

Integrating taxonomy, sequence data, phylogeny, and genomes to build a comprehensive and scalable Hemiptera phylogeny

Chris Owen
Computational Biology Institute
The George Washington University

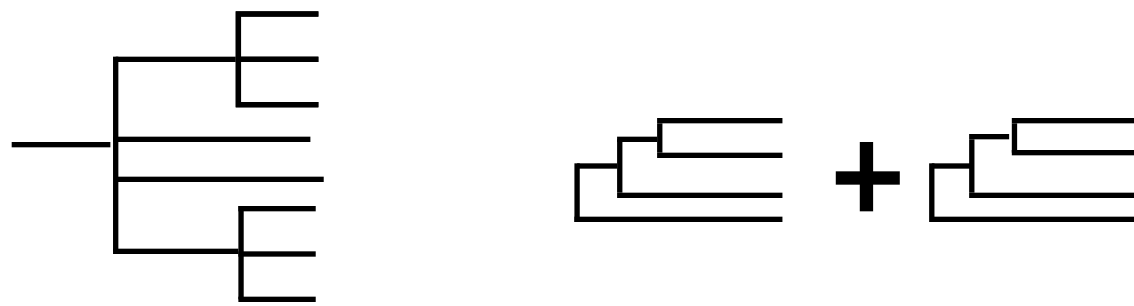
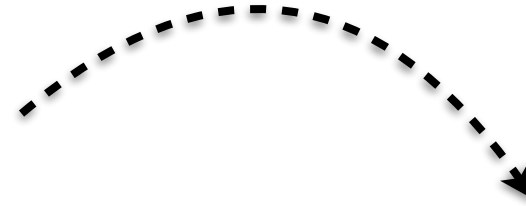
AVAToL: Open Tree of Life

Published Taxonomies

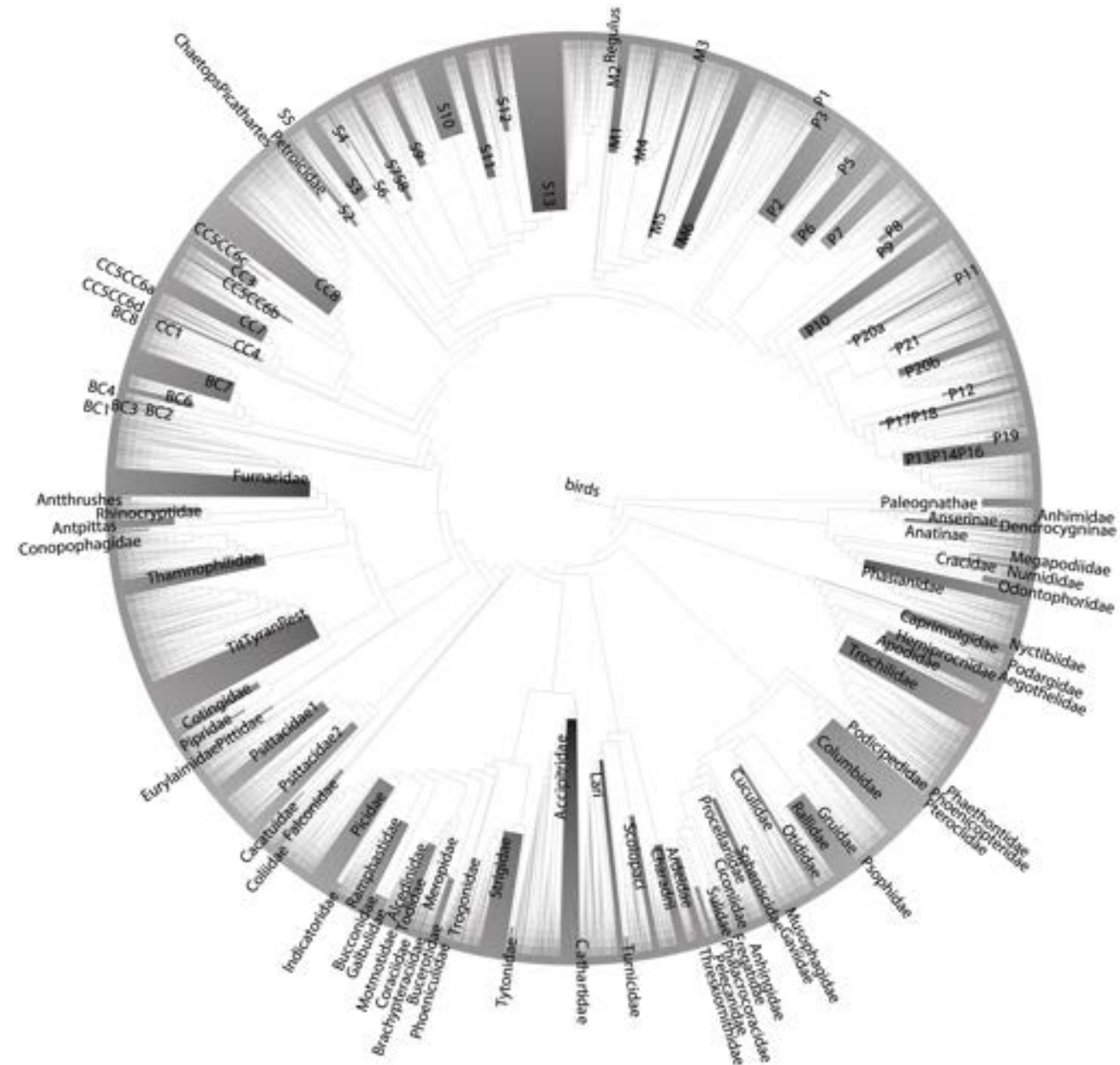
+

Published Phylogenies

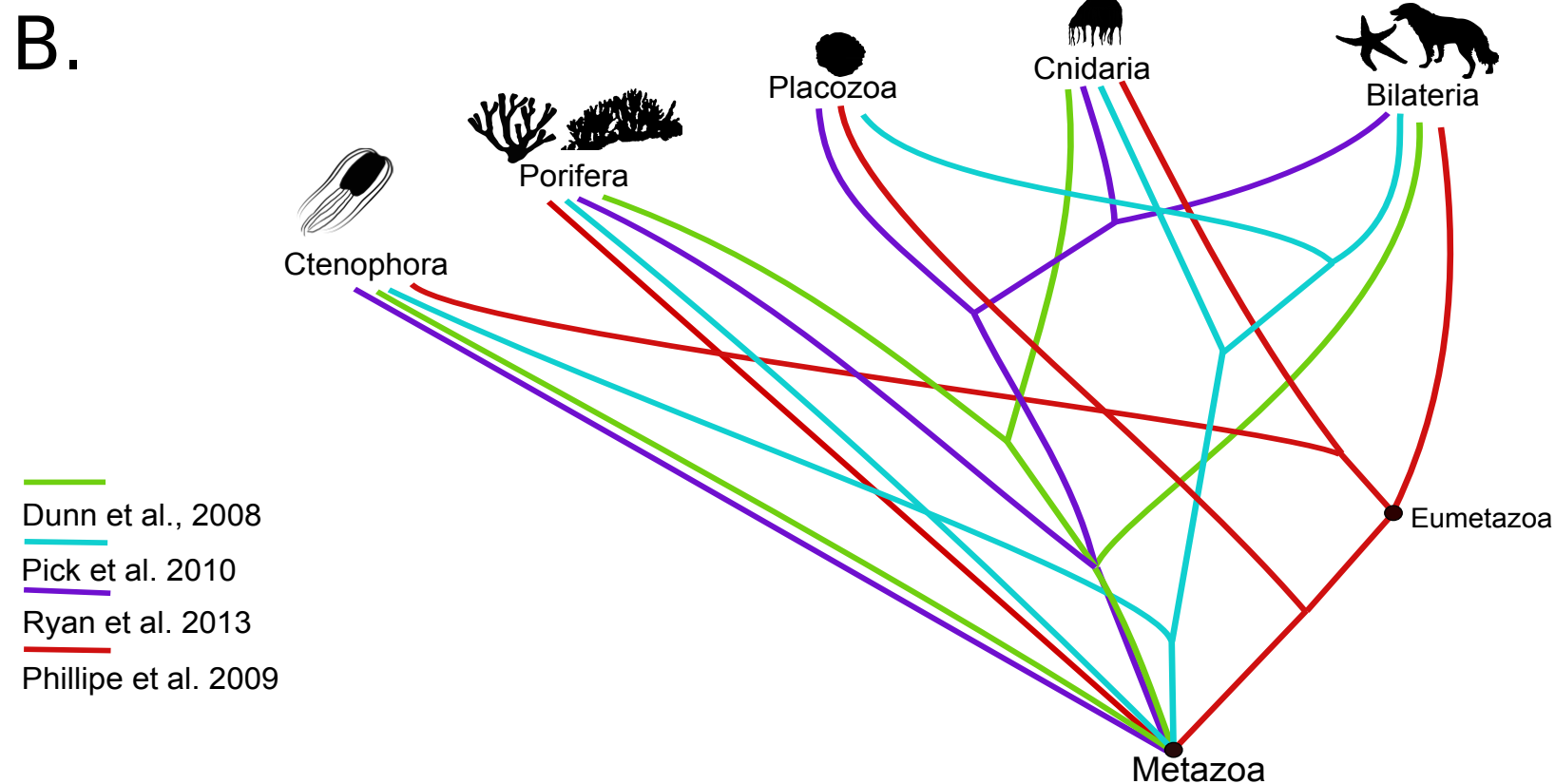
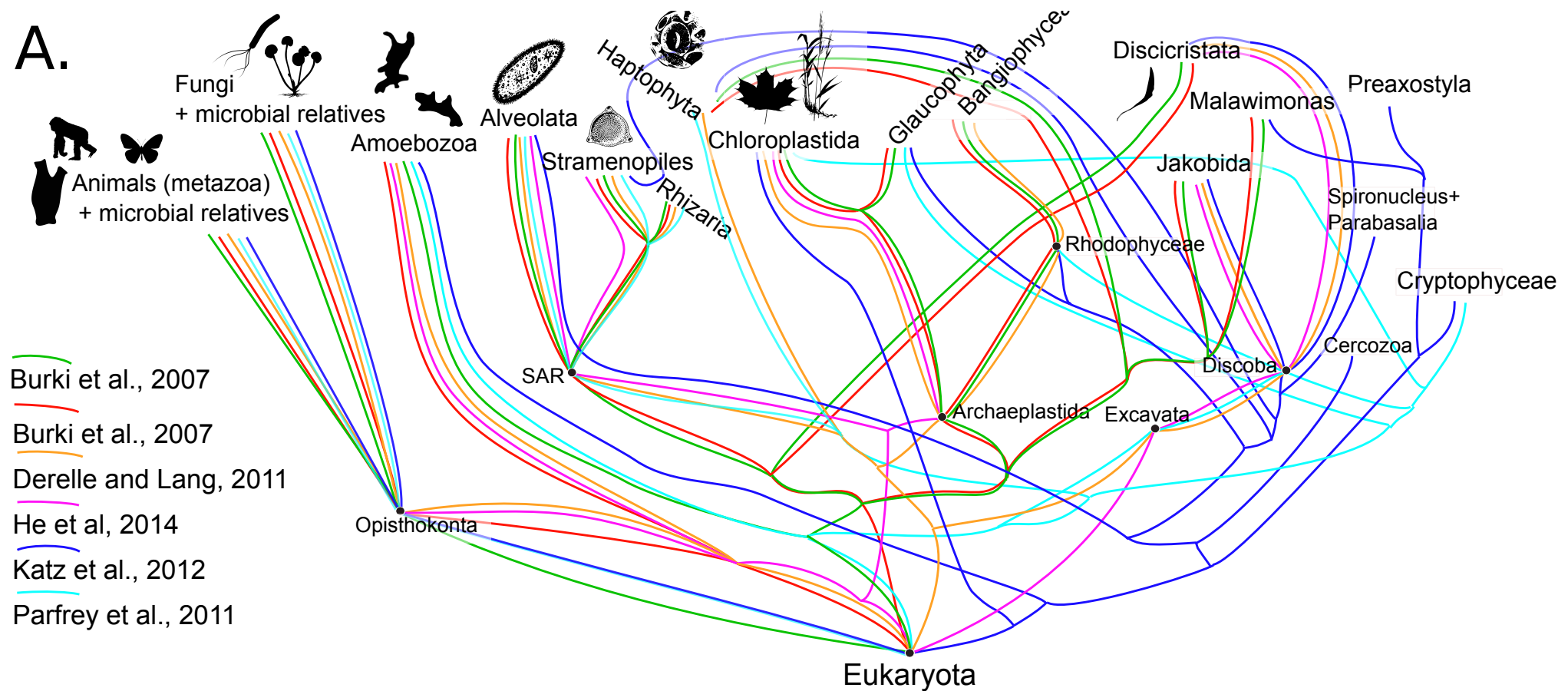
www.opentreeoflife.org



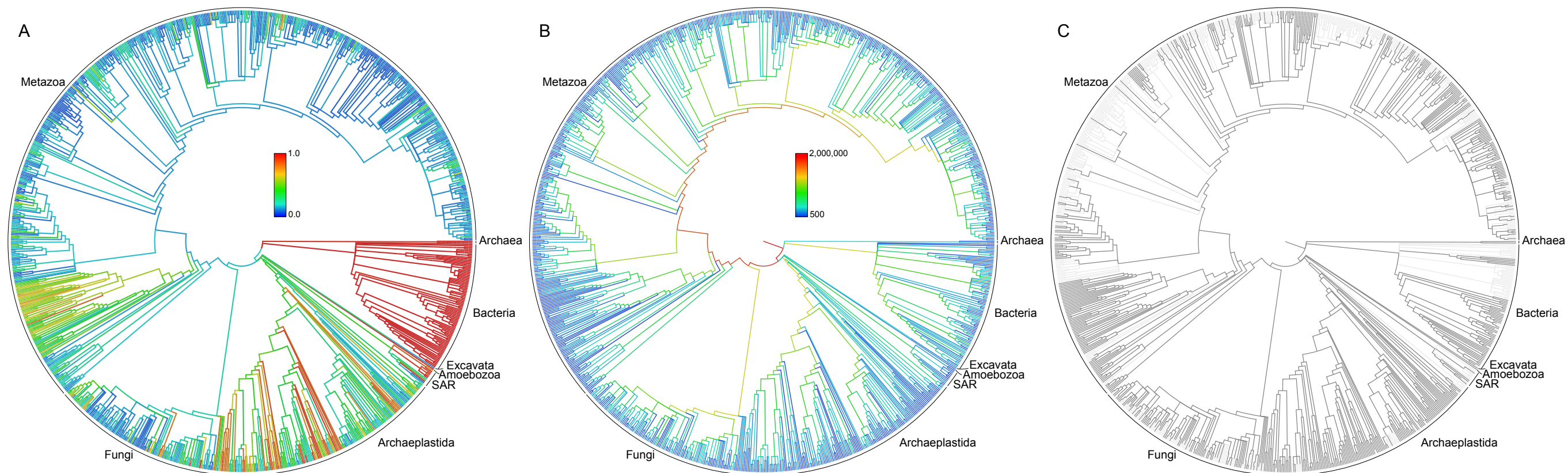
- Phylogeny of all life
 - no new relationships
- Curated taxonomy
- Accessible to the public
- Add/Extract phylogenies
- Two additional AVAToL groups



AVAToL: example phylogeny graphs



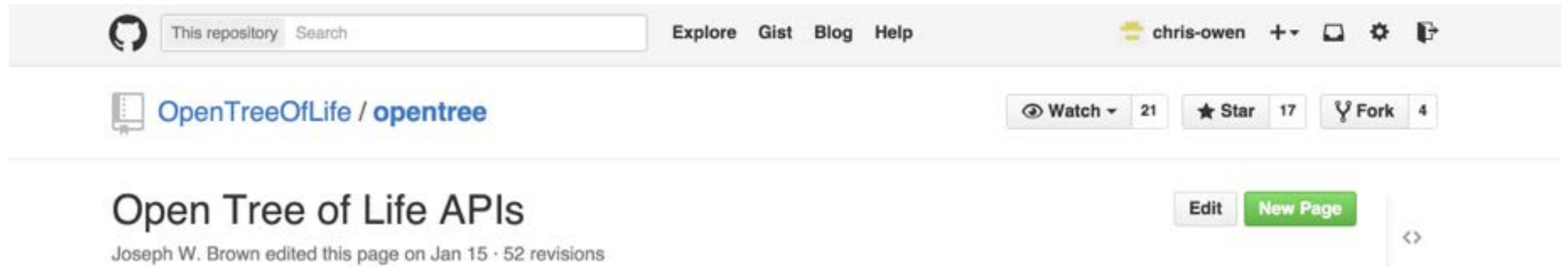
The Tree of Life



Hinchliff et al. *in review*

- GenBank taxonomy has ~411,000 binomials (<5% of est.)
- AVAToL taxonomy = 2,227,481 terminals
- Tips represented by phylogenies = 49,487
- <http://blog.opentreeoflife.org/>, www.opentreeoflife.org

Open Tree APIs



- command-line access to versions of the Open Tree of Life
- access the Graph of Life
- taxonomic name resolution services
- taxonomy
- studies containing source tree
- <https://github.com/OpenTreeOfLife/opentree/wiki/Open-Tree-of-Life-APIs>

Python API wrappers



[Getting Started](#)

[PhyloSystem](#)

[Transformations](#)


[API wrappers](#)

[Design](#)

[Home](#)

- additional Open Tree access via python API wrappers
- <http://opentreeoflife.github.io/peyotl/>

Open Tree Curator App

 Open Tree of Life

Curator Home

Studies ▾

Login

Study Curator

This tool allows you to contribute studies and trees to our database for synthesis. Once your data is submitted, you can return and make improvements, share your data more widely, and watch for changes in your area of interest.

All data is stored as documents in [GitHub](#), a collaborative coding website. This curation tool makes it easy to edit and contribute data, but you'll need a (free) GitHub account to participate.

Once you've created an account on GitHub, just [login](#) using the link above. Once you've agreed to give access to the OpenTree tool, you can start adding data immediately.

Add new study

Filter by reference text, DOI, tag, curator... Newest publication first ▾

Reference (click to view study)	Focal clade	Curator
Coburn R.A., 2015 [show details]	—	redavids
Malyshova V.F., 2015 [show details]	—	isaacovercast
Dos Remedios, 2015 [show details]	—	Joseph W. Brown
Tan J., 2015 [show details]	—	MGNute
Dai S., 2015 [show details]	—	redavids
Pedro Selvatti, 2015 [show details]	Passeriformes	Joseph W. Brown
Shi J.J., 2015 [show details]	—	isaacovercast

Viewing study Tschopp A., 2013

<http://dx.doi.org/10.1186/1471-2148-13-74>

Study quality 92% [show details]

Metadata

Trees 1

Files

OTU Mapping

Tools

History

Return to study list

Login to Edit Study

Publication reference Tschopp A., Riedel M., Kropf C., Nentwig W., & Klopstein S. 2013. The evolution of host associations in the parasitic wasp genus *Ichneumon* (Hymenoptera: Ichneumonidae): convergent adaptations to host pupation sites. *BMC Evolutionary Biology*, 13: 74.

Publication DOI (or URL) <http://dx.doi.org/10.1186/1471-2148-13-74>

Data deposit DOI/URL Data for this study is archived as [Treebase study 13911](#)

Study Year 2013

Focal clade [Ichneumonidae](#)

Tags None

Submitted by Chris Owen

Waiver or license None

This study should contribute to synthesis.

This tool is a friendly visual editor for curating [NexSON](#) files. Like the [NeXML](#) and [NEXUS](#) formats from which it evolved, each [NexSON](#) file contains information about a single published study, with a particular focus on its trees and OTUs.

Each tab in this curation tool manages a different aspect of the study data. Watch for prompts—they look like **2**—in each tab that list areas ripe for improvement.


Here in the **Metadata** tab, we're primarily concerned with clearly identifying this study and making it easy to find using standard publication references, DOIs, and free-form tags.

Working with study data in other tools

You can download this study in several formats:

[NexSON](#) [NeXML](#) [NEXUS](#) [Trees as Newick](#)

You can also use the [Open Tree API](#) (the same one that powers this editor) to retrieve and make changes to study data.

 Open Tree of Life

Curator Home

Studies ▾

Viewing study ANDREAS ZWICK, 2011

<http://dx.doi.org/10.1111/j.1365-3113.2010.00543.x>

Study quality

Metadata

Trees 1

Files

OTU Mapping

Tools

History

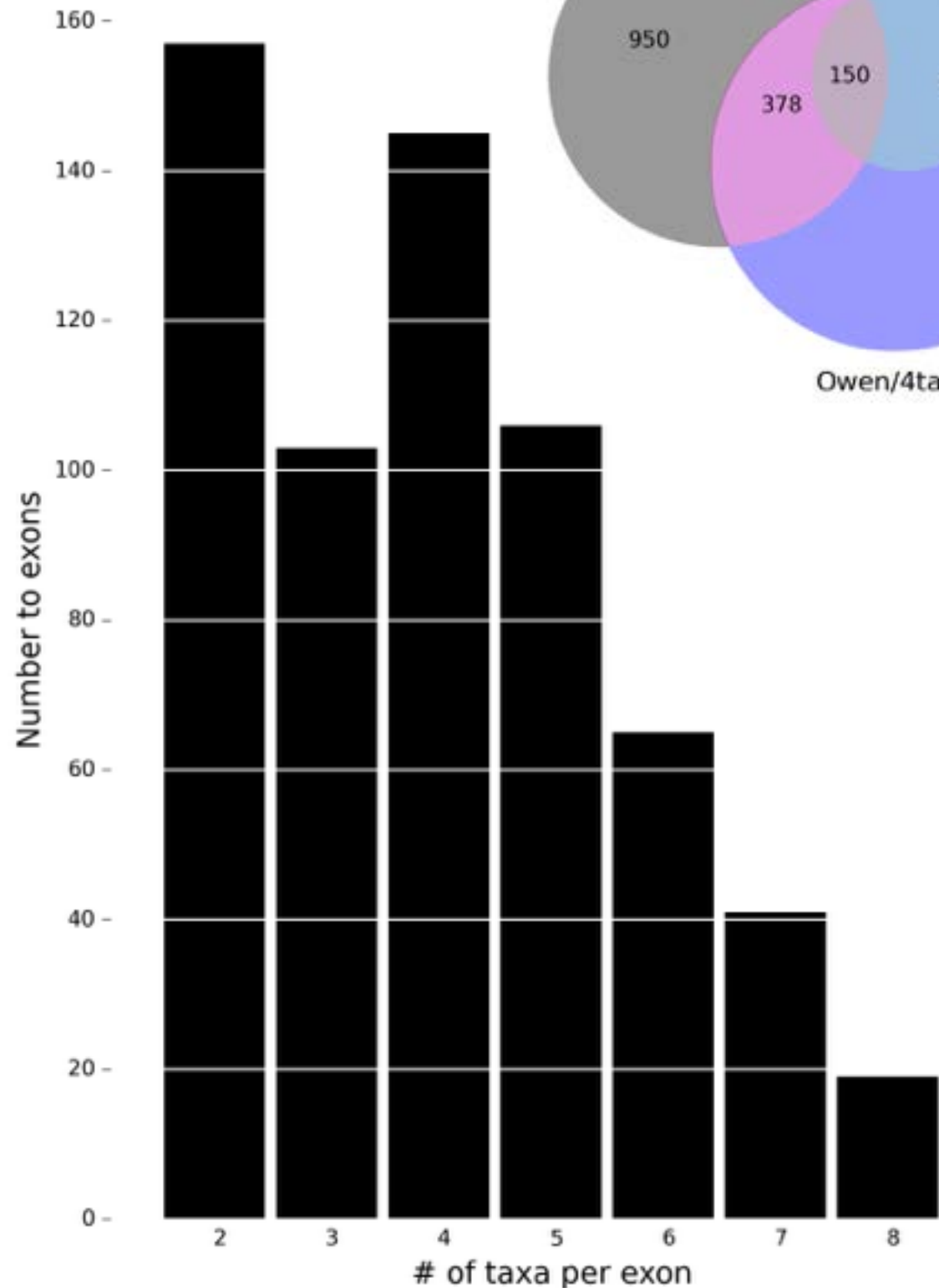
Filter by original or mapped label In all trees ▾ Unmapped OTUs first ▾

Original name	Mapped to taxon
Apha aequalis AAEQ	Apha aequalis
Brahmaea certhia BHEAR	Brahmaea certhia
Carthaea satumioides TB032177	Carthaea saturnioides
Eupterote nr. naessigi EUPNN	Eupterote aff. naessigi JCR-2007
Apatelodes torrefacta ATOR	Apatelodes torrefacta
Prismosticta fenestrata PFENE	Prismosticta fenestrata
Hemaris diffinis HTHY	Hemaris diffinis
Malacosoma californicum MCALIFORN	Malacosoma californicum
Bombyx mandarina BMAND	Bombyx mandarina
Endromis versicolora EVERSICOL	Endromis versicolora
Gonometa rufobrunnea GRUFO	Gonometa rufobrunnea
Eutachyptera psidii EPSIDII	Eutachyptera psidii
Janiodes laverna nigripuncta JCER	Janiodes laverna nigropuncta
Malacosoma americanum MAME2	Malacosoma americanum
Oxytenis modestia OMOD	Oxytenis modestia
Colla glaucescens CGLA	Colla glaucescens
Poecilocampa populi PPOP	Poecilocampa populi
Sorocaba sp. ROLE	Sorocaba sp. Janzen01

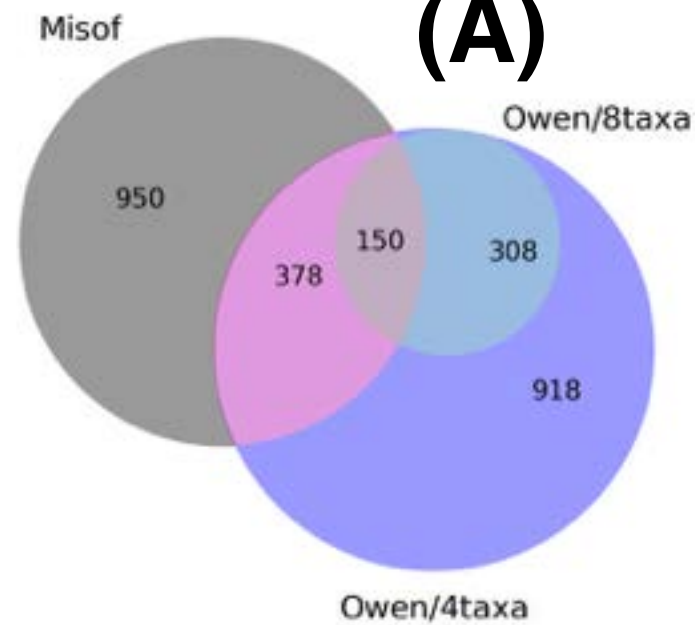
Why build the Hemiptera/
Hemipteroid Tree of Life using
a supertree-like method?

Different ortholog sets

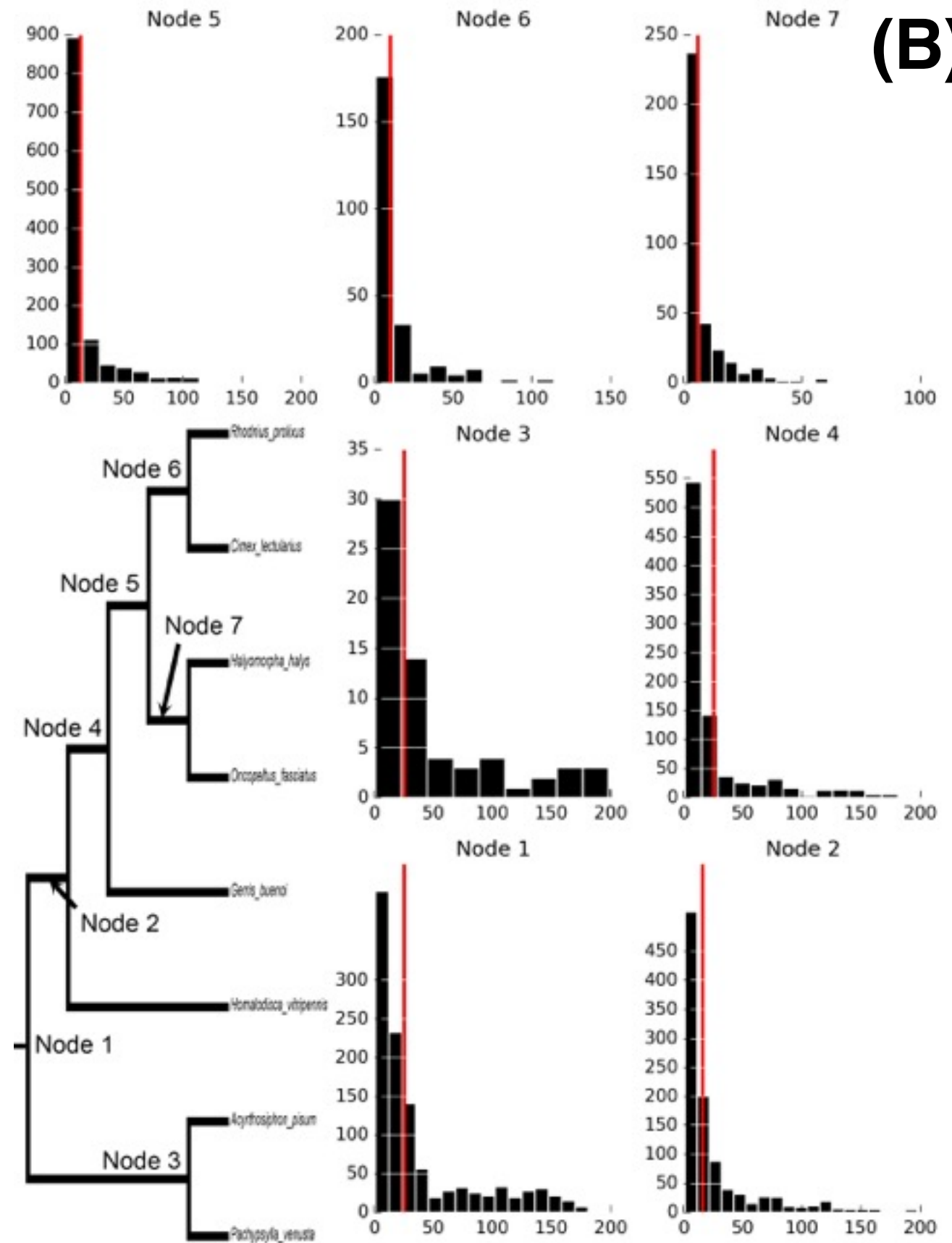
(C)



(A)



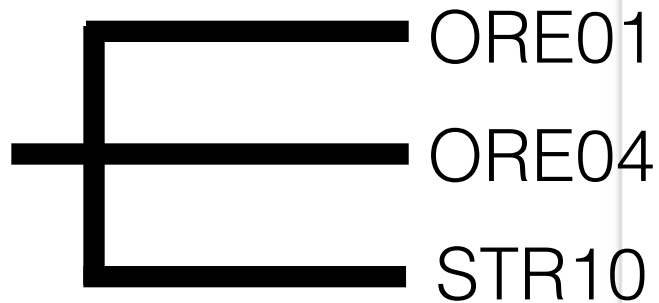
(B)



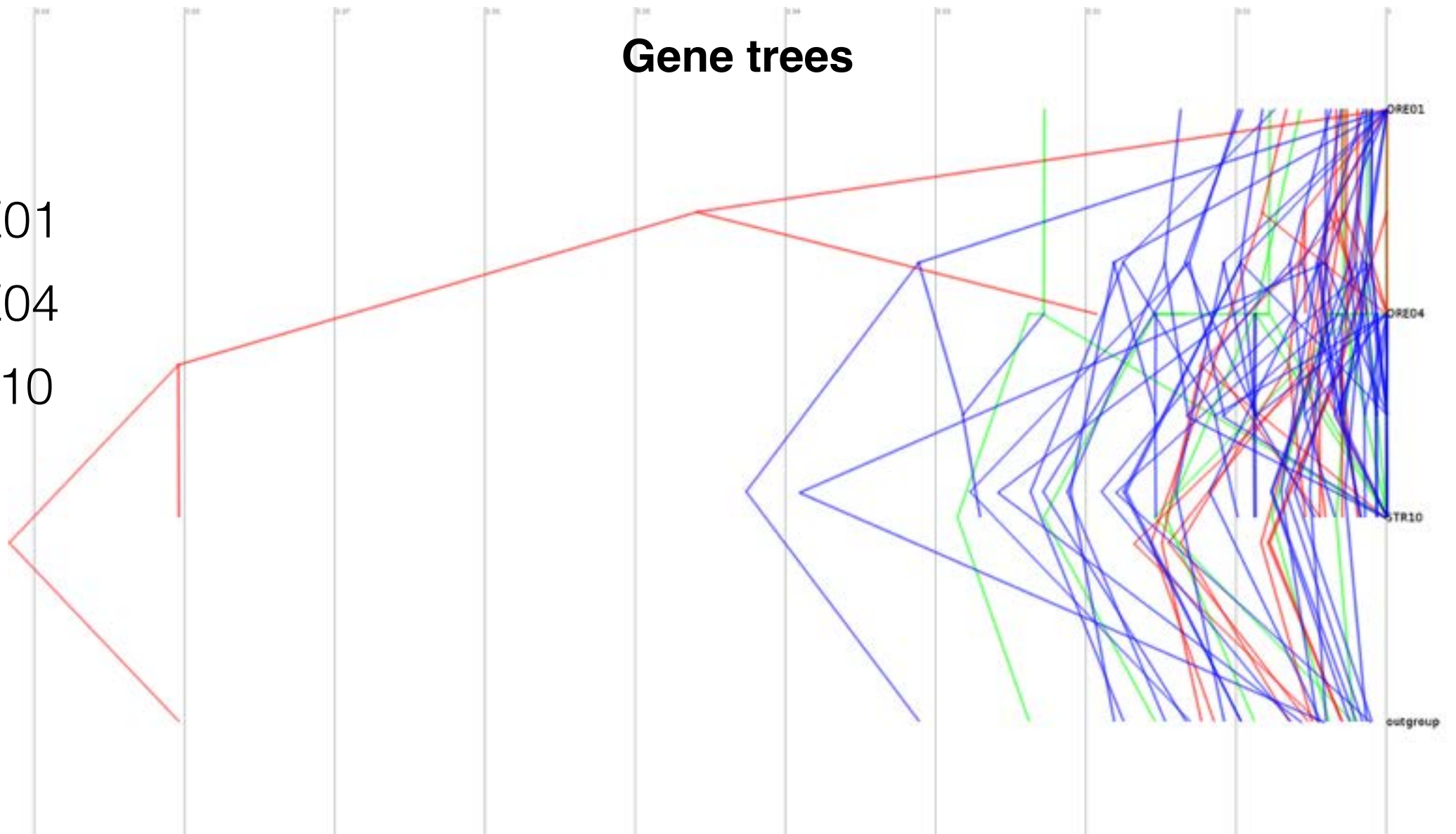
Different ortholog sets

mtDNA phylogeny

- 1500bp COI
- Garli ML phylo



Gene trees



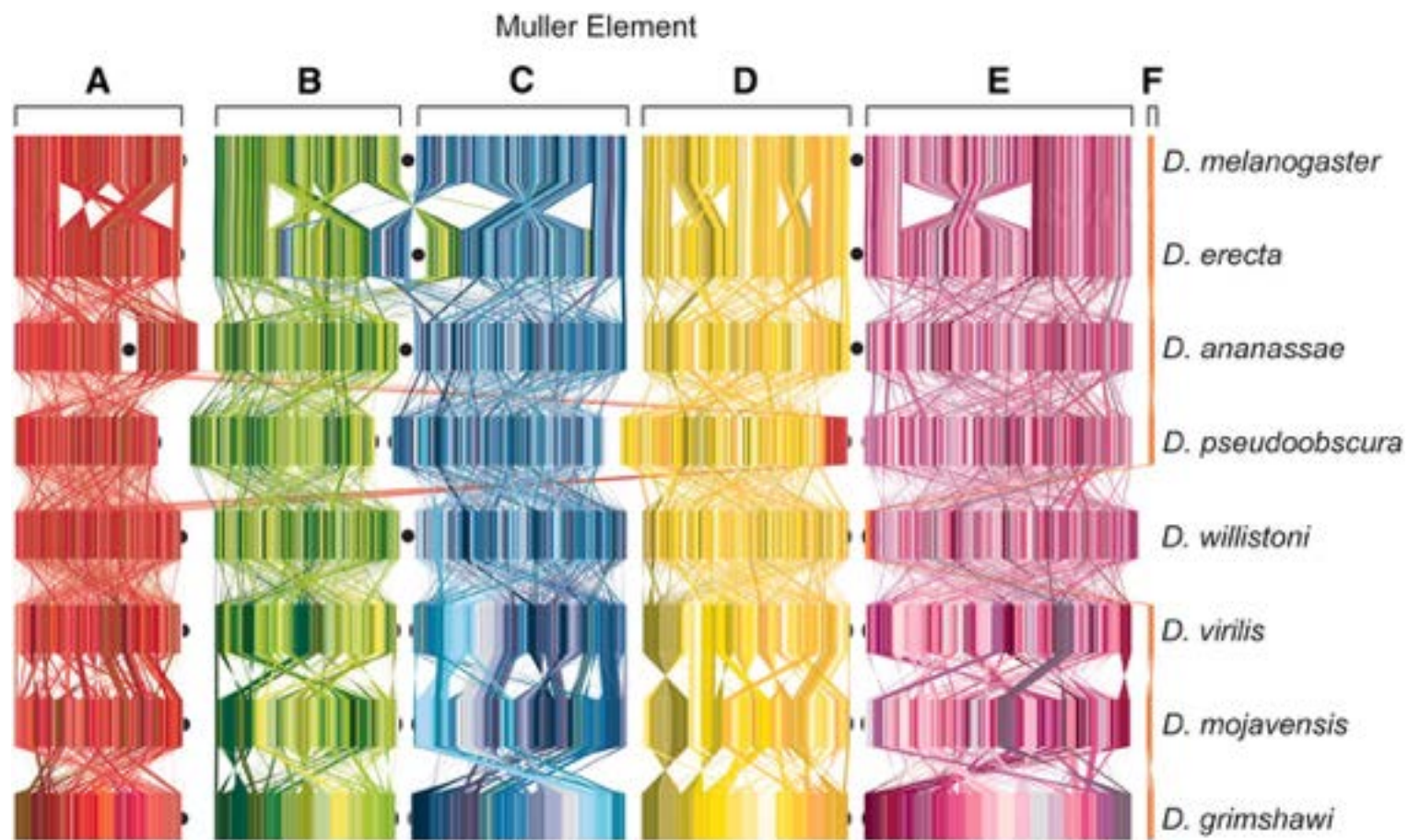
- ORE04, STR10 = 59.46%
- ORE01, ORE04 = 24.32%
- ORE01, STR10 = 16.22%

- songs distinct; mtDNA polytomy
- 40 loci from hybrid capture data

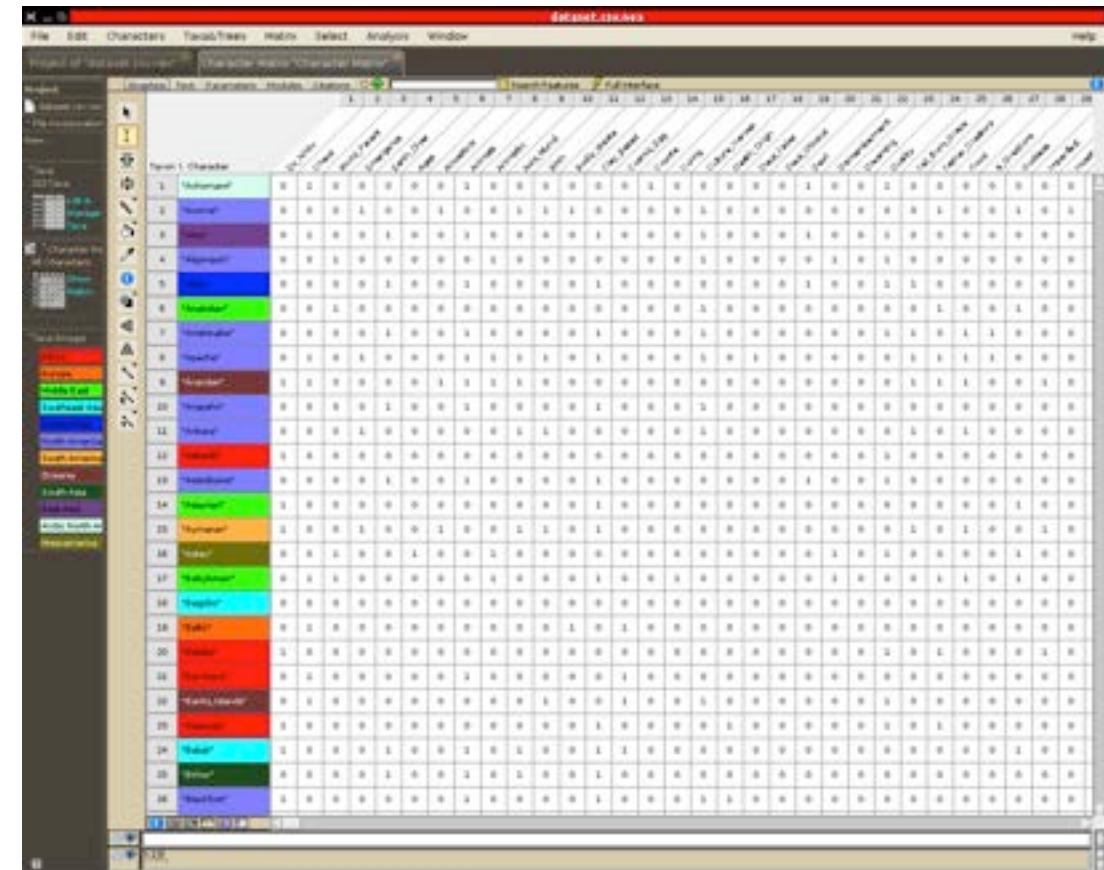
Different types of characters

Genome rearrangements

Morphological character matrices

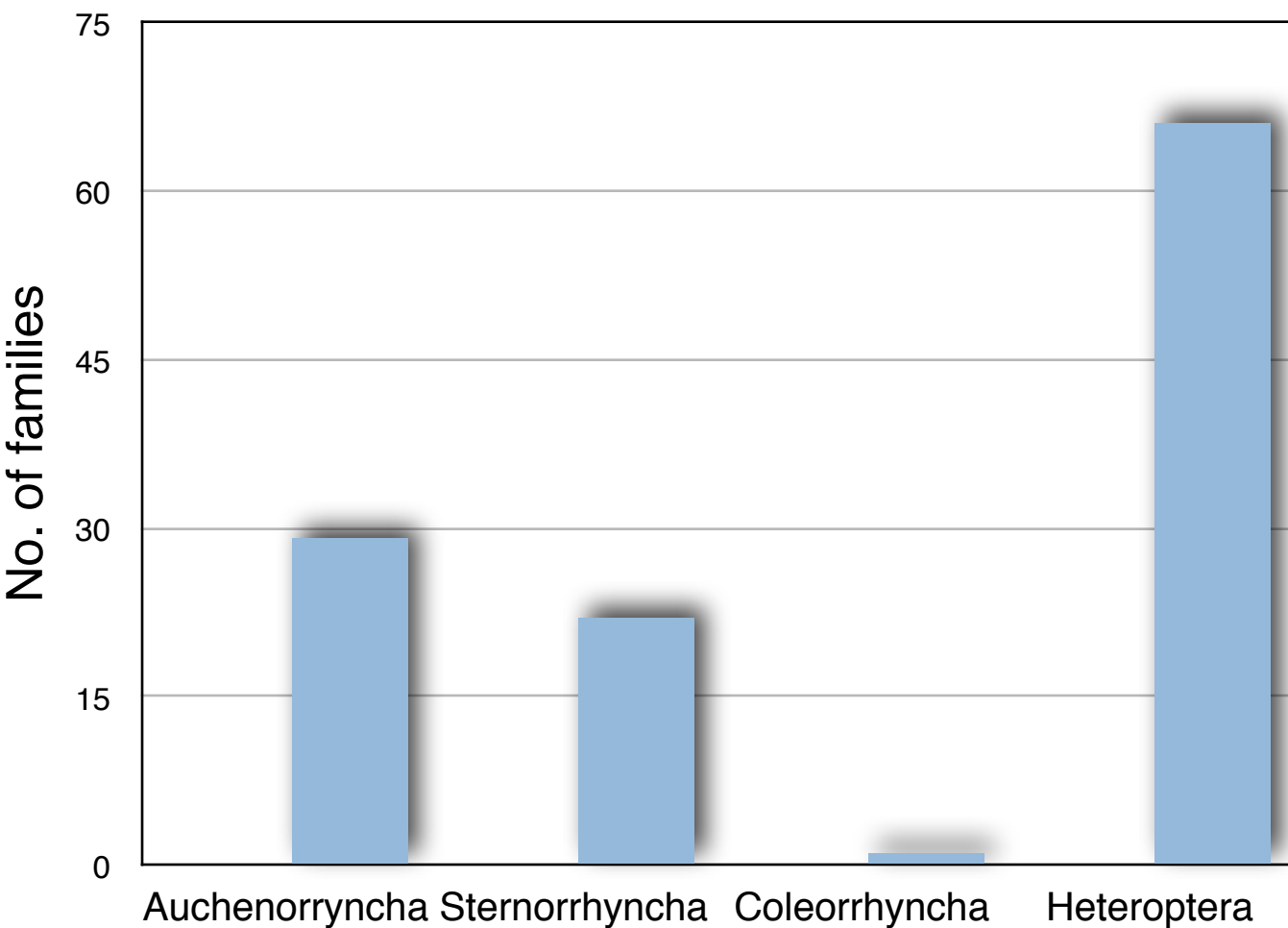


Bhutkar et al. 2008

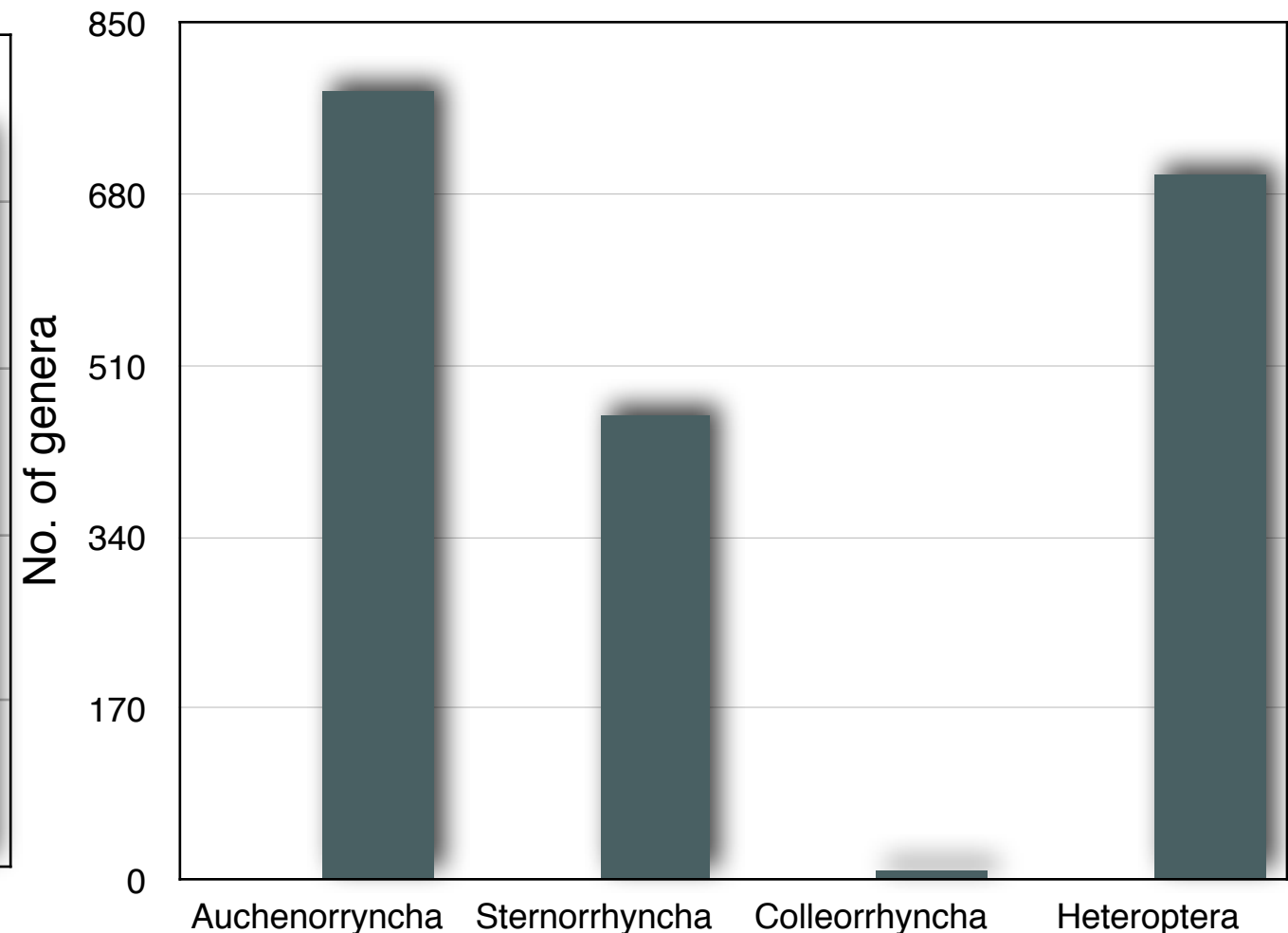


Taxonomic diversity of GenBank Sanger data

Family diversity sampled within suborders



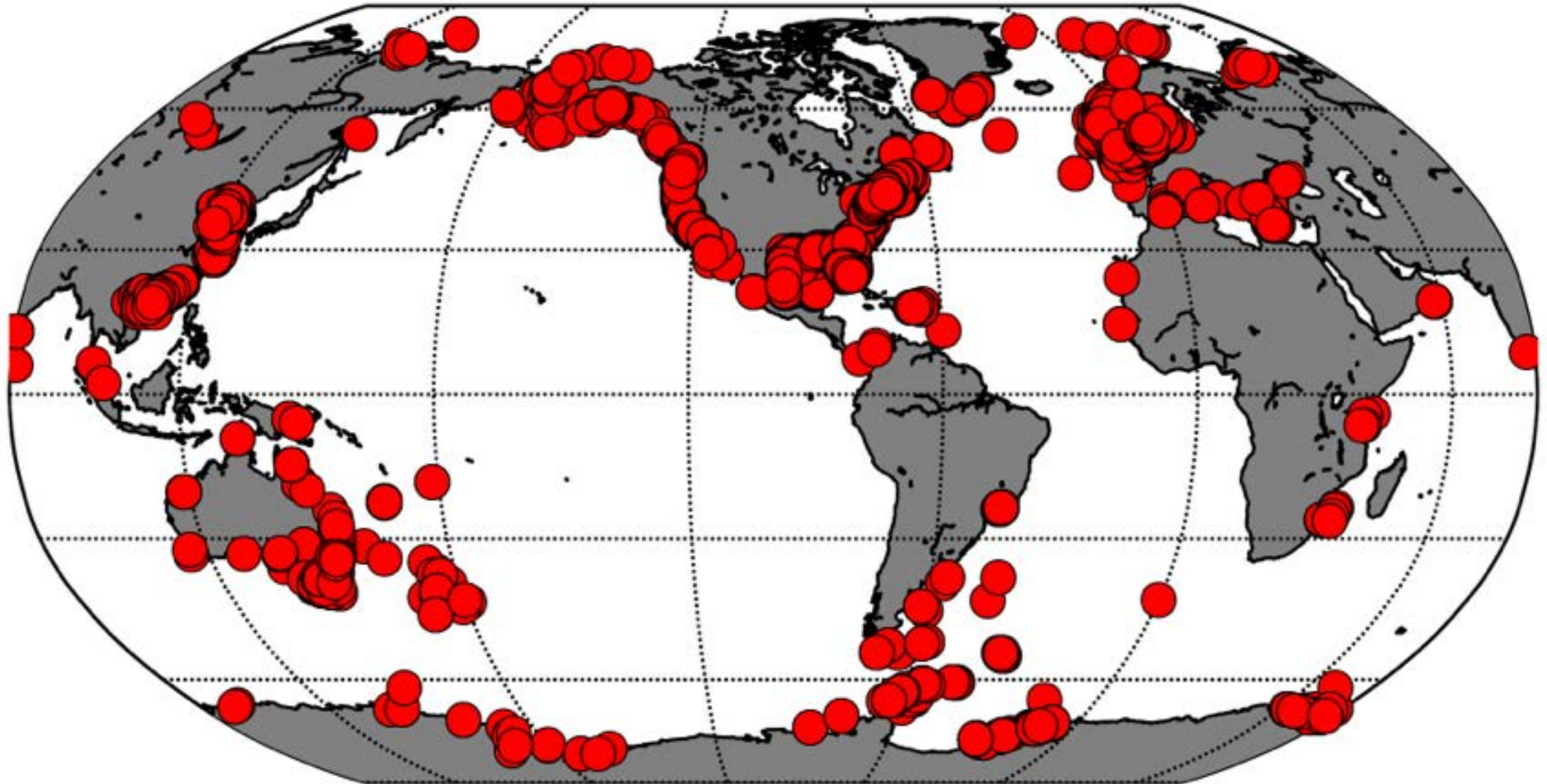
Generic diversity sampled within suborders



Open Tree Hemiptera Taxonomy Stats

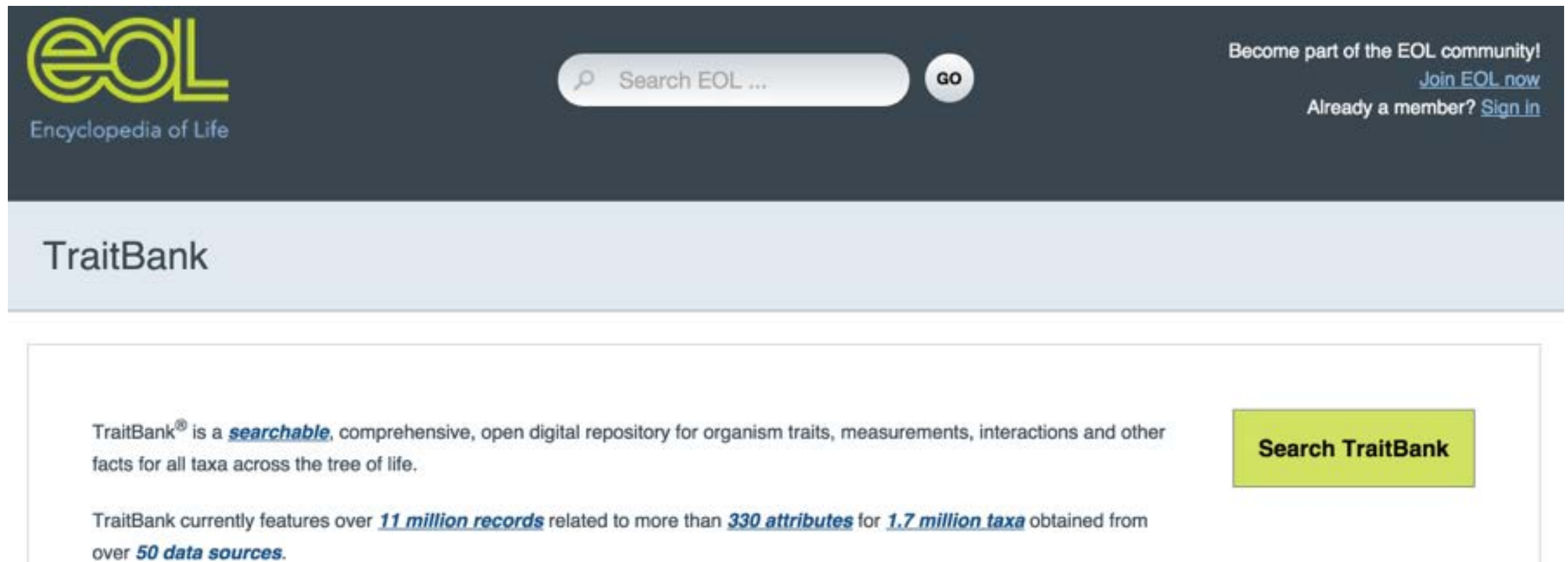
- 68,095 terminal taxa
- 67 tribes
- 174 families
- 9,976 genera

Downstream analyses: species occurrence data



- link taxonomy to GBIF, SI, Living Australia, etc.
- map sampled/unsampled taxa distributions
- extend/confirm ranges for climate and biogeography analyses

Downstream analyses: trait and character evolution



The screenshot shows the EOL (Encyclopedia of Life) website header with the logo and a search bar. Below the header is a light blue banner for TraitBank. The main content area describes TraitBank as a searchable digital repository for organism traits, measurements, interactions, and other facts for all taxa across the tree of life. It also mentions that TraitBank currently features over 11 million records related to more than 330 attributes for 1.7 million taxa obtained from over 50 data sources. A yellow button labeled 'Search TraitBank' is visible on the right side of the main content area.

eOL
Encyclopedia of Life

Search EOL ... GO

Become part of the EOL community!
[Join EOL now](#)
Already a member? [Sign In](#)

TraitBank

TraitBank® is a [searchable](#), comprehensive, open digital repository for organism traits, measurements, interactions and other facts for all taxa across the tree of life.

TraitBank currently features over [11 million records](#) related to more than [330 attributes](#) for [1.7 million taxa](#) obtained from over [50 data sources](#).

Search TraitBank

- link trait data to taxon names / query data for matrices
- co-phylogenies

Conclusions

1. Framework is in place to answer broad questions concerning Hemiptera/Hemipteroid evolution
2. Taxonomy needs expert curation
3. Need to make published data more accessible

**Send me your published phylogenies
in newick or nexus format!!!
clowen@gwu.edu**

Questions?