

Phylogenomics I

NEW TECHNOLOGIES FOR DEVELOPMENTAL EVOLUTIONARY STUDIES

intensive postgrad course combining theoretical and hands-on sessions

15th- 19th December 2025

Isabel Almudí
Manuel Irimia
Iker Irisarri
Christoph Liedtke
Ignacio Maeso
Marta Portela
Ana Riesgo
María Roselló
María Eleonora Rossi
Jordi Solana
Juan Tena
Aida Verdes

- . comparative transcriptomics
- . single cell transcriptomics
- . bulk transcriptomics
- . spatial transcriptomics
- . ATAC-seq technique
- . phylogenomics
- . comparative genomics
- . gene regulation
- . plastic phenotypes
- . evolutionary novelties

Iker Irisarri

mncn
museo nac de ciencias naturales

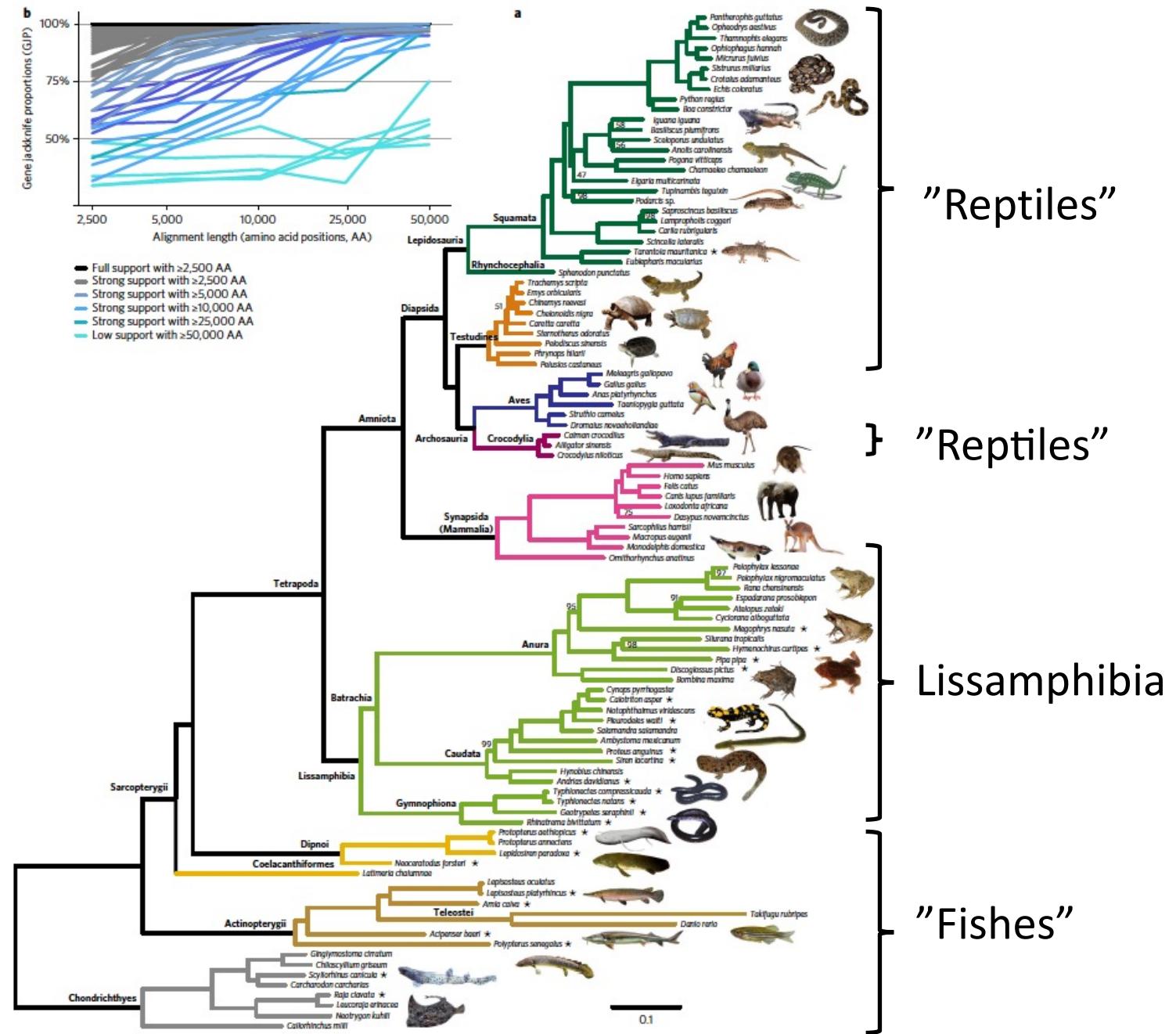
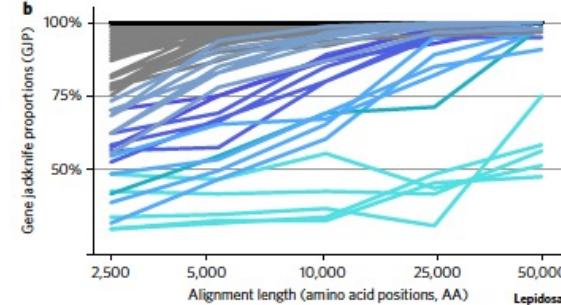


+INFO & REGISTRATION

<https://sites.google.com/view/evodevo2025/home>
application due: november 15th, 2025



Tree thinking



"Reptiles"

Polyphyletic

"Reptiles"

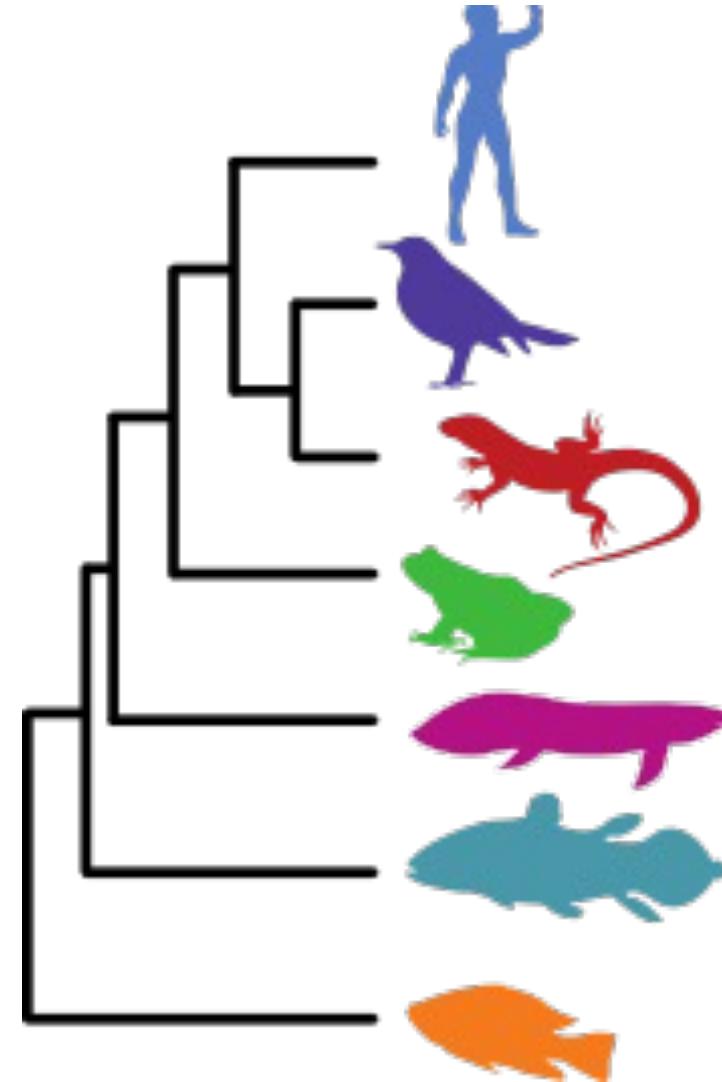
Monophyletic

Lissamphibia

Paraphyletic

"Fishes"

- Trees are bifurcating acyclical graphs
- Evolutionary relationships = nested sister groups/lineages/taxa (clades)
- Nodes are (hypothetical) ancestors
- Root provides polarity (time)
- Ancestral / derived character states



Stuff that we (unfortunately) read in scientific literature

Andreas Hejnol @Hejnol_Lab · Mar 5
The "Great Chain of Being" saying "Hello again".... 🤖

While most studies have focused on neuronal cell types in mammalian systems [13–17,18*,19], others have pioneered cell type comparison in other vertebrates [20*], chordates [21*], other bilaterians [22*,23*,24,25,26*], and several lower metazoans [27,28*].

9 9 7 32

Alejandro Damian-Serrano @antropoteuthis

Replying to @Hejnol_Lab

this one wins

Cnidarians (Coelenterates), evolutionarily the most primitive metazoans, evolved a special weaponry called cnidocysts (nematocysts). These organelles are

4:17 PM - 5 Mar 2019

Nature Plants @NaturePlants

As we all know, humans (and Arabidopsis) are early diverging, ancient, lowly, basal, primitive organisms that have stopped evolving hundreds of millions years ago. Obviously demonstrated by the below phylogenetic tree. Moss is the pinnacle of evolution.

Nice paper !

b

Stuart McDaniel @mcdaniellab · Feb 1

Bryophytes are not early diverging land plants
nph.onlinelibrary.wiley.com/doi/abs/10.1111...
[Show this thread](#)



Nature Plants @NaturePlants · 20h
New:

"Molecular landscape of etioplast inner membranes in higher plants" rdcu.be/ciYun

"Photocatalytic LPOR forms helical lattices that shape membranes for chlorophyll synthesis" rdcu.be/ciYut

"Chlorophyll biogenesis sees the light" rdcu.be/ciYuD

4 24 57

PMDelaux @PierreMarcDelau · 20h
@NaturePlants Please edit this title. Please! "higher plants", you can do better !!

2 1 29

Nature Plants @NaturePlants

Replying to @PierreMarcDelau

@RensingStefan @ReskiLab @Grandpa_Hiro @facundoromani @simonrdg @CellEvo @KlausRiede

Mea Culpe. We discourage the use of the term 'higher plants' and encourage authors about alternatives. But it is so frequently used that sometimes we miss. We will try to do better.

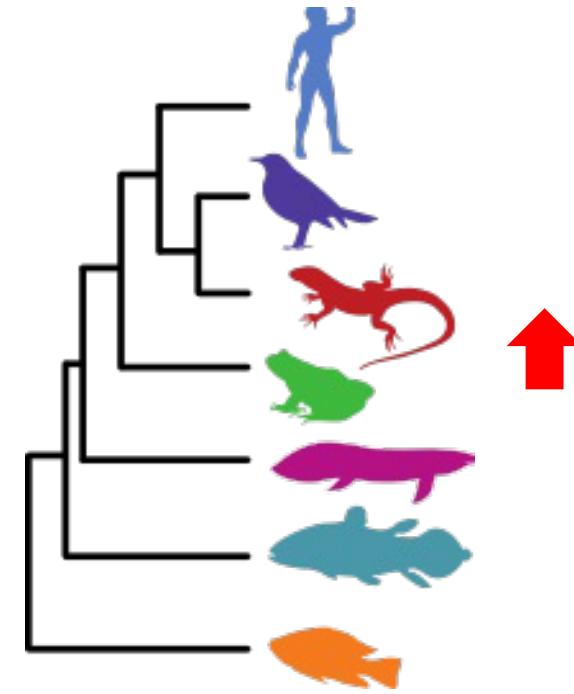
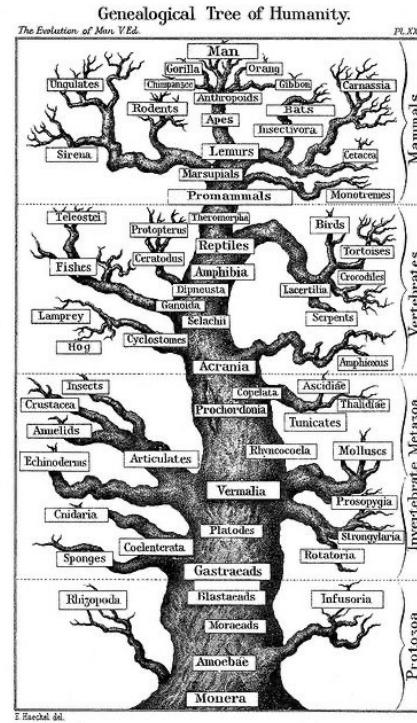
3:49 PM · Apr 20, 2021 · TweetDeck

Tree thinking is not trivial!

- All living organisms are equally evolved (same time since their common ancestor)
- No organism is “**lower/higher**” or “**ancient**”
- Species are neither “**primitive**” or “**derived**” (character states are)
- Careful with “**basal/early branching**”: it refers to proximity to the tree base (root). Often applied to extant species (wrong!) or ignoring the equally-basal sister branch
- Solution: Describe trees in terms of **sister taxa**!

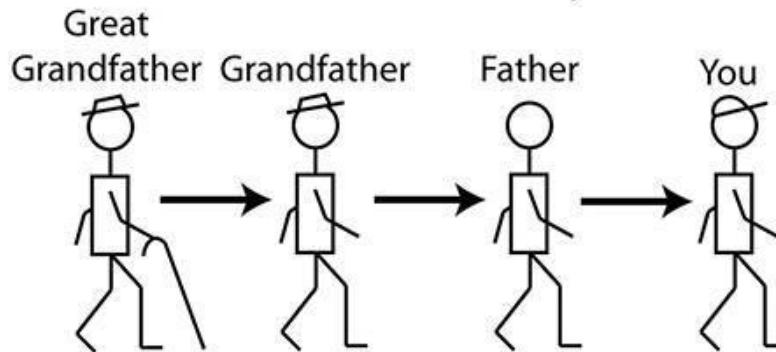
Tree thinking is not trivial!

- Watch out with linear thinking in evolution!! Directionality is time (root→tips NOT along the tips)
- “Conserved from flies to humans” → “Conserved since the MRCA of flies and humans”

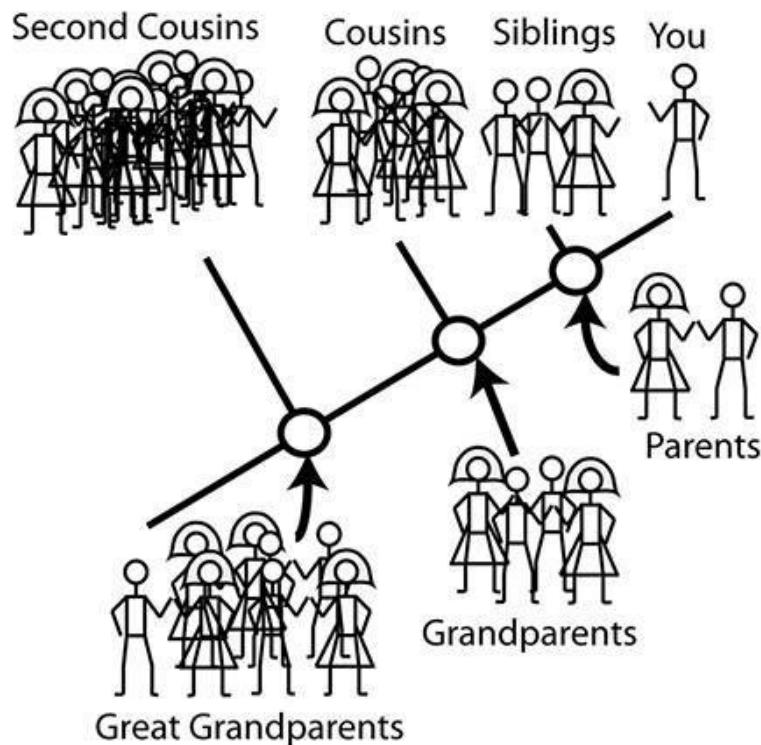


Linear thinking =
“Scala Natura”

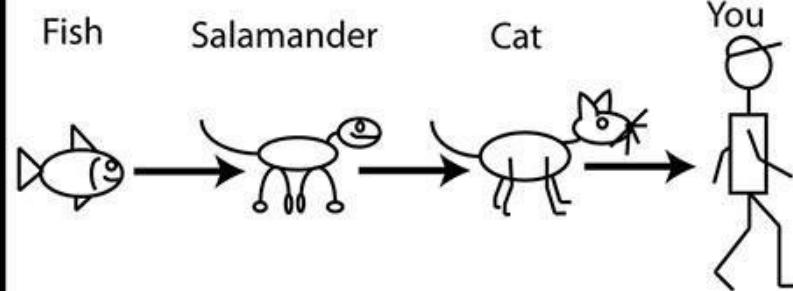
This is NOT Your Family Tree



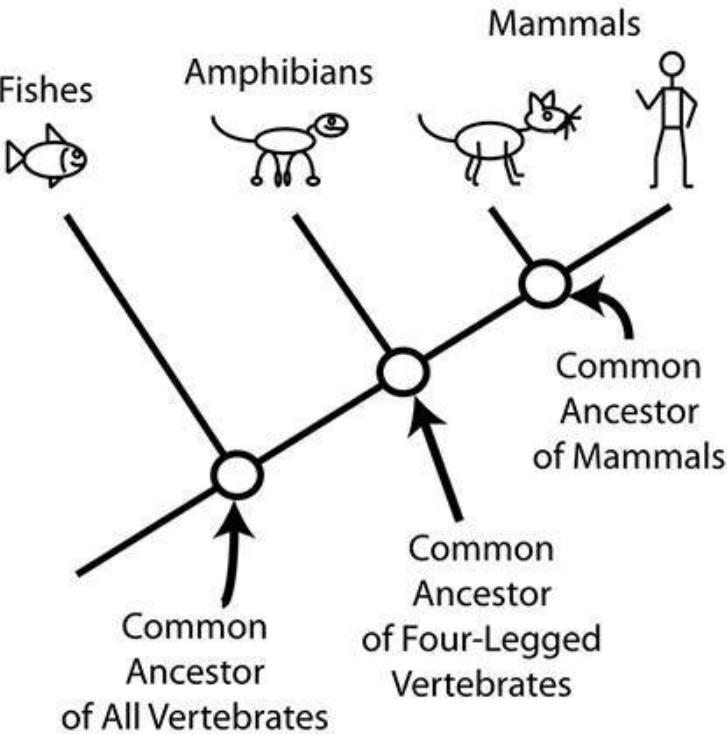
This is Your Family Tree



This is NOT Evolution

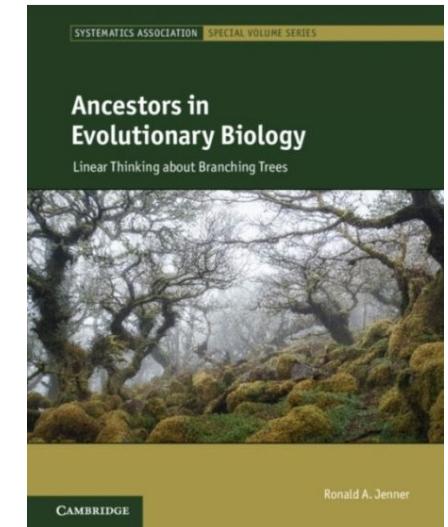


This is Evolution



Tree thinking is not trivial!

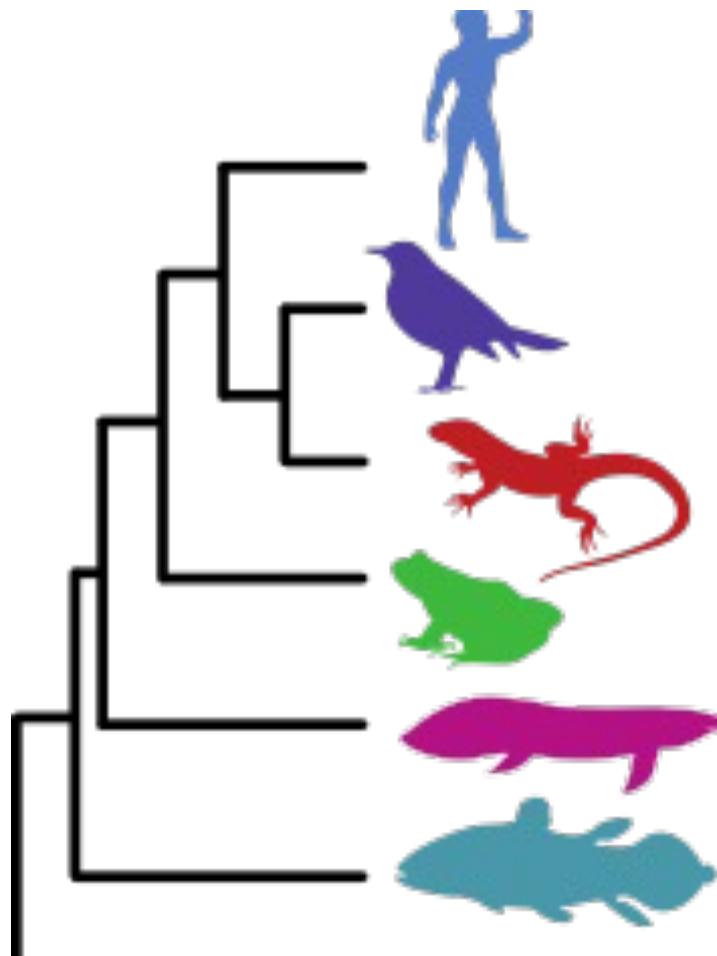
- Watch out with linear thinking in evolution!! Directionality is time (root→tips NOT along the tips)
- “Conserved **from** flies **to** humans” → “Conserved **since** the MRCA of flies and humans”
- Solution: think in terms of ancestors, talk about “sister groups”
- “Linear thinking about branching trees”



This is not just semantics!

Incorrect tree thinking often leads to wrong evolutionary inferences
(phylogenetic patterns, gene function, diversification patters, etc.)

Example 1. "Lungfishes are primitive/basal"

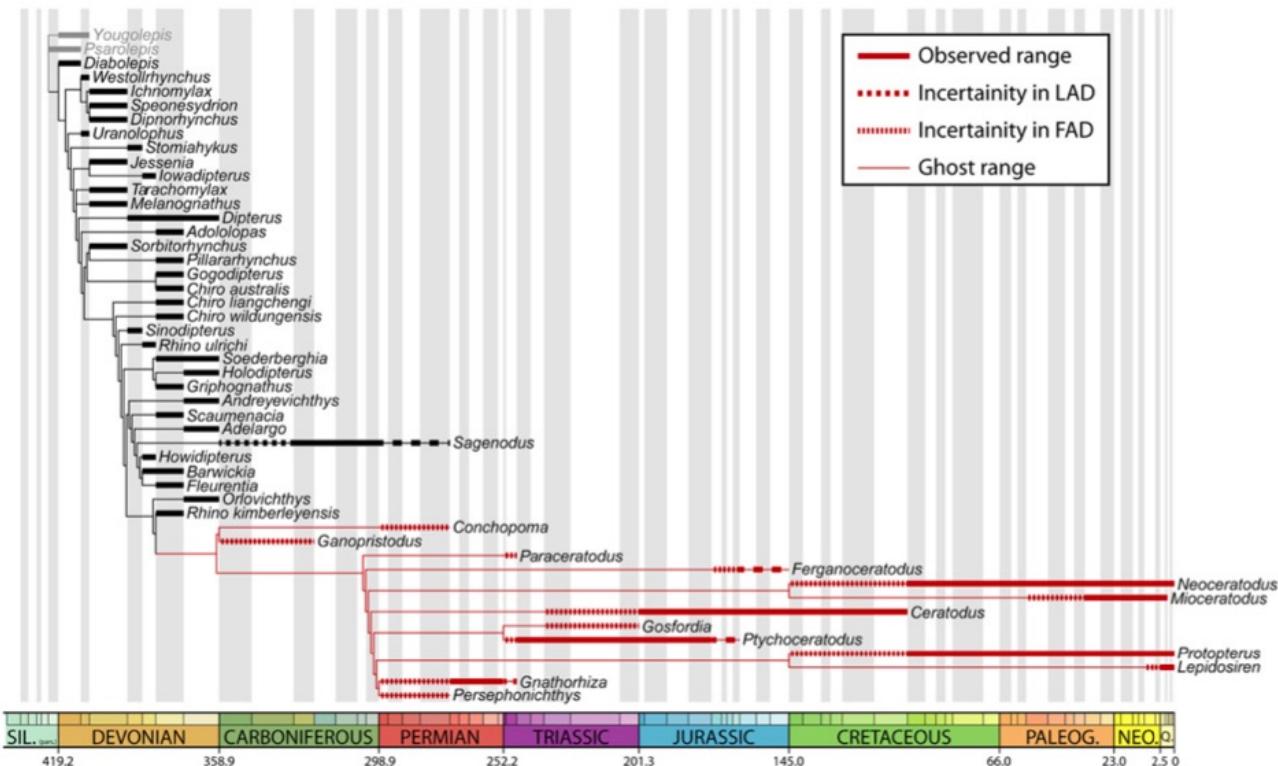


lungfish species. The reconstructed phylogenetic relationships provide evidence that lungfish is the sister group of terrestrial vertebrates and that *Neoceratodus forsteri* is the most primitive lungfish. Moreover, the divergence time between the most primitive lungfish and other lungfish species is between 186.11 and 195.36 MYA. Finally, 43

Ceratodus (Triassic, 228-70 Ma)
vs. *Neoceratodus* (extant)



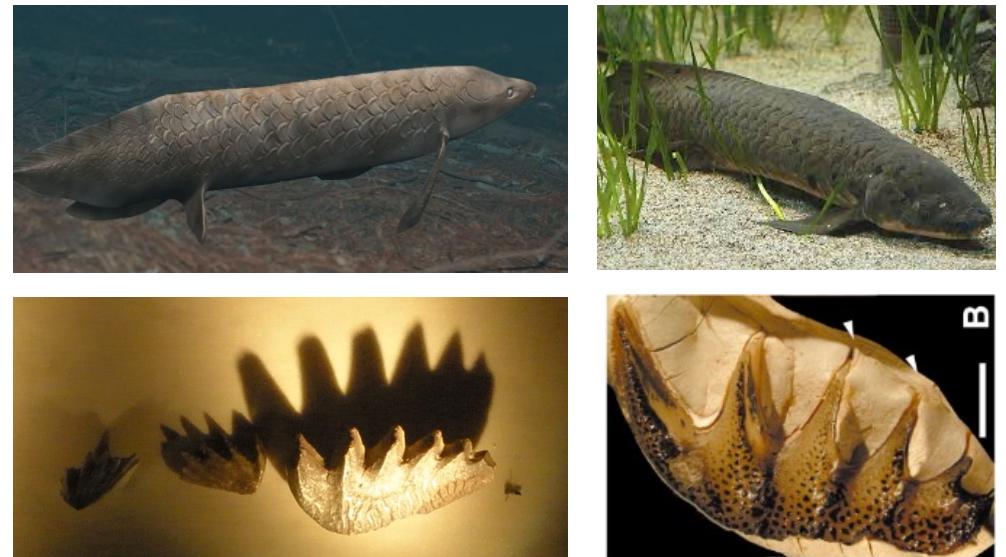
Example 1. "Lungfishes are primitive/basal"



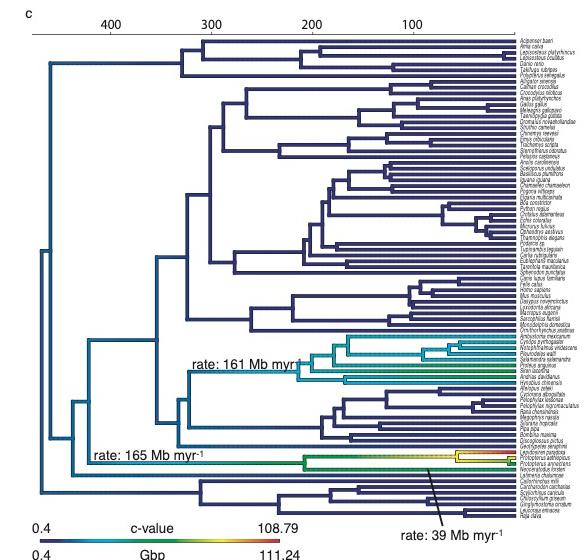
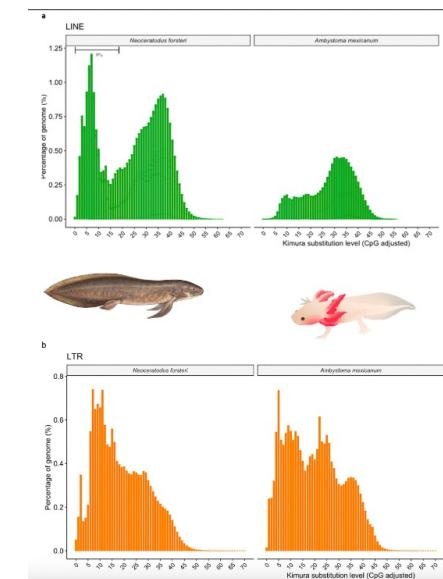
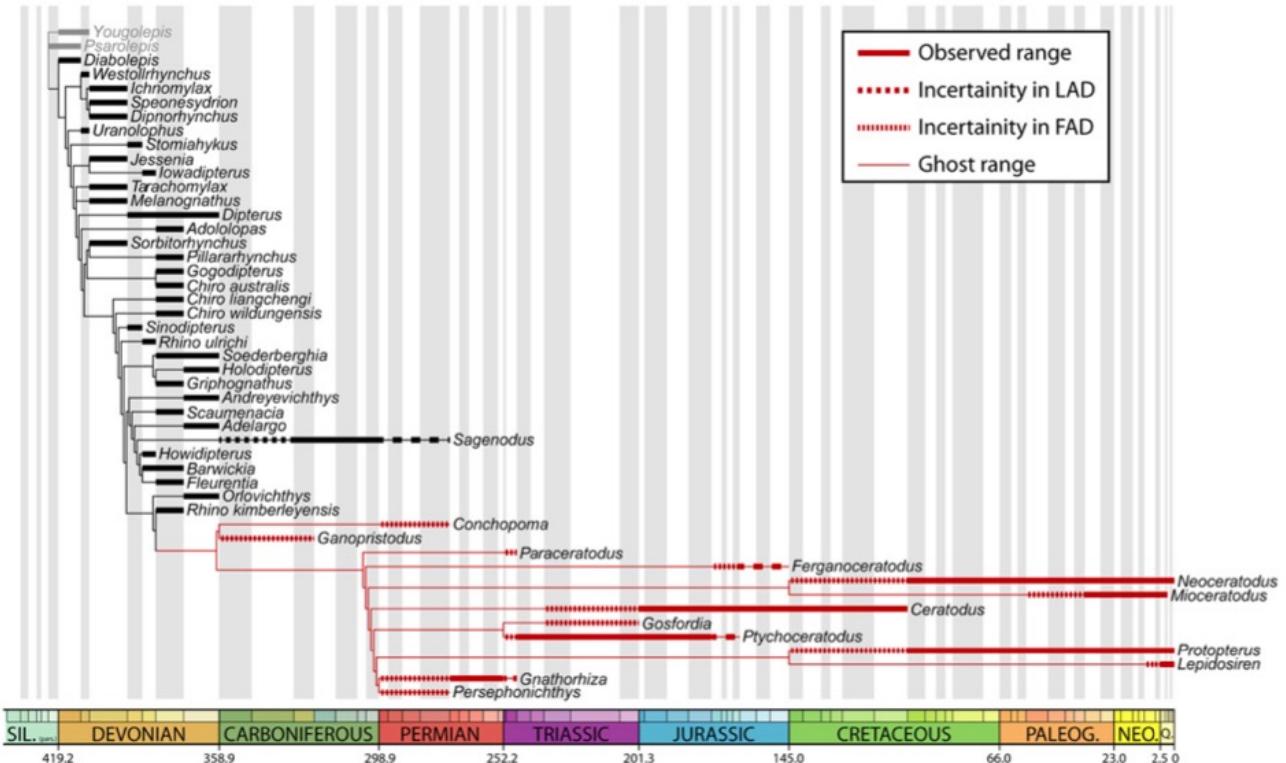
Kemp et al. 2017

lungfish species. The reconstructed phylogenetic relationships provide evidence that lungfish is the sister group of terrestrial vertebrates and that *Neoceratodus forsteri* is the most primitive lungfish. Moreover, the divergence time between the most primitive lungfish and other lungfish species is between 186.11 and 195.36 MYA. Finally, 43

Ceratodus (Triassic, 228-70 Ma) vs. *Neoceratodus* (extant)



Morphological stasis does not imply genomic stasis/lack of evolution!



Meyer et al. 2021

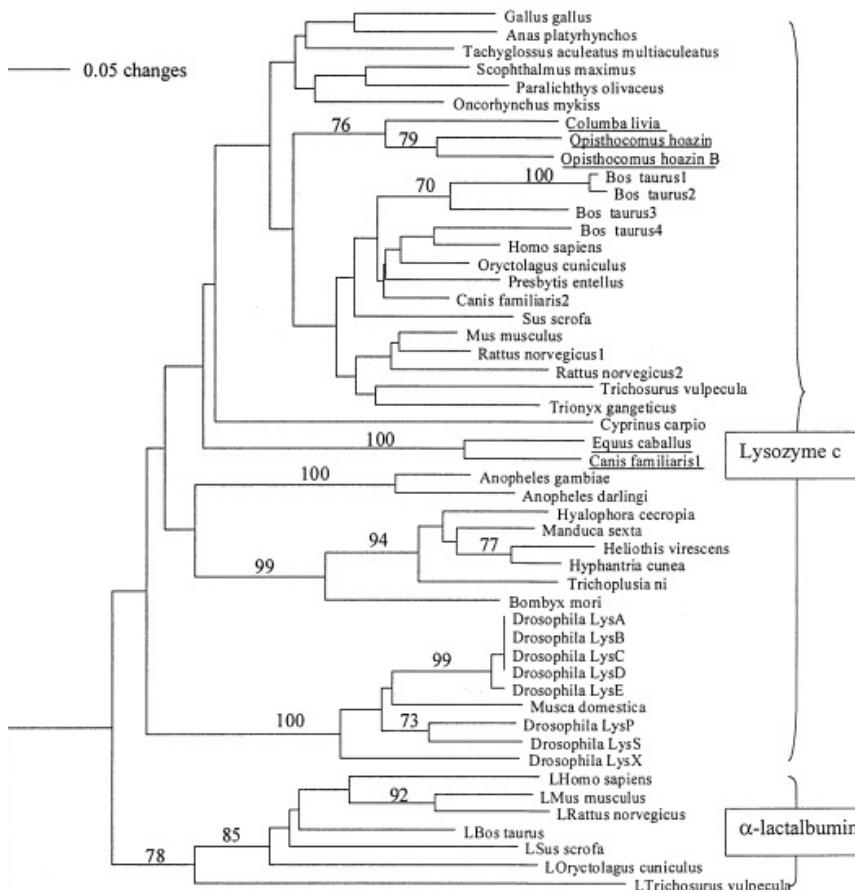
Kemp et al. 2017

Example 2. Inferring ancestral gene functions

Phylogenetic Analysis of Invertebrate Lysozymes and the Evolution of Lysozyme Function

Sana Bachali,¹ Muriel Jager,² Alexandre Hassanin,² Françoise Schoentgen,³ Pierre Jollès,⁴ Aline Fiala-Medioni,⁵ Jean S. Deutsch^{1,2}

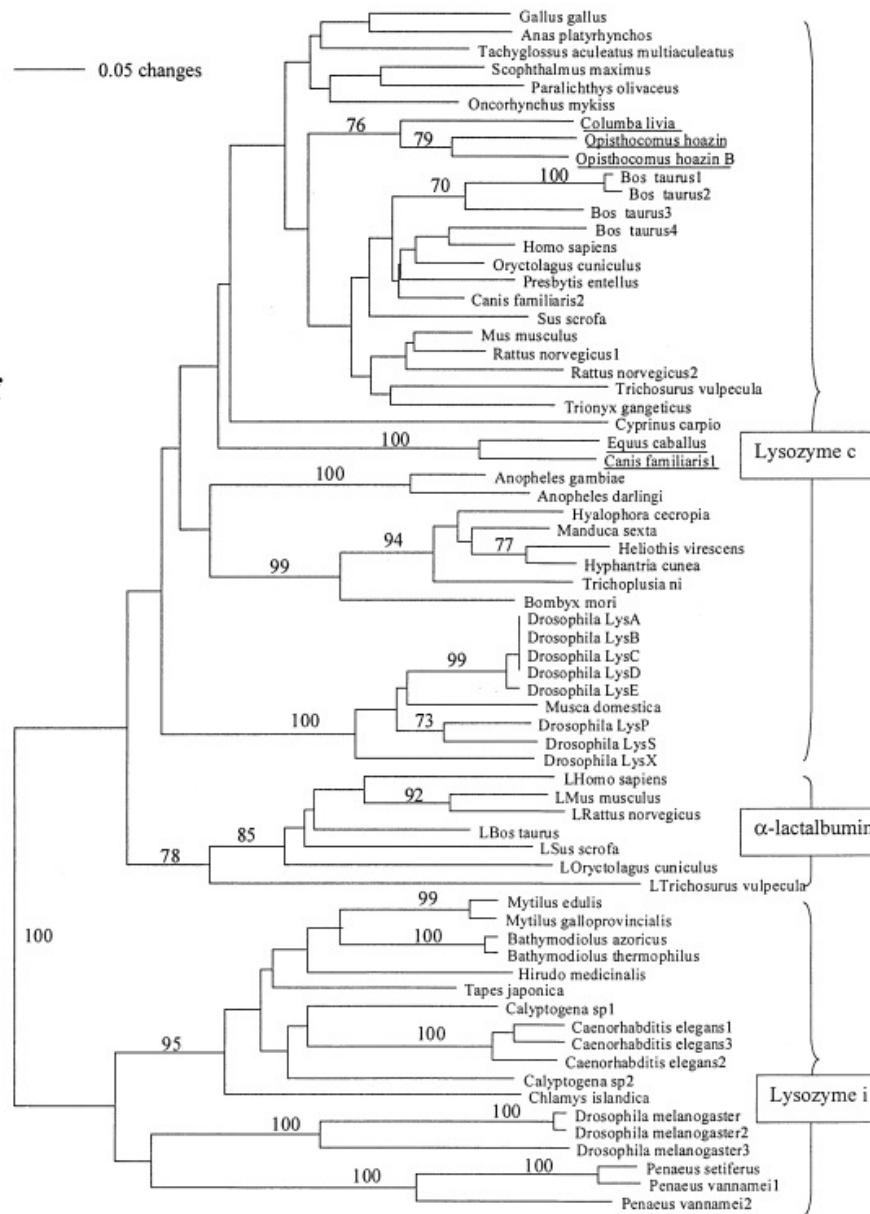
What do you think was the ancestral function of this protein?



Phylogenetic Analysis of Invertebrate Lysozymes and the Evolution of Lysozyme Function

Sana Bachali,¹ Muriel Jager,² Alexandre Hassanin,² Françoise Schoentgen,³ Pierre Jollès,⁴ Aline Fiala-Medioni,⁵ Jean S. Deutsch^{1,2}

What do you think was the ancestral function of this protein?



Tree thinking for all biology: the problem with reading phylogenies as ladders of progress

Kevin E. Omland,^{1,2*} Lyn G. Cook,³ and Michael D. Crisp²

Mammalian Biology 81 (2016) 185–188

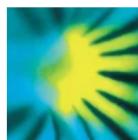
Contents lists available at ScienceDirect



Mammalian Biology



journal homepage: www.elsevier.com/locate/mambio



New Phytologist

Viewpoints | Free Access

Bryophytes are not early diverging land plants

Stuart F. McDaniel

First published: 01 February 2021 | [New Phytologist homepage](https://doi.org/10.1111/nph.17241) | <https://doi.org/10.1111/nph.17241> | Citations: 10



Phylogenetics & Tree-Thinking

Author(s): David A. Baum and Susan Offner

Source: The American Biology Teacher, 70(4):222-229.

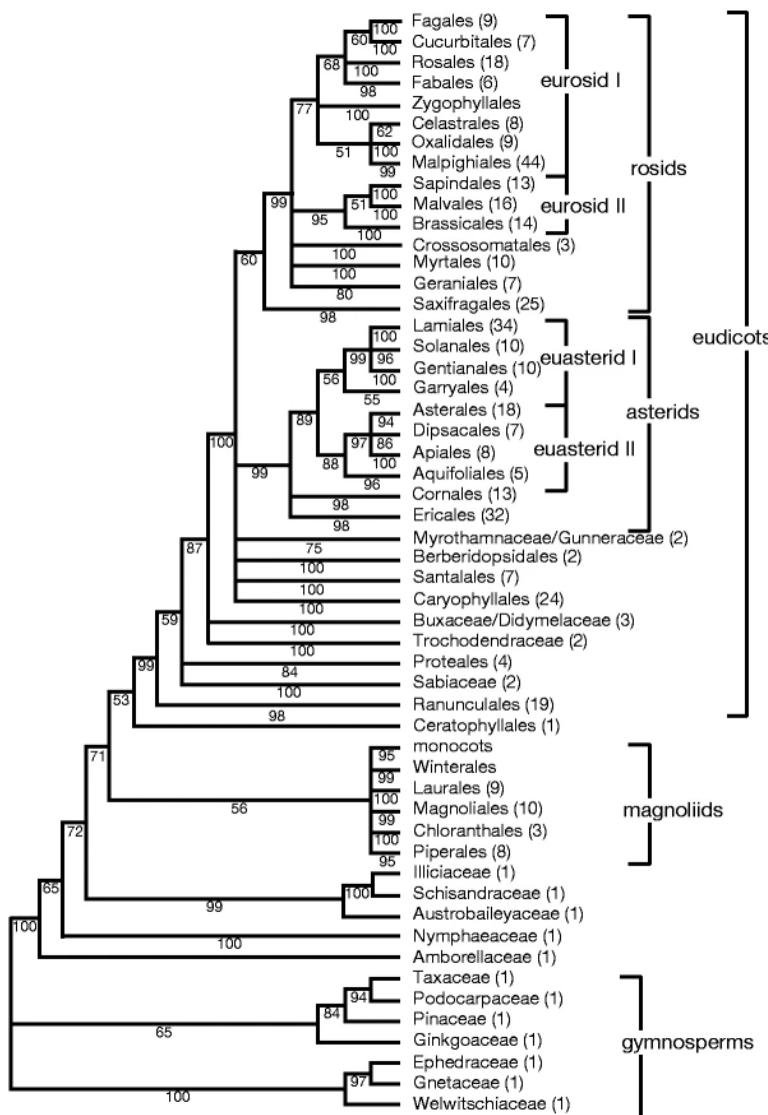
Published By: National Association of Biology Teachers

[https://doi.org/10.1662/0002-7685\(2008\)70\[222:PT\]2.0.CO;2](https://doi.org/10.1662/0002-7685(2008)70[222:PT]2.0.CO;2)



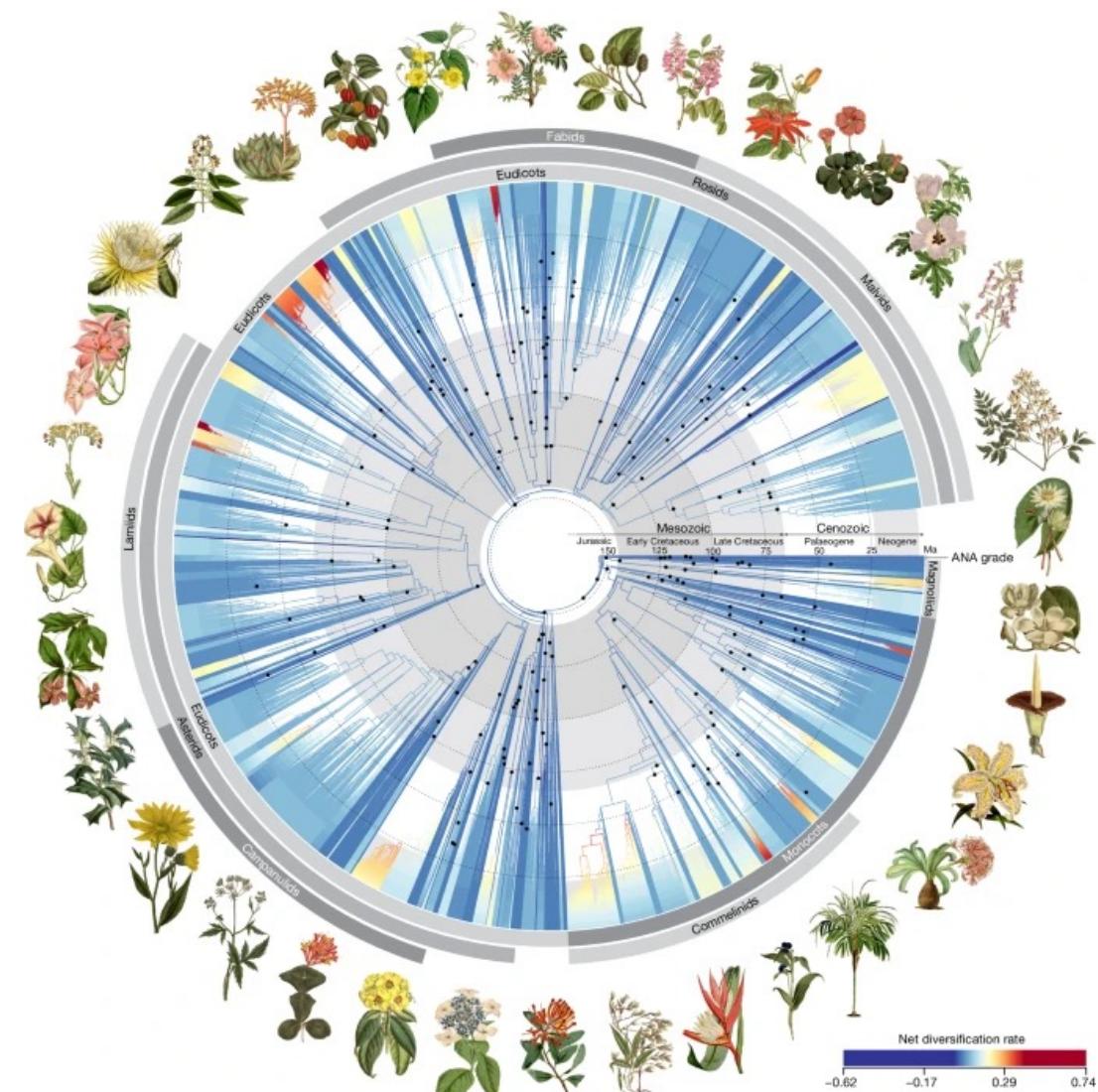
Phylogenetics vs. phylogenomics

Phylogenetics



Soltis et al. 1999, 3 markers

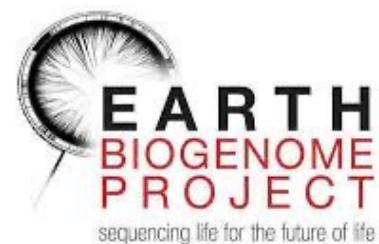
Phylogenomics



Zuntini et al. 2024, 353 markers

Why phylogenomics?

- Increased resolving power for trickly phylogenetic questions (e.g. fast radiations, deep nodes, confounded by gene flow...)
- Genomes for all species to become available soon(ish)



Ending incongruence

N&V

Henry Gee

Recovering the true evolutionary history of any group of organisms has seemed impossible. The availability of large amounts of genomic data promises an era in which the uncertainties are better constrained.

Genome-scale approaches to resolving incongruence in molecular phylogenies

Antonis Rokas*, Barry L. Williams*, Nicole King & Sean B. Carroll

Howard Hughes Medical Institute, Laboratory of Molecular Biology, R. M. Bock Laboratories, University of Wisconsin-Madison, 1525 Linden Drive, Madison, Wisconsin 53706, USA

* These authors contributed equally to this work

One of the most pervasive challenges in molecular phylogenetics is the **incongruence between phylogenies obtained using different data sets, such as individual genes**. To systematically investigate the degree of incongruence, and potential methods for resolving it, we screened the genome sequences of eight yeast species and selected **106** widely distributed **orthologous genes** for phylogenetic analyses, singly and by concatenation. Our results suggest that data sets consisting of single or a small number of concatenated genes have a significant probability of supporting conflicting topologies. By contrast, analyses of the entire data set of concatenated genes yielded a single, fully resolved species tree with maximum support. Comparable results were obtained with a concatenation of a minimum of 20 genes; substantially more genes than commonly used but a small fraction of any genome. **These results have important implications for resolving branches of the tree of life.**

Rokas et al. 2003 Nature

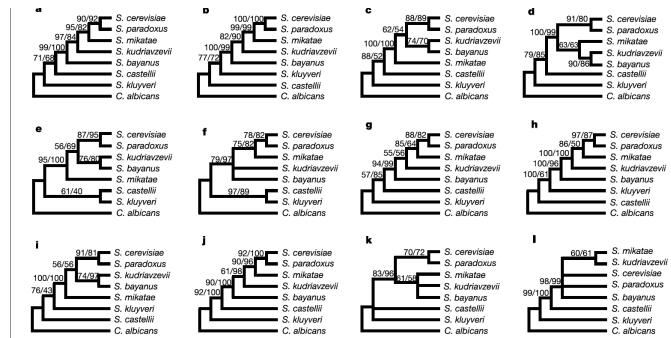


Figure 1 Single-gene data sets generate multiple, robustly supported alternative topologies. Representative alternative trees recovered from analyses of nucleotide data of 106 selected single genes and six commonly used genes are shown. The trees are the 50% majority-rule consensus trees from the genes YBL091C (**a**), YDL031W (**b**), YER005W (**c**), YGL001C (**d**), YNL155W (**e**) and YOL097C (**f**), as well as those from the commonly used genes actin (**g**), hsp70 (**h**), β -tubulin (**i**), RNA polymerase II elongation factor 1- α (**j**) and 18S rDNA (**k**). Numbers above branches indicate bootstrap values (ML on nucleotides/MP on nucleotides).

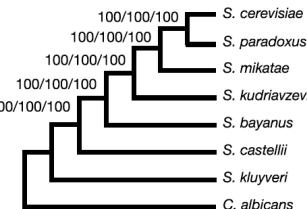
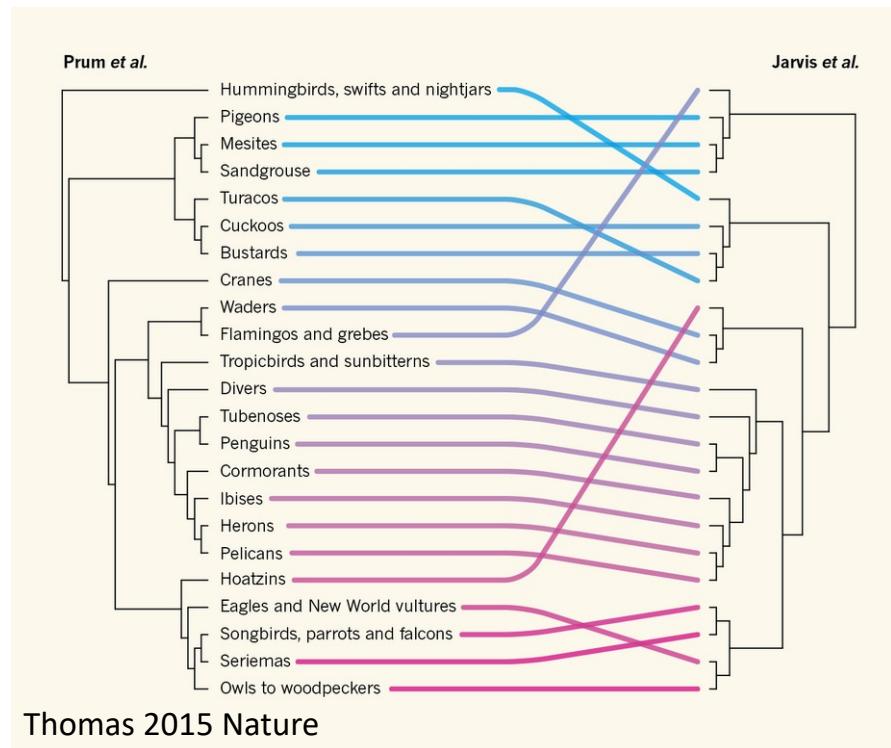


Figure 4 Phylogenetic analyses of the concatenated data set composed of 106 genes yield maximum support for a single tree, irrespective of method and type of character evaluated. Numbers above branches indicate bootstrap values (ML on nucleotides/MP on nucleotides/MP on amino acids).

Phylogenomics: the beginning of incongruence?

Olivier Jeffroy, Henner Brinkmann, Frédéric Delsuc and Hervé Philippe

Canadian Institute for Advanced Research, Centre Robert-Cedergren, Département de Biochimie, Université de Montréal,
Succursale Centre-Ville, Montréal, Québec, Canada, H3C3J7



Same principles, just more data?

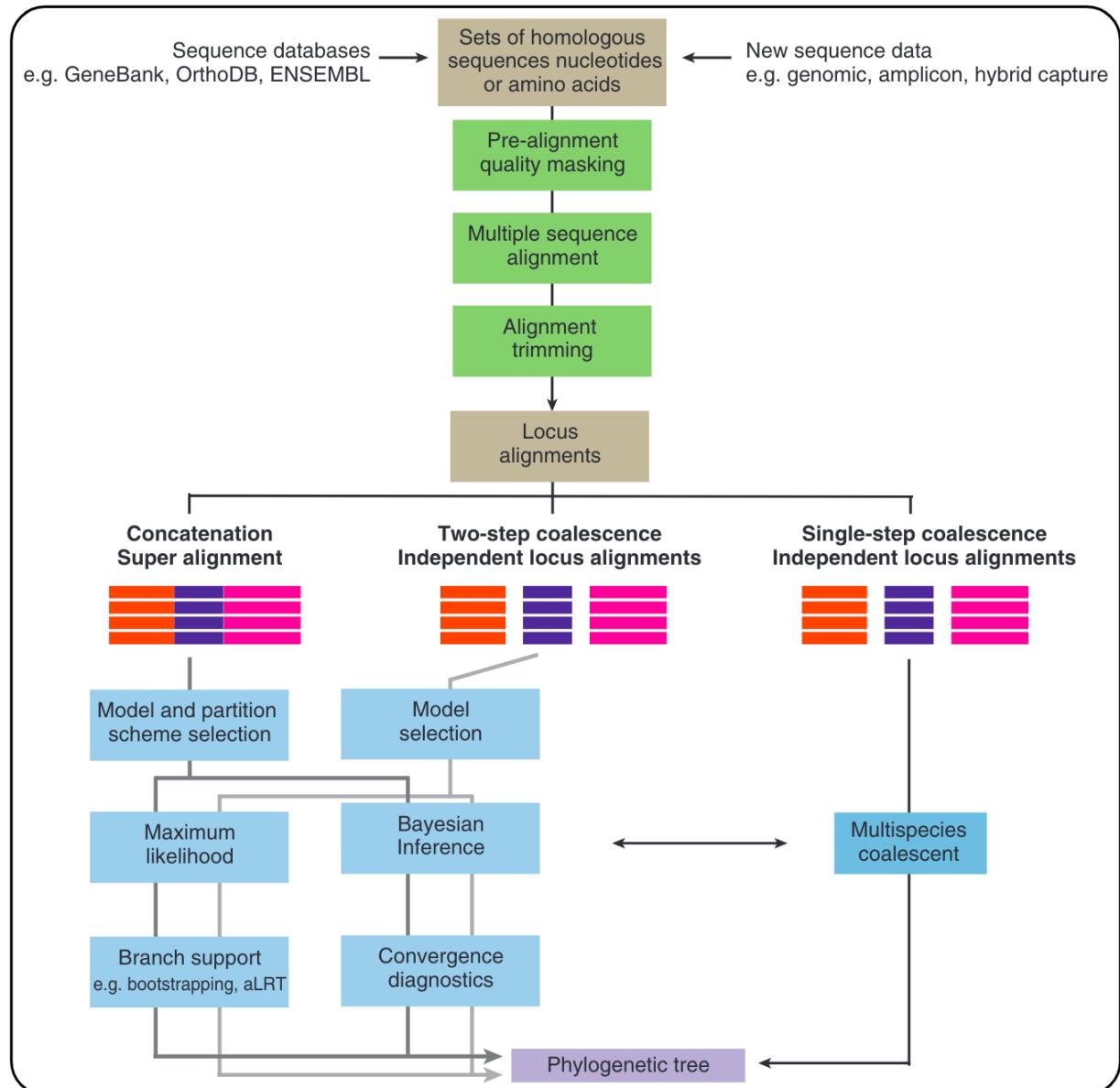
- In principle yes, not quite in practice
- Effort in sequencing vs. effort in assembling datasets => data quality
- Impossibility to manually check every single alignment => automatic tools
- Phylogenomics are more prone to model misspecification => systematic error
- Phylogenetic incongruence is the norm: concatenation vs. coalescence

Automatization

Multiple tools exist for:

- Ortholog identification
- Paralog/outlier removal
- Quality filtering
- Alignment
- Alignment trimming
- Concatenation
- ...

Workflow managers:



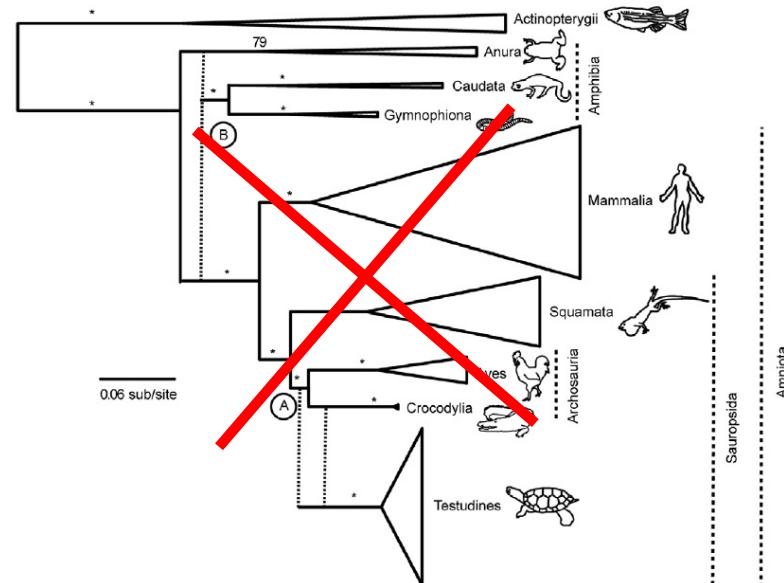
Data quality is key

OPEN  ACCESS Freely available online



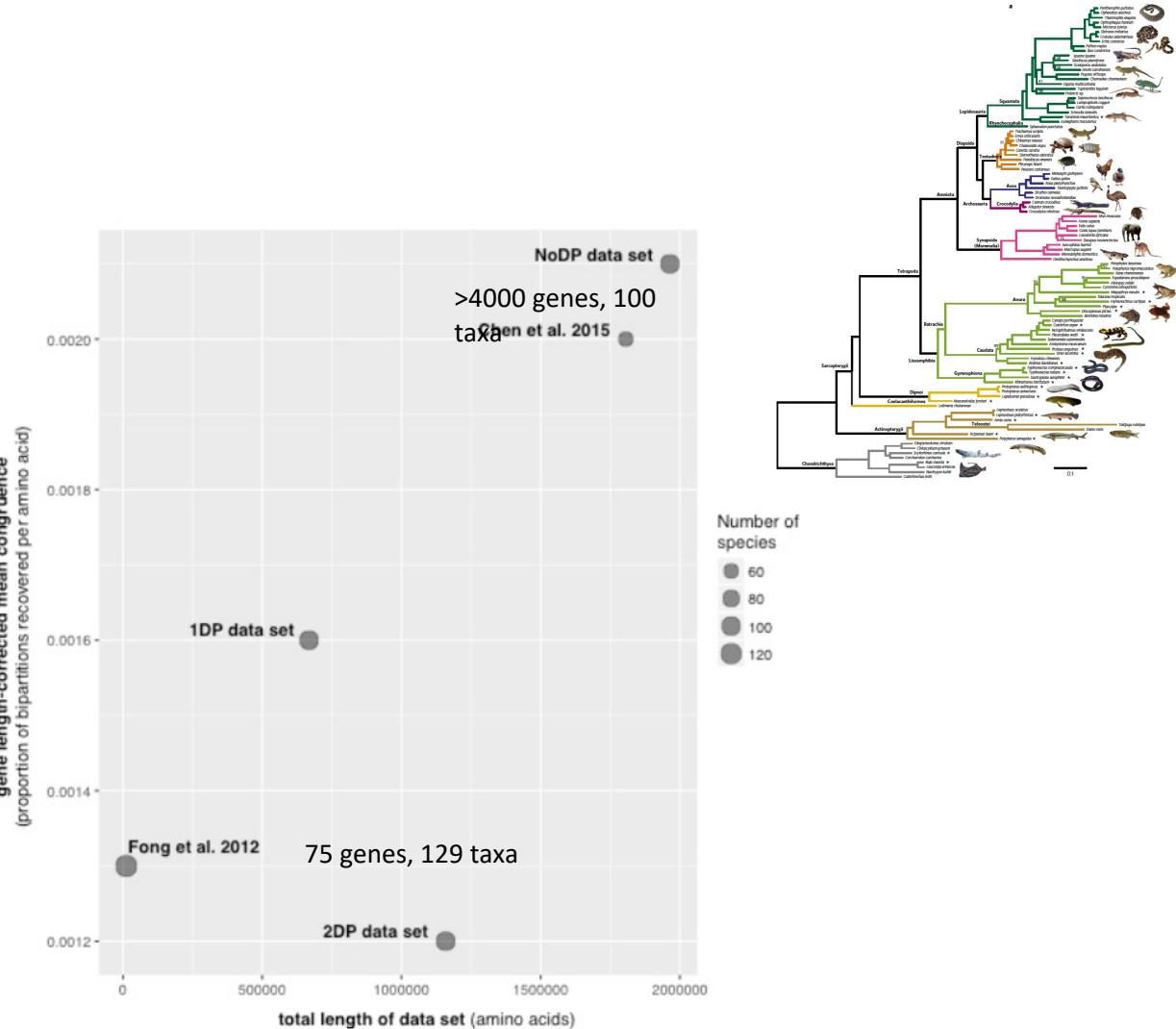
A Phylogenomic Approach to Vertebrate Phylogeny Supports a Turtle-Archosaur Affinity and a Possible Paraphyletic Lissamphibia

Jonathan J. Fong^{1,2,3*}, Jeremy M. Brown^{2,4}, Matthew K. Fujita^{1,2,5,6}, Bastien Boussau^{2,7}



Phylotranscriptomic consolidation of the jawed vertebrate timetree

Iker Irisarri^{1,11*}, Denis Baurain^{1,2}, Henner Brinkmann³, Frédéric Delsuc^{1,4}, Jean-Yves Sire⁵, Alexander Kupfer⁶, Jörn Petersen³, Michael Jarek⁷, Axel Meyer¹, Miguel Vences⁸ and Hervé Philippe^{9,10*}



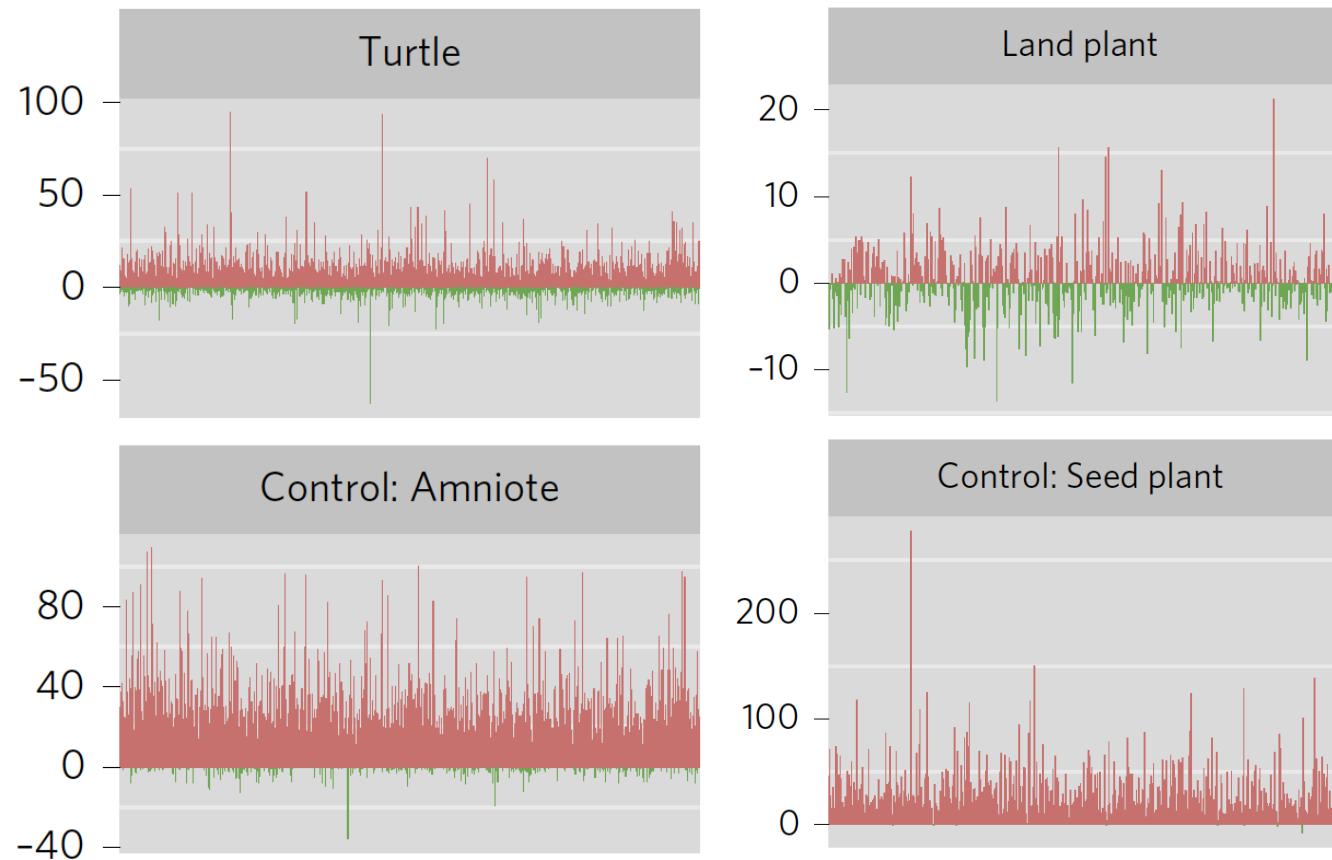
(Hidden) paralogy, HGT, ILS, hybridization, contamination

- They might be buffered in large datasets
- Or they can negatively affect tree inference (for “difficult” phylogenetic questions)



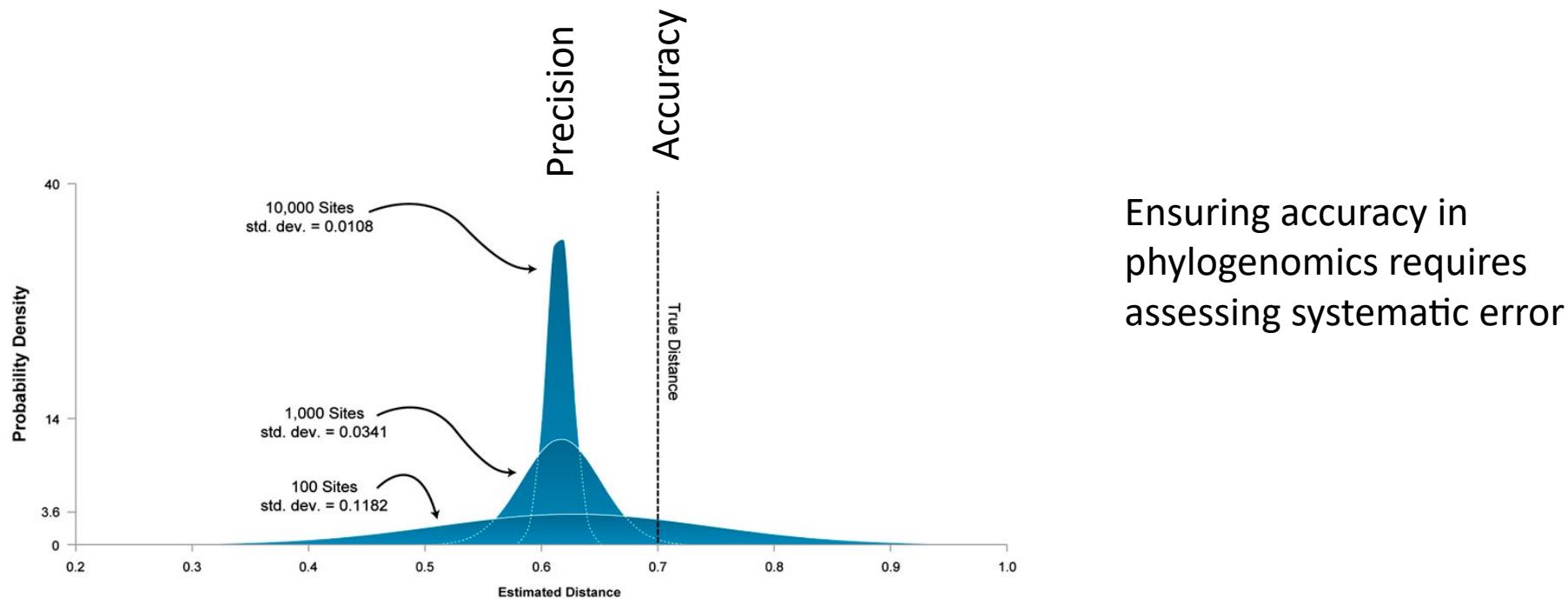
Contentious relationships in phylogenomic studies can be driven by a handful of genes

Xing-Xing Shen¹, Chris Todd Hittinger² and Antonis Rokas^{1*}



Precision vs. accuracy

Stochastic errors are reduced by adding data, increasing precision (but not necessarily accuracy)



Systematic error: increases with alignment length

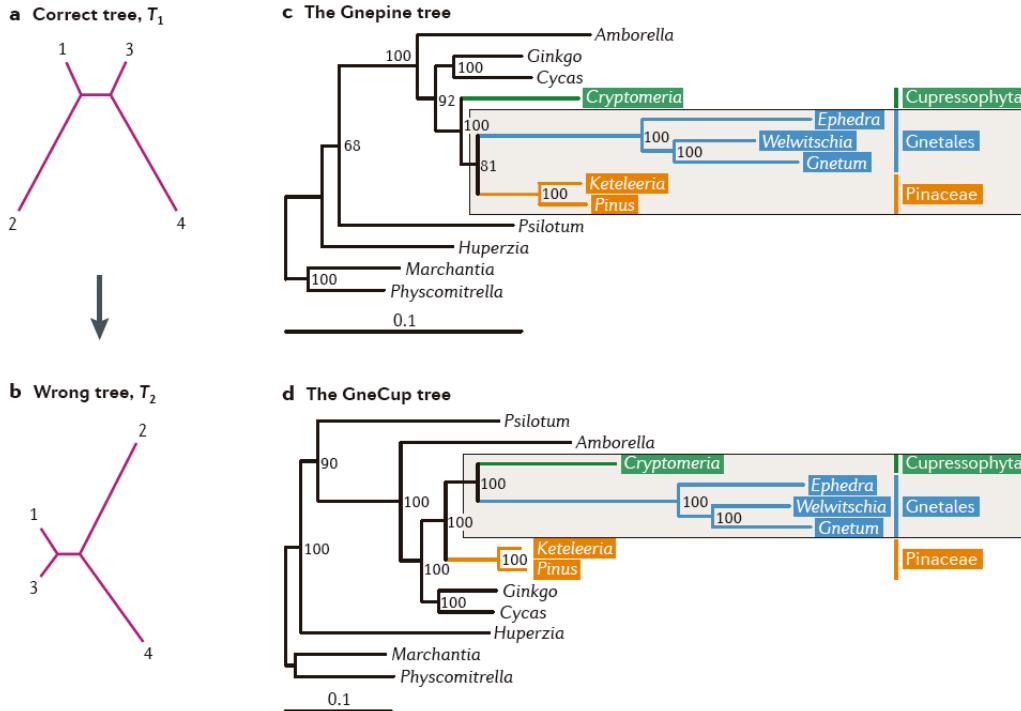


Figure 3 | Long-branch attraction in theory and in practice. Panels a and b show the four-species case analysed by Felsenstein⁴³. If the correct tree (T_1 in a) has two long branches separated by a short internal branch, parsimony (as well as model-based methods such as likelihood and Bayesian methods under simplistic models) tends to recover a wrong tree (T_2 in b), in which the two long branches are grouped together. Panels c and d show a similar phenomenon in a real data set, concerning the phylogeny of seed plants¹³⁴. The Gnetales is a morphologically and ecologically diverse group of Gymnosperms including three genera (Ephedra, Gnetum and Welwitschia), but its phylogenetic position has been controversial. Maximum likelihood analysis of 56 chloroplast proteins produced the GneCup tree (d), in which the Gnetales are grouped with Cupressophytá, apparently owing to a long-branch attraction artefact. However, the Gnepine tree (c), in which the Gnetales joins the Pinaceae, was inferred by excluding the fastest-evolving 18 proteins as well as three proteins (namely, psbC, rpl2 and rps7) that had experienced many parallel substitutions between the Cryptomeria branch and the branch ancestral to the Gnetales. The Gnepine tree (c) is also supported by two proteins from the nuclear genome and appears to be the correct tree. Branch lengths and bootstrap proportions are all calculated using RAxML. See REF. 134 for details.

- Compositional heterogeneity among lineages
- Different rates among lineages (LBA)
- Variation in rate over time (heterotachy)

Phylogenomic analyses should assess systematic errors!

OPEN  ACCESS Freely available online

PLOS BIOLOGY

Perspective

Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough

Hervé Philippe^{1*}, Henner Brinkmann¹, Dennis V. Lavrov², D. Timothy J. Littlewood³, Michael Manuel⁴, Gert Wörheide^{5,6}, Denis Baurain⁷

1 Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal, Québec, Canada, **2** Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, United States of America, **3** Department of Zoology, The Natural History Museum, London, United Kingdom, **4** Université Paris 6, UMR 7138 "Systématique, Adaptation, Evolution" UPMC CNRS IRD MHNH, Paris, France, **5** Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität München, München, Germany, **6** GeoBio-Center, Ludwig-Maximilians-Universität München, München, Germany, **7** Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège, Liège, Belgium

Philippe and Roure *BMC Biology* 2011, 9:91
<http://www.biomedcentral.com/1741-7007/9/91>



COMMENTARY

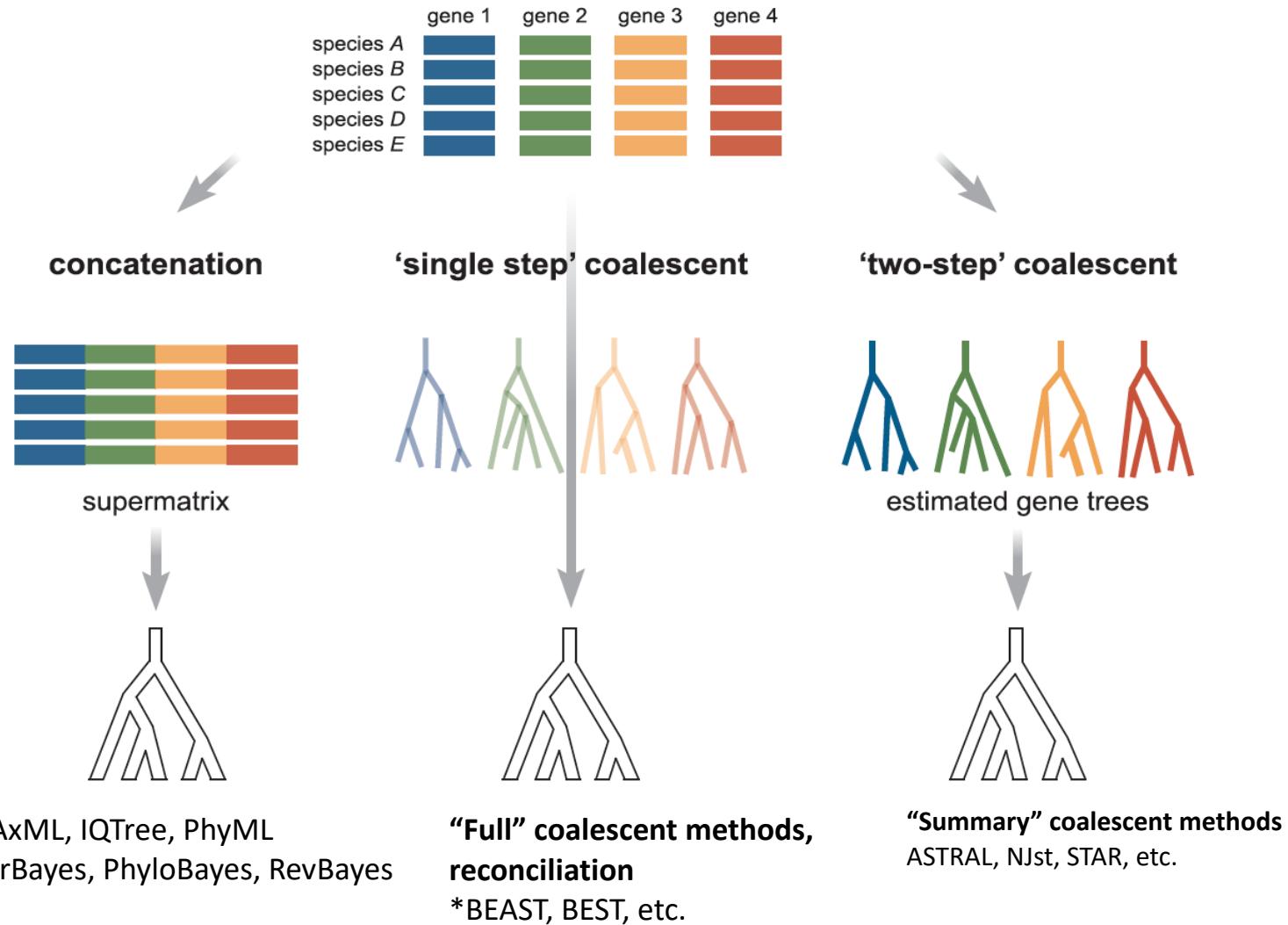
Open Access

Difficult phylogenetic questions: more data, maybe; better methods, certainly

Hervé Philippe* and Béatrice Roure

See research article: <http://www.biomedcentral.com/1741-7007/9/87>

Concatenation vs. coalescence



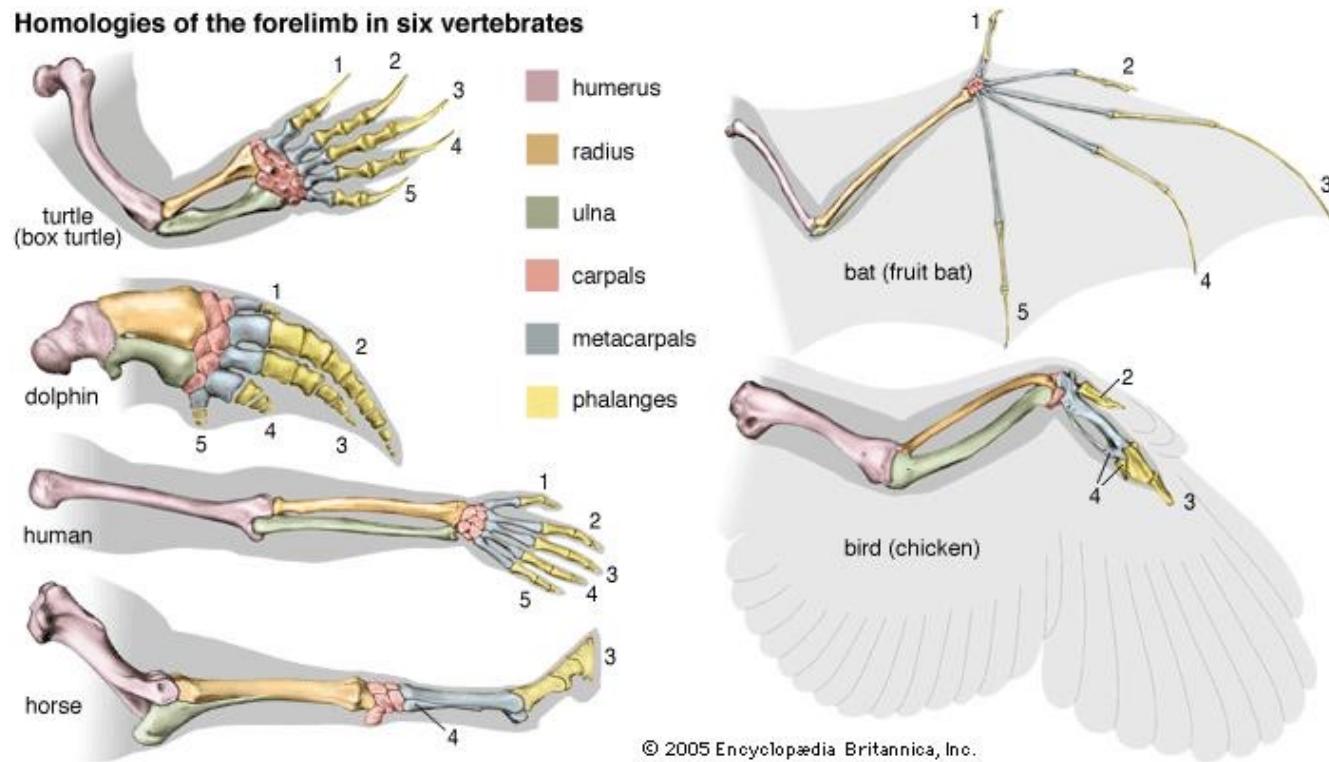
Take-home messages

- **Phylogenomics** allow us to resolve relationships with confidence and explore genome-wide divergence patterns (e.g., ILS, gene flow)
- Data quality is key
- Systematic error must be assessed
- Concatenation vs. coalescence

Orthology inference

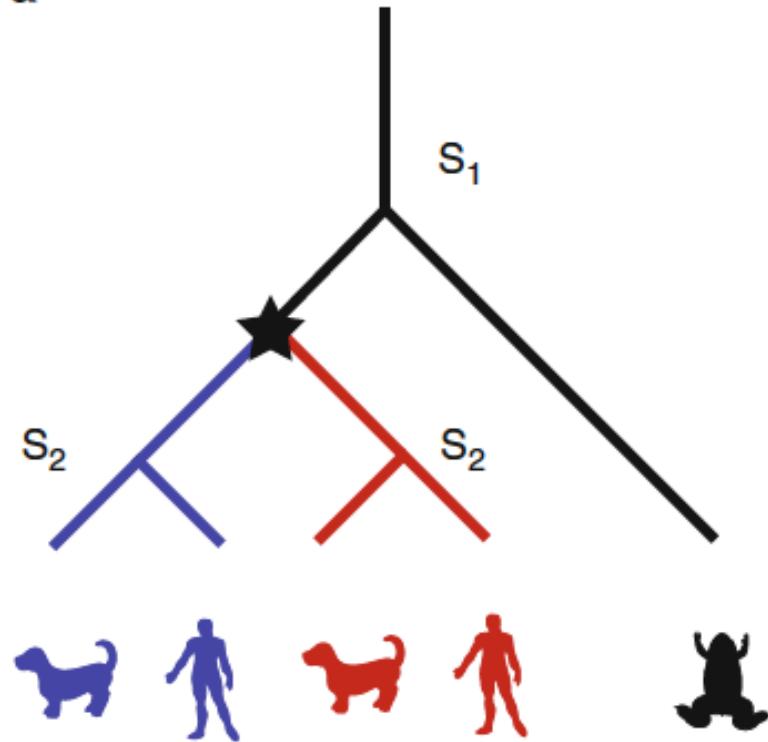
Homology

- Homology hypothesis = common ancestry
- Homology is relative
- Homology can be tricky



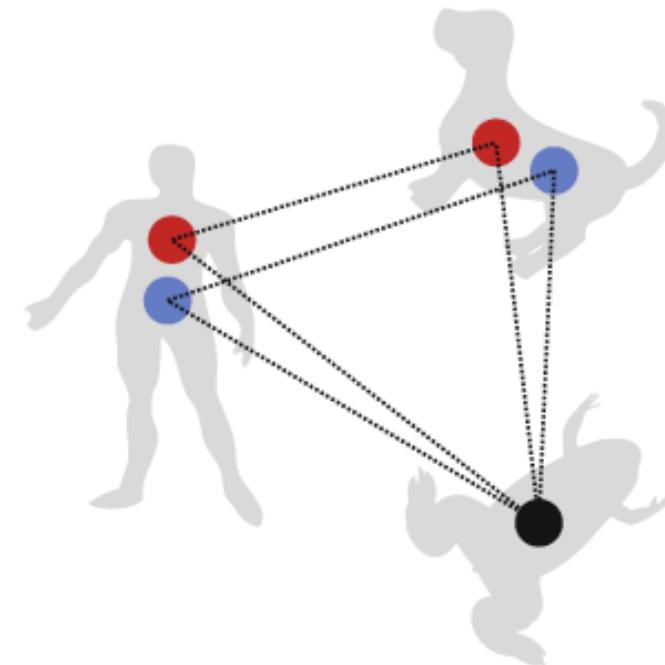
Orthology and paralogy

a



b

Multispecies comparisons: sequence share common ancestry with multiple sequences through speciation and duplication

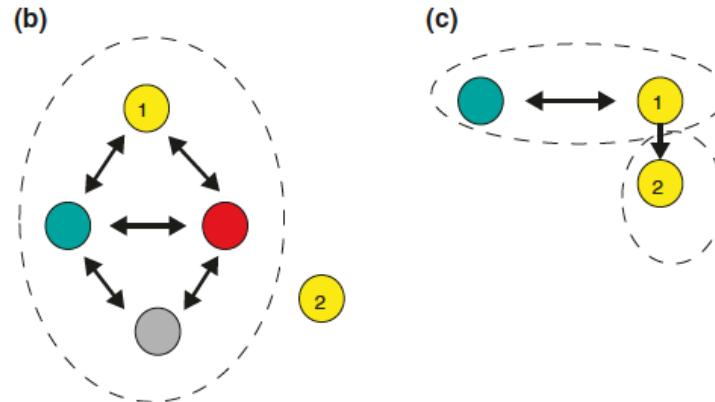
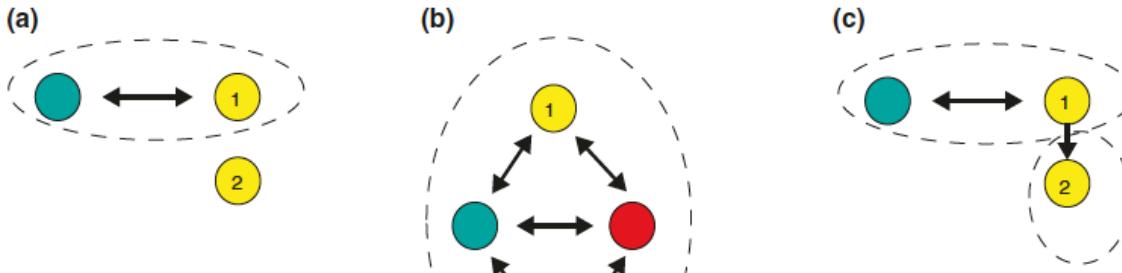


Inferring orthology

Reciprocal best BLAST hits

Graph-based

- OrthoFinder
- OMA
- INPARANOID
- OrthoMCL
- eggNOG
- Ortholinspector

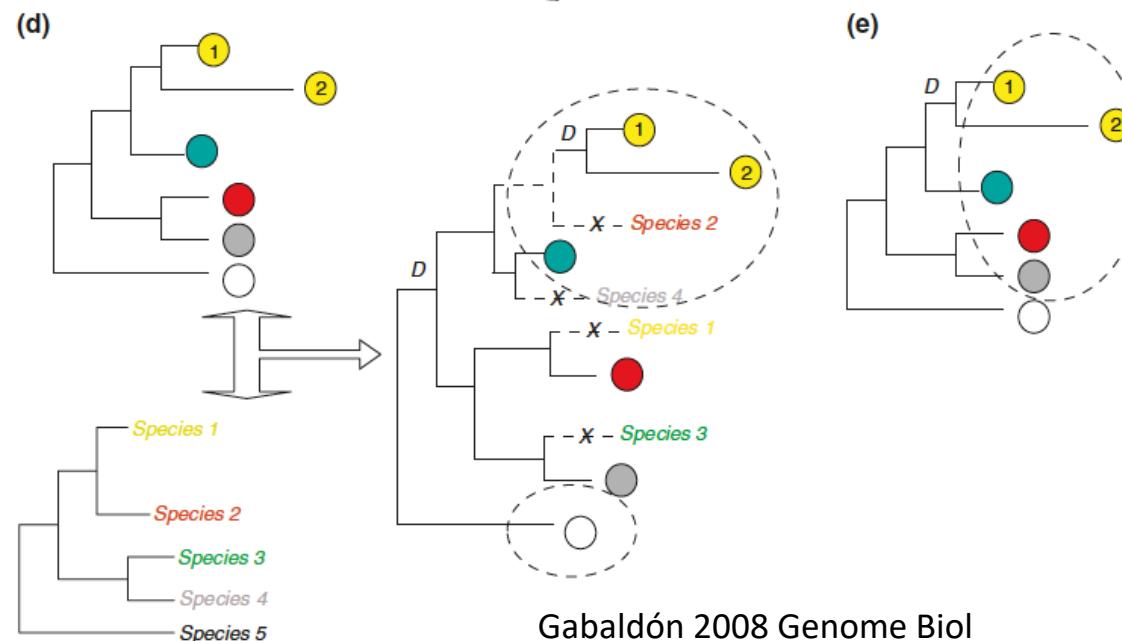


Tree-based

- ENSEMBL Compara
- PANTHER
- PhylomeDB

Hybrids

- MetaPhors



Multiple sequence alignment,
filtering and trimming

PRE-alignment QUALity filtering

Bioinformatics, 34(22), 2018, 3929–3930

doi: 10.1093/bioinformatics/bty448

Advance Access Publication Date: 1 June 2018

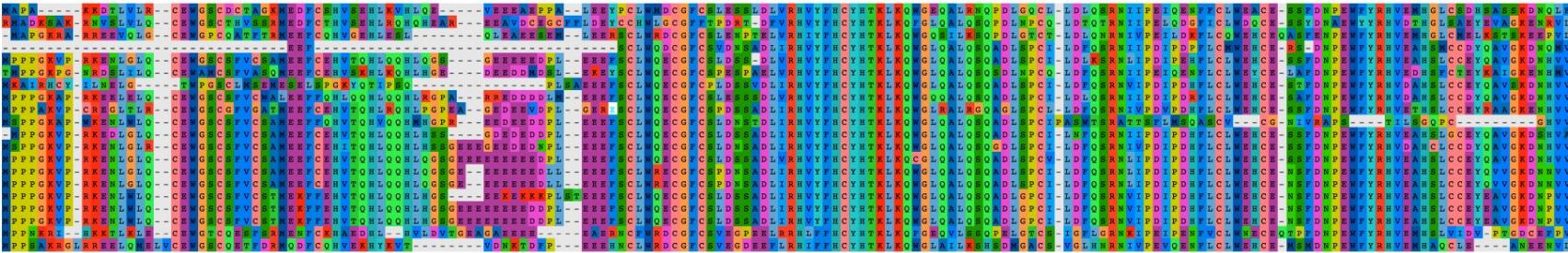
Applications Note

Sequence analysis

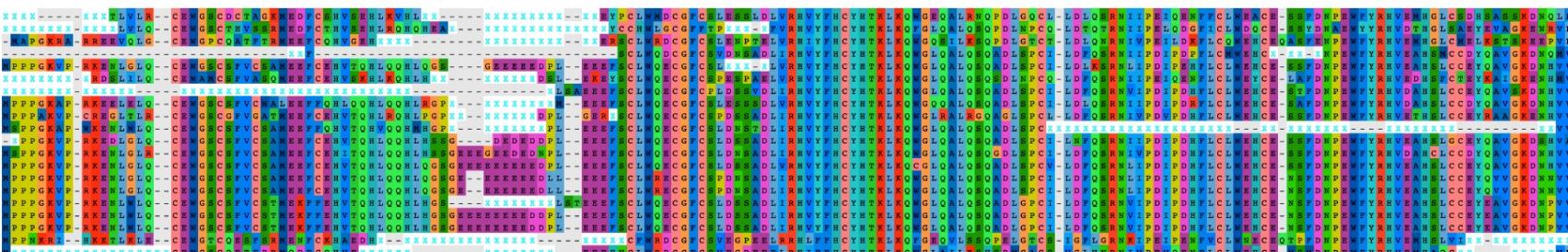
PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences

Simon Whelan^{1,*}, Iker Irisarri² and Fabien Burki^{2,3,*}

A Original alignment



B PREQUAL



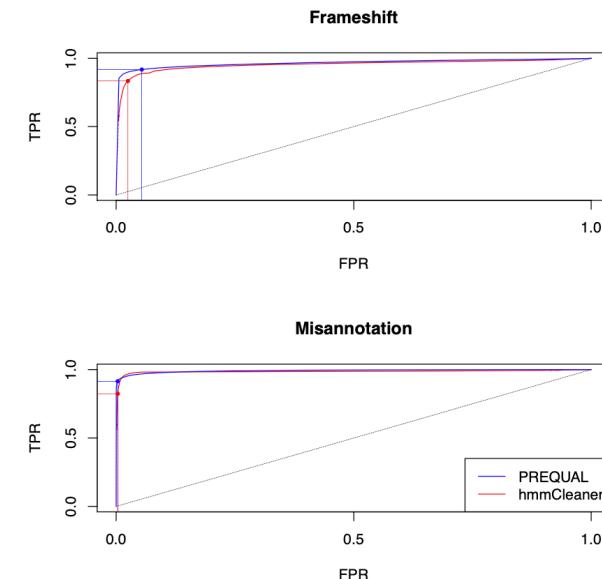
PRE-alignment QUALity filtering

Automated Removal of Non-homologous Sequence Stretches with PREQUAL

Iker Irisarri, Fabien Burki, and Simon Whelan

In: Katoh Ed. Methods in Molecular Biology 2231 (2021)

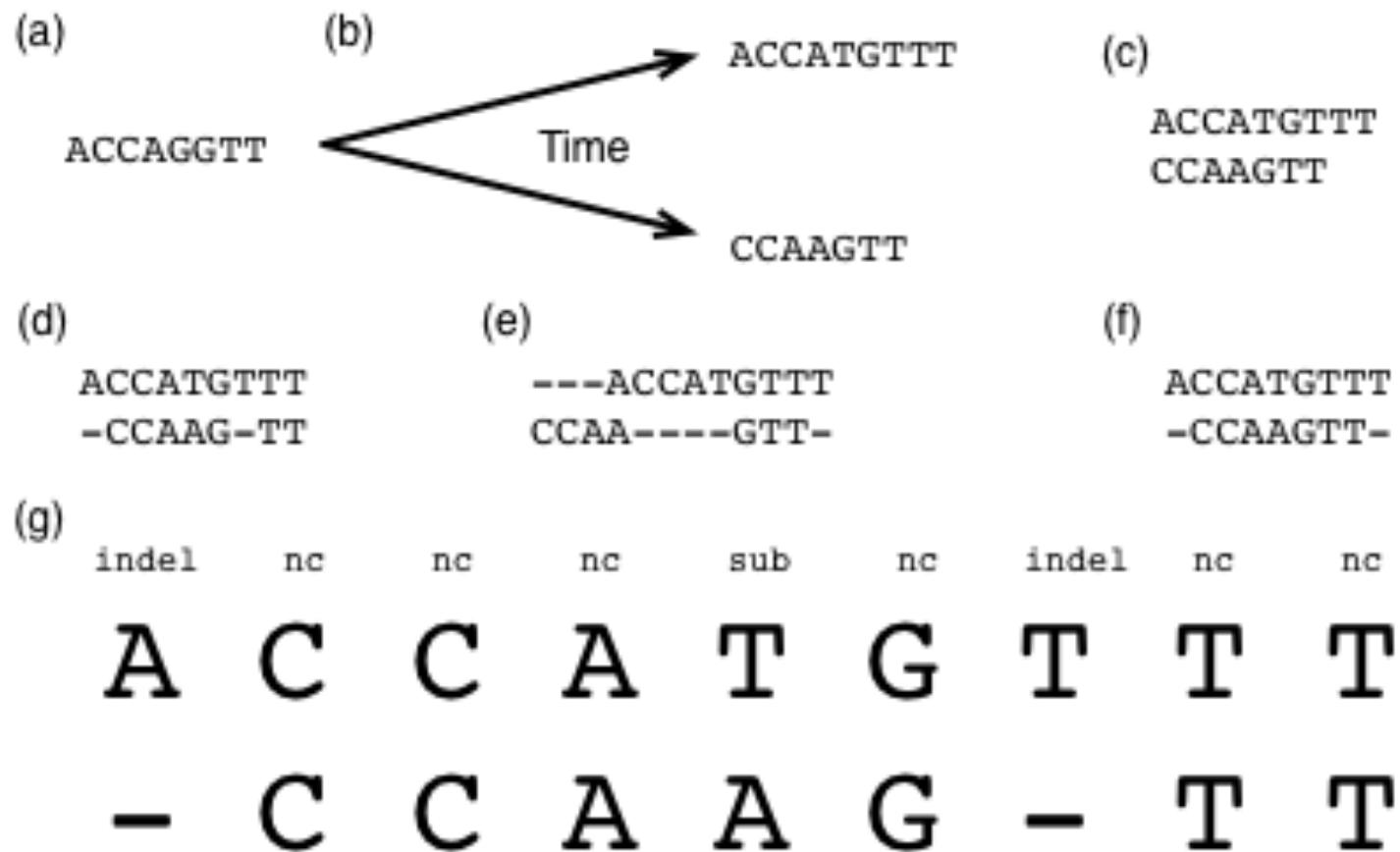
```
-----  
PREQUAL v.1.02 by Simon Whelan  
  
A There are 251 sequences of max length 746 !  
Prepping pairHMM ... done  
Collecting subset of posterior probabilities based on closest 10 sequences determined by Kmers  
This may take time for larger data sets: B1  
  
B Creating collection sets of PPs based on Kmer distances  
\ 251 / 251 ... done B2  
Getting posterior probabilities:  
... Done Get Posteriors  
Outputting posterior probabilities to example.fasta.filtered.PP... done B3  
  
C Performing filtering:  
Applying standard threshold 0.994 resulting in 3494 residues removed C1  
Extending filtered regions with width of 10 ... 19 additional regions removed C2  
Applying front/back trimming for runs of 3 resulting in 90 sections removed C3  
  
D Outputting results:  
Outputting filtered amino acid sequences to example.fasta.filtered  
  
Computation complete  
  
E ===== Summary ======  
Original Filtered %Retained  
#Sequences 251 251 100%  
#Residues 137880 134290 97.4%  
  
F Analysis may have some problems. Warnings output to example.fasta.warning  
  
Filtering complete!
```



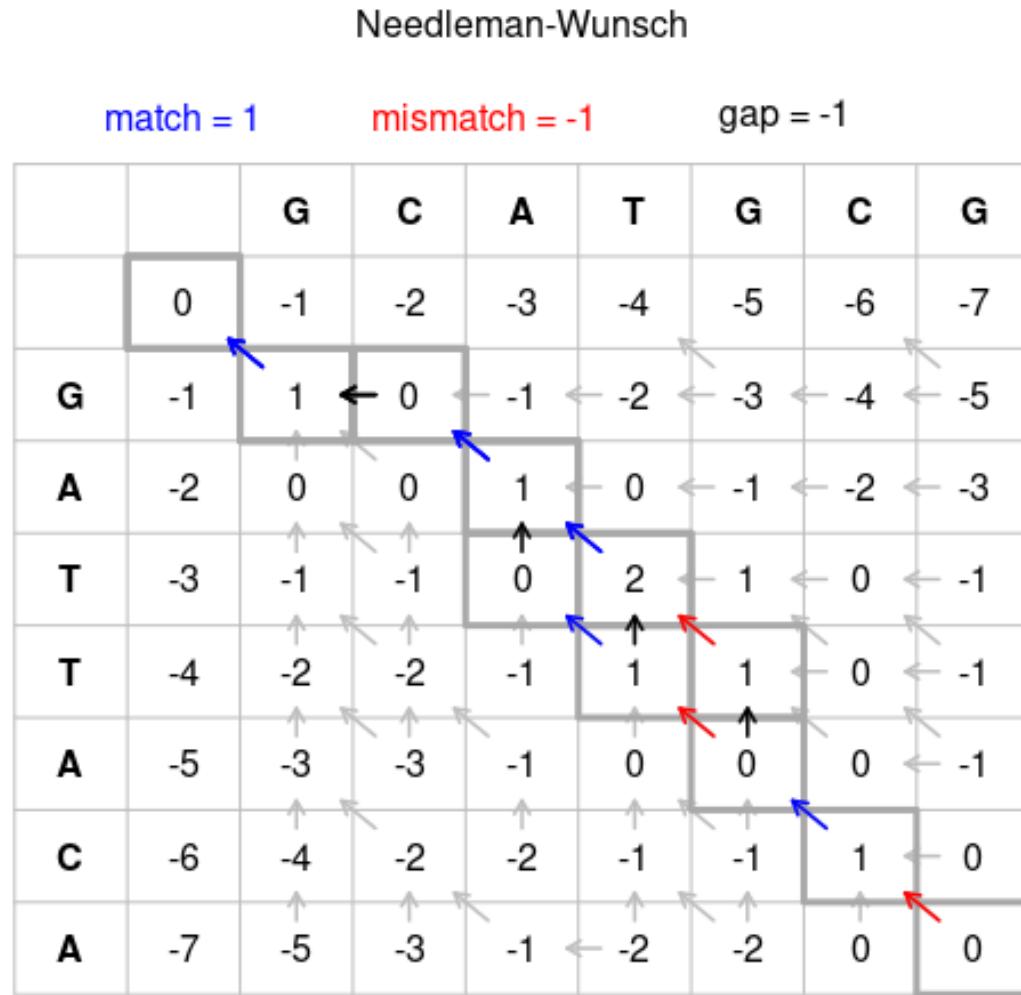
Supplementary Figure 1. ROC curves for PREQUAL (blue) and HMMcleaner (red) for our frameshift and misannotation simulation schemes. These plots show the trade-off between true positive rate (TPR) and false positive rate (FPR), with the best performing method being that which reaches closest to the top left-hand corner. The dots on the curves show the performance for the methods under the recommended thresholds. AUC values calculated using the trapezoidal approximation for PREQUAL are 0.965 (frameshifts) and 0.992 (misannotations), and for HMMcleaner 0.951 (frameshifts) and 0.985 (misannotations). PREQUAL offers an increase to classifier performance in terms of AUC for both frameshifts and misannotations of HMMcleaner.

Sequence alignment

Hypothesis of positional (NT/CODON/AA) homology



Needleman-Wunsch algorithm



GCAT-GCG
G-ATTACA

Local Alignment

Pairwise Sequence Alignment

Target Sequence

5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'

||||| ||||| ||||| ||||| ||||| |||||

Query Sequence

5' TACTCACGGATGAGGTACTTTAGAGGC 3'

BLAST

Global Alignment

Target Sequence

5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'

||||| ||||| ||||| ||||| ||||| ||||| |||||

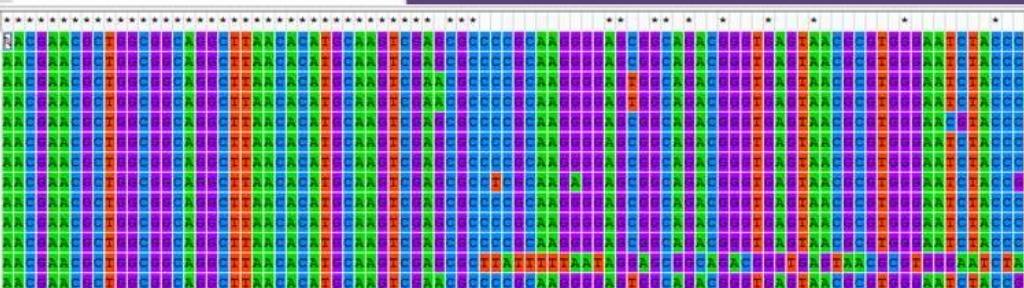
5' ACTACTAGATT----ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'

Query Sequence

Multiple Sequence Alignment (MSA)

MSA

Species/Abbrv
1. Rhizobium_leguminosarum_bv._viciae_3841_g115254414
2. Sinorhizobium_medicae_WSM419_g150026743
3. Agrobacterium_fabrum_str._C58_g159139455
4. Agrobacterium_fabrum_str._C58_g159140696
5. Rhizobium_etli_CIAT_652_g190694918
6. Rhizobium_leguminosarum_bv._trifoli_WSM2304_g209533368
7. Agrobacterium_radiobacter_K84_g221721649
8. Agrobacterium_vitis_S4_g221737306
9. Sinorhizobium_fredii_NGR234_g227339586
10. Rhizobium_leguminosarum_bv._trifoli_WSM1325_g240856645
11. Sinorhizobium_meliiloti_1021_g30407155
12. Candidatus_Liberibacter_solanacearum_CLso-ZC1_g313495152
13. Agrobacterium_sp._H13-3_g325062059



Multiple sequence alignment

Often progressive alignment (dynamic programming) guide tree
(distance): Iterative refinement (mafft, Muscle), Consistency-based (T-Coffee), HMM, Secondary structure, etc.

MSA of protein-coding genes (ORF integrity)

A color-coded multiple sequence alignment of protein-coding genes. The sequences are aligned vertically, and each nucleotide position is represented by a colored square. The colors represent the four bases: Adenine (blue), Thymine (green), Cytosine (red), and Guanine (yellow). The alignment shows a high degree of conservation, with many positions having the same color across all sequences, indicating identical or very similar nucleotides at those sites.

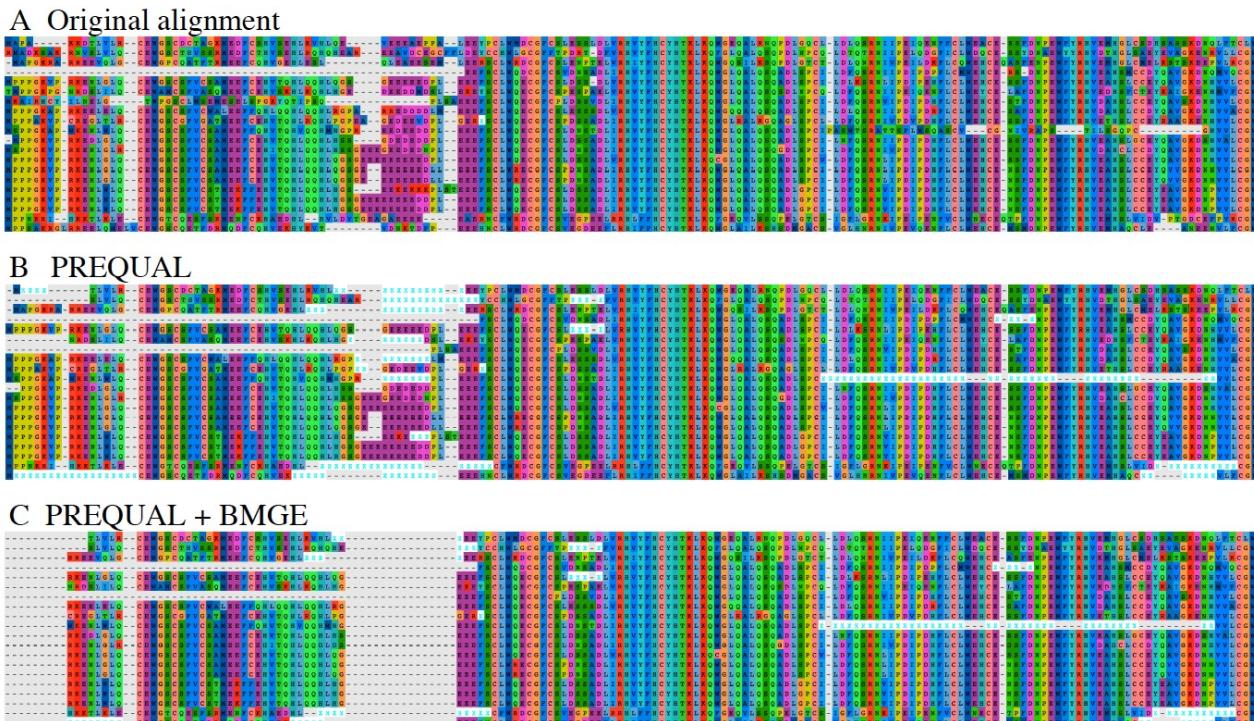
A color-coded multiple sequence alignment of protein-coding genes, similar to the one above. The sequences are aligned vertically, and each nucleotide position is represented by a colored square. The colors represent the four bases: Adenine (blue), Thymine (green), Cytosine (red), and Guanine (yellow). The alignment shows a high degree of conservation, with many positions having the same color across all sequences, indicating identical or very similar nucleotides at those sites.

To trim or not to trim?

Syst. Biol. 56(4):564–577, 2007
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150701472164

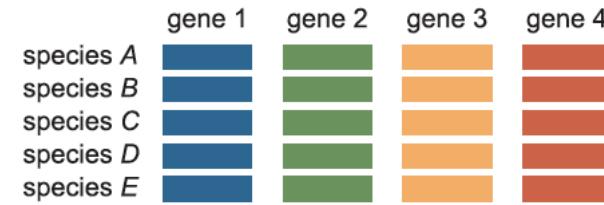
Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments

GERARD TALAVERA AND JOSE CASTRESANA

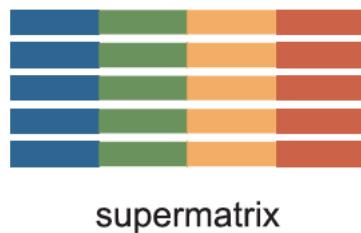


- Gblocks
- BMGE
- TrimAL
- ClipKit

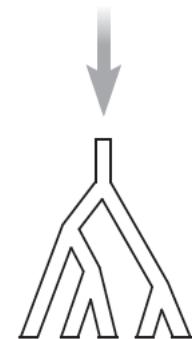
Dataset assembly



concatenation



supermatrix



RAxML, IQTree, PhyML
MrBayes, PhyloBayes, RevBayes

'single step' coalescent



"Full" coalescent methods,
reconciliation
*BEAST, BEST, etc.

'two-step' coalescent



estimated gene trees



"Summary" coalescent methods
ASTRAL, NJst, STAR, etc.

Thank you!

NEW TECHNOLOGIES FOR DEVELOPMENTAL EVOLUTIONARY STUDIES

intensive postgrad course combining theoretical and hands-on sessions

15th- 19th December 2025

Isabel Almudi
Manuel Irimia
Iker Irisarri
Christoph Liedtke
Ignacio Maeso
Marta Portela
Ana Riesgo
María Roselló
Maria Eleonora Rossi
Jordi Solana
Juan Tena
Aida Verdes

- . comparative transcriptomics
- . single cell transcriptomics
- . bulk transcriptomics
- . spatial transcriptomics
- . ATAC-seq technique
- . phylogenomics
- . comparative genomics
- . gene regulation
- . plastic phenotypes
- . evolutionary novelties



FACULTAD
DE CIENCIAS

Departamento
de Biología

SESBe:
Sociedad
Española de
Biología
Evolutiva

+INFO & REGISTRATION

<https://sites.google.com/view/evodevo2025/home>
application due: november 15th, 2025

