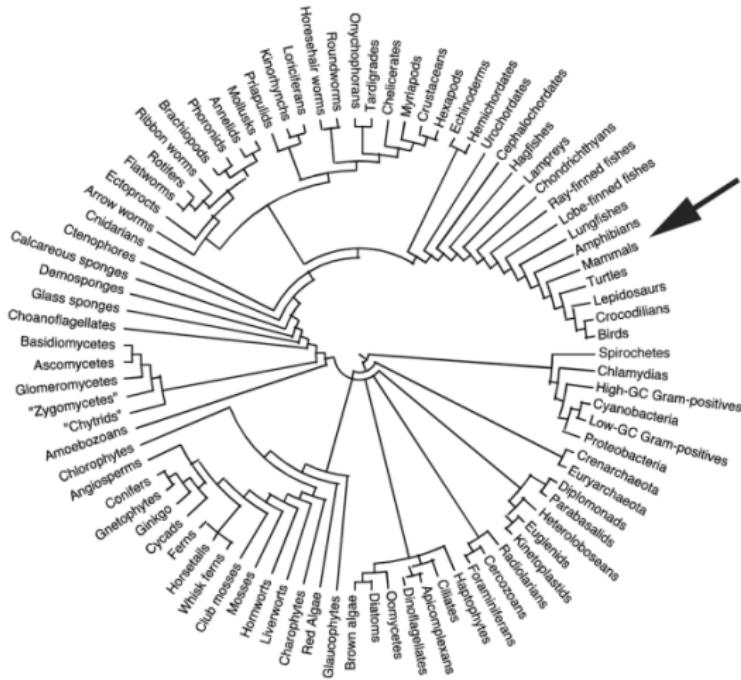


# The Open Tree of Life

Emily Jane McTavish

University of California, Merced  
[ejmctavish@ucmerced.edu](mailto:ejmctavish@ucmerced.edu), Twitter:@snacktavish





you  
are  
here

Image Ethan Hein

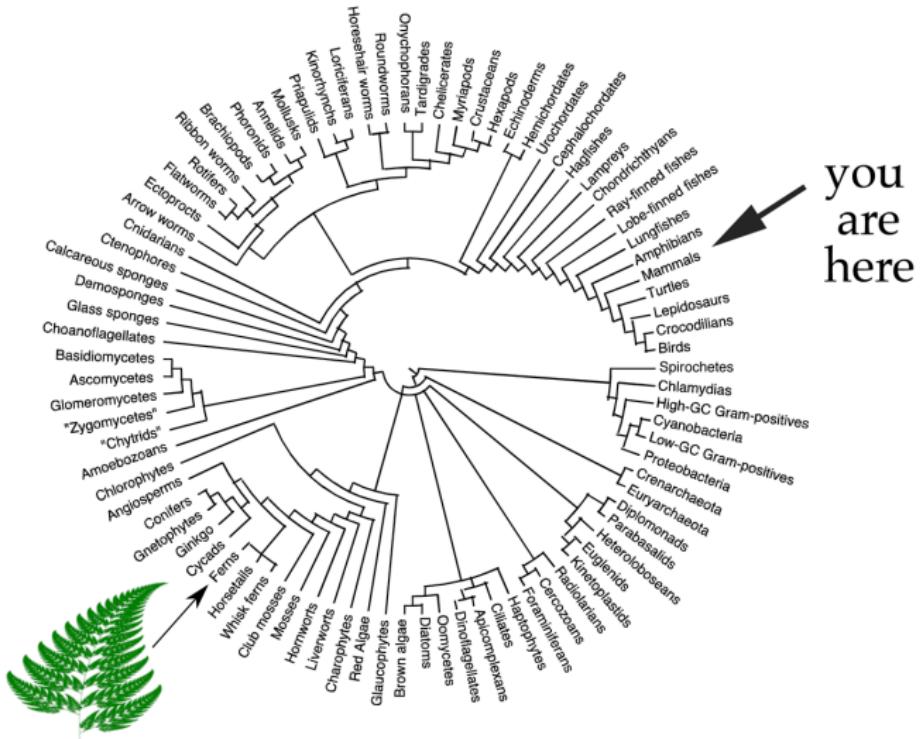
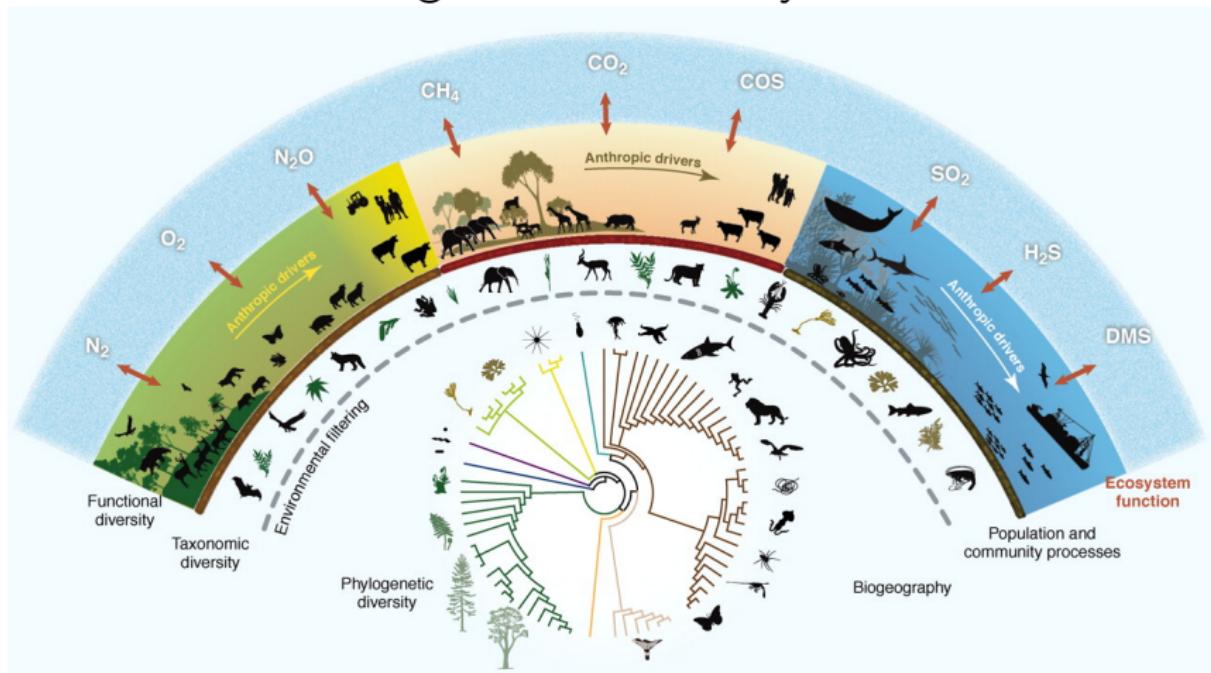


Image Ethan Hein

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



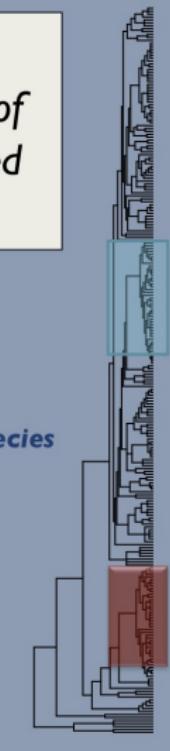


## Phylogenetic homogenization of human dominated landscapes

More *lineages*



More *species*



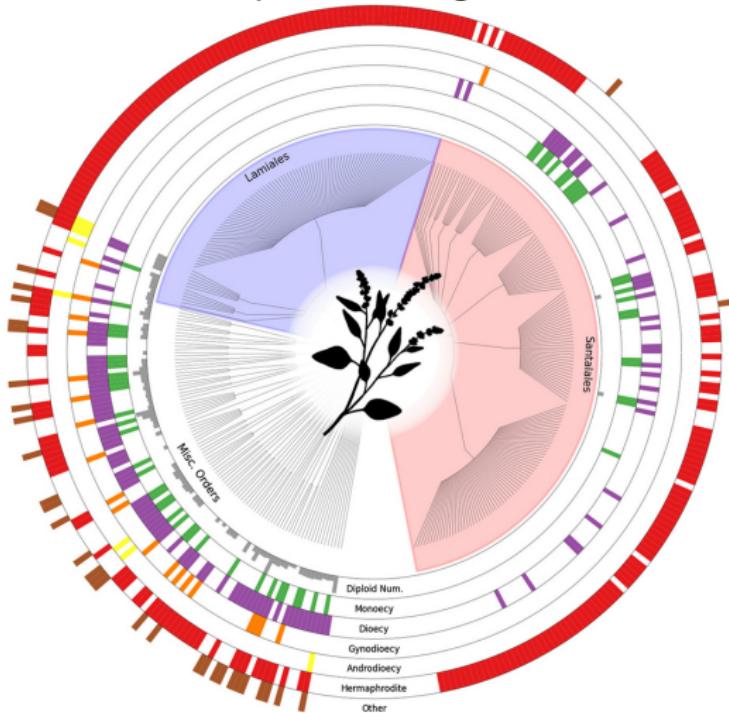
Urban gardens



(example from the Cavender-Bares lab webpage)

Taxonomy is often used as a proxy for shared evolutionary history

# Evolutionary transitions in plant mating



?

**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



?



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

*Why do we 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.



*Goal:* Build a tree of all life.

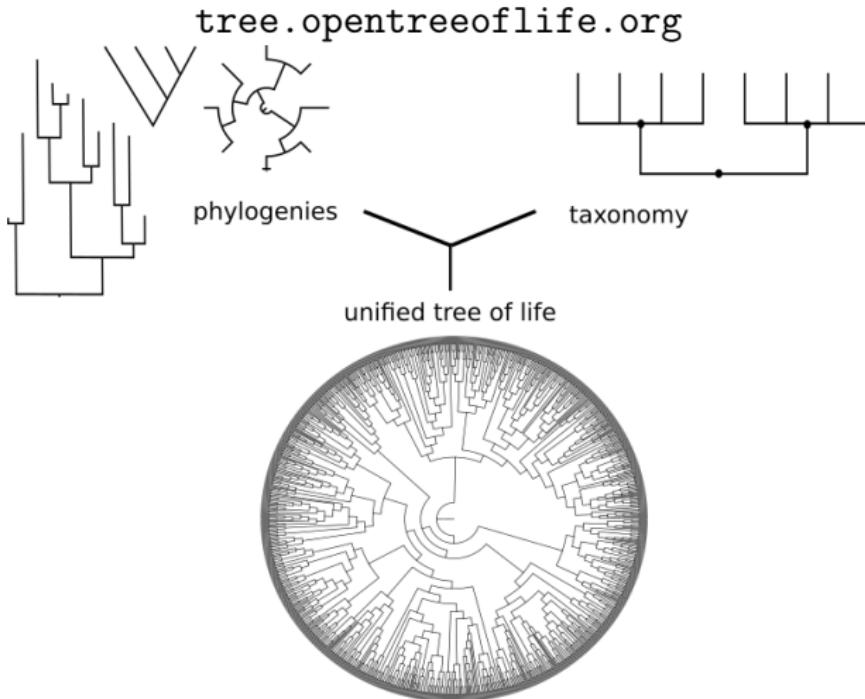


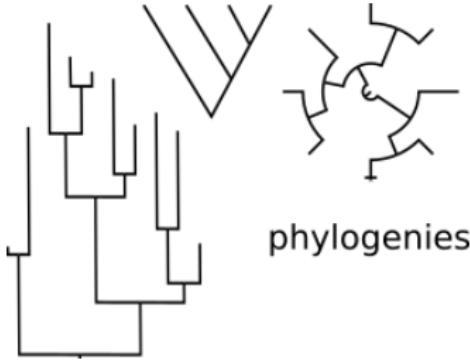
*Goal:* Build a tree of all life.

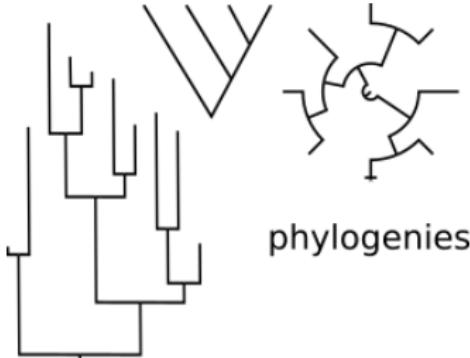
Every named species

Updated as new data becomes available

Freely and easily accessible







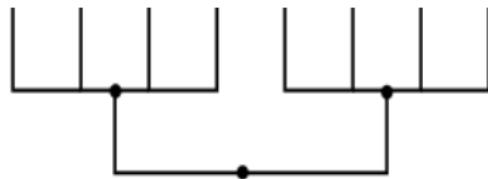
## Current synthetic tree

1,245 representative phylogenies

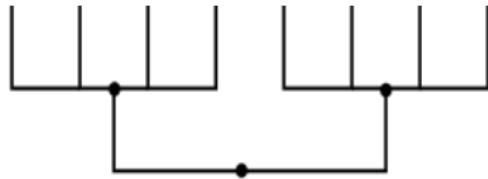
94,028 tips from phylogeny

Current draft posted June, 2021, will be updated with more  
input trees soon

Redelings and Holder, PeerJ 2017



taxonomy



## taxonomy

4.9 million named taxa, around 2.7 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, 2021

Rees and Cranston, Biodiversity Data Journal 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. [Hide](#)

Selected label 1 of 1 [Previous label](#) [Next label](#) \* required fields

Original label  [Use as taxon name](#)

New taxon name \*  No duplicates found.

Taxonomic rank \*

Parent taxon \* [Zygodontomys — Open in OTT browser](#)  
 in   
 Use this parent taxon for all labels (un-check to edit)

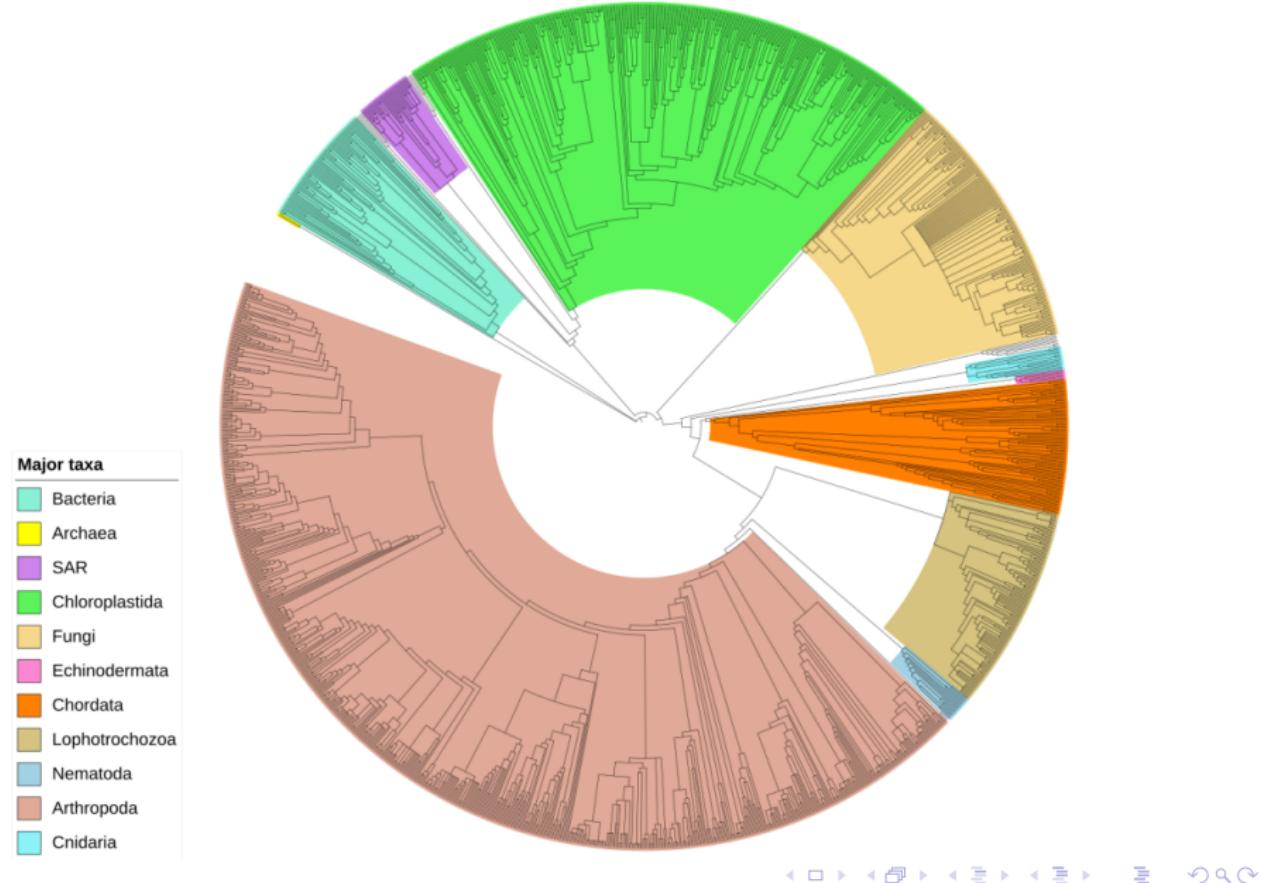
Source(s) for this taxon \*

The taxon is described in this study

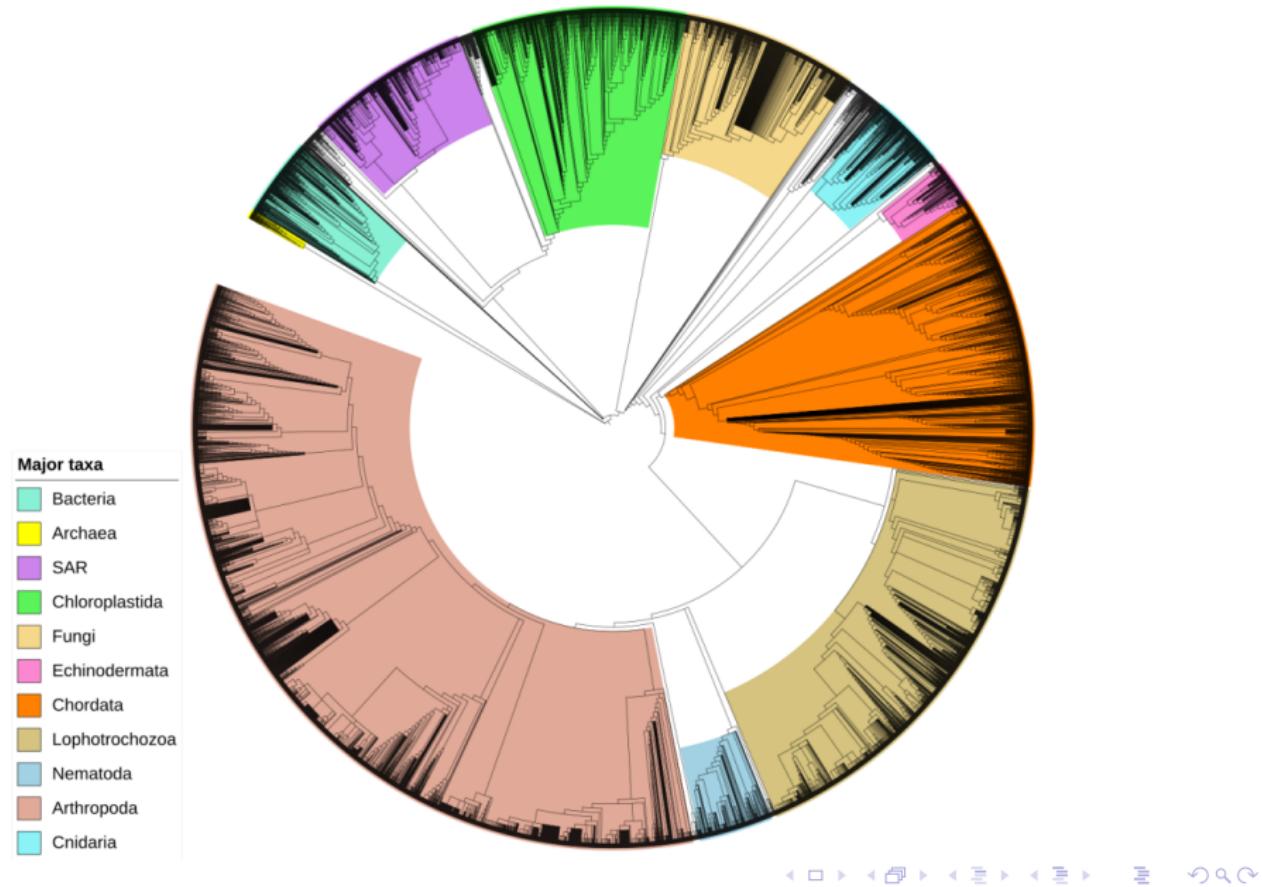
Use this source information for all labels (un-check to edit) [Add another source](#)

[Cancel](#) [Submit](#)

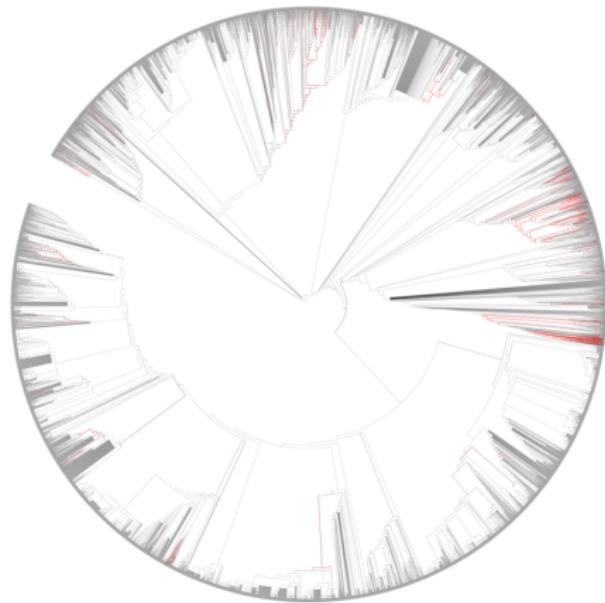
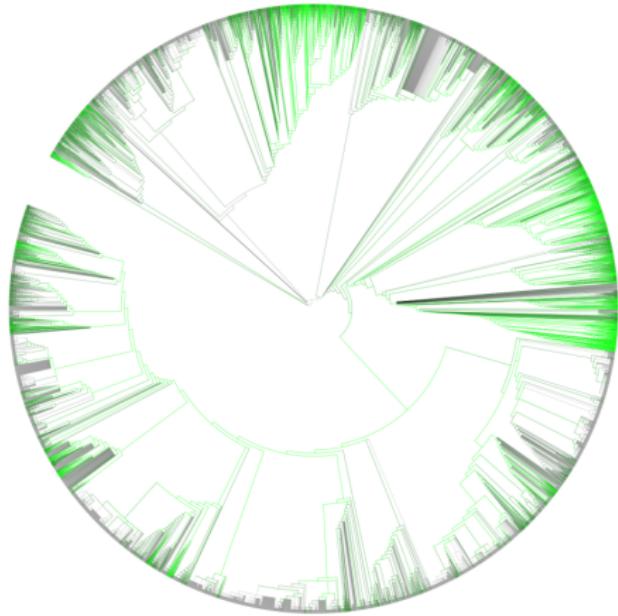
# Tree of tips with 500 or more descendants in OpenTree (1,408 tips, supported by 442 published studies)



# Tree of all families in OpenTree (9,306 tips, supported by 455 published studies)



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



Annotated tree available at

<https://itol.embl.de/tree/10823010783159601596222895>

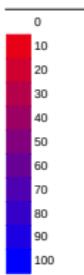
Cross-link data from other resources

## Proportion of

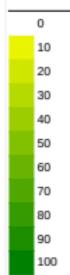
lineages with

data:

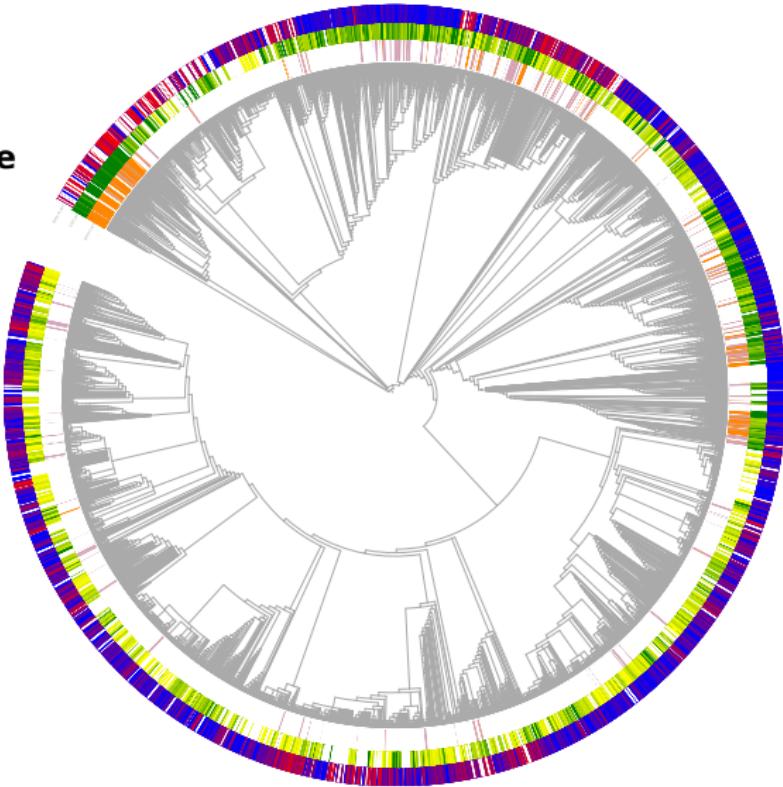
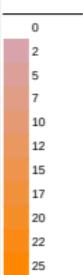
**GenBank**



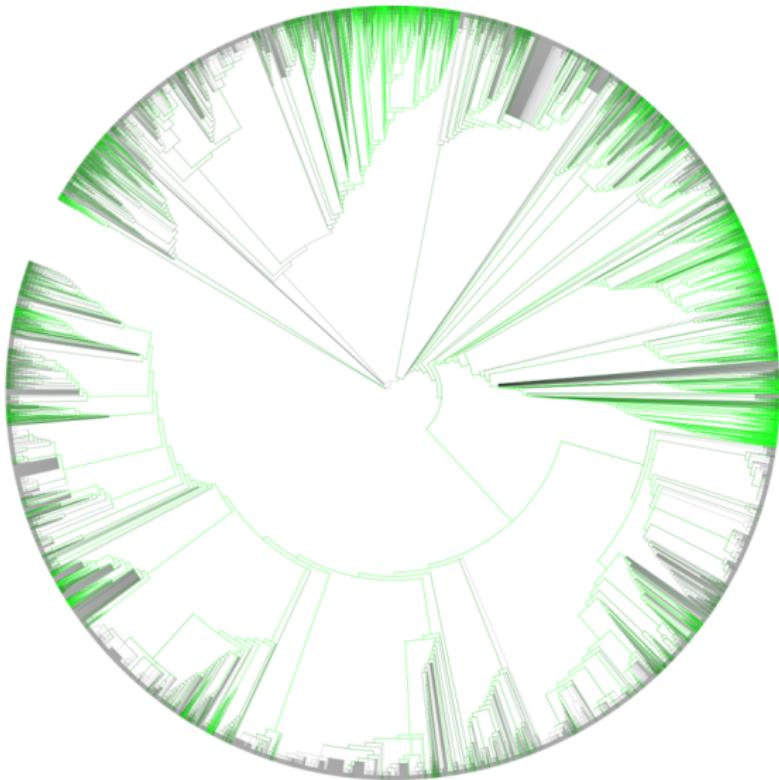
**GBIF**



**Genome**



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



Need to build on existing phylogenetic information.

Need to build on existing phylogenetic information.

OPEN  ACCESS Freely available online



Perspective

## Lost Branches on the Tree of Life

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

**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. PLoS Biology 2013

# only 16% of phylogenies published 2000-2013 are digitally available

Drew et al. PLoS Biology 2013

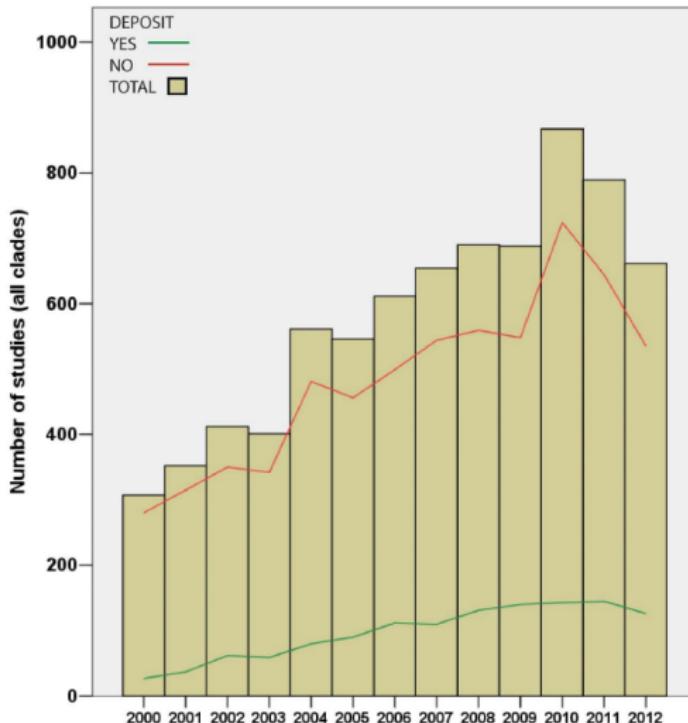


Figure 1. Overview of total number of publications surveyed from animal, fungus, seed plant, microbial eukaryote, archaea, and bacteria literature (indicated in red), and the number of those publications that archived their trees and alignments in either Dryad or TreeBASE (indicated in green).

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. BioEssays 2018)

Tragic recent demise (???) of treebase makes future of phylogenetic data sharing even more concerning

## Adding phylogenetic data to OpenTre 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](https://github.com/OpenTreeOfLife/phylesystem-1)

McTavish et al. Bioinformatics 2015



## **Community Curation**

290 individual curators of 4,550 uploaded studies

## **Community Curation**

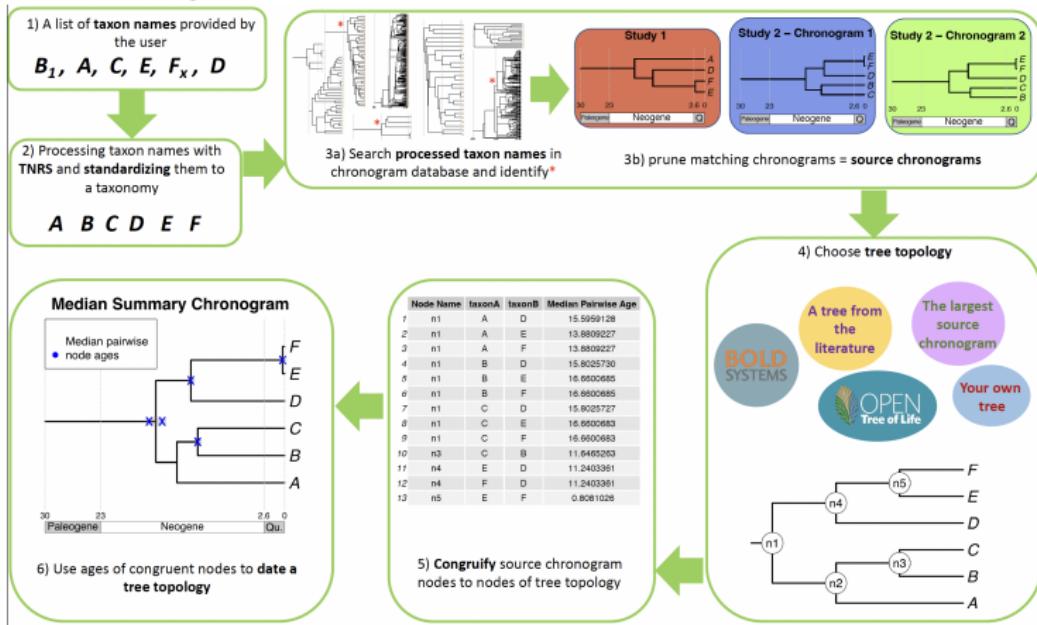
290 individual curators of 4,550 uploaded studies

Currently a several month lag for incorporation into synthetic tree, will begin monthly builds in the next year

## Date estimates

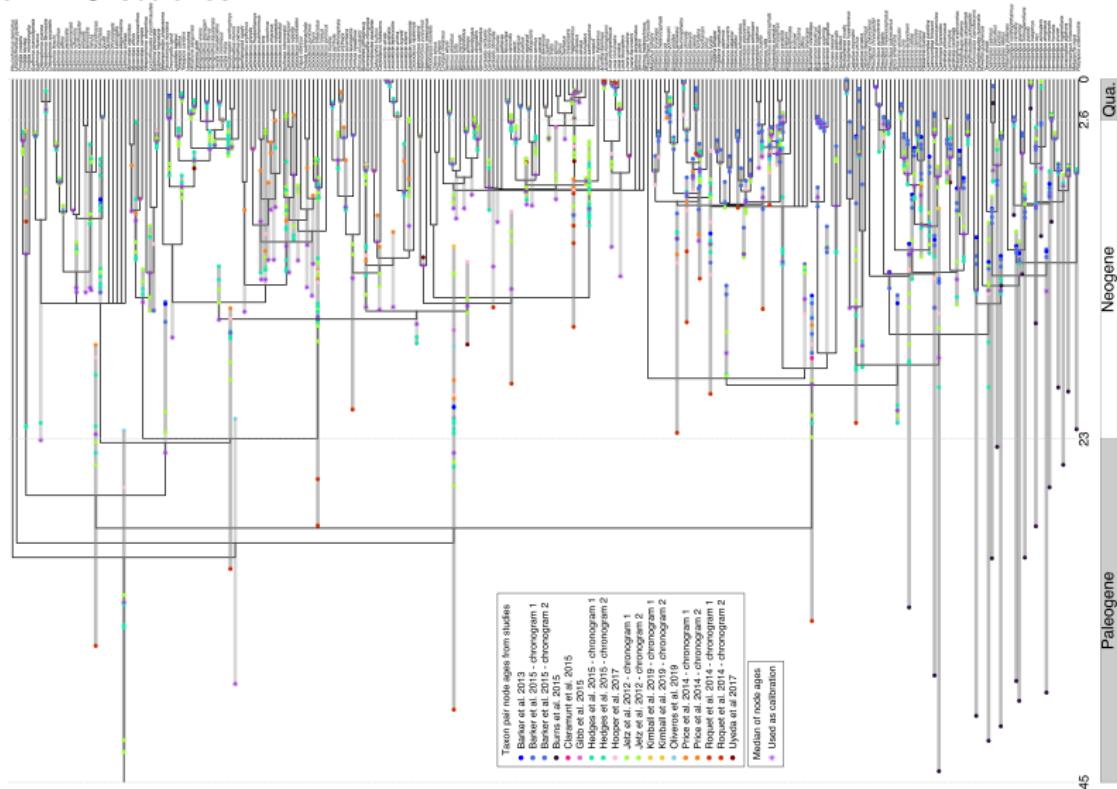
- The synthetic tree does not currently have any branch lengths (when combining across DNA, morphology and taxonomy, branch lengths are not obvious!!)
- However! We can translate dates from input trees to the synthetic tree

# DATELIFE



Webserver at [datelife.opentreeoflife.org](http://datelife.opentreeoflife.org)  
 Sanchez-Reyes, McTavish, O'Meara, to be submitted this week!

# Dated finch tree - topology estimated based on 33 studies, dates from 13 studies

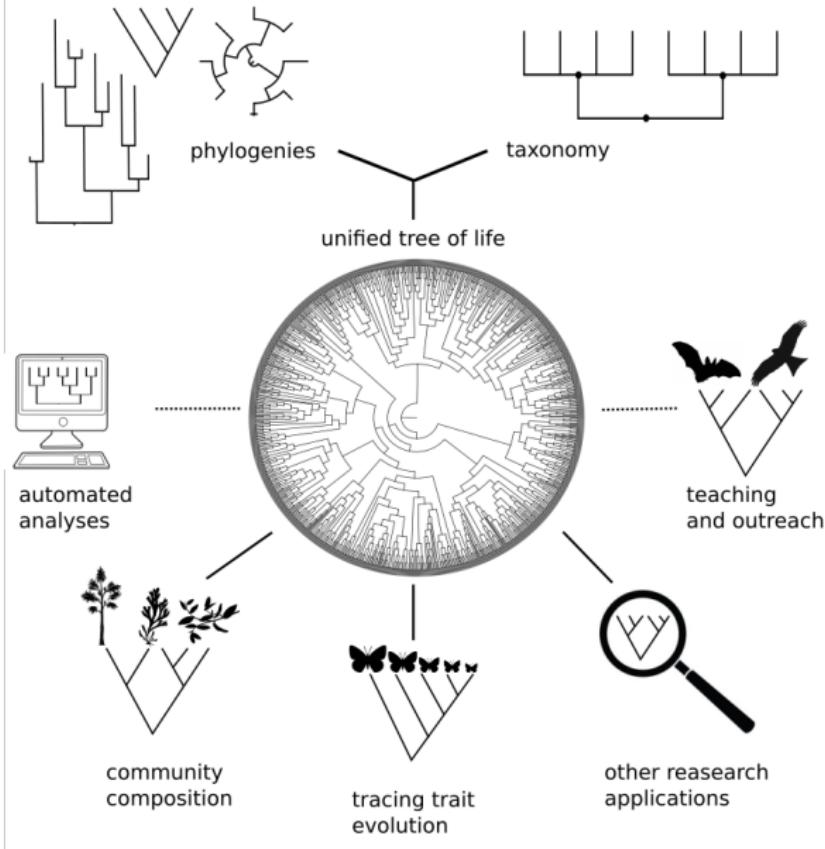


Sanchez-Reyes, McTavish, O'Meara, to be submitted this week!

## **Direct API calls for node in the synthetic tree**

Beta functionality - [dates.opentreeoflife.org](https://dates.opentreeoflife.org)

Date estimates for 46,639 internal nodes from 320 input  
chronograms

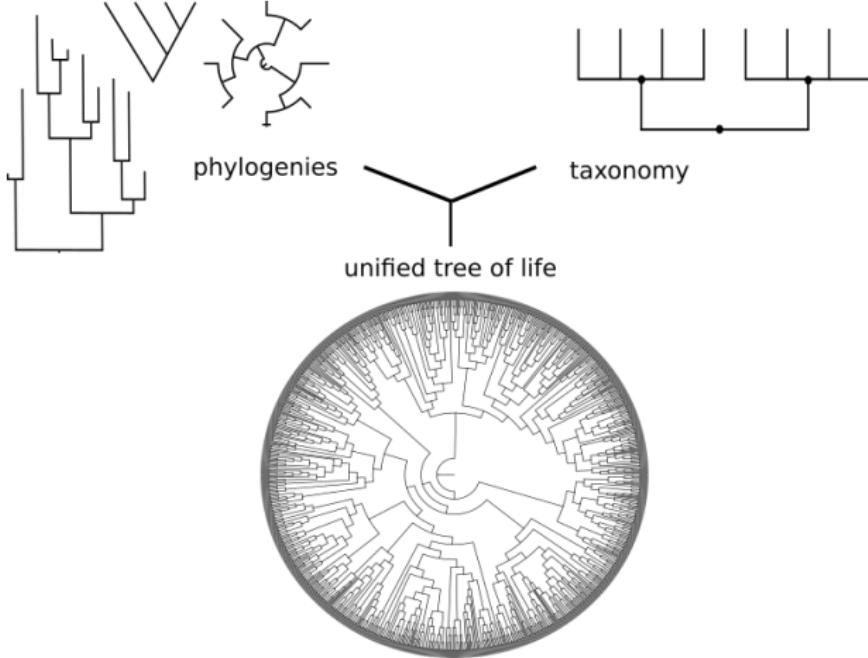


(McTavish et al. Bioessays 2017)

Open Tree resources are available via a range of implementations

- Browser interface, `tree.opentreeoflife.org`
- Open Tree of Life API
-  **Peyotl** python wrapper
- R Open Tree of Life (rotl) 

## Custom Synthesis - in beta testing now!



Synthesis on demand

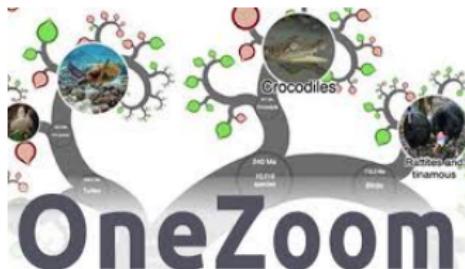
Personal phylogeny rankings

Any root node

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:



# Case study: What is the dated phylogeny of all bird species observed in Merced?

eBird Submit Explore My eBird Science About News Help

Change location All years

Merced California, US Map

Overview Illustrated Checklist 302 Species observed 13.2K Complete checklists 2445 eBinders 118 Hotspots

Sightings Last seen First seen High counts

SPECIES NAME	COUNT	DATE	OWNER	LOCATION
1. Greater White-fronted Goose	250	12 Oct 2020	Cora Barnhill	
2. Northern Shoveler	75	12 Oct 2020	Cora Barnhill	
3. Gadwall	6	12 Oct 2020	Cora Barnhill	
4. Northern Pintail	10	12 Oct 2020	Cora Barnhill	
5. Mourning Dove	1	12 Oct 2020	Cora Barnhill	
6. American Coot	500	12 Oct 2020	Cora Barnhill	
7. Sandhill Crane	150	12 Oct 2020	Cora Barnhill	
8. Black-necked Stilt	45	12 Oct 2020	Cora Barnhill	
9. Killdeer	4	12 Oct 2020	Cora Barnhill	
10. Double-crested Cormorant	1	12 Oct 2020	Cora Barnhill	
11. American White Pelican	22	12 Oct 2020	Cora Barnhill	
12. Great Blue Heron	3	12 Oct 2020	Cora Barnhill	
13. Great Egret	2	12 Oct 2020	Cora Barnhill	
14. White-faced Ibis	150	12 Oct 2020	Cora Barnhill	
15. White-tailed Kite	2	12 Oct 2020	Cora Barnhill	

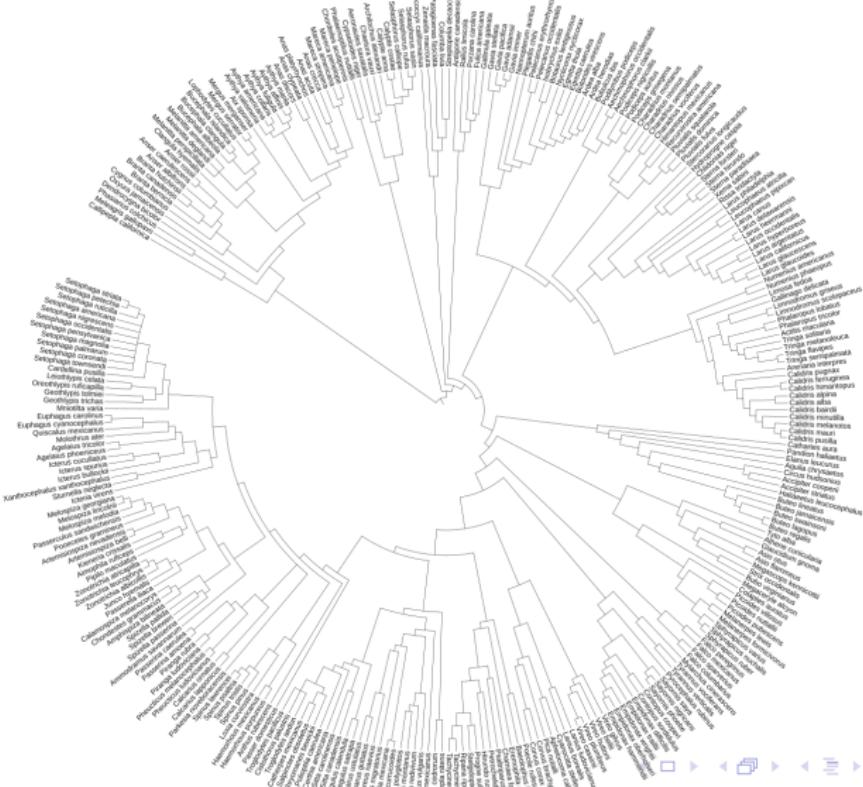
Updated -2 days ago Show all details

Top media UPLOADED IN LAST 7 DAYS

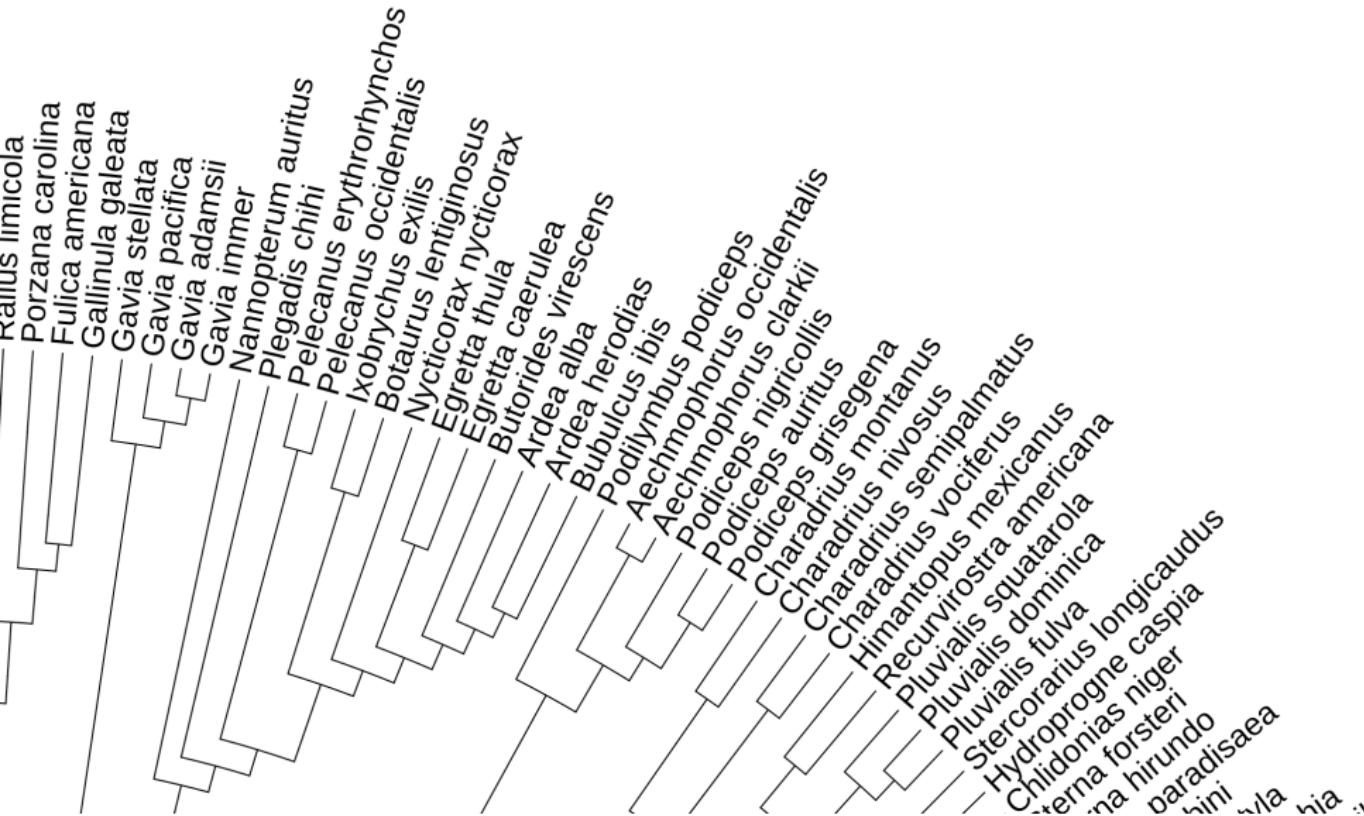
☰ ↻ 🔍

Undated tree:

TNRS on species names to match to unique identifiers  
Request induced subtree from OpenTree APIs  
Tree with 302 tips, informed by 102 published studies



## Taxonomy lags behind phylogeny (e.g. *Ardea*)



<https://itol.embl.de/tree/1082301078373711602860218>

Adding dates (work in progress!)

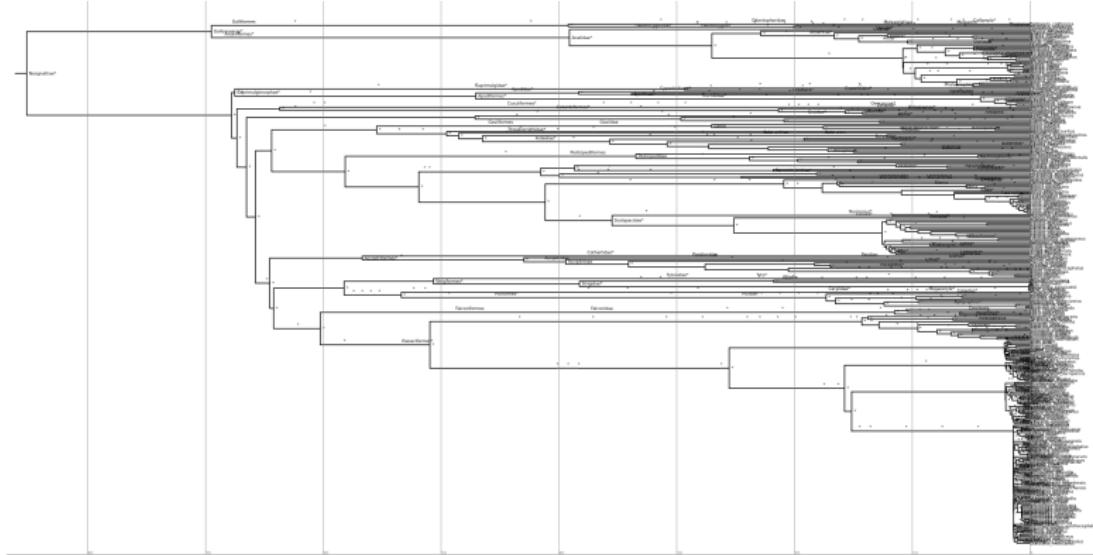
Match internal nodes of dated inputs to nodes in the synthetic tree (phyloreferenceing approach)

Apply dates to nodes induced subtree

Smooth undated nodes using bladj (Webb et all, phylocom)

<https://github.com/OpenTreeOfLife/ChronoSynth>

We have date estimates aligned to 991/1,261 nodes in this tree.



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

Use existing estimates to compare and contextualize new inferences



## Conclusions

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

## Lab today:

- Browser interface, [tree.opentreeoflife.org](http://tree.opentreeoflife.org)
- Standardizing taxon names
- Getting existing trees for arbitrary sets of taxa
- Gathering date estimates for nodes

Contribute your knowledge!  
[tree.opentreeoflife.org/curator](http://tree.opentreeoflife.org/curator)



# Thank You



NSF ABI 1759846

Mark Holder

Karen Cranston



NSF AVATOL 1208809

AVATOL PI'S: Burleigh,  
Crandall, Cranston, Gude,  
Hibbett, Holder, Katz, Ree,  
Smith, Soltis, Williams

**Dendropy** Jeet Sukumaran

Lab group:

Luna Luisa Sanches Reyes

Lesly Lopez Fang

Jasper Toscani-Field

Lucia Bazan Williamson

