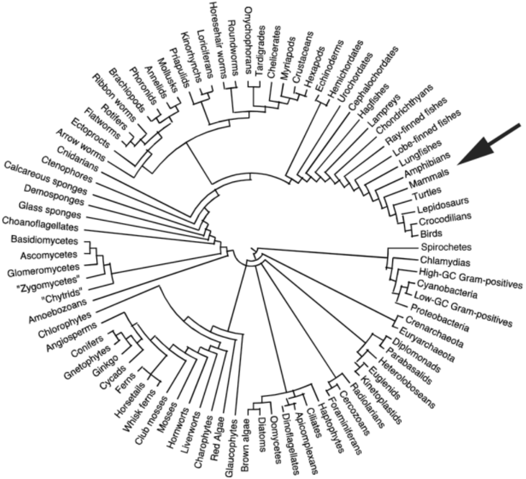
Open Tree of Life resources connect phylogenetic data

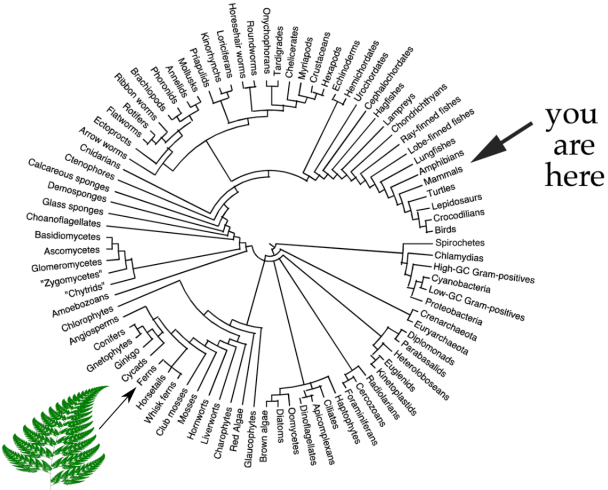
Emily Jane McTavish

University of California, Merced
`ejmctavish@ucmerced.edu`,



you
are
here

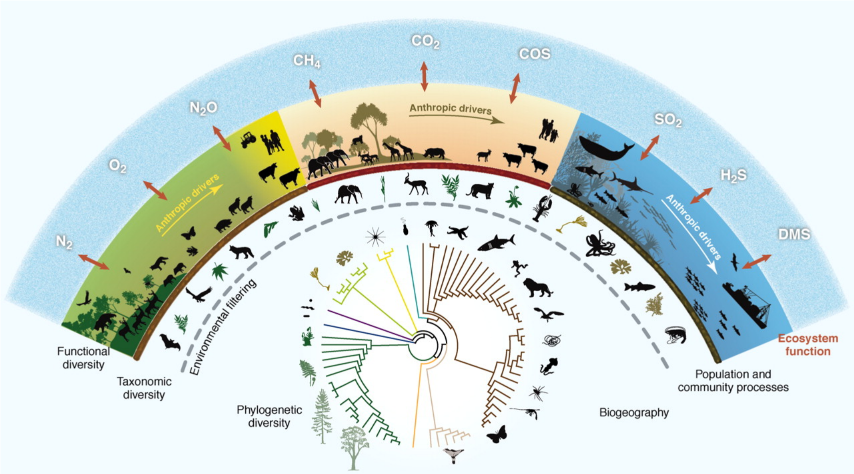
Image Ethan Hein



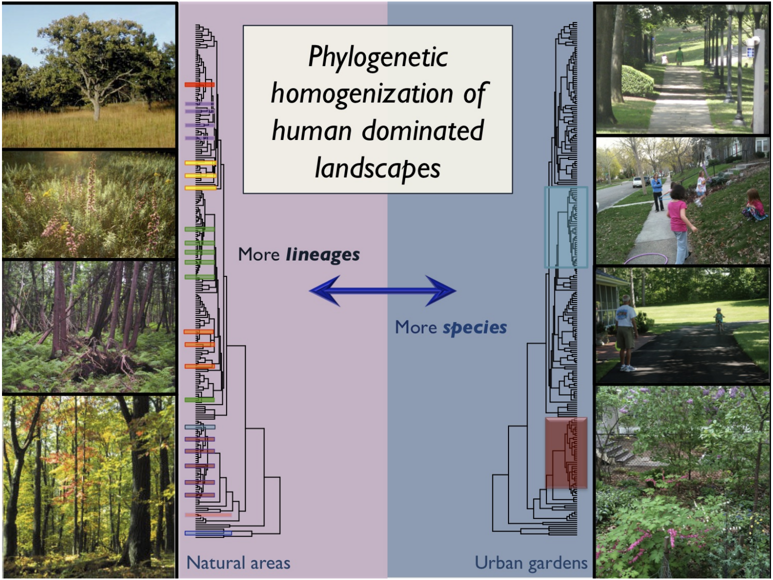
you
are
here

Image Ethan Hein

Evolutionary context provides a framework for **understanding and conserving** global biodiversity



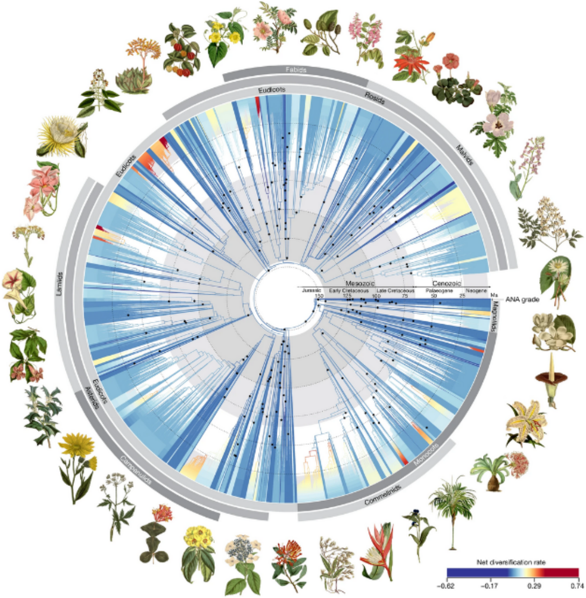
Naeem et al. (2012)



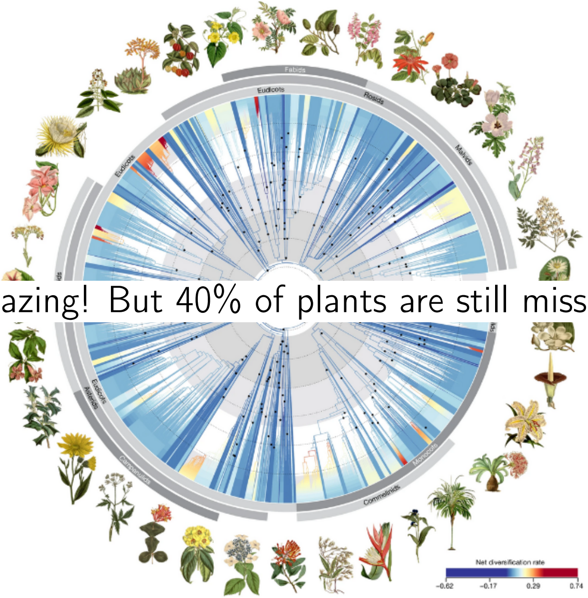
(example from the Cavender-Bares lab webpage)

New and improved evolutionary trees are published all the time!

Recent publication analyzed relationships across 60% of all plant taxa - 8,000 species! (Zuntini et al. 2024)



Recent publication analyzed relationships across 60% of all plant taxa - 8,000 species! (Zuntini et al. 2024)



Amazing! But 40% of plants are still missing.

Taxonomy is often used as a proxy for shared evolutionary history

at best taxonomy is a coarse representation of evolutionary history

at worst taxonomy is a mis-representation of evolutionary history

at worst taxonomy is a mis-representation of evolutionary history

Traditional Taxonomic Groupings Mask Evolutionary History: A Molecular Phylogeny and New Classification of the Chromodorid Nudibranchs

Rebecca Fay Johnson ✉, Terrence M. Gosliner

Johnson and Gosliner (2012)

Why do people use taxonomy as a proxy for shared evolutionary history?

Why do people use taxonomy as a proxy for shared evolutionary history?

Phylogenies:

don't include all the species we are interested in,
keep changing,
are hard to access.

OPEN
Tree of Life

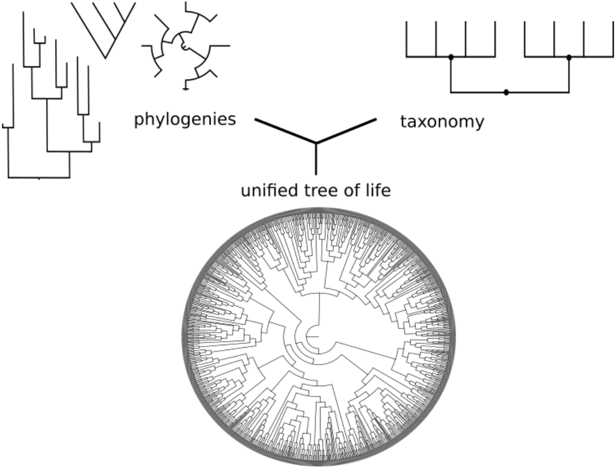
Goal: Build a tree of all life.

OPEN
Tree of Life

Goal: Build a tree of all life.

- Every named species
- Updated as new data becomes available
- Freely and easily accessible

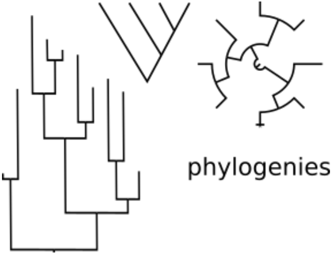
`tree.opentreeoflife.org`



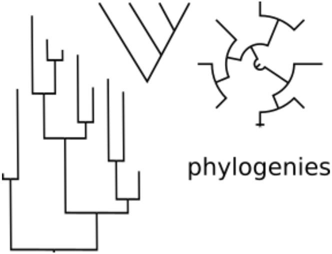
phylogenies

taxonomy

unified tree of life



phylogenies



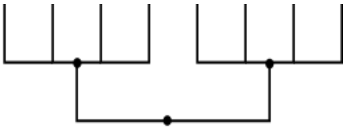
Current synthetic tree

1,362 representative phylogenies

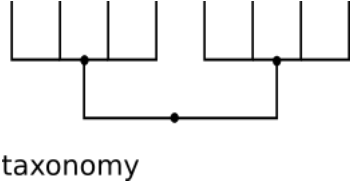
129,778 tips from phylogeny, 2,384,572 total tips.

Current draft posted July 2024, will be updated with more input trees regularly

Synthesis algorithm published in Redelings and Holder (2017), recent major speed improvements



taxonomy



- 4.5 million named taxa, around 2.4 million tips
- Merges NCBI, Index fungorum, Silva, IRMNG, GBIF and other taxonomies
- Scaffold for combining ranked phylogenetic estimates
- New drafts released as inputs change, most recent update June 2023

Rees and Cranston (2017)

Open Tree taxonomy: Limulus polyphemus

The current taxonomy version is ott3.3 (click for more information). See the OTT documentation for an explanation of the taxon flags used below, e.g., `extinct`

Taxon details

species **Limulus polyphemus** ncbi:6850 (worms:150514, gbif:1010610, irmng:10838080) (OTT id 511973)
View this taxon in the current synthetic tree

Synonym(s)

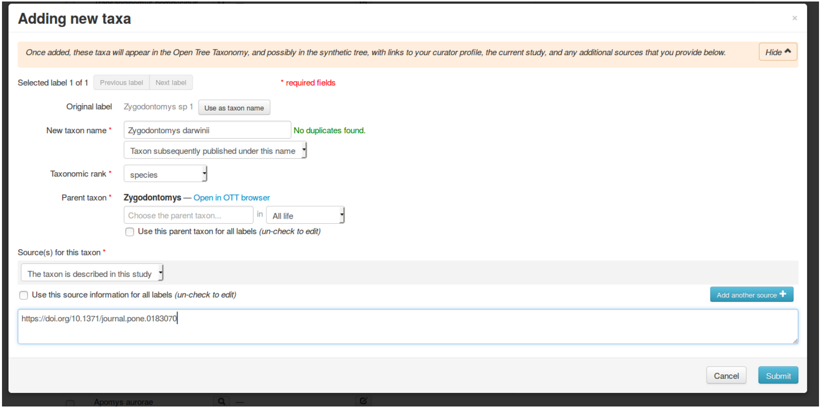
? polyphemus, **Limulus albus**, **Limulus americanus**, **Limulus cyclops**, **Limulus occidentalis**, **Limulus sowerbii**, **Xiphosura polyphemus**

Lineage

life > **cellular organisms** > **Eukaryota** > **Opisthokonta** > **Holozoa** > **Metazoa** > **Eumetazoa** > **Bilateria** > **Protostomia** > **Ecdysozoa** > **Panarthropoda** > **Arthropoda** > **Chelicerata** > **Merostomata** > **Xiphosura** > **Limulidae** > **Limulus**

Taxonomic amendments

- New taxa can be added from uploaded trees, and will be included in future synthetic trees
- Opportunity to feed-back to input taxonomic resources



Adding new taxa

Once added, these taxa will appear in the Open Tree Taxonomy, and possibly in the synthetic tree, with links to your curator profile, the current study, and any additional sources that you provide below.

Selected label 1 of 1

* required fields

Original label: Zygodontomys sp 1

New taxon name *: Zygodontomys darwinii — No duplicates found.

Taxon subsequently published under this name

Taxonomic rank *: species

Parent taxon *: **Zygodontomys** — Open in OTT browser

Choose the parent taxon... in All life

☐ Use this parent taxon for all labels *(un-check to edit)*

Source(s) for this taxon *

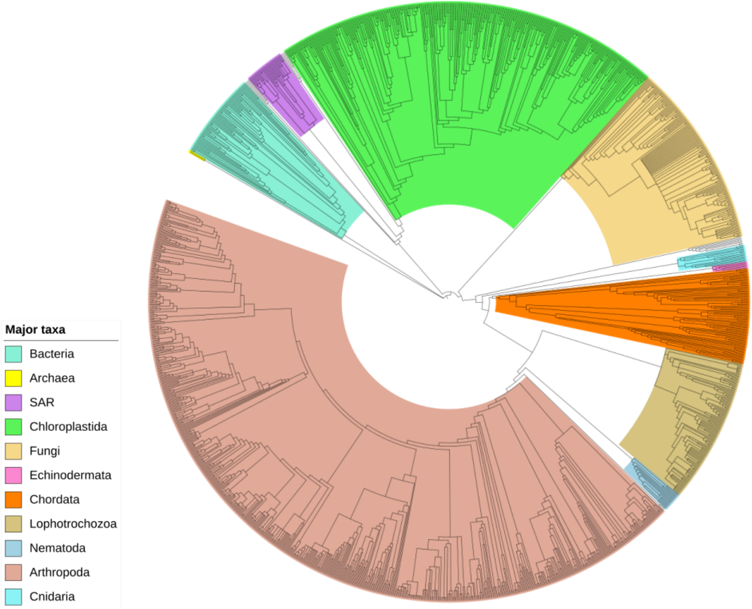
The taxon is described in this study

☐ Use this source information for all labels *(un-check to edit)*

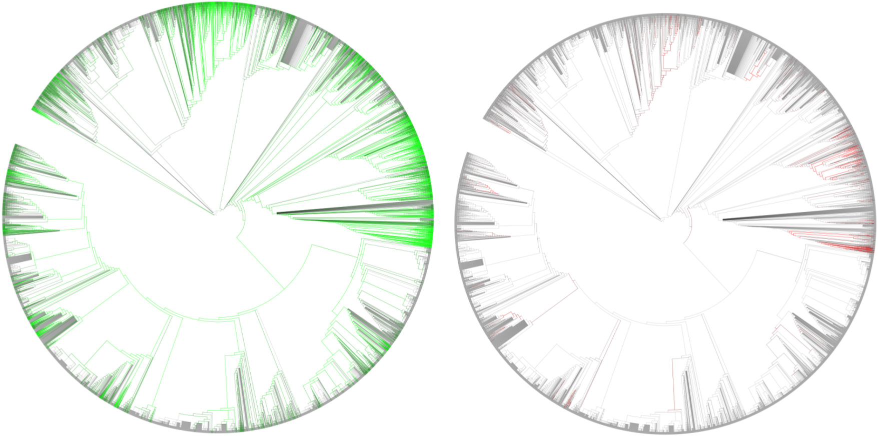
https://doi.org/10.1371/journal.pone.0183070

Apomys aurorae

Tree of tips with 500 or more descendants in OpenTree (1,408 tips, supported by 442 published studies)



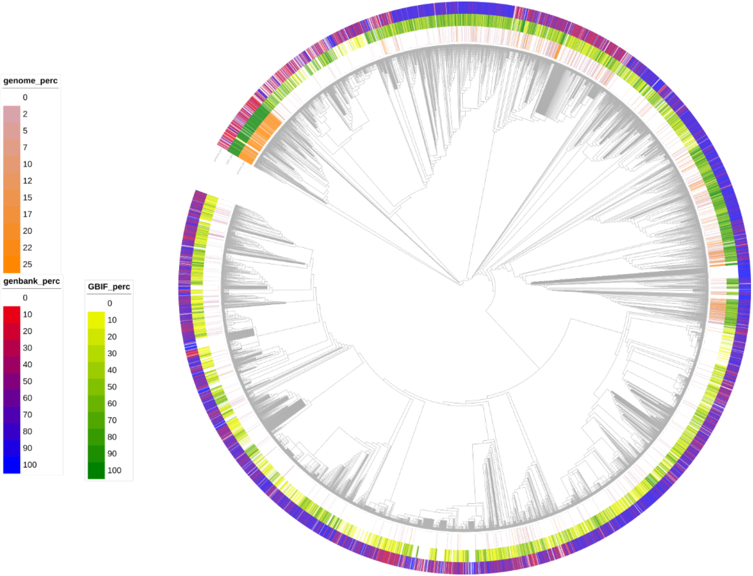
Phylogenetic information (green) and conflict (red) across the tree



Annotated tree available at

`https://itol.embl.de/tree/1082301078315960159622895`

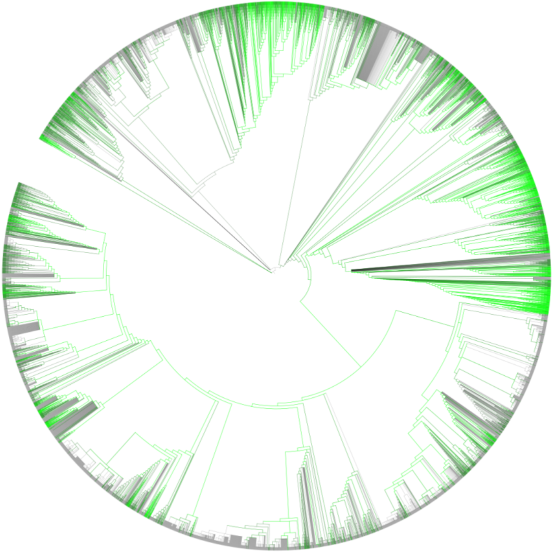
Cross-link data from other resources



tree viewer demo

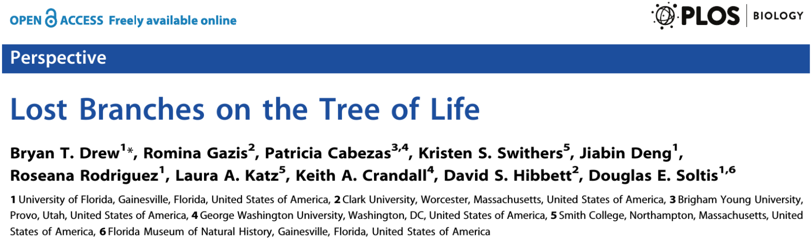
`https://tree.opentreeoflife.org/opentree/argus/opentree15.1@mrcaott30845ott82415/Amazonetta--Sibirionetta`

We are missing phylogenetic data for many lineages. How will we fill in the gaps?



We need to build on existing phylogenetic information.

We need to build on existing phylogenetic information.



OPEN ACCESS Freely available online

PLOS | BIOLOGY

Perspective

Lost Branches on the Tree of Life

Bryan T. Drew[1]*, Romina Gazis[2], Patricia Cabezas[3,4], Kristen S. Swithers[5], Jiabin Deng[1], Roseana Rodriguez[1], Laura A. Katz[5], Keith A. Crandall[4], David S. Hibbett[2], Douglas E. Soltis[1,6]

1 University of Florida, Gainesville, Florida, United States of America, **2** Clark University, Worcester, Massachusetts, United States of America, **3** Brigham Young University, Provo, Utah, United States of America, **4** George Washington University, Washington, DC, United States of America, **5** Smith College, Northampton, Massachusetts, United States of America, **6** Florida Museum of Natural History, Gainesville, Florida, United States of America

Drew et al. (2013)

only 16% of phylogenies published 2000-2013 are digitally available Drew et al. PLoS Biology 2013

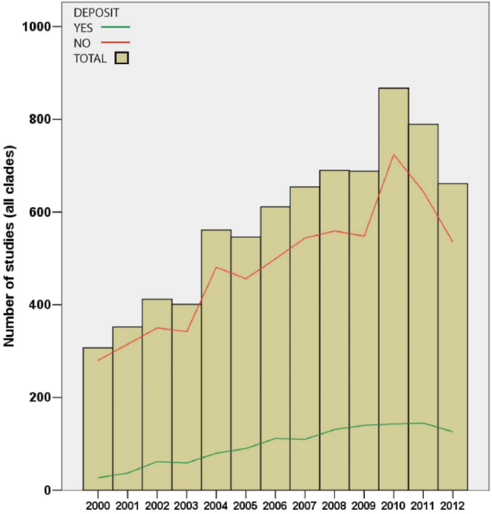


Figure 1. Overview of total number of publications surveyed from animal, fungus, seed plant, microbial eukaryote, archaea, and bacteria literature (indicated in red), and the number of those publications that archived their trees and alignments in either Dryad or TreeBASE (indicated in green).
doi:10.1371/journal.pbio.1001636.g001

only 16% of phylogenies published 2000-2013 are digitally available (Drew et al. PLoS Biology 2013)

20% of phylogenies published 2013-2018

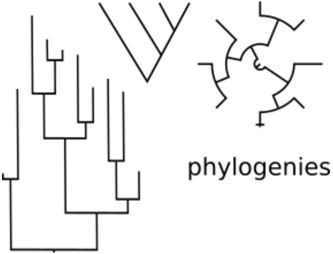
McTavish et al. (2017)

Recent multi-month outage of treebase makes future of phylogenetic data sharing even more concerning

Adding phylogenetic data to OpenTree data store

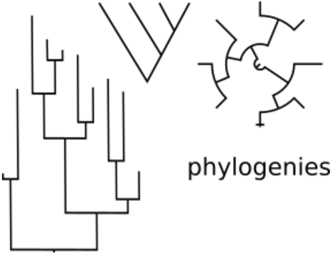
- Trees can be uploaded from any source, does not have to be own data.
- Easy to use browser based interface
- Track curation attribution by name or pseudonym
- Files are json representation of NeXML phylogenetic data format
- Data store is hosted publicly on GitHub

`github.com/OpenTreeOfLife/phylesystem-1`
McTavish et al. (2015)



Community Curation

363 individual curators of 5,241 uploaded studies



Community Curation

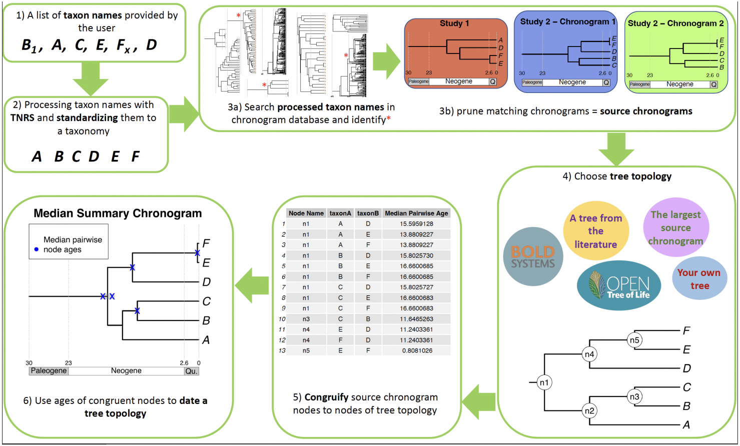
363 individual curators of 5,241 uploaded studies

Currently a several month lag for incorporation into synthetic tree

Date estimates

- The synthetic tree does not currently have any branch lengths (when combining across DNA, morphology and taxonomy, branch lengths are not obvious!!)
- However! We can translate dates from input trees to the synthetic tree

DATElife



1) A list of **taxon names** provided by the user

B_1, A, C, E, F_x, D

2) Processing taxon names with **TNRS** and **standardizing** them to a taxonomy

A B C D E F

3a) Search **processed taxon names** in chronogram database and identify*

3b) prune matching chronograms = **source chronograms**

4) Choose **tree topology**

	Node Name	taxonA	taxonB	Median Pairwise Age
1	n1	A	D	15.5959128
2	n1	A	E	13.8809227
3	n1	A	F	13.8809227
4	n1	B	D	15.8025730
5	n1	B	E	16.6600685
6	n1	B	F	16.6600685
7	n1	C	D	15.8025727
8	n1	C	E	16.6600683
9	n1	C	F	16.6600683
10	n3	C	B	11.6465263
11	n4	E	D	11.2403361
12	n4	F	D	11.2403361
13	n5	E	F	0.8081026

5) **Congruify** source chronogram nodes to nodes of tree topology

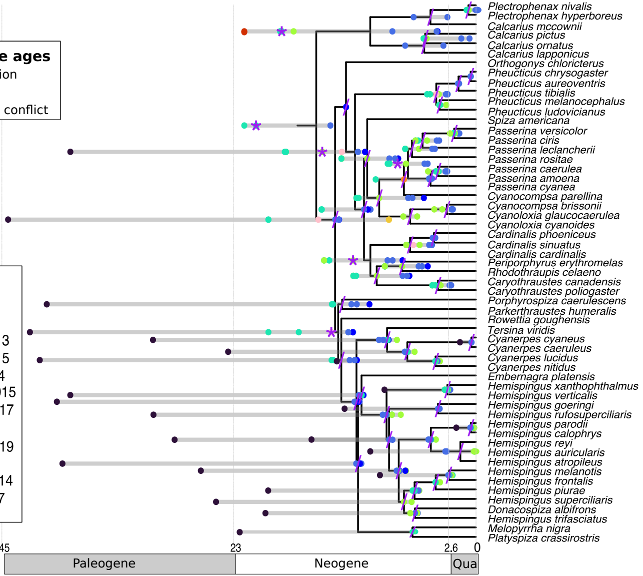
Median Summary Chronogram

6) Use ages of congruent nodes to **date a tree topology**

Webserver at `datelife.opentreeoflife.org`

Sanchez-Reyes, McTavish, O'Meara, Systematic Biology, 2024

Dated finch tree - topology estimated based on 33 studies, dates from 11 studies



Median of node ages

/ used as calibration

* not used due to parent-child age conflict

Input study node age estimates

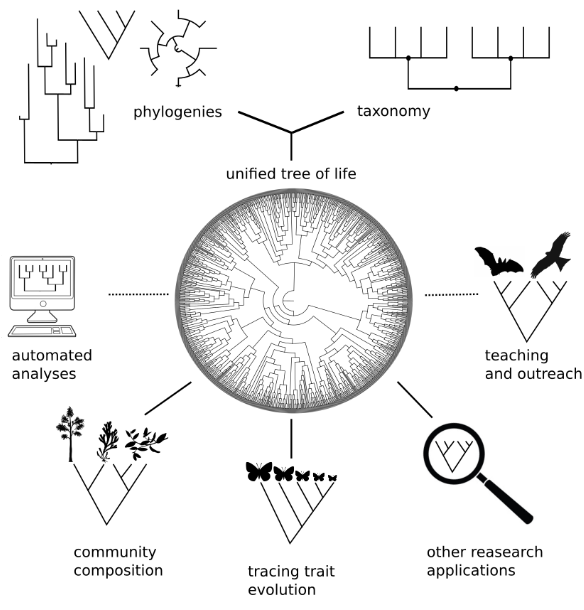
- ● Barker et al. 2013
- ● Barker et al. 2015
- ● Burns et al. 2014
- ● Hedges et al. 2015
- ● Hooper et al. 2017
- ● Jetz et al. 2012
- ● Kimball et al. 2019
- ● Price et al. 2014
- ● Roquet et al. 2014
- ● Uyeda et al 2017

Sanchez-Reyes, McTavish, O'Meara, Systematic Biology, 2024

OpenTree Dates API

- Match internal nodes of dated inputs to nodes in the synthetic tree (phyloreferenceing approach)
- Apply dates to nodes induced subtree
- Smooth undated nodes using bladj (Webb et al. (2008))
- Beta functionality available at `dates.opentreeoflife.org`
- Date estimates for 46,639 internal nodes from 320 input chronograms

`https://github.com/OpenTreeOfLife/ChronoSynth`



McTavish et al. (2017)

Open Tree resources are available via a range of implementations

- Browser interface, `tree.opentreeoflife.org`
- Open Tree of Life API
- Python-opentree python wrapper McTavish et al. (2021)
- R Open Tree of Life (rotl) Michonneau et al. (2016)

CC0 license provides fully open access for downstream re-usability

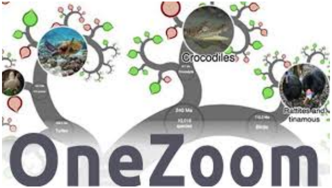
CC0 license provides fully open access for downstream re-usability

Open Tree provides the tree backend for:

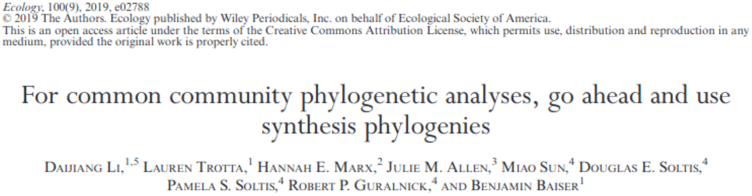
PHYLOtastic

Lifemap
OPEN Tree of Life

DATElife



Why would you need OpenTree if you can build your own trees?



Ecology, 100(9), 2019, e02788
© 2019 The Authors. Ecology published by Wiley Periodicals, Inc. on behalf of Ecological Society of America.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

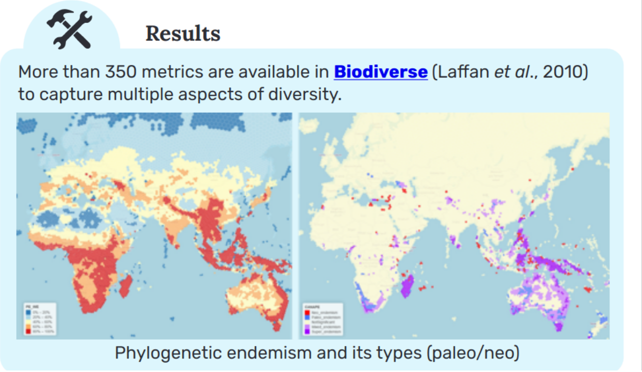
For common community phylogenetic analyses, go ahead and use synthesis phylogenies

DAIJIANG LI,[1,5] LAUREN TROTTA,[1] HANNAH E. MARX,[2] JULIE M. ALLEN,[3] MIAO SUN,[4] DOUGLAS E. SOLTIS,[4]
PAMELA S. SOLTIS,[4] ROBERT P. GURALNICK,[4] AND BENJAMIN BAISER[1]

"our results justify taking advantage of recently developed and continuously improving synthesis trees, especially the Open Tree of Life." Li et al. (2019)

Assess phylogenetic diversity at large scales

PhyloNext: a pipeline for phylogenetic diversity analysis of GBIF-mediated data in the cloud:



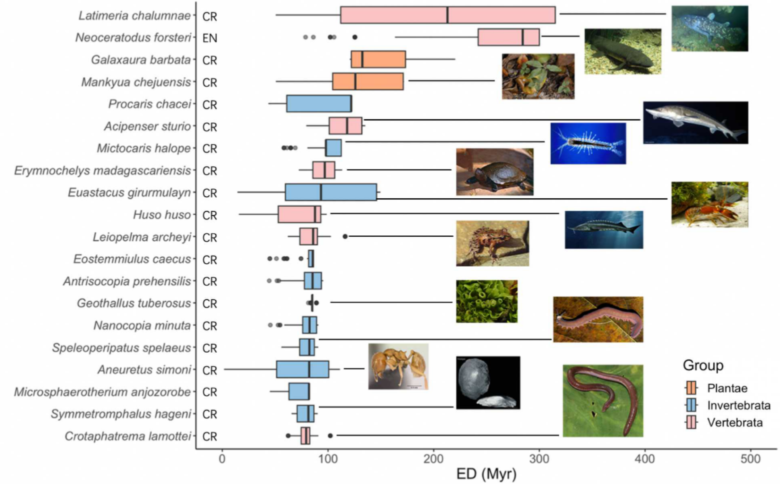
Results

More than 350 metrics are available in **Biodiverse** (Laffan *et al.*, 2010) to capture multiple aspects of diversity.

Phylogenetic endemism and its types (paleo/neo)

(Mikryukov et al., 2024)

Measure potential for phylogenetic diversity loss due to extinctions



Species at risk of extinction (per IUCN red list) which could result in greatest loss of evolutionary distinct history

(Guo, McTavish, Rosindell, preprint at `https://www.researchsquare.com/article/rs-4397985/v1`)

Easily get accurate relationships (and citations!) for arbitrary sets of species

Easily get accurate relationships (and citations!) for arbitrary sets of species

Which species with an existing reference genome is most closely related to the Society finch?

Lonchura striata domestica

Zebra finch
(*Taeniopygia guttata*)

Ground finch
(*Geospiza fortis*)

Access OPEN Tree of Life via `rotl` package from ROpenSci

Michonneau et al. (2016))

Access OPEN Tree of Life via `rotl` package from rOpenSci

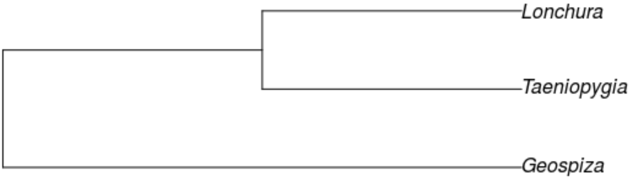
Michonneau et al. (2016))

```
install.packages("rotl")
library(rotl)
spp <- c("Geospiza", "Taeniopygia", "Lonchura")
taxa <- tnrs_match_names(spp, context="Animals")
tr <- tol_induced_subtree(ott_id(taxa), label="name")
plot(tr)
```

Access OPEN Tree of Life via `rotl` package from ROpenSci

Michonneau et al. (2016))

```
install.packages("rotl")
library(rotl)
spp <- c("Geospiza", "Taeniopygia", "Lonchura")
taxa <- tnrs_match_names(spp, context="Animals")
tr <- tol_induced_subtree(ott_id(taxa), label="name")
plot(tr)
```

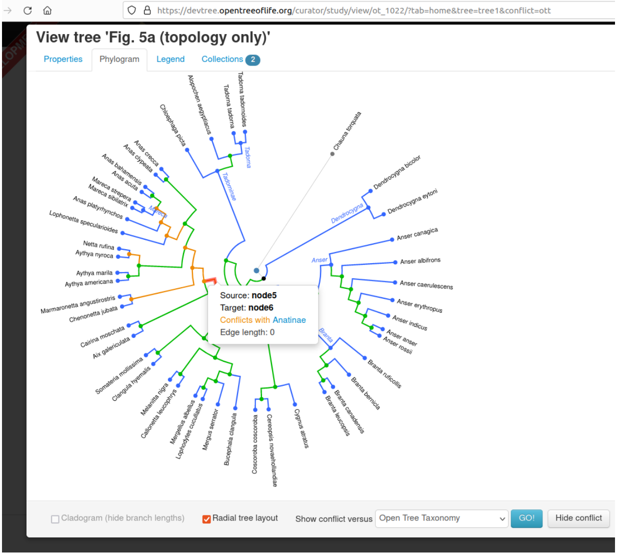


Use the zebra finch as reference!

Phylogenies generating subtree:

(Barker et al. 2004 PNAS; Selvatti et al. 2015 MPE; Burns et al. 2014 MPE; Barker et al. 2015 The Auk)

Compare new phylogenetic estimates to taxonomy, synthetic tree, or other phylogeneies



https://devtree.opentreeoflife.org/curator/study/view/ot_1022/?tab=home&tree=tree1&conflict=ott

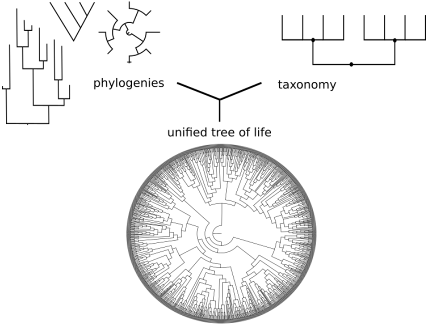
View tree 'Fig. 5a (topology only)'

Source: **node5**
Target: **node6**
Conflicts with Anatinae
Edge length: 0

☐ Cladogram (hide branch lengths)
☑ Radial tree layout
Show conflict versus Open Tree Taxonomy

https://tree.opentreeoflife.org/curator/study/view/ot_1022/?tab=home&tree=tree1&conflict=ott
https://tree.opentreeoflife.org/curator/study/view/ot_500/?tab=home&tree=tree1&conflict=pg_420@tree522

Generate a custom synthetic tree for your taxa



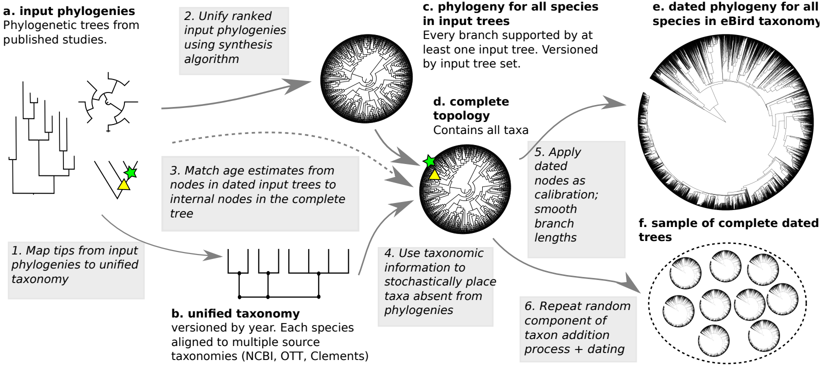
Synthesis on demand

Personal phylogeny rankings

Any root node

Synthesis phylogeny of all birds (McTavish et al., 2025)
Collaboration with Eliot Miller and others at Cornell Lab of Ornithology

- 321 published trees with a total of 48,188 tips (already updated since April publication!)
- Tree files collected from treebase, online data deposits, and many, many, MANY, personal emails haranguing authors
- All input trees are deposited, with metadata, in public Open Tree phylogenetic datastore, *phylesystem*
- These trees include 9,789 of 11,017 (87%) species in the 2024 Clements taxonomy



a. input phylogenies
Phylogenetic trees from published studies.

1. Map tips from input phylogenies to unified taxonomy

b. unified taxonomy
versioned by year. Each species aligned to multiple source taxonomies (NCBI, OTT, Clements)

2. Unify ranked input phylogenies using synthesis algorithm

3. Match age estimates from nodes in dated input trees to internal nodes in the complete tree

c. phylogeny for all species in input trees
Every branch supported by at least one input tree. Versioned by input tree set.

d. complete topology
Contains all taxa

4. Use taxonomic information to stochastically place taxa absent from phylogenies

5. Apply dated nodes as calibration; smooth branch lengths

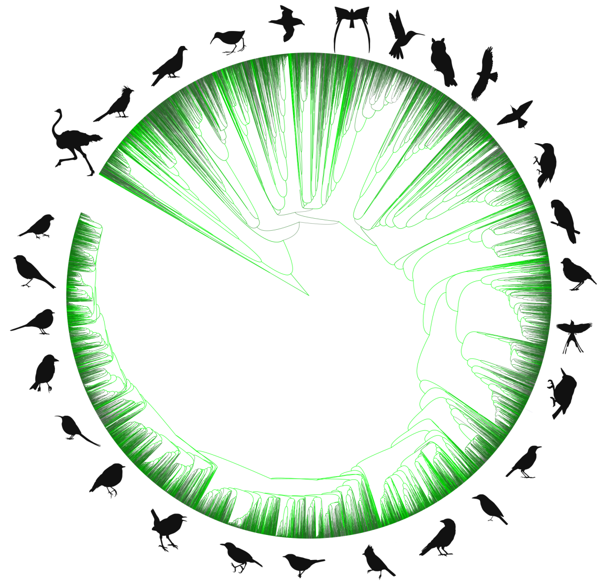
e. dated phylogeny for all species in eBird taxonomy

f. sample of complete dated trees

6. Repeat random component of taxon addition process + dating

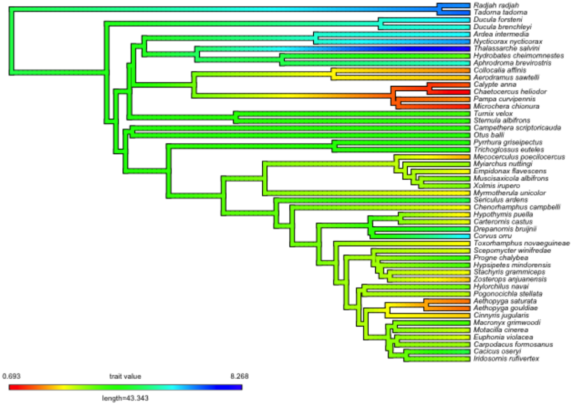
Data products in `https://github.com/McTavishLab/AvesData`

Synthesis phylogeny of all birds - 87% phylogenetic coverage



McTavish et al. (2025)

R package to access bird trees available on CRAN *clootl*
Easy to access the phylogeny, and connect to large scale data sets.



Log body mass of sampled birds, using AVONET data (Tobias et al., 2022) Tutorial at `https://github.com/eliotmiller/clootl/blob/master/examples/avonet.md`

We're applying this apporach to other taxa - currently fish (actinopterygii), ants, and amphibians. Get in touch if you're interested in being involved!

Conclusions

- Phylogenetic estimates should be freely accessible and reusable
- Open Tree tools cross-link phylogenetic and taxonomic information
- A variety of tools and approaches provides wide access to Open Tree resources

Contribute your knowledge!

`tree.opentreeoflife.org/curator`

OPEN
Tree of Life

Try it out! A variety of tutorials are available at:
`https://opentreeoflife.github.io/use`

Ask for help on our gitter chat:
`https://gitter.im/OpenTreeOfLife/public`

Todays tutorial:

Accessing, comparing, and dating phylogenies using Open Tree of Life

`https://github.com/snacktavish/TreeUpdatingComparison/blob/master/TreeComparison.md`

Thank You

OPEN
Tree of Life

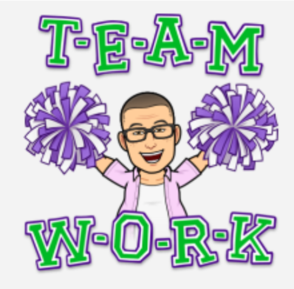
NSF ABI 1759846
Mark Holder
Karen Cranston

Bird tree:
Eliot T. Miller
Jeff A. Gerbracht
Mark T. Holder
Marshall J. Iliff
Denis Lepage
Pam Rasmussen

Dendropy Jeet Sukumaran

Developer team:
Luna Luisa Sanchez Reyes
Ben Redelings
Jim Allman

UNIVERSITY OF CALIFORNIA
UCMERCED



Drew, B. T., Gazis, R., Cabezas, P., Swithers, K. S., Deng, J., Rodriguez, R., Katz, L. A., Crandall, K. A., Hibbett, D. S., and Soltis, D. E. (2013). Lost Branches on the Tree of Life. *PLoS Biology*, 11(9):e1001636. Number: 9.

Johnson, R. F. and Gosliner, T. M. (2012). Traditional Taxonomic Groupings Mask Evolutionary History: A Molecular Phylogeny and New Classification of the Chromodorid Nudibranchs. *PLOS ONE*, 7(4):e33479.

Li, D., Trotta, L., Marx, H. E., Allen, J. M., Sun, M., Soltis, D. E., Soltis, P. S., Guralnick, R. P., and Baiser, B. (2019). For common community phylogenetic analyses, go ahead and use synthesis phylogenies. *Ecology*, 100(9).

McTavish, E. J., Drew, B. T., Redelings, B., and Cranston, K. A. (2017). How and Why to Build a Unified Tree of Life. *BioEssays*, 39(11).

McTavish, E. J., Gerbracht, J. A., Holder, M. T., Iliff, M. J., Lepage, D., Rasmussen, P. C., Redelings, B. D., Sánchez Reyes, L. L., and Miller, E. T. (2025). A complete and dynamic tree of birds. *Proceedings of the National Academy of Sciences*, 122(18):e2409658122. Publisher: Proceedings of the National Academy of Sciences.

McTavish, E. J., Hinchliff, C. E., Allman, J. F., Brown, J. W., Cranston, K. A., Holder, M. T., Rees, J. A., and Smith, S. A. (2015). Phylesystem: a git-based data store for community-curated phylogenetic estimates. *Bioinformatics*, 31(17):2794–2800.

McTavish, E. J., Sánchez-Reyes, L. L., and Holder, M. T. (2021). OpenTree: A Python Package for Accessing and Analyzing Data from the Open Tree of Life. *Systematic Biology*, 70(6):1295–1301.

Michonneau, F., Brown, J. W., and Winter, D. J. (2016). rotl: an R package to interact with the Open Tree of Life data. *Methods in Ecology and Evolution*, pages n/a–n/a.

Mikryukov, V., Abarenkov, K., Laffan, S., Robertson, T., McTavish, E. J., Jeppesen, T. S., Waller, J., Blissett, M., Kõljalg, U., and Miller, J. T. (2024). PhyloNext: a pipeline for phylogenetic diversity analysis of GBIF-mediated data. *BMC Ecology and Evolution*, 24(1):76.

Naeem, S., Duffy, J. E., and Zavaleta, E. (2012). The Functions of Biological Diversity in an Age of Extinction. *Science*, 336(6087):1401–1406.

Redelings, B. D. and Holder, M. T. (2017). A supertree pipeline for summarizing phylogenetic and taxonomic information for millions of species. *PeerJ*, 5:e3058.

Rees, J. A. and Cranston, K. (2017). Automated assembly of a reference taxonomy for phylogenetic data synthesis. *Biodiversity Data Journal*.

Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J. M., Yang, J., Sayol, F., Neate-Clegg, M. H. C., Alioravainen, N., Weeks, T. L., Barber, R. A., Walkden, P. A., MacGregor, H. E. A., Jones, S. E. I., Vincent, C., Phillips, A. G., Marples, N. M., Montaño-Centellas, F. A., Leandro-Silva, V., Claramunt, S., Darski, B., Freeman, B. G., Bregman, T. P., Cooney, C. R., Hughes, E. C., Capp, E. J. R., Varley, Z. K., Friedman, N. R., Korntheuer, H., Corrales-Vargas, A., Trisos, C. H., Weeks, B. C., Hanz, D. M., Töpfer, T., Bravo, G. A., Remeš, V., Nowak, L., Carneiro, L. S., Moncada R., A. J., Matysioková, B., Baldassarre, D. T., Martínez-Salinas, A., Wolfe, J. D., Chapman, P. M., Daly, B. G., Sorensen, M. C., Neu, A., Ford, M. A., Mayhew, R. J., Fabio Silveira, L., Kelly, D. J., Annorbah, N. N. D., Pollock, H. S., Grabowska-Zhang, A. M., McEntee, J. P., Carlos T. Gonzalez, J., Meneses, C. G., Muñoz, M. C., Powell, L. L., Jamie, G. A., Matthews, T. J., Johnson, O., Brito, G. R. R., Zyskowski, K., Crates, R., Harvey, M. G., Jurado Zevallos, M., Hosner, P. A., Bradfer-Lawrence, T., Maley, J. M., Stiles, F. G., Lima, H. S., Provost, K. L., Chibesa, M., Mashao, M., Howard, J. T., Mlamba, E., Chua, M. A. H., Li, B., Gómez, M. I., García, N. C., Päckert, M., Fuchs, J., Ali, J. R., Derryberry, E. P., Carlson, M. L., Urriza, R. C., Brzeski, K. E., Prawiradilaga, D. M., Rayner, M. J., Miller, E. T., Bowie, R. C. K., Lafontaine, R.-M., Scofield, R. P., Lou, Y., Somarathna, L., Lepage, D., Illif, M., Neuschulz, E. L., Templin, M., Dehling, D. M., Cooper, J. C., Pauwels, O. S. G., Analuddin, K., Fjeldså, J., Seddon, N., Sweet, P. R., DeClerck, F. A. J., Naka, L. N., Brawn, J. D., Aleixo, A., Böhning-Gaese, K., Rahbek, C., Fritz, S. A., Thomas, G. H., and Schleuning, M. (2022). Avonet: morphological, ecological and geographical data for all birds. *Ecology Letters*, 25(3):581–597.

Webb, C. O., Ackerly, D. D., and Kembel, S. W. (2008). Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18):2098–2100.