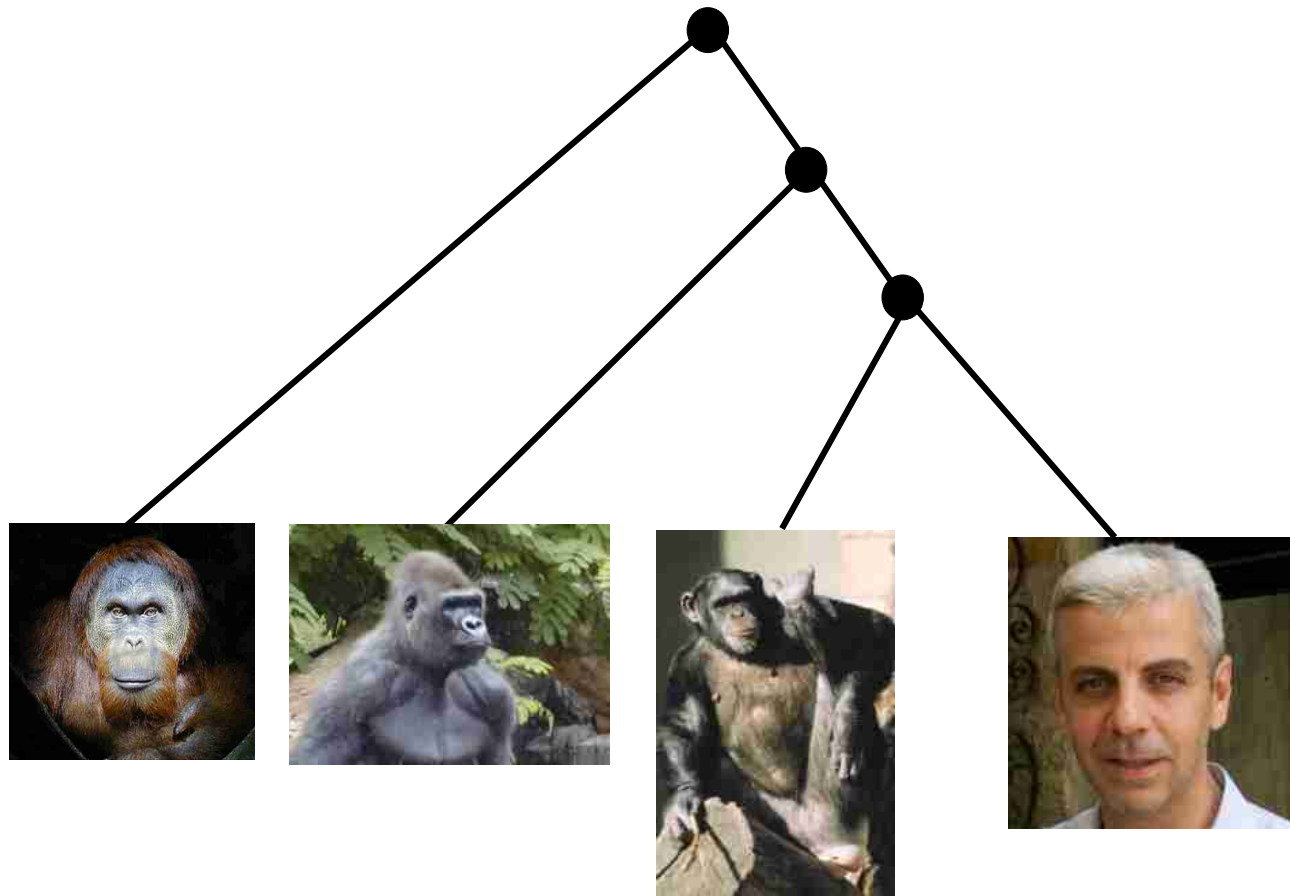
Ελληνικά και Αρμενικά



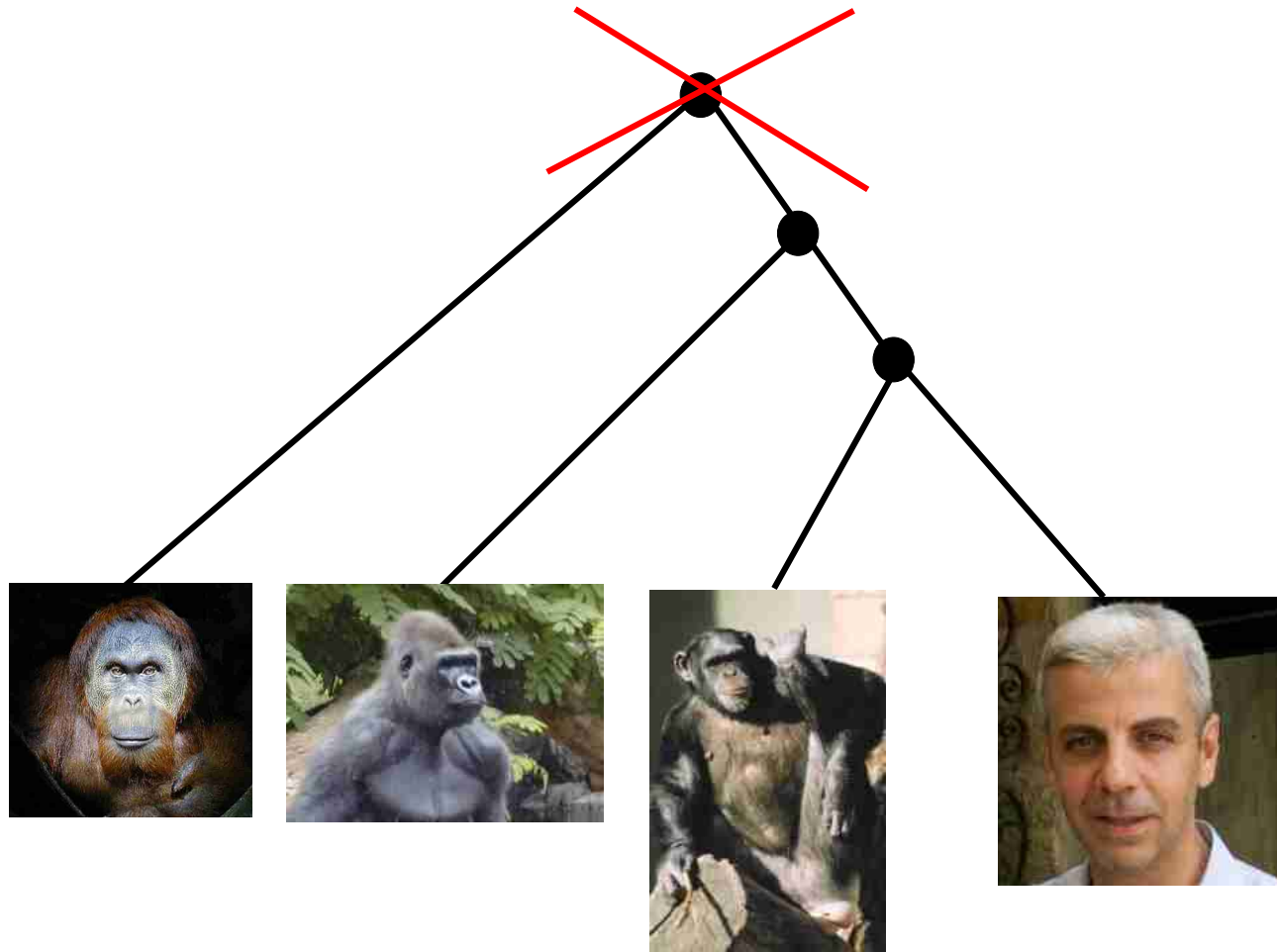
Δομή Ομιλίας

- Τι υπολογίζει το RAxML;
- Εφαρμογές
- **Υπολογιστική Πολυπλοκότητα**
- Γιατί ανοιχτός κώδικας;
- Προβλήματα και Προκλήσεις

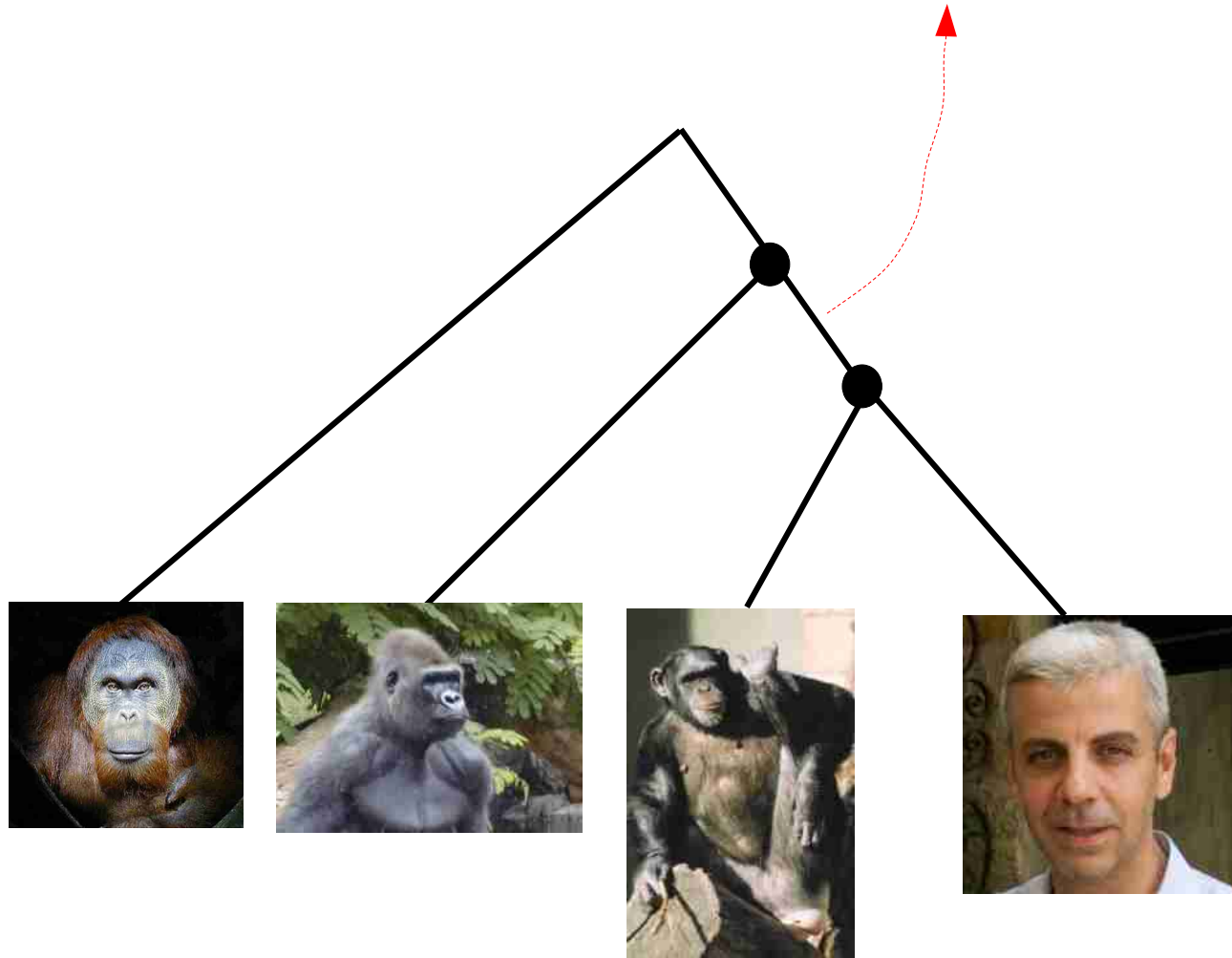
Υπολογίζοντας Δέντρα



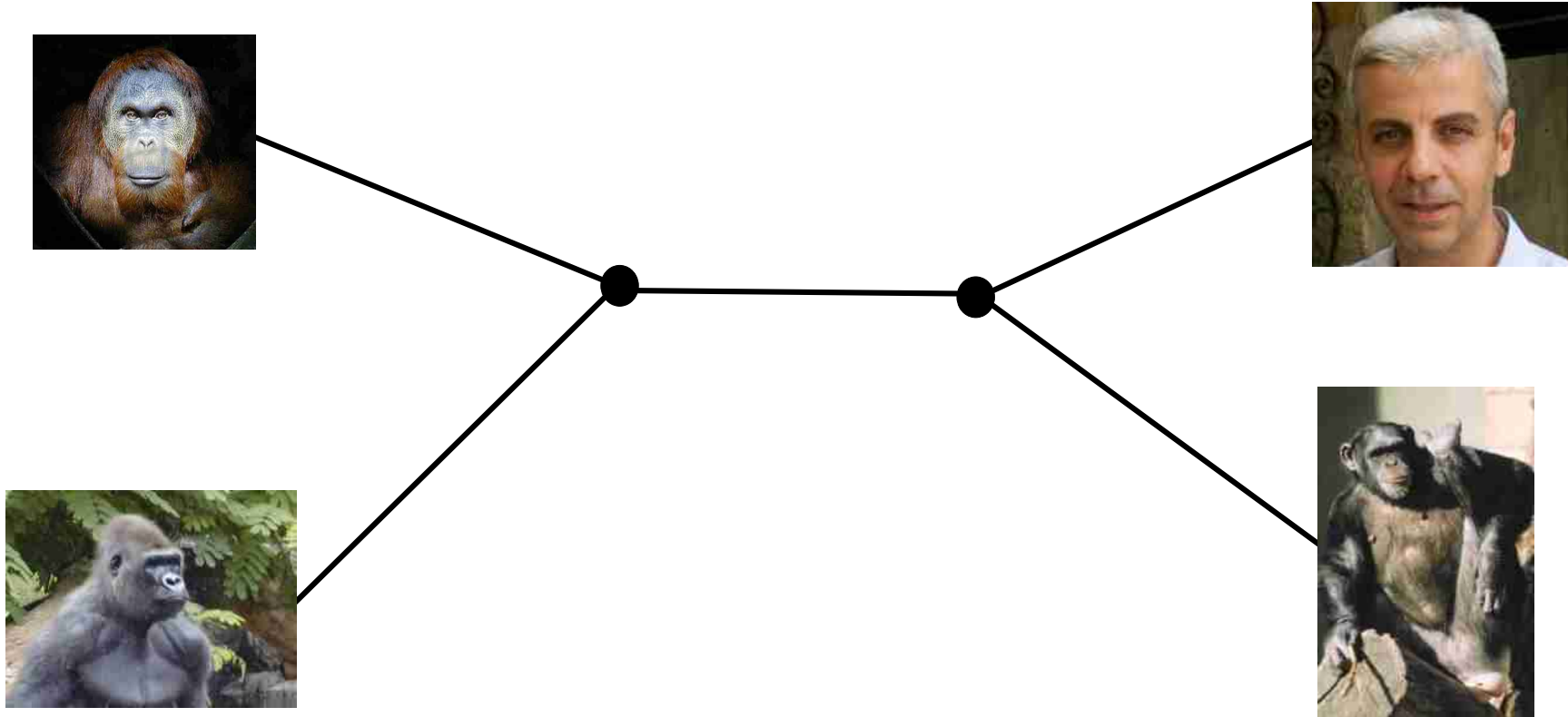
Υπολογίζοντας Δέντρα



Υπολογίζοντας Δέντρα

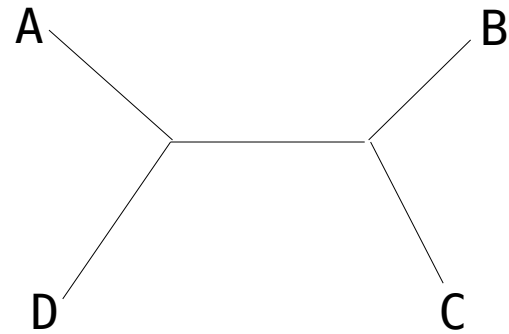
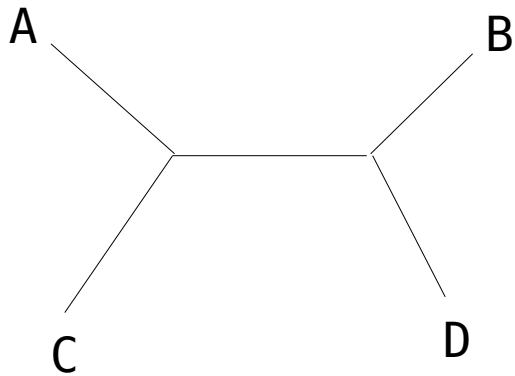
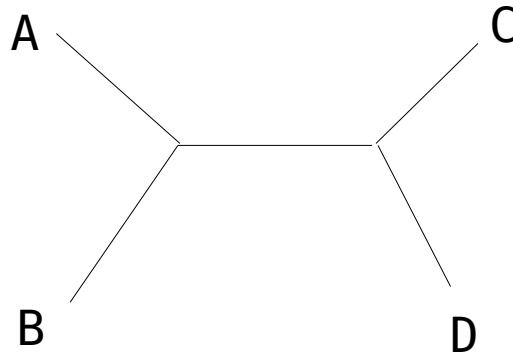


Υπολογίζοντας Δέντρα

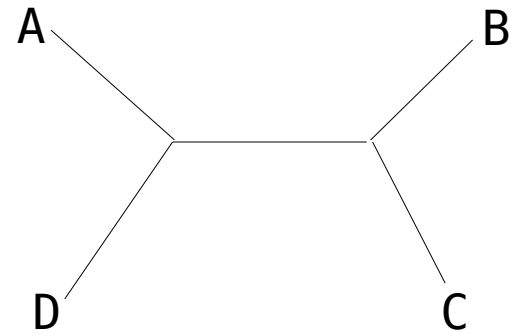
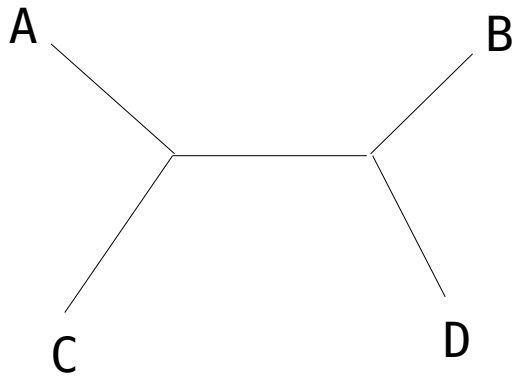
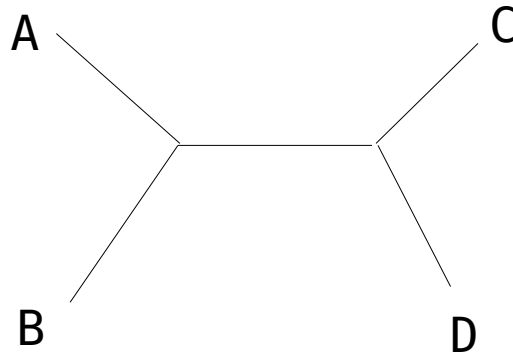


Πόσα φυλογενετικά δέντρα
υπάρχουν με 4 είδη A, B, C, D ;

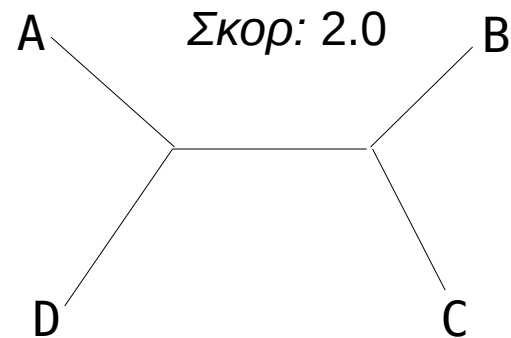
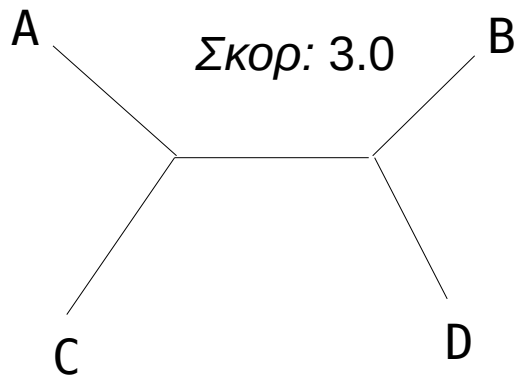
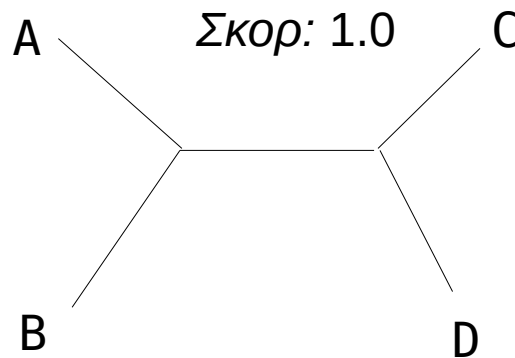
Πόσα φυλογενετικά δέντρα υπάρχουν με 4 είδη A, B, C, D ;



Πώς διαλέγουμε το πιο πιθανό;

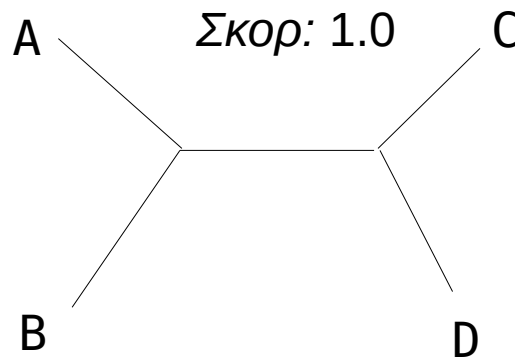


Πώς διαλέγουμε το πιο πιθανό;

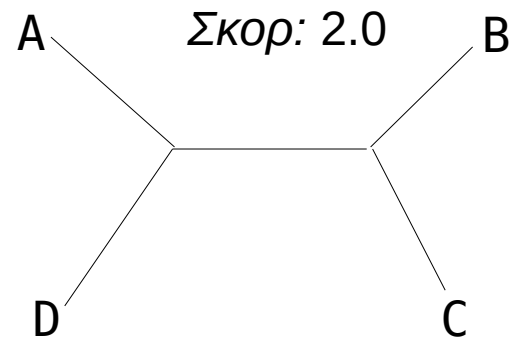
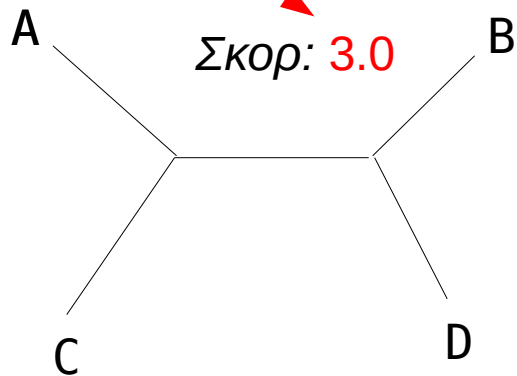


Χρειαζόμαστε κριτήρια (μαθηματικά **μοντέλα**) για να επιλέξουμε το πιο πιθανό δέντρο – δηλαδή το δέντρο που ταιριάζει/εξηγεί πιο καλά τα δεδομένα DNA που παρατηρούμε

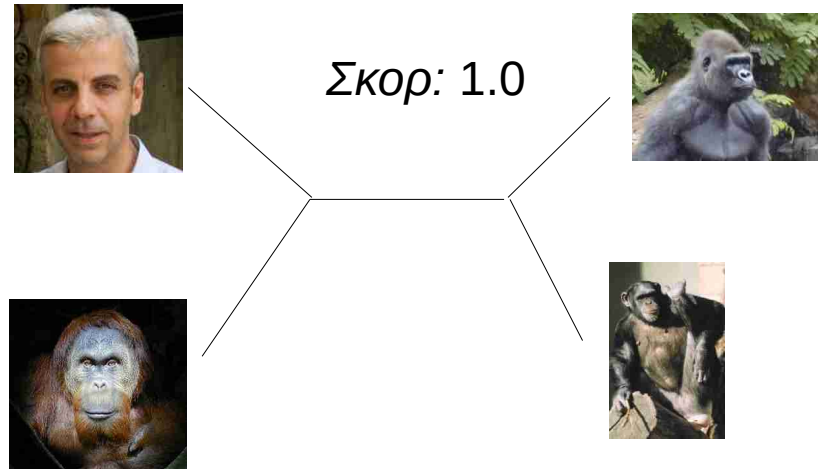
Άρα, πρέπει να βρούμε το δέντρο με το βέλτιστο σκορ



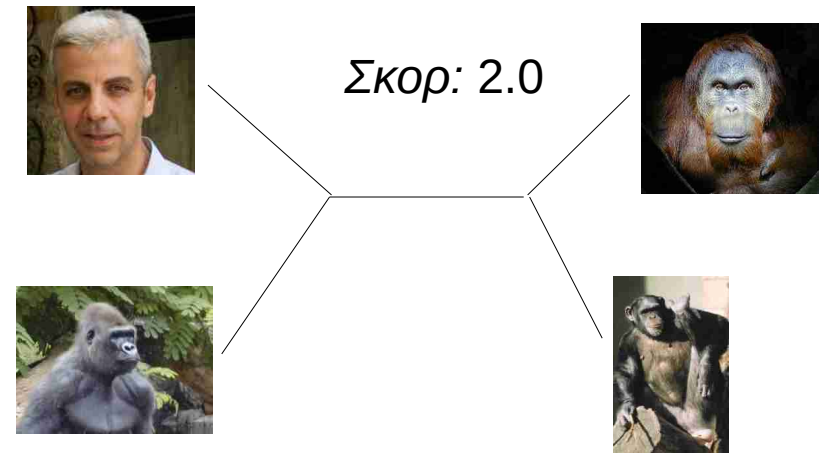
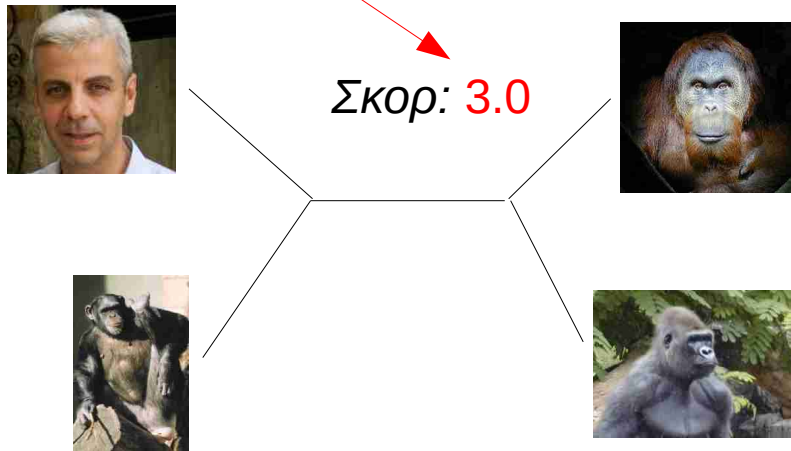
Βέλτιστο δέντρο



Άρα, πρέπει να βρούμε το δέντρο με το βέλτιστο σκορ

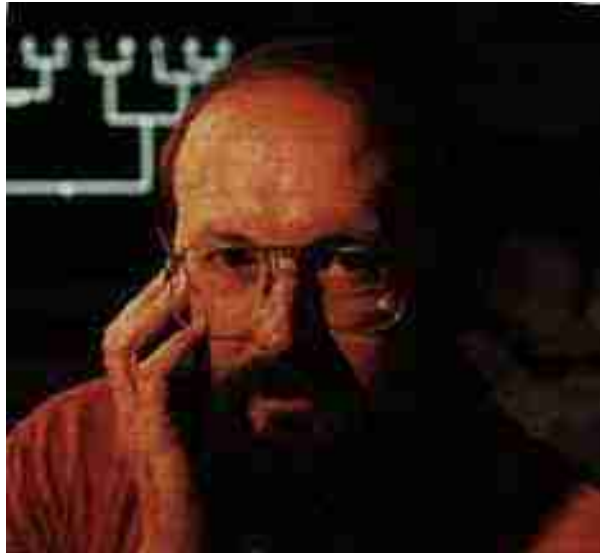


Βέλτιστο δέντρο



Το σκορ στο RAxML

- Maximum Likelihood – Μέγιστη πιθανοφάνεια
 - πόσο πιθανό είναι ένα συγκεκριμένο δέντρο να δημιουργήσει τις αλληλουχίες που παρατηρούμε στα φύλα δεδομένου ενός στατιστικού μοντέλου εξέλιξης των αλληλουχιών;



[Published: November 1981](#)

Evolutionary trees from DNA sequences: A maximum likelihood approach

[Joseph Felsenstein](#)

[Journal of Molecular Evolution](#) 17, 368–376 (1981) | [Cite this article](#)

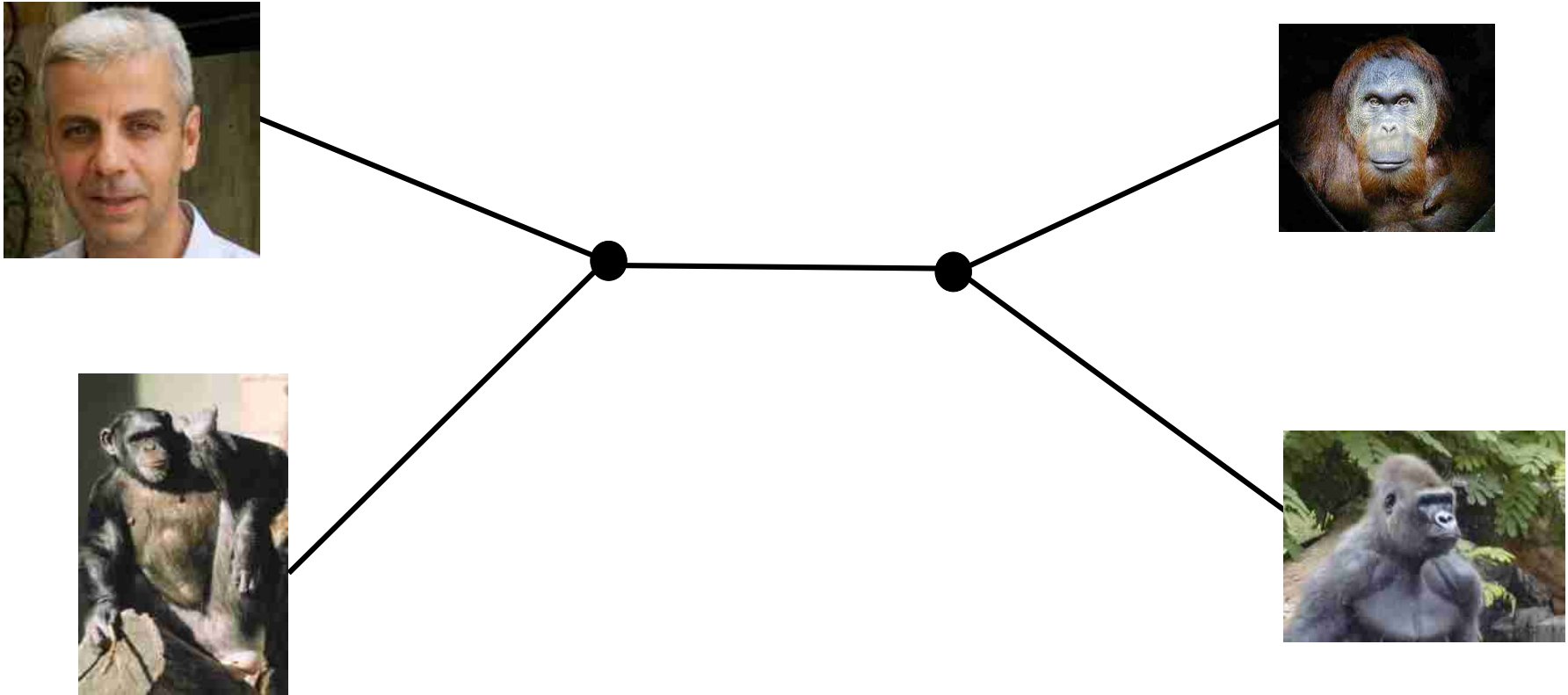
10k Accesses | 11k Citations | 54 Altmetric | [Metrics](#)

Summary

The application of maximum likelihood techniques to the estimation of evolutionary trees from nucleic acid sequence data is discussed. A computationally feasible method for finding such maximum likelihood estimates is developed, and a computer program is available. This method has advantages over the traditional parsimony algorithms, which can give misleading results if rates of evolution differ in different lineages. It also allows the testing of hypotheses about the constancy of evolutionary rates by likelihood ratio tests, and gives rough indication of the error of the estimate of the tree.

Joe Felsenstein:
αλγόριθμος δυναμικού προγραμματισμού 1981

Βέλτιστο Δέντρο - Πού είναι το DNA;

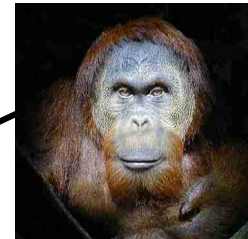


Βέλτιστο Δέντρο - Πού είναι το DNA;

...CCCGCC...



...AACGTT...



...CCCAAC...

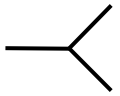


...CCCATG...

Συμπέρασμα

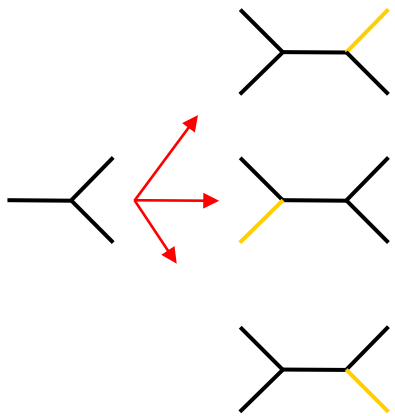
- Για 4 είδη έχουμε 3 διαφορετικά δέντρα
- Άρα, πρέπει απλά να υπολογίσουμε το σκορ μας (**στον υπολογιστή**) για αυτά τα τρία δέντρα για να βρούμε το δέντρο με το βέλτιστο σκορ
- Για 5, 6, 7, ... είδη πόσα δέντρα πρέπει να σκοράρουμε όμως – πόσα δέντρα υπάρχουν;

Ας κατασκευάσουμε δέντρα



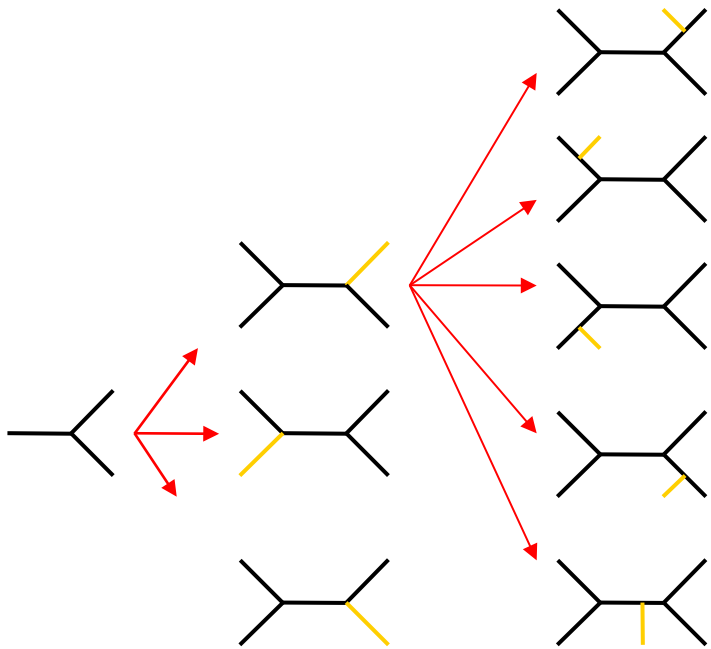
3 είδη \rightarrow 1 δέντρο

Ας κατασκευάσουμε δέντρα



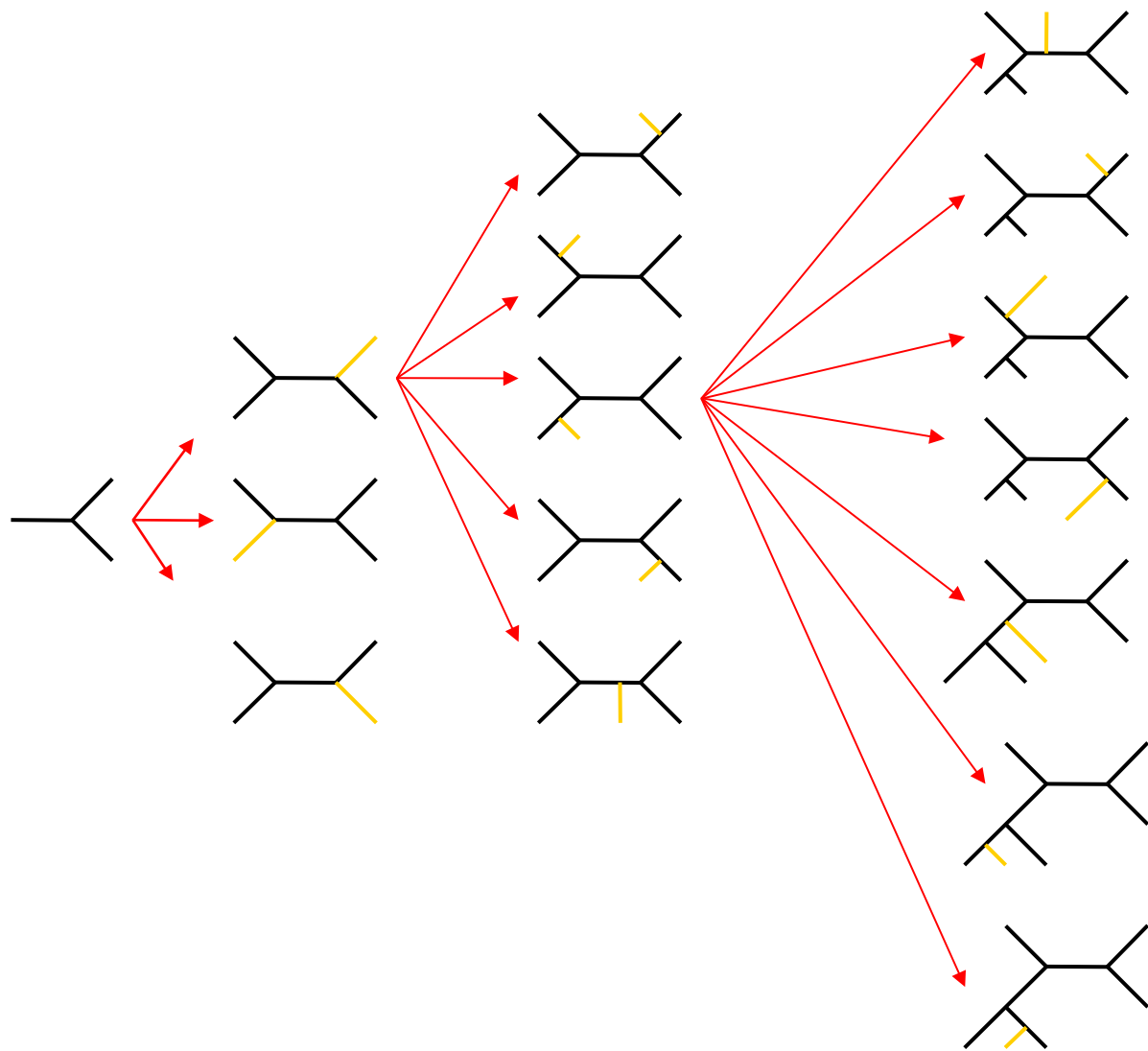
4 είδη \rightarrow 3 δέντρα

Ας κατασκευάσουμε δέντρα



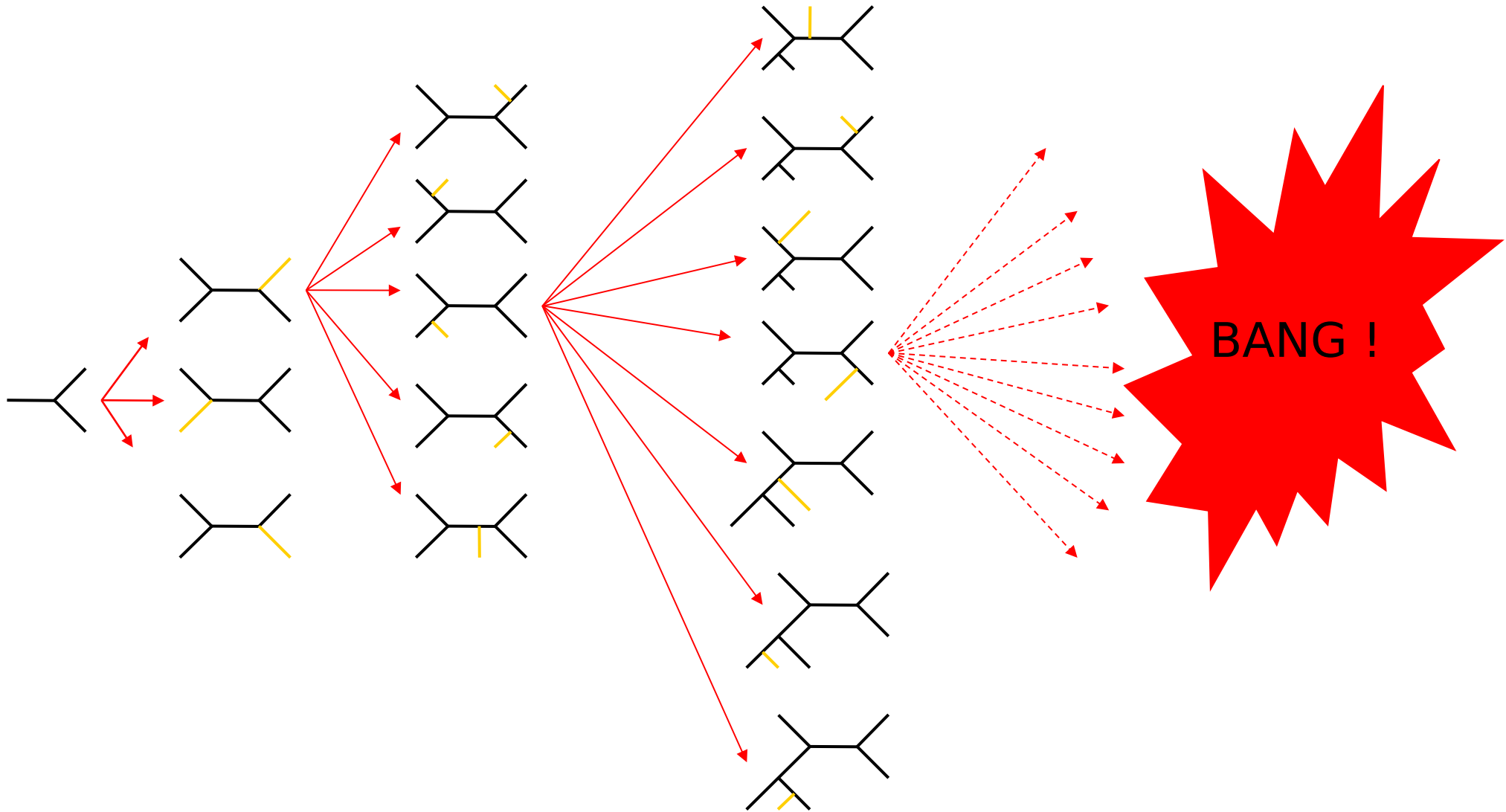
5 είδη \rightarrow 15 δέντρα

Ας κατασκευάσουμε δέντρα

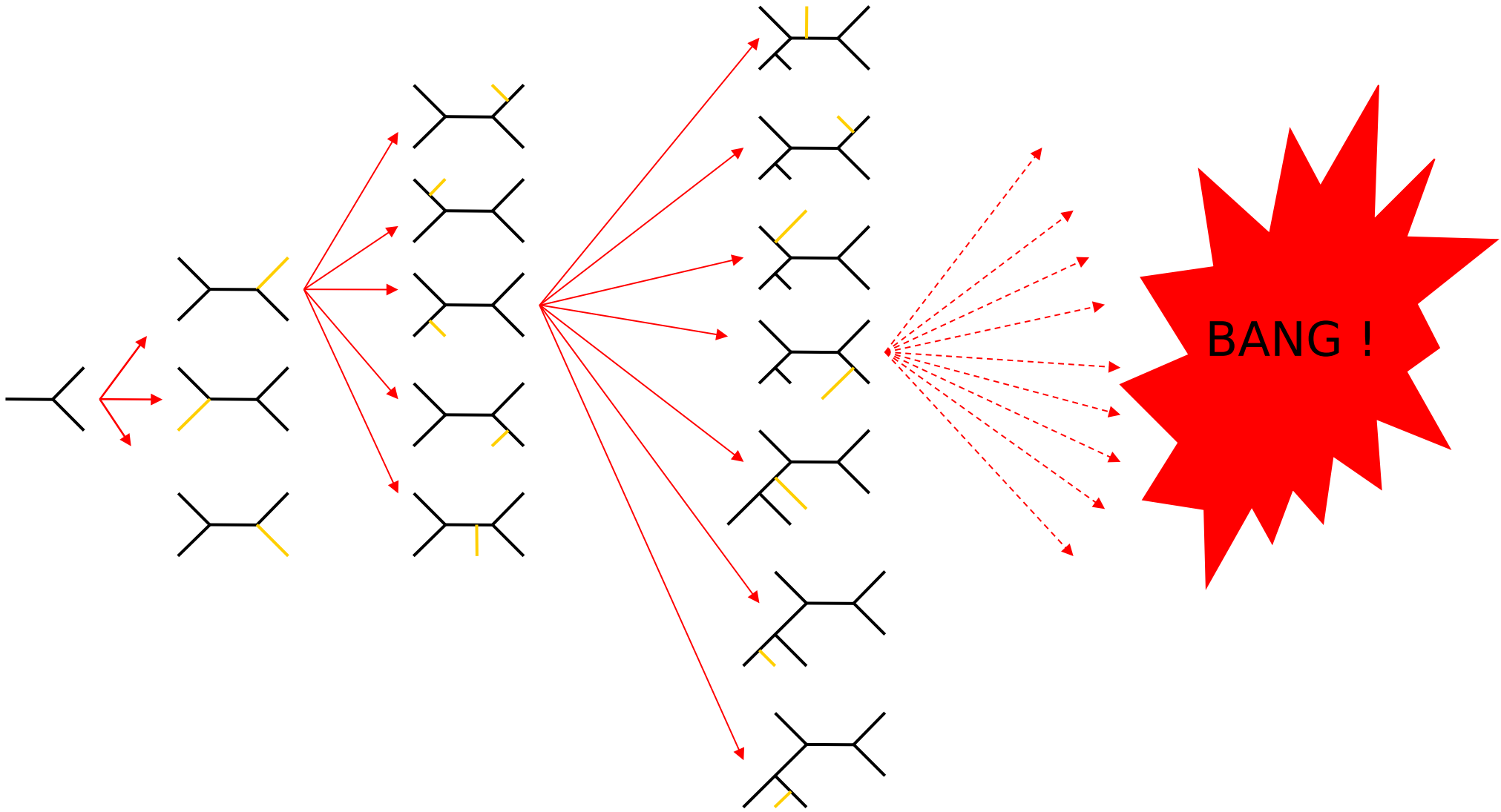


6 είδη → 105 δέντρα

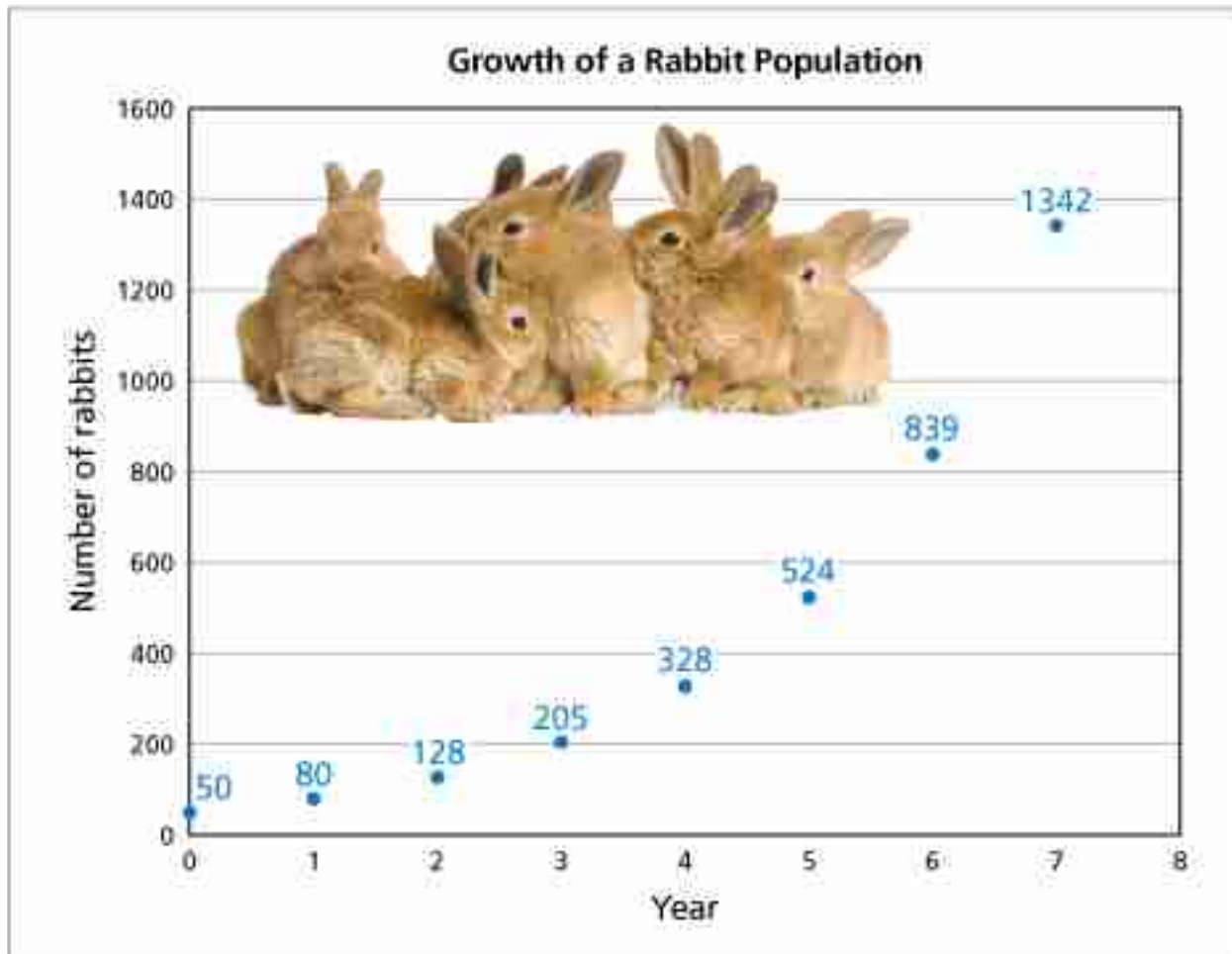
Ο αριθμός των δέντρων αυξάνεται
υπερεκθετικά με τον αριθμό των ειδών



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 – 2×10^6 δέντρα
 - 11 είδη – 3×10^7 δέντρα
 - 15 είδη – 8×10^{12} δέντρα
 - 20 είδη – 2×10^{20} δέντρα
 - 50 είδη – 3×10^{74} δέντρα

Χρόνος Υπολογισμού

- Αν υποθέσουμε, αρκετά ρεαλιστικά, ότι ο υπολογισμός του σκορ για ένα δέντρο παίρνει 1 δευτερόλεπτο
 - 10 είδη → 555 ώρες
 - 11 είδη → 8333 ώρες η 347 μέρες
 - 15 είδη → 92592592 μέρες η 260091 χρόνια
 - 20 είδη → περίπου 2×10^{15} μέρες η 5.5×10^{12} χρόνια
 - 50 είδη → περίπου 3×10^{69} μέρες η 9.5×10^{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
3439405123669055544932486165082539957795036326704947844293498853172797348177797146567175151178876396434069332458076346110734214328195049909680874027397688914704517472055543
8969396668742601477241894693212902453317334188286773194653544113302100866575081713240342647580489218623663461607955943720516395154096949806486242309796947211169665961580041
7883992232264628498942352276926391124360767508997671789683459337890742346354557193530665615379900277916265636361861974048590938234062235459769733137213659343717585590664439
6461328300113672601934068706442339489199215304385281541659630118549423634863524585746642834609062916279564926584723003608555598989761916293248140094592489899468468862322586
0170551468905649837276003927487069550463788819741699429049102853180410776516007263384721638903782270013843599599730265727198804324664312359758555217196960513921010226596324
7887830977405333131551762978152880718652603176327264828094499370456258099938053058497699570080289379808014902901005293898472279947167804821689424159118284257696464786505731
5325178302336307298251692210346584265894447464916123854689718507968172909139032182834111184821384767728316548653212317382004131990510518967022201887049585687180509590730360
6930402937216038968917605587676955382318093705826257083898387409098468656634271397500013291835105943321729879825243707508272087959859437157667660155782699660343197752623308
8989962587800628009560944416932377944955441033696586261556256010669390303203878970983673786087056641433585106111658314520424513208508589994932364831689671194951671619567622
7070906973889588855579562466641536561723549301807394004760529801721771391686788000277851966173070061284517307582503735643102065112443730825229625040453160590741343881872563
4779138306605909318802522310085340176840261401539616989192075147108033757708849740141834599753972059878682064879116064969858177601153972058498222698907181349432691801821173
3188063653910893689811714891357456680542807485170175858266639633570189354449832669762835092657922201746372190273119641751489944010079636876017826747107019945473218887832742
6088966724371574713420600093704251309893630537459784279980403132989417266492290425730958368534416215640557290282066224003863237526380910233269897838860423759625601567975262
6950798639868104294832333160267216555178120899264677804935741326387137408423885546538336158643451305439624281397279559725995110706314305992615495622958320232708057681156690
4895866105220300573725298472118747827136713666058669271094875563974858489475910819727033878284439864486743456200958161930314727345961900499318424337975243662489363321244850
5971992523668529249305346252764137853413208943128901523738092556045987090912766662329678703328882059134949580074074473143388800724532321747309659741967114441453127132790205
1010047670104350638857953478447255389801541923317027519896180635152682543173193832925891931530164130548972311128666465492971930479296432829556719092881692091042334122007454
2420499008725850462080511048758830594959903111887366685094148821725734576355233964038481318213167408359006916400053262258184783765067804451177717328658189899215358309447765
350341796875
```

Approximately 3.00 times 10⁶6328

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

Πίσω στα δέντρα – λύσεις στην πράξη



Πίσω στα δέντρα – λύσεις στην πράξη



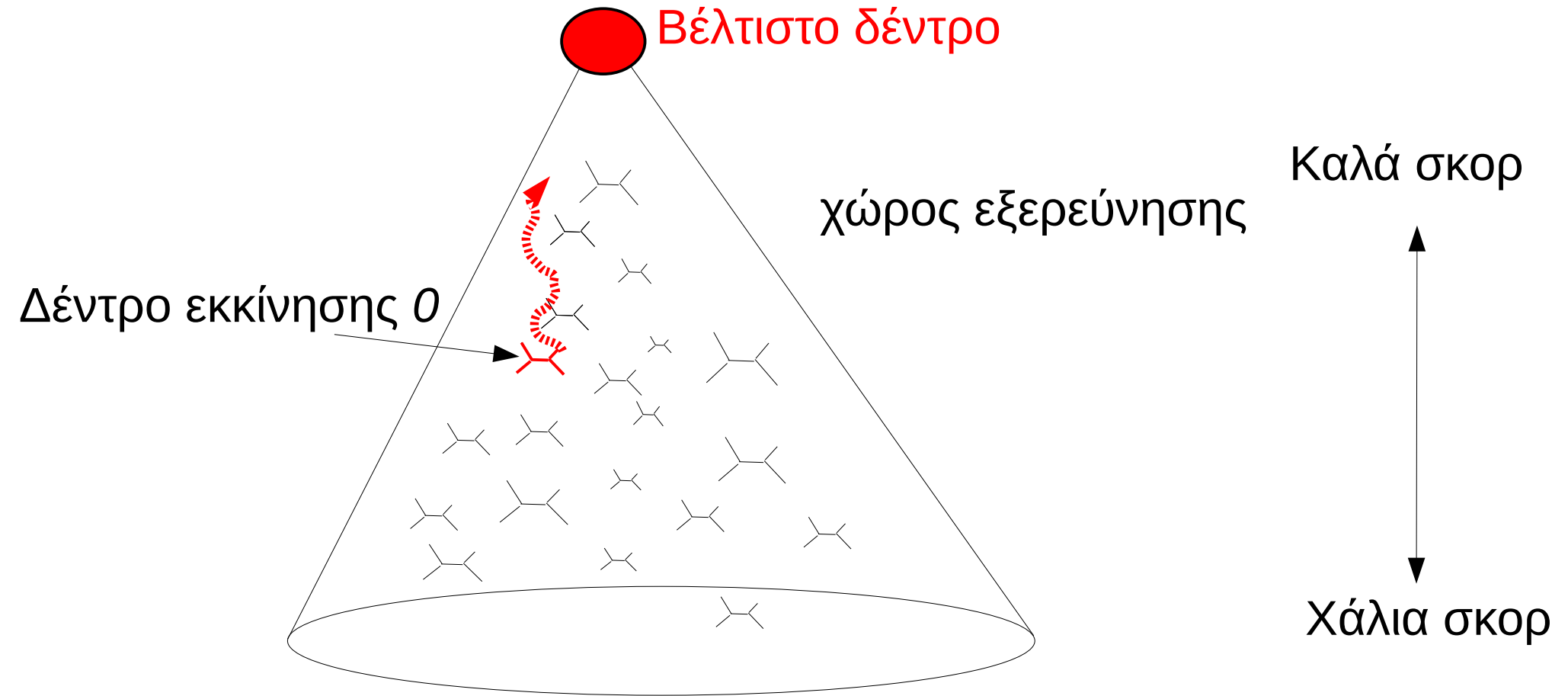
Το πρόβλημα να βρούμε το βέλτιστο δέντρο είναι NP-σκληρό για όλα τα γνωστά κριτήρια

Πίσω στα δέντρα – λύσεις στην πράξη

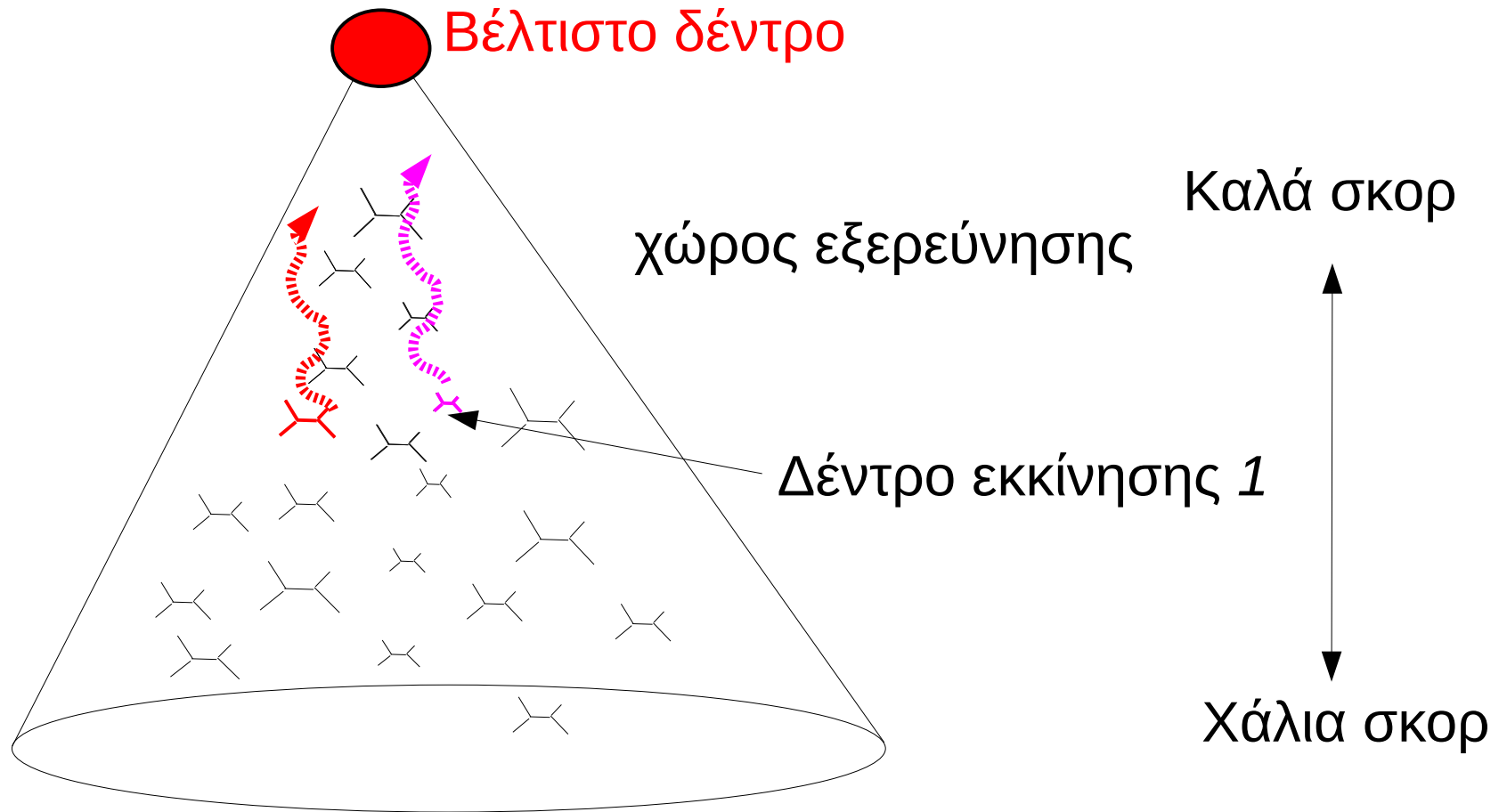


Το πρόβλημα να βρούμε το βέλτιστο δέντρο είναι NP-σκληρό για όλα τα γνωστά κριτήρια

Δέντρα εκκίνησης



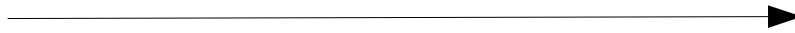
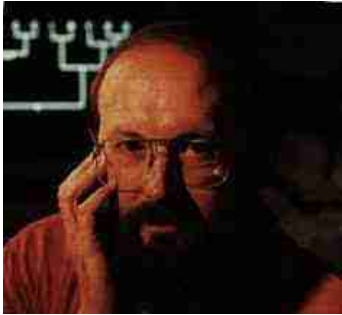
Δέντρα εκκίνησης



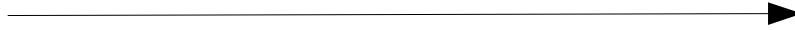
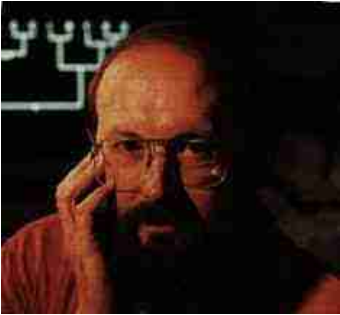
Δομή Ομιλίας

- Τι υπολογίζει το RA_xML;
- Εφαρμογές
- Υπολογιστική Πολυπλοκότητα
- **Γιατί ανοιχτός κώδικας;**
- Προβλήματα και Προκλήσεις

Πάμε πάλι στον Joe Felsenstein



Πάμε πάλι στον Joe Felsenstein



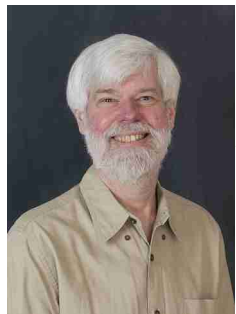
> [Comput Appl Biosci](#). 1994 Feb;10(1):41-8. doi: 10.1093/bioinformatics/10.1.41.

fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood

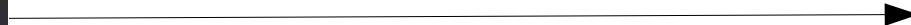
[G J Olsen](#)¹, [H Matsuda](#), [R Hagstrom](#), [R Overbeek](#)

Affiliations + expand

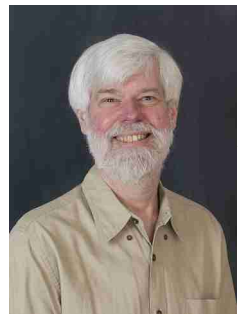
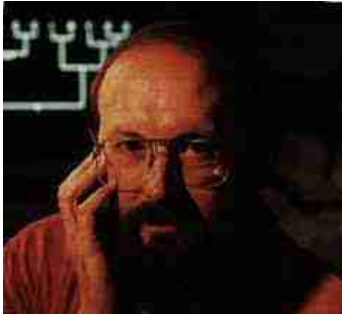
PMID: 8193955 DOI: 10.1093/bioinformatics/10.1.41



Gary Olsen



Πάμε πάλι στον Joe Felsenstein



Gary Olsen

> *Comput Appl Biosci*. 1994 Feb;10(1):41-8. doi: 10.1093/bioinformatics/10.1.41.

fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood

G J Olsen ¹, H Matsuda, R Hagstrom, R Overbeek

Affiliations + expand

PMID: 8193955 DOI: 10.1093/bioinformatics/10.1.41



JOURNAL ARTICLE

RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies

Alexandros Stamatakis [Author Notes](#)

Bioinformatics, Volume 30, Issue 9, May 2014, Pages 1312–1313,

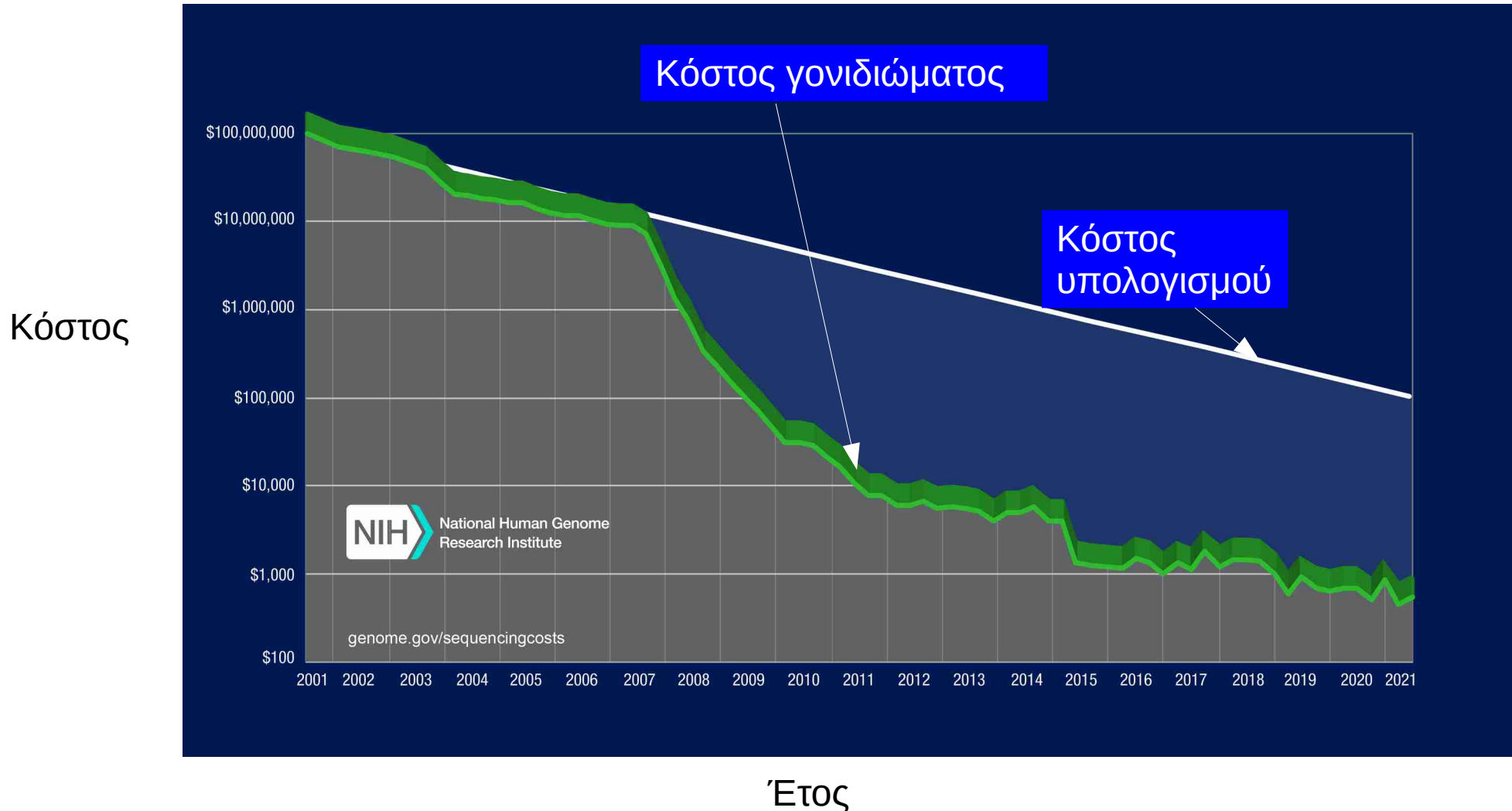
<https://doi.org/10.1093/bioinformatics/btu033>

Published: 21 January 2014 [Article history](#) ▼

Δομή Ομιλίας

- Τι υπολογίζει το RAxML;
- Εφαρμογές
- Υπολογιστική Πολυπλοκότητα
- Γιατί ανοιχτός κώδικας;
- Προβλήματα και Προκλήσεις
θα μπορούσα να μιλάω για ώρες

Από τη μία η πολυπλοκότητα, από την άλλη ο όγκος των δεδομένων



Δεδομένα με την πάροδο 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
 - 40 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

500,000 ώρες υπολογισμού;

- Χρησιμοποιούμε παράλληλους επεξεργαστές



Αξίζει να καταναλώνουμε τόση ενέργεια για ένα κ...δέντρο ;

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

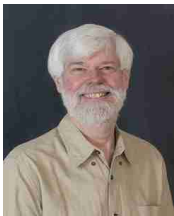
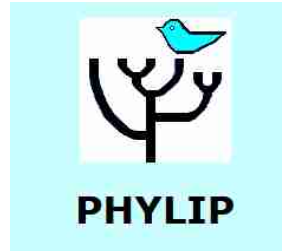
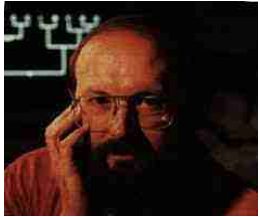
Εξέλιξη επεξεργαστών

- x86
- FPGAs
- Επεξεργαστής Playstation
- Intel Xeon PHI
- Διανυσματικές εντολές SSE3, AVX, AVX512 σε x86
- OpenMP/Pthreads: παράλληλος προγραμματισμός σε σύστημα κοινής μνήμης
- MPI: Message Passing Interface – παράλληλος προγραμματισμός
- GPUs
- ARM
- Υβριδικά συστήματα (ARM/GPU/x86)

Γιατί ξαναγράψαμε το RA_xML από την αρχή

- Προβλήματα με την συντήρηση – συνεχείς επεκτάσεις – χάναμε την μπάλα
- Διαχωρίσιμος σε ExaML (μεγάλοι παράλληλοι επεξεργαστές) και RA_xML
- Γραμμένο σε C
- Δεν είχα πια χρόνο να γράφω κώδικα
- Ξανά από την αρχή
 - Σε C++
 - Βελτιωμένο Software Engineering
 - Ενιαίο λογισμικό που τρέχει αποδοτικά σε laptop αλλά και σε μεγάλους παράλληλους υπολογιστές
 - RA_xML-NG – **N**ext **G**eneration

Πάμε πάλι στον Joe Felsenstein

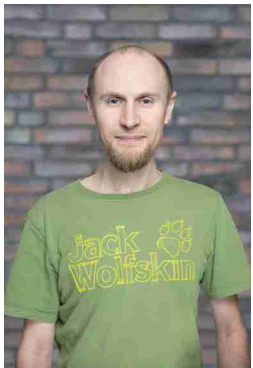


Gary Olsen

> *Comput Appl Biosci.* 1994 Feb;10(1):41-8. doi: 10.1093/bioinformatics/10.1.41.
fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood
G J Olsen ¹, H Matsuda, R Hagstrom, R Overbeek
Affiliations + expand
PMID: 8193955 DOI: 10.1093/bioinformatics/10.1.41



JOURNAL ARTICLE
RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies
Alexandros Stamatakis [Author Notes](#)
Bioinformatics, Volume 30, Issue 9, May 2014, Pages 1312–1313,
<https://doi.org/10.1093/bioinformatics/btu033>
Published: 21 January 2014 [Article history](#) ▼



Alexey Kozlov

JOURNAL ARTICLE
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, Volume 35, Issue 21, November 2019, Pages 4453–4455,
<https://doi.org/10.1093/bioinformatics/btz305>
Published: 09 May 2019 [Article history](#) ▼

Next Gen Εργαλεία

- Ξαναγράψαμε πολλά εργαλεία (δικά μας και τρίτων) από την αρχή
 - RAxML-NG
 - ModelTest-NG
 - EPA-NG
 - Lagrange-NG
- Βελτιωμένη απόδοση
- Καλύτερο SW Engineering
- Καλύτερη παραλληλοποίηση
- Μεγαλύτερη αριθμητική σταθερότητα

Software Engineering

- Είναι πιο σημαντικό απ' ότι νόμιζα παλιά
- Μπορούμε να μετρήσουμε την ποιότητα ανοιχτού κωδικά σε C και C++;

SoftWipe

- Ανάπτυξη του SoftWipe - Ένα αυτοματοποιημένο εργαλείο και σημείο αναφοράς για τη **σχετική ποιοτική κατάταξη** επιστημονικού λογισμικού (και όχι μόνο)
- Κριτήρια; → cyclomatic complexity, code sanitizers, compiler warnings, code duplication, usage of assertions κτλ.
- Κατάταξη 53 εργαλείων ανοιχτού κώδικα γραμμένα σε C ή C++ από ένα ευρύ φάσμα ερευνητικών περιοχών
 - Αστροφυσική
 - Πληροφορική
 - Βιοπληροφορική

[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 11 May 2021](#)

The SoftWipe tool and benchmark for assessing coding standards adherence of scientific software

[Adrian Zapletal](#), [Dimitri Höhler](#), [Carsten Sinz](#) & [Alexandros Stamatakis](#) 

[Scientific Reports](#) **11**, Article number: 10015 (2021) | [Cite this article](#)

3414 Accesses | 1 Citations | 92 Altmetric | [Metrics](#)

Σημείο Αναφοράς

- Διαθέσιμο στο <https://github.com/adrianzap/softwipe/wiki/Code-Quality-Benchmark>

Code Quality Benchmark

angftf edited this page on Apr 28, 2022 - 54 revisions

To generate a benchmark, we have executed softwipe on a collection of programs, most of which are bioinformatics tools from the area of evolutionary biology. Some of the below tools (genesis, raxml-ng, repeatscounter, hyperphylo) have been developed in our [lab](#). You will find a table containing the code quality scores below. Note that this is subject to change as we are refining our scoring criteria and including more tools.

Softwipe scores for each category are assigned such that the "best" program in each category that is not an outlier obtains a 10 out of 10 score, and the "worst" program in each category that is not an outlier is assigned a 0 out of 10 score. An outlier is defined to be a value that lies outside of [Tukey's fences](#).

All code quality categories use relative scores. For instance, we calculate the number of compiler warnings per total Lines Of Code (LOC). Hence, we can use those relative scores to compare and rank the different programs in our benchmark. The overall score that is used for our ranking is simply the average over all score categories. You can find a detailed description of the scoring categories and the tools included in our benchmark below.

program	overall	relative score	compiler_and_sanitizer	assertions	cppcheck	clang_tidy	cyi
genesis-0.24.0	9.0	9.1	9.9	8.7	8.4	9.2	9.0
fastspar	8.3	8.6	9.6	2.0	9.9	9.9	8.8
axe-0.3.3	7.6	7.6	9.4	1.2	6.6	9.3	6.2
pstl	7.5	7.1	10.0	0.4	8.0	5.6	9.3
raxml-ng_v1.0.1	7.5	7.8	9.9	4.2	6.6	9.0	7.9

Ας δούμε το σκορ του **SoftWipe**

program name	absolute score	relative score
genesis	8.6	8.8
hyperphyllo	8.6	8.6
kahypar	8.4	8.5
candy-kingdom	8.2	8.2
bindash-1.0	8.0	7.9
fastspar	7.8	7.9
repeatscounter	7.5	7.7
axe-0.3.3	7.5	7.5
virulign-1.0.1	7.4	7.4
naf-1.1.0/unnaf	7.4	7.5
naf-1.1.0/ennaf	7.4	7.4
ExpansionHunter	7.3	7.5
glucose-3-drup	7.1	7.0
raxml-ng	7.0	7.0
dawg	6.8	6.9
ntEdit-1.2.3	6.4	6.2
defor	6.3	6.4
swarm	6.2	6.2
lemon	6.1	6.0
treerecs	6.1	6.1
IQ-TREE-2.0-rc1	6.1	5.7
BGSA_CPU-1.0	5.9	5.4
emerald	5.8	5.5
dr_sasa_n	5.7	6.0
copmem-0.2	5.7	5.7
samtools	5.6	5.6
seq-gen	5.6	5.6
dna-nn-0.1	5.3	5.2
sf	5.2	5.2
cryfa-18.06	5.1	5.1
ngsLD	5.1	5.0
HLA-LA	4.9	4.5
iqtree l.6.10	4.9	4.9
vsearch	4.6	4.6
prank	4.6	4.5
prequal	4.5	4.4
minimap	4.5	4.4
phym1	4.4	4.4
clustal	4.2	4.3
mrBayes	4.1	4.1
tcoffee	4.1	4.2
gadget	4.1	4.0
crisflash	4.0	4.0
PopLDdecay	3.8	3.8
celloal	3.8	3.6
bpp	3.8	3.6
ms	3.7	3.7
mafft	3.3	3.1
athena	2.9	2.8
covid-sim-0.13.0	2.5	2.4
indelible	1.4	1.0

SoftWipe

program name	absolute score	relative score
genesis	8.6	8.8
hyperphylo	8.6	8.6
kahypar	8.4	8.5
candy-kingdom	8.2	8.2
bindash-1.0	8.0	7.9
fastspar	7.8	7.9
repeatscounter	7.5	7.7
axe-0.3.3	7.5	7.5
virulign-1.0.1	7.4	7.4
naf-1.1.0/unnaf	7.4	7.5
naf-1.1.0/ennaf	7.4	7.4
ExpansionHunter	7.3	7.5
glucose-3-drup	7.1	7.0
raxml-ng	7.0	7.0
dawg	6.8	6.9
ntEdit-1.2.3	6.4	6.2
defor	6.3	6.4
swarm	6.2	6.2
lemon	6.1	6.0
treerecs	6.1	6.1
IQ-TREE-2.0-rc1	6.1	5.7
BGSA_CPU-1.0	5.9	5.4
emerald	5.8	5.5
dr_sasa_n	5.7	6.0
copmem-0.2	5.7	5.7
samttools	5.6	5.6
seq-gen	5.6	5.6
dna-nn-0.1	5.3	5.2
sf	5.2	5.2
cryfa-18.06	5.1	5.1
ngsLD	5.1	5.0
HLA-1.A	4.9	4.5
iqtree1.6.10	4.9	4.9
vsearch	4.6	4.6
prank	4.6	4.5
prequal	4.5	4.4
minimap	4.5	4.4
phym1	4.4	4.4
clustal	4.2	4.3
mrBayes	4.1	4.1
tcoffee	4.1	4.2
gadget	4.1	4.0
crisflash	4.0	4.0
PopLDdecay	3.8	3.8
cellcoal	3.8	3.6
bpp	3.8	3.6
ms	3.7	3.7
mafft	3.3	3.1
athena	2.9	2.8
covid-sim-0.13.0	2.5	2.4
indelible	1.4	1.0

SoftWipe

Το απόλυτο σκορ δεν αλλάζει καθώς προσθέτουμε εργαλεία / λογισμικά

SoftWipe

program name	absolute score	relative score
genesis	8.6	8.8
hyperphylo	8.6	8.6
kahypar	8.4	8.5
candy-kingdom	8.2	8.2
bindash-1.0	8.0	7.9
fastspar	7.8	7.9
repeatscounter	7.5	7.7
axe-0.3.3	7.5	7.5
virulign-1.0.1	7.4	7.4
naf-1.1.0/unnaf	7.4	7.5
naf-1.1.0/ennaf	7.4	7.4
ExpansionHunter	7.3	7.5
glucose-3-drup	7.1	7.0
raxml-ng	7.0	7.0
dawg	6.8	6.9
ntEdit-1.2.3	6.4	6.2
defor	6.3	6.4
swarm	6.2	6.2
lemon	6.1	6.0
treerecs	6.1	6.1
IQ-TREE-2.0-rc1	6.1	5.7
BGSA_CPU-1.0	5.9	5.4
emerald	5.8	5.5
dr_sasa_n	5.7	6.0
copmem-0.2	5.7	5.7
samttools	5.6	5.6
seq-gen	5.6	5.6
dna-nn-0.1	5.3	5.2
sf	5.2	5.2
cryfa-18.06	5.1	5.1
ngsLD	5.1	5.0
HLA-1.A	4.9	4.5
iqtree1.6.10	4.9	4.9
vsearch	4.6	4.6
prank	4.6	4.5
prequal	4.5	4.4
minimap	4.5	4.4
phym1	4.4	4.4
clustal	4.2	4.3
mrBayes	4.1	4.1
tcoffee	4.1	4.2
gadget	4.1	4.0
crisflash	4.0	4.0
PopLDdecay	3.8	3.8
cellcoal	3.8	3.6
bpp	3.8	3.6
ms	3.7	3.7
mafft	3.3	3.1
athena	2.9	2.8
covid-sim-0.13.0	2.5	2.4
indelible	1.4	1.0

Το σχετικό σκορ αλλάζει καθώς προσθέτουμε εργαλεία

program name	absolute score	relative score
genesis	8.6	8.8
hyperphyllo	8.6	8.6
kahypar	8.4	8.5
candy-kingdom	8.2	8.2
bindash-1.0	8.0	7.9
fastspar	7.8	7.9
repeatscounter	7.5	7.7
axe-0.3.3	7.5	7.5
virulign-1.0.1	7.4	7.4
naf-1.1.0/unnaf	7.4	7.5
naf-1.1.0/ennaf	7.4	7.4
ExpansionHunter	7.3	7.5
glucose-3-drup	7.1	7.0
raxml-ng	7.0	7.0
dawg	6.8	6.9
ntEdit-1.2.3	6.4	6.2
defor	6.3	6.4
swarm	6.2	6.2
lemon	6.1	6.0
treerecs	6.1	6.1
IQ-TREE-2.0-rc1	6.1	5.7
BGSA_CPU-1.0	5.9	5.4
emeraldD	5.8	5.5
dr_sasa_n	5.7	6.0
copmem-0.2	5.7	5.7
samtools	5.6	5.6
seq-gen	5.6	5.6
dna-nn-0.1	5.3	5.2
sf	5.2	5.2
cryfa-18.06	5.1	5.1
ngsLD	5.1	5.0
HLA-1.A	4.9	4.5
iqtree1.6.10	4.9	4.9
vsearch	4.6	4.6
prank	4.6	4.5
prequal	4.5	4.4
minimap	4.5	4.4
phym1	4.4	4.4
clustal	4.2	4.3
mrBayes	4.1	4.1
tcoffee	4.1	4.2
gadget	4.1	4.0
crisflash	4.0	4.0
PopLDdecay	3.8	3.8
cellcoal	3.8	3.6
bpp	3.8	3.6
ms	3.7	3.7
mafft	3.3	3.1
athena	2.9	2.8
covid-sim-0.13.0	2.5	2.4
indelible	1.4	1.0

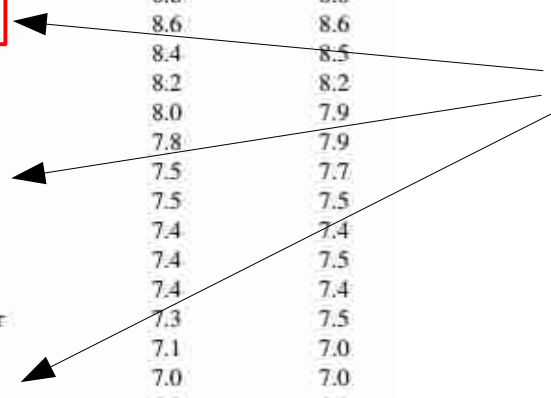
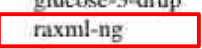
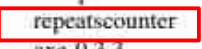
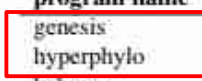
Διδακτορικοί φοιτητές
πληροφορικής

SoftWipe

program name	absolute score	relative score
genesis	8.6	8.8
hyperphylo	8.6	8.6
kahypar	8.4	8.5
candy-kingdom	8.2	8.2
bindash-1.0	8.0	7.9
fastspar	7.8	7.9
repeatscounter	7.5	7.7
axe-0.3.3	7.5	7.5
virulign-1.0.1	7.4	7.4
naf-1.1.0/unnaf	7.4	7.5
naf-1.1.0/ennaf	7.4	7.4
ExpansionHunter	7.3	7.5
glucose-3-drup	7.1	7.0
raxml-ng	7.0	7.0
dawg	6.8	6.9
ntEdit-1.2.3	6.4	6.2
defor	6.3	6.4
swarm	6.2	6.2
lemon	6.1	6.0
treerecs	6.1	6.1
IQ-TREE-2.0-rc1	6.1	5.7
BGSA_CPU-1.0	5.9	5.4
emerald	5.8	5.5
dr_sasa_n	5.7	6.0
copmem-0.2	5.7	5.7
samtools	5.6	5.6
seq-gen	5.6	5.6
dna-nn-0.1	5.3	5.2
sf	5.2	5.2
cryfa-18.06	5.1	5.1
ngsLD	5.1	5.0
HLA-1.A	4.9	4.5
iqtree1.6.10	4.9	4.9
vsearch	4.6	4.6
prank	4.6	4.5
prequal	4.5	4.4
minimap	4.5	4.4
phym1	4.4	4.4
clustal	4.2	4.3
mrBayes	4.1	4.1
tcoffee	4.1	4.2
gadget	4.1	4.0
crisflash	4.0	4.0
PopLDdecay	3.8	3.8
cellcoal	3.8	3.6
bpp	3.8	3.6
ms	3.7	3.7
mafft	3.3	3.1
athena	2.9	2.8
covid-sim-0.13.0	2.5	2.4
indelible	1.4	1.0

Το εργαστήριό μου :-)

SoftWipe



SoftWipe

program name	absolute score	relative score
genesis	8.6	8.8
hyperphylo	8.6	8.6
kahypar	8.4	8.5
candy-kingdom	8.2	8.2
bindash-1.0	8.0	7.9
fastspar	7.8	7.9
repeatscounter	7.5	7.7
axe-0.3.3	7.5	7.5
virulign-1.0.1	7.4	7.4
naf-1.1.0/unnaf	7.4	7.5
naf-1.1.0/ennaf	7.4	7.4
ExpansionHunter	7.3	7.5
glucose-3-drup	7.1	7.0
raxml-ng	7.0	7.0
dawg	6.8	6.9
ntEdit-1.2.3	6.4	6.2
defor	6.3	6.4
swarm	6.2	6.2
lemon	6.1	6.0
treerecs	6.1	6.1
IQ-TREE-2.0-rc1	6.1	5.7
BGSA_CPU-1.0	5.9	5.4
emerald	5.8	5.5
dr_sasa_n	5.7	6.0
copmem-0.2	5.7	5.7
samtools	5.6	5.6
seq-gen	5.6	5.6
dna-nn-0.1	5.3	5.2
sf	5.2	5.2
cryfa-18.06	5.1	5.1
ngsLD	5.1	5.0
HLA-1.A	4.9	4.5
iqtree1.6.10	4.9	4.9
vsearch	4.6	4.6
prank	4.6	4.5
prequal	4.5	4.4
minimap	4.5	4.4
phym1	4.4	4.4
clustal	4.2	4.3
mrBayes	4.1	4.1
tcoffee	4.1	4.2
gadget	4.1	4.0
crisflash	4.0	4.0
PopLDdecay	3.8	3.8
celloal	3.8	3.6
bpp	3.8	3.6
ms	3.7	3.7
mafft	3.3	3.1
athena	2.9	2.8
covid-sim-0.13.0	2.5	2.4
indelible	1.4	1.0

Αστροφυσική



SoftWipe

program name	absolute score	relative score
genesis	8.6	8.8
hyperphyllo	8.6	8.6
kahypar	8.4	8.5
candy-kingdom	8.2	8.2
bindash-1.0	8.0	7.9
fastspar	7.8	7.9
repeatscounter	7.5	7.7
axe-0.3.3	7.5	7.5
virulign-1.0.1	7.4	7.4
naf-1.1.0/unnaf	7.4	7.5
naf-1.1.0/ennaf	7.4	7.4
ExpansionHunter	7.3	7.5
glucose-3-drup	7.1	7.0
raxml-ng	7.0	7.0
dawg	6.8	6.9
ntEdit-1.2.3	6.4	6.2
defor	6.3	6.4
swarm	6.2	6.2
lemon	6.1	6.0
treerecs	6.1	6.1
IQ-TREE-2.0-rc1	6.1	5.7
BGSA_CPU-1.0	5.9	5.4
emerald	5.8	5.5
dr_sasa_n	5.7	6.0
copmem-0.2	5.7	5.7
samttools	5.6	5.6
seq-gen	5.6	5.6
dna-nn-0.1	5.3	5.2
sf	5.2	5.2
cryfa-18.06	5.1	5.1
ngsLD	5.1	5.0
HLA-1.A	4.9	4.5
iqtree1.6.10	4.9	4.9
vsearch	4.6	4.6
prank	4.6	4.5
prequal	4.5	4.4
minimap	4.5	4.4
phym1	4.4	4.4
clustal	4.2	4.3
mrBayes	4.1	4.1
tcoffee	4.1	4.2
gadget	4.1	4.0
crisflash	4.0	4.0
PopLDdecay	3.8	3.8
cellcoal	3.8	3.6
bpp	3.8	3.6
ms	3.7	3.7
mafft	3.3	3.1
athena	2.9	2.8
covid-sim-0.13.0	2.5	2.4
indelible	1.4	1.0

Εργαλεία με ανάλογες λειτουργίες

dawg

seq-gen

indelible

SoftWipe

program name	absolute score	relative score
genesis	8.6	8.8
hyperphyllo	8.6	8.6
kahypar	8.4	8.5
candy-kingdom	8.2	8.2
bindash-1.0	8.0	7.9
fastspar	7.8	7.9
repeatscounter	7.5	7.7
axe-0.3.3	7.5	7.5
virulign-1.0.1	7.4	7.4
naf-1.1.0/unnaf	7.4	7.5
naf-1.1.0/ennaf	7.4	7.4
ExpansionHunter	7.3	7.5
glucose-3-drup	7.1	7.0
raxml-ng	7.0	7.0
dawg	6.8	6.9
ntEdit-1.2.3	6.4	6.2
defor	6.3	6.4
swarm	6.2	6.2
lemon	6.1	6.0
treerecs	6.1	6.1
IQ-TREE-2.0-rc1	6.1	5.7
BGSA_CPU-1.0	5.9	5.4
emerald	5.8	5.5
dr_sasa_n	5.7	6.0
copmem-0.2	5.7	5.7
samtools	5.6	5.6
seq-gen	5.6	5.6
dna-nn-0.1	5.3	5.2
sf	5.2	5.2
cryfa-18.06	5.1	5.1
ngsLD	5.1	5.0
HLA-1.A	4.9	4.5
iqtree1.6.10	4.9	4.9
vsearch	4.6	4.6
prank	4.6	4.5
prequal	4.5	4.4
minimap	4.5	4.4
phyml	4.4	4.4
clustal	4.2	4.3
mrBayes	4.1	4.1
tcoffee	4.1	4.2
gadget	4.1	4.0
crisflash	4.0	4.0
PopLDdecay	3.8	3.8
cellcoal	3.8	3.6
bpp	3.8	3.6
ms	3.7	3.7
mafft	3.3	3.1
athena	2.9	2.8
covid-sim-0.13.0	2.5	2.4
indelible	1.4	1.0

Εργαλεία με ανάλογες λειτουργίες

SoftWipe

program name	absolute score	relative score
genesis	8.6	8.8
hyperphylo	8.6	8.6
kahypar	8.4	8.5
candy-kingdom	8.2	8.2
bindash-1.0	8.0	7.9
fastspar	7.8	7.9
repeatscounter	7.5	7.7
axe-0.3.3	7.5	7.5
virulign-1.0.1	7.4	7.4
naf-1.1.0/unnaf	7.4	7.5
naf-1.1.0/ennaf	7.4	7.4
ExpansionHunter	7.3	7.5
glucose-3-drup	7.1	7.0
raxml-ng	7.0	7.0
dawg	6.8	6.9
ntEdit-1.2.3	6.4	6.2
defor	6.3	6.4
swarm	6.2	6.2
lemon	6.1	6.0
treerecs	6.1	6.1
IQ-TREE-2.0-rc1	6.1	5.7
BGSA_CPU-1.0	5.9	5.4
emeralD	5.8	5.5
dr_sasa_n	5.7	6.0
copmem-0.2	5.7	5.7
samtools	5.6	5.6
seq-gen	5.6	5.6
dna-nn-0.1	5.3	5.2
sf	5.2	5.2
cryfa-18.06	5.1	5.1
ngsLD	5.1	5.0
HLA-1.A	4.9	4.5
iqtree l.6.10	4.9	4.9
vsearch	4.6	4.6
prank	4.6	4.5
prequal	4.5	4.4
minimap	4.5	4.4
phym1	4.4	4.4
clustal	4.2	4.3
mrBayes	4.1	4.1
tcoffee	4.1	4.2
gadget	4.1	4.0
crisflash	4.0	4.0
PopLDdecay	3.8	3.8
celloal	3.8	3.6
bpp	3.8	3.6
ms	3.7	3.7
mafft	3.3	3.1
athena	2.9	2.8
covid-sim-0.13.0	2.5	2.4
indelible	1.4	1.0

NEWS WEBSITE OF THE YEAR

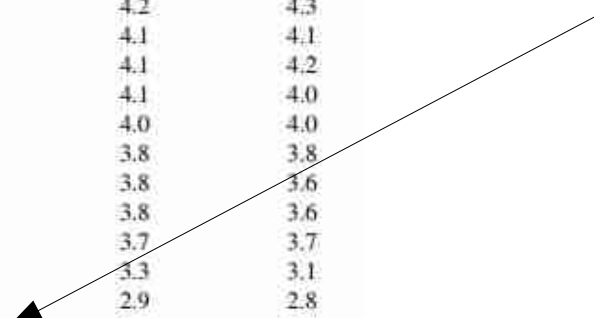
The Telegraph Coronavirus News Politics Sport Business Money Opinion Tech Life Style Travel Culture

Gadgets ▾ Innovation ▾ Big tech ▾ Start-ups ▾ Politics of tech ▾ Gaming ▾

Coding that led to lockdown was 'totally unreliable' and a 'buggy mess', say experts

The code, written by Professor Neil Ferguson and his team at Imperial College London, was impossible to read, scientists claim

Εξομοιωτής εξάπλωσης Covid



SoftWipe στην πράξη

- Οδηγεί σε υγιή ανταγωνισμό μεταξύ των μελών του εργαστηρίου → όλοι θέλουν να γράψουν τον πιο καθαρό κώδικα
- Χρησιμοποιείται από ερευνητές εντός και εκτός εργαστηρίου κατά τη διαδικασία ανάπτυξης λογισμικού
 - εντοπισμός και αποφυγή πιθανών σφαλμάτων (π.χ. bug που απέφερε εύλογα αποτελέσματα και δεν θα είχε εντοπιστεί αλλιώς)
 - απέδωσε βελτιωμένη απόδοση (διορθώθηκαν προειδοποιήσεις inlining)
 - χρησιμοποιείται στο εργαλείο για Continuous Integration
- Χρησιμοποιείται ως εργαλείο διδασκαλίας σε εργαστήρια προγραμματισμού
- SoftWipe σκορ χρησιμοποιείται ήδη από εμάς και άλλους σε υποβολές άρθρων που περιγράφουν λογισμικό Βιοπληροφορικής

Απο το Abstract: *Finally, Lagrange-NG exhibits substantially higher adherence to coding quality standards. It improves a respective software quality indicator as implemented in the SoftWipe tool from average (5.5; Lagrange) to high (7.8; Lagrange-NG)*
- **Όραμα:** Καθιέρωση δεικτών ποιότητας λογισμικού ως απαραίτητη προϋπόθεση για υποβολή άρθρου που περιγράφει λογισμικό (Βιοπληροφορικής)

Πως πουλάμε ανοιχτό κωδικά;

- Πρόγραμμα PEAR για ανάλυση αλληλουχιών DNA

JOURNAL ARTICLE

PEAR: a fast and accurate Illumina Paired-End read mergeR

Jiajie Zhang, Kassian Kobert, Tomáš Flouri ✉, Alexandros Stamatakis [Author Notes](#)

Bioinformatics, Volume 30, Issue 5, March 2014, Pages 614–620,

<https://doi.org/10.1093/bioinformatics/btt593>

Published: 18 October 2013 **Article history** ▼

- Δωρεάν για την επιστημονική κοινότητα, αλλά όχι για εταιρείες
- Έσοδα για την ομάδα μου → χρηματοδότηση ερευνάς
- Δεν ελέγχεται στο `github` ποιος το κατεβάζει

Πως πουλάμε ανοιχτό κωδικά;

- Αφαίρεση από το `github`
- Εξειδικευμένο `web-site` για ακαδημαϊκά και εμπορικά *downloads* με *IP-Tracking* και ανθρώπινο έλεγχο των *downloads*
- Με κάθε εταιρία κουραστικές διαπραγματεύσεις σχετικά με τους ορούς της σύμβασης χρήσης
- Με άλλα λόγια, θέλει αρκετή δουλειά

Θέματα που δεν κάλυψα: υπολογισμός φυλογενετικών δέντρων

- Τεχνικές εξοικονόμησης ενέργειας
- Πολλαπλοί εφιάλτες με floating-point
- NP-σκληρό πρόβλημα βέλτιστης κατανομής δεδομένων DNA σε παράλληλους υπολογιστές
- Βελτιστοποίηση παράλληλου I/O αρχείων
- Ναι, κάνουμε και εμείς machine learning τώρα
- Θέματα και λύσεις για την (παράλληλη) αναπαραγωγιμότητα
- ...

Σας ευχαριστώ πολύ



SoftWipe Criteria I

- Per criterion, calculate & assign score 0-10 such that
 - "best" program under criterion *that is not an outlier* gets 10/10
 - "worst" program under criterion *that is not an outlier* gets 0/10
 - Outliers: values that are outside of Tukey's fences.
- Then just take the unweighted average over all criteria to get an overall SoftWipe score
- We apply some corrections such that the global score does not change when more tools are added to the benchmark (details omitted)

SoftWipe Criteria II

- **compiler and sanitizer:** use `clang` compiler and count the number of warnings - we activate almost all warnings for this. Warnings are weighted - each warning has a *weight* of 1, 2, or 3, where 3 is most dangerous (**this is totally subjective**).

We also use `clang sanitizers` (ASan and UBSan) - if they yield warnings, we add them to the weighted warning sum above with weight 3. The compiler and sanitizer score is calculated from the weighted sum of warnings per total *LOC*.

- **assertions:** The count of assertions (C-Style `assert()`, `static_assert()`, or custom `assert macros`, if defined) per total *LOC*.
- **cppcheck:** #warnings found by the static code analyzer `cppcheck` per total *LOC*. `cppcheck` also categorizes warnings → analogous weighting as for compiler warnings.
- **clang-tidy:** #warnings found by the static code analyzer `clang-tidy` per total *LOC*. `clang-tidy` also categorizes warnings → analogous weighting as for compiler warnings.

SoftWipe Criteria III

- **cyclomatic complexity:** software metric to quantify the complexity/modularity of a program. We use `lizard` to assess the cyclomatic complexity of a source code.
- **lizard warnings:** Number of functions that are considered too complex, relative to the total number of functions - `lizard` considers a function as "too complex" if its cyclomatic complexity, its length, or its parameter count exceeds a certain threshold value.
- **unique rate:** amount of unique code; a higher amount of duplicated code yields a lower value. Also computed with `lizard`.
- **kwstyle:** #warnings found by the static code style analyzer `KWStyle` per total *LOC*. We configure `KWStyle` using the `KWStyle.xml` file that ships with `SoftWipe`.
- **infer:** we weight the warnings found by the static analyzer `Infer` and use the weighted warnings per *LOC* rate to calculate a score.
- **test count:** We try to relate the unit test *LOC* in with overall *LOC* count by computing: $\text{test_code_loc} / \text{overall_loc}$. The detection of unit test *LOC* should be improved – at present we interpret source files containing the keyword "test" in their path as unit test files.

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/btz025>

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, 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, btz025, <https://doi.org/10.1093/bioinformatics/btz025>

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 June 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/btz205>

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, btz205

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 1629–1638, <https://doi.org/10.1093/bioinformatics/btx025>

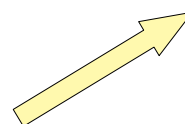
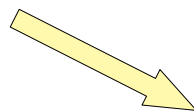
Published: 20 June 2017

1000x επιτάχυνση

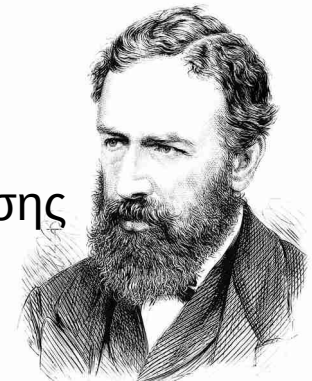
W. S. Jevons „The Coal Question“ (1865)

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



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>

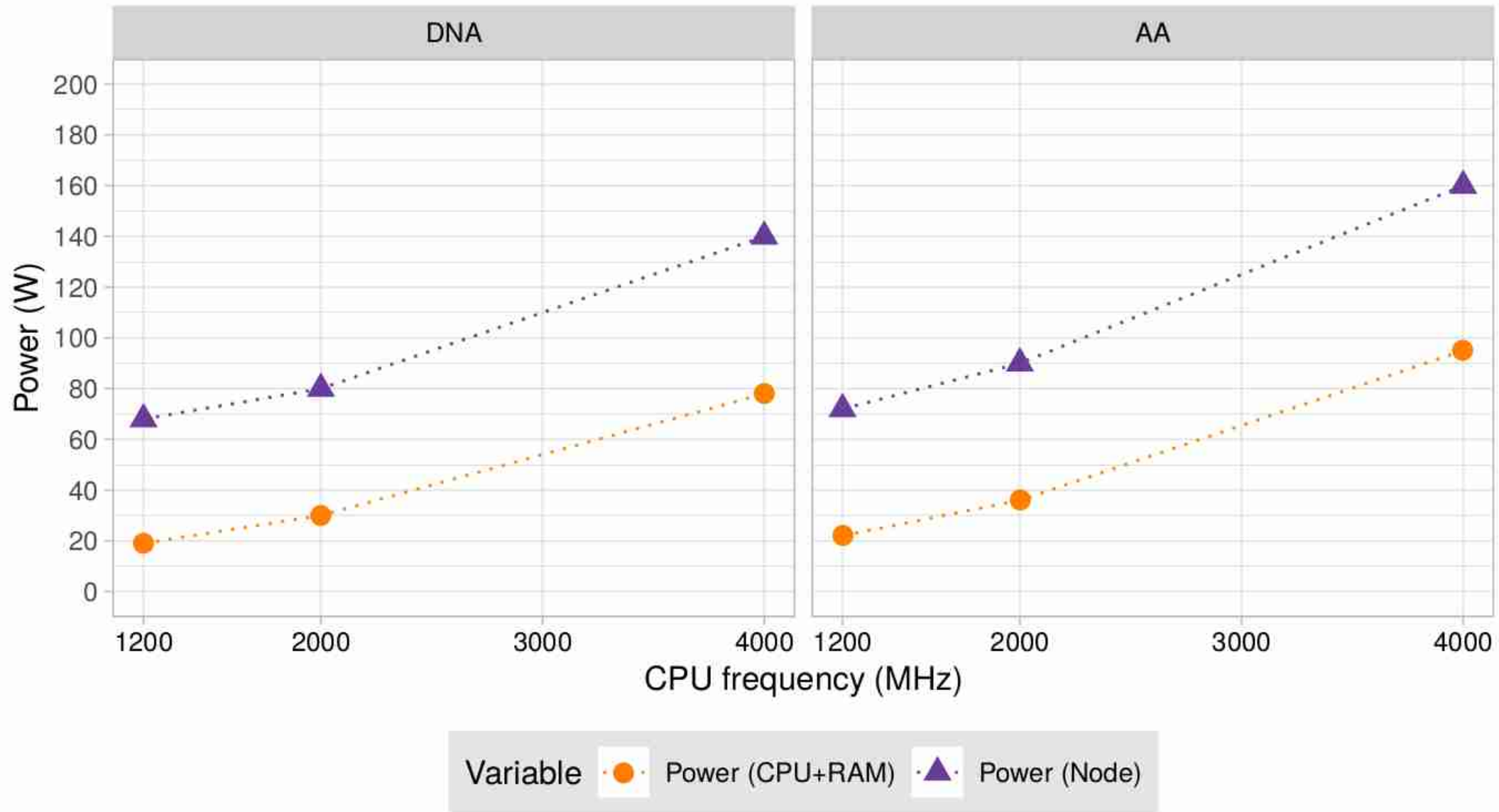
Published: 17 November 2022

Μέθοδος τεχνητής νοημοσύνης

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

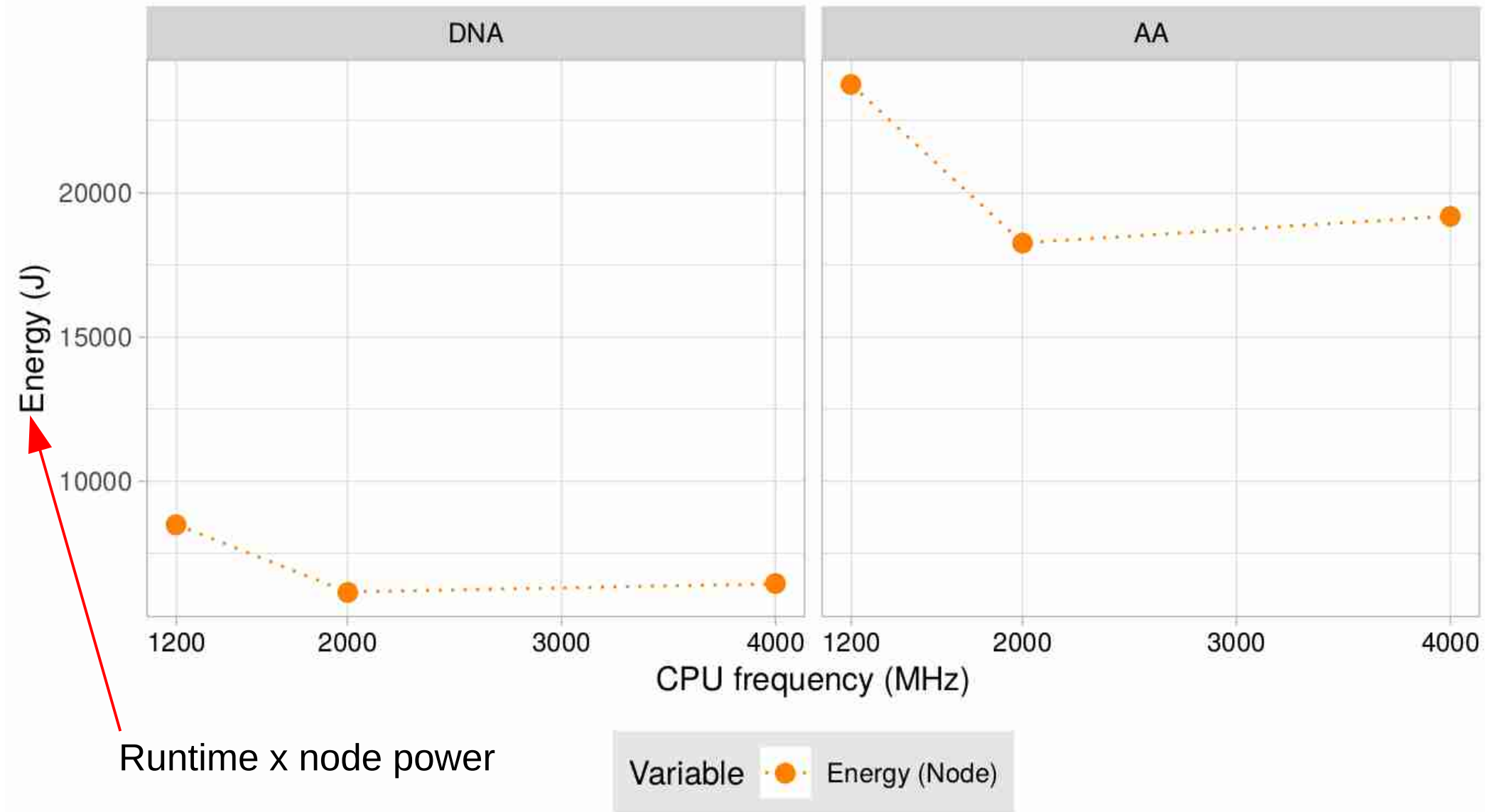
Η πρόγνωση του βαθμού δυσκολίας μας επιτρέπει να προσαρμόσουμε ανάλογα την στρατηγική ανάλυσης των δεδομένων μας

Phylogenetic Inference

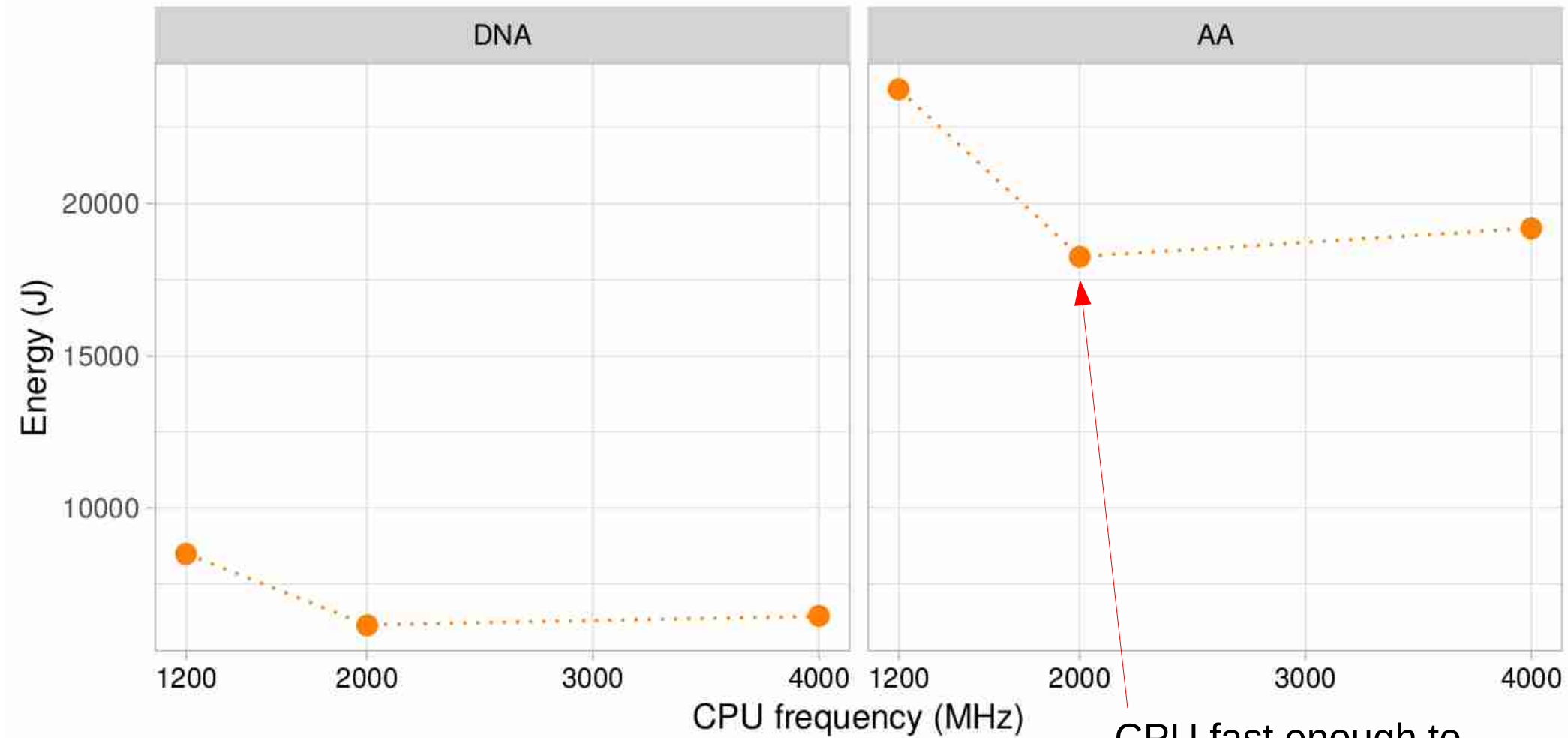


CPU frequency correlates well with power for RA_xML-NG

Phylogenetic Inference



Phylogenetic Inference



Runtime x node power

Variable  Energy (Node)

CPU fast enough to match memory access speed, everything to the right means that we are waiting for data

RAxML-NG v0.9 vs. v1.0

energy saving

