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

Chris Owen
Computational Biology Institute
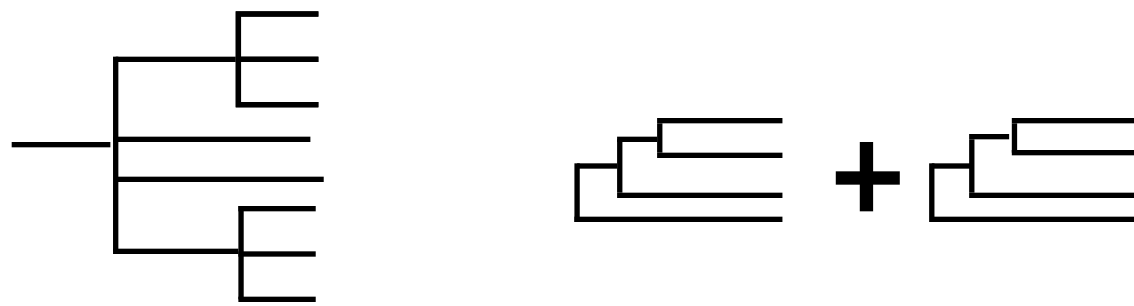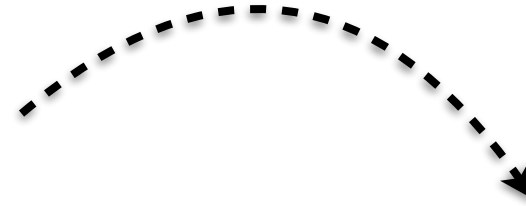The George Washington University

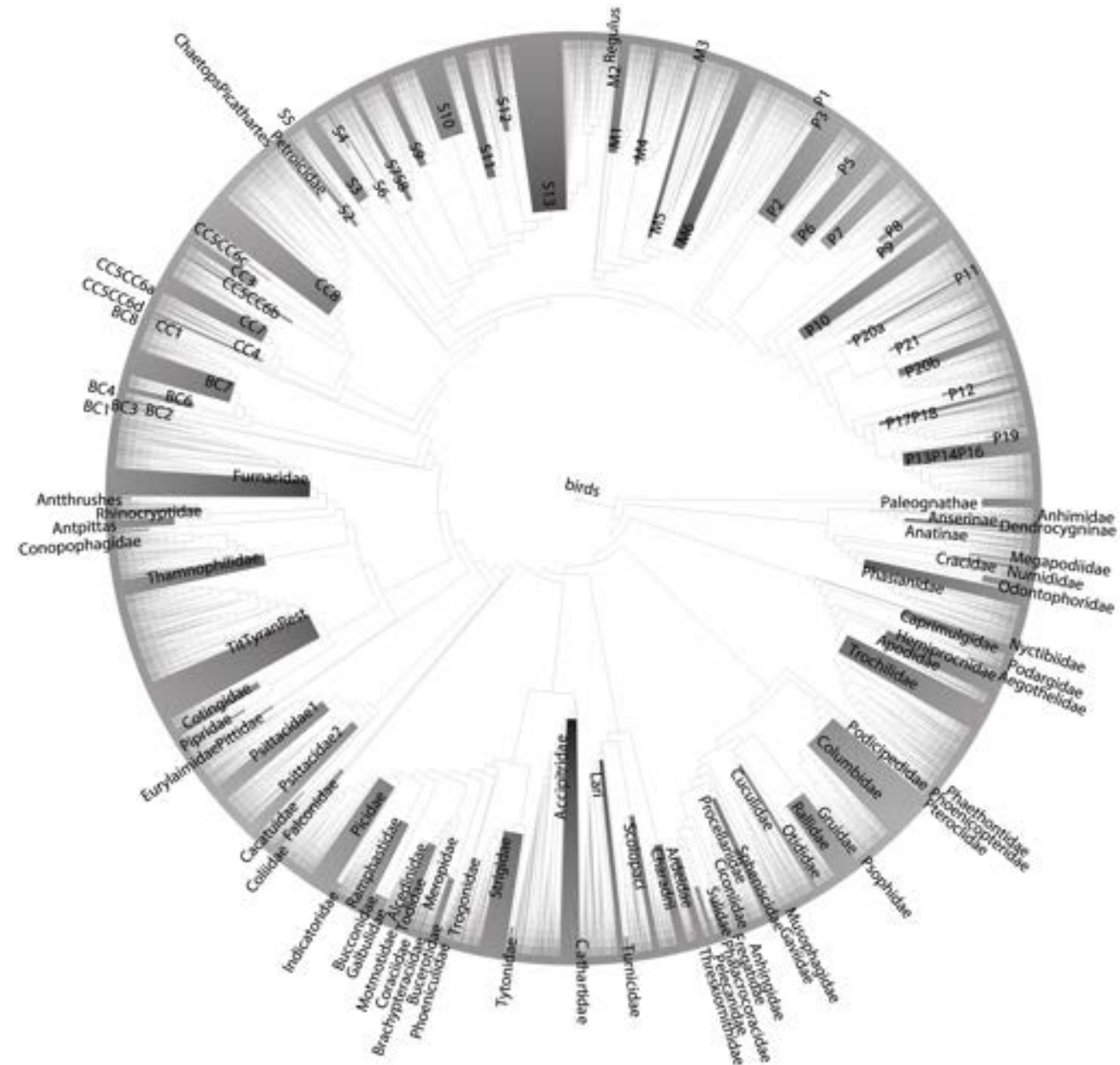# AVAToL: Open Tree of Life

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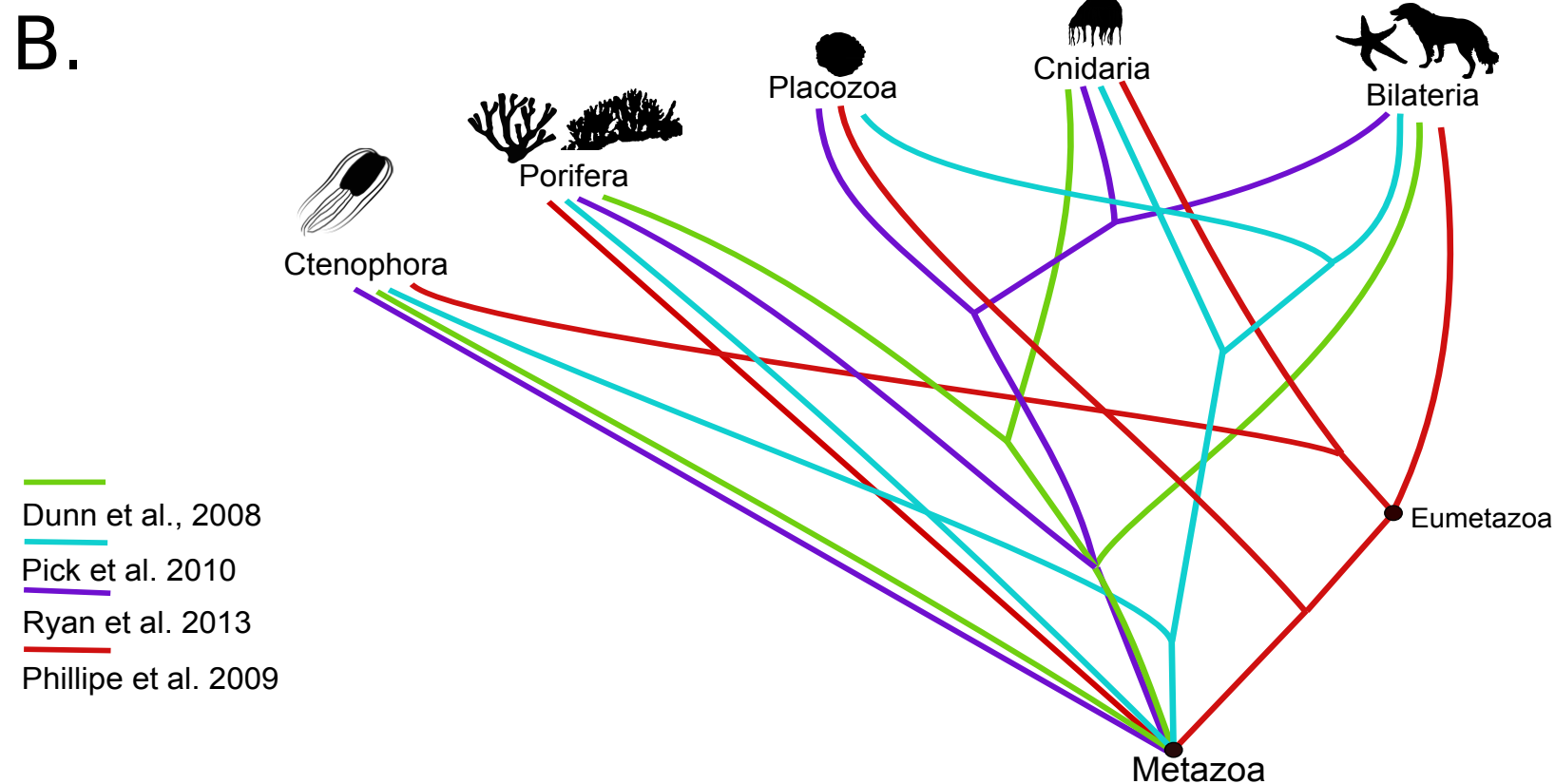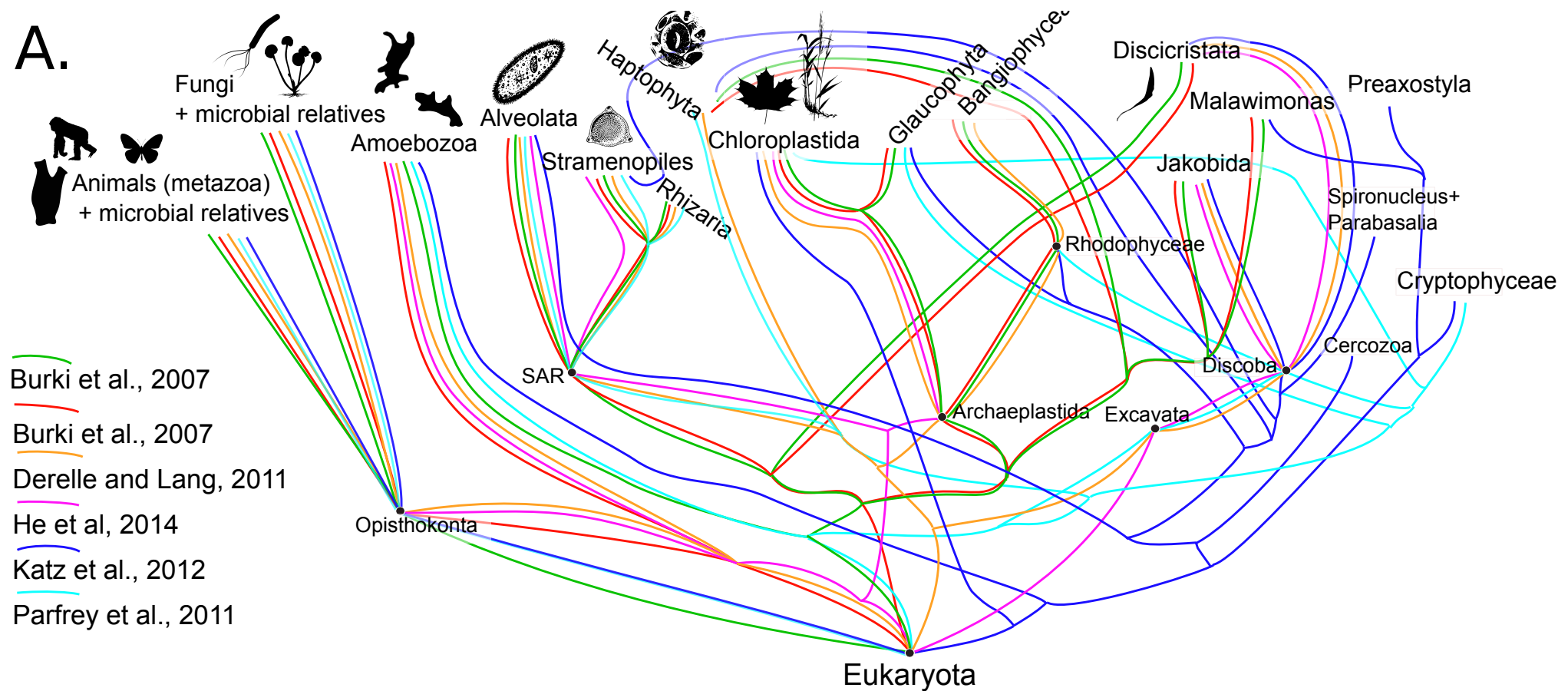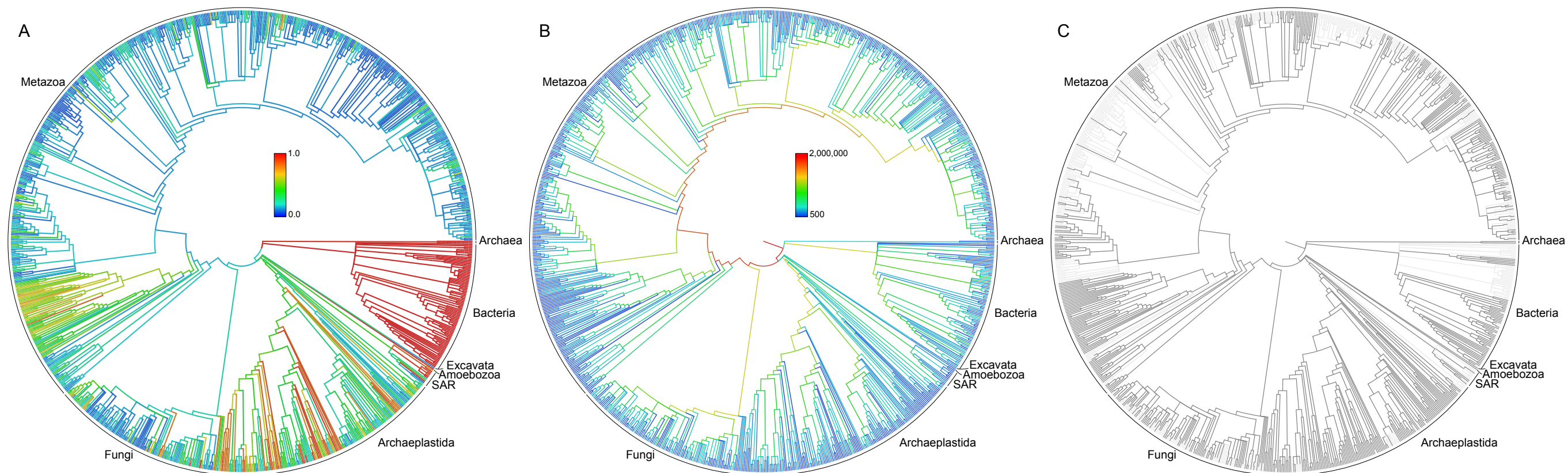
- Phylogeny of all life
  - no new relationships
- Curated taxonomy
- Accessible to the public
- Add/Extract phylogenies
- Two additional AVAToL groups

AVAToL: example phylogeny graphs

A.

Fungi + microbial relatives

Animals (metazoa) + microbial relatives

Amoebozoa

Alveolata

Stramenopiles

Rhizaria

Haptophyta

Chloroplastida

Glaucophyta

Bangiophyceae

Discicristata

Malawimonas

Preaxostyla

Jakobida

Spironucleus+ Parabasalia

Rhodophyceae

Cryptophyceae

Cercozoa

Discoba

Archaeplastida

Excavata

SAR

Opisthokonta

Eukaryota

- Burki et al., 2007
- Burki et al., 2007
- Derelle and Lang, 2011
- He et al, 2014
- Katz et al., 2012
- Parfrey et al., 2011

B.

Placozoa

Cnidaria

Bilateria

Porifera

Ctenophora

Eumetazoa

Metazoa

- Dunn et al., 2008
- Pick et al. 2010
- Ryan et al. 2013
- Phillipe et al. 2009
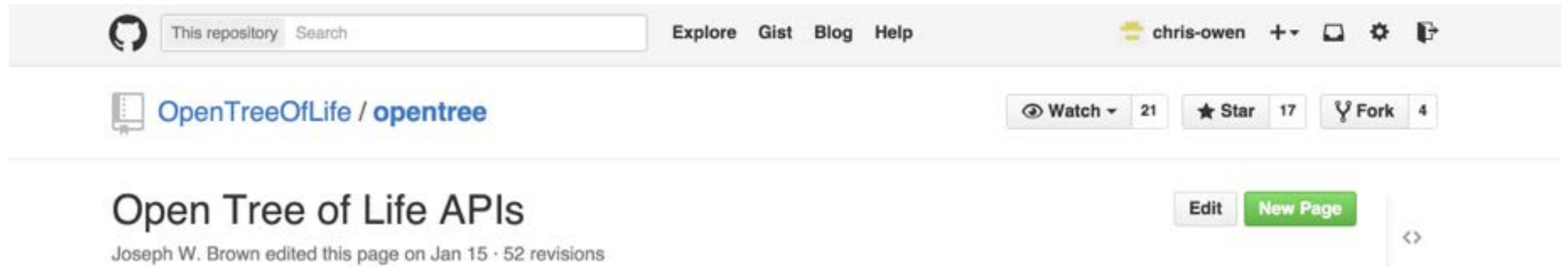
Hinchliff et al. *in review*

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   Phylesystem   Transformations   API wrappers   Design   Home

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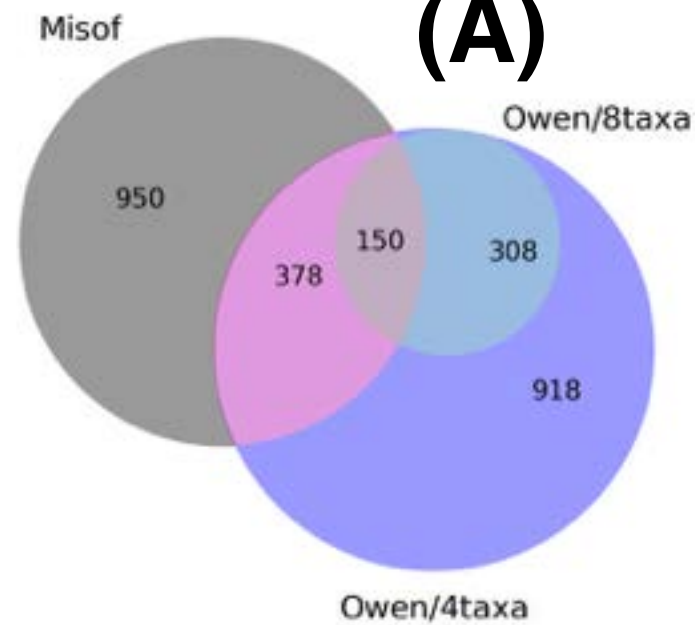- http://opentreeoflife.github.io/peyotl/

Open Tree Curator App

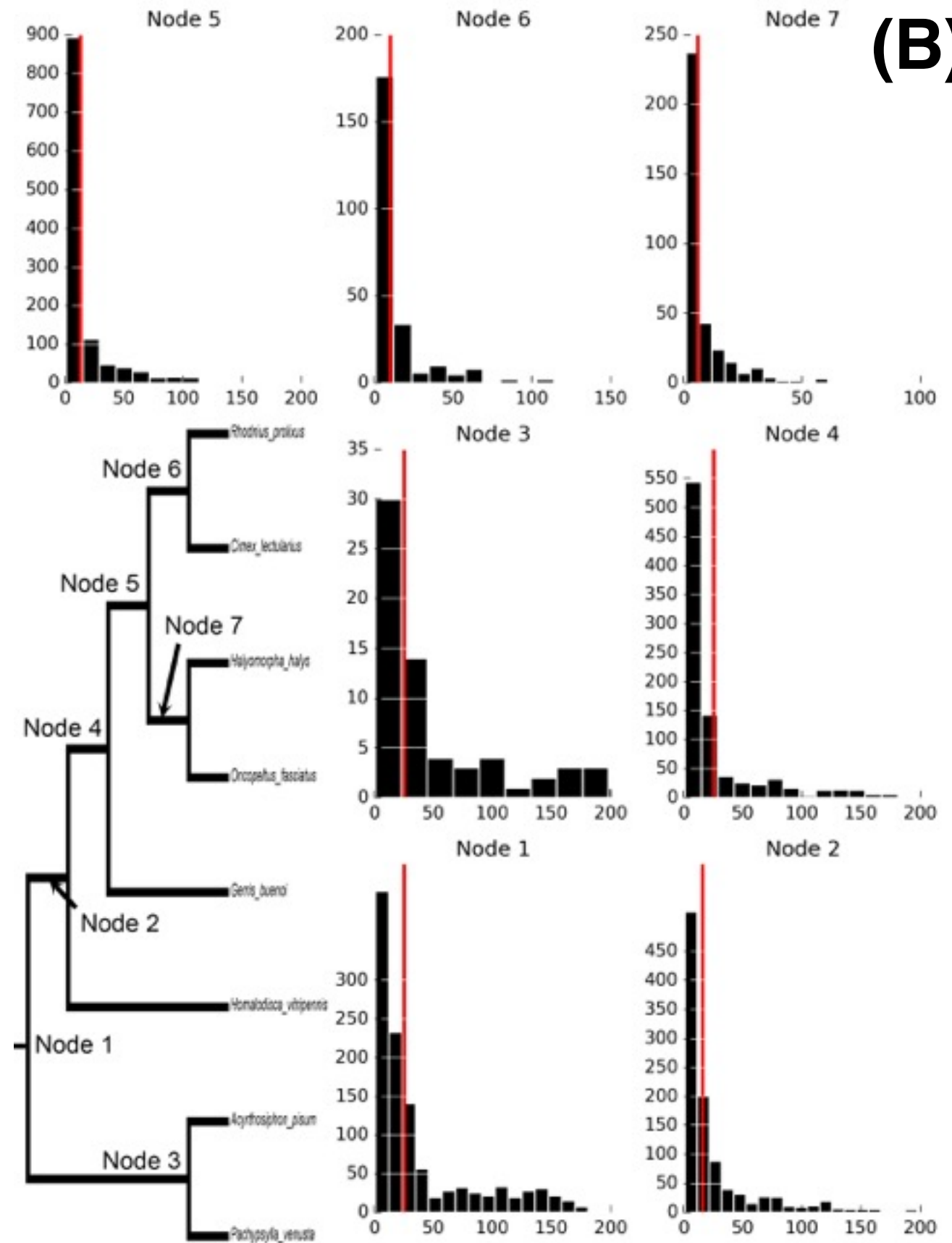# Why build the Hemiptera/ Hemipteroid Tree of Life using a supertree-like method?

Different ortholog sets
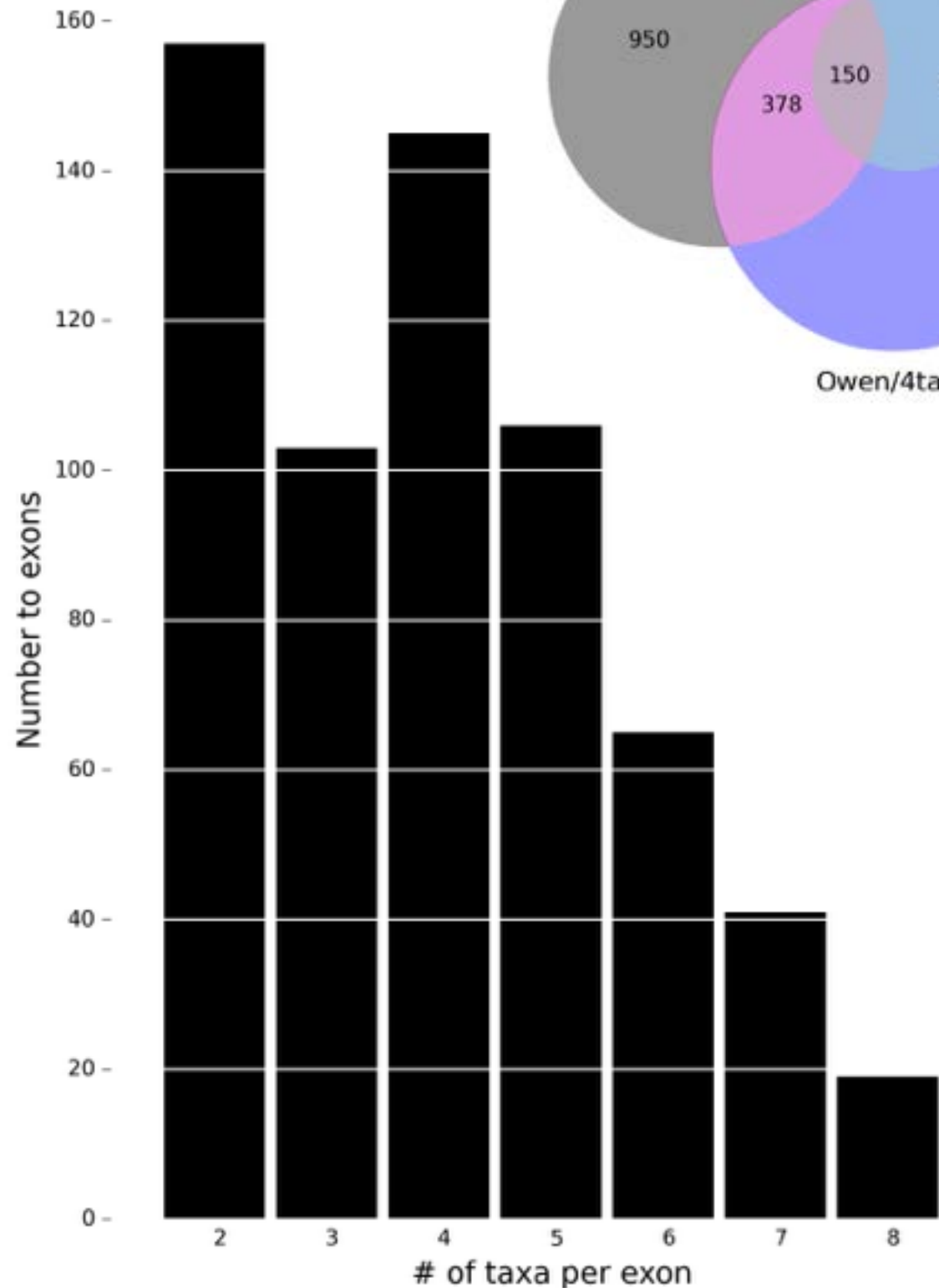
Different ortholog sets

mtDNA phylogeny
- 1500bp COI
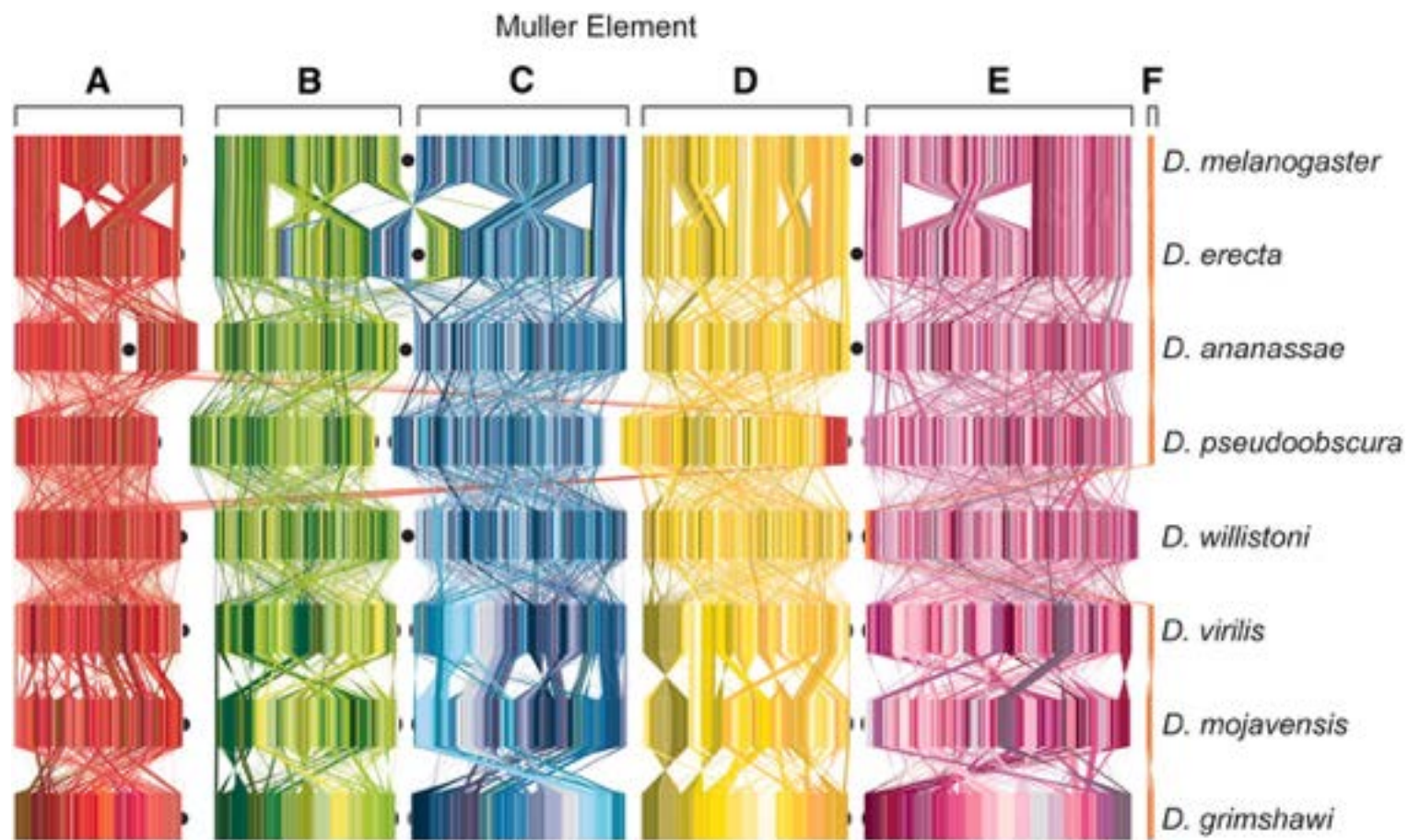- Garli ML phylo
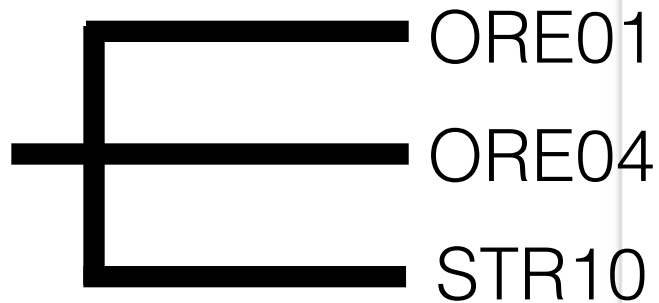
ORE01
ORE04
STR10

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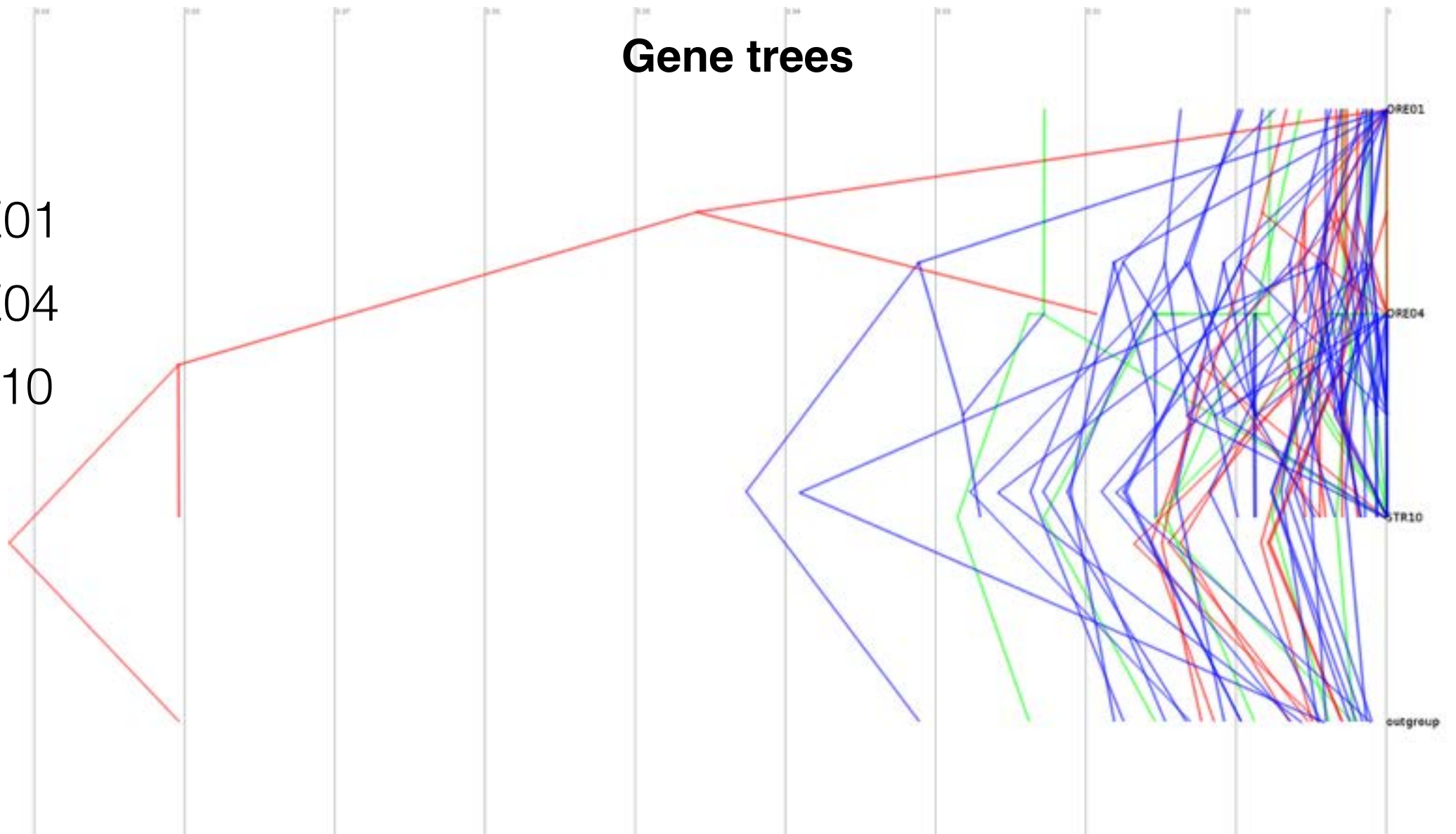
- ORE04, STR10 = 59.46%
- ORE01, ORE04 = 24.32%
- ORE01, STR10 = 16.22%

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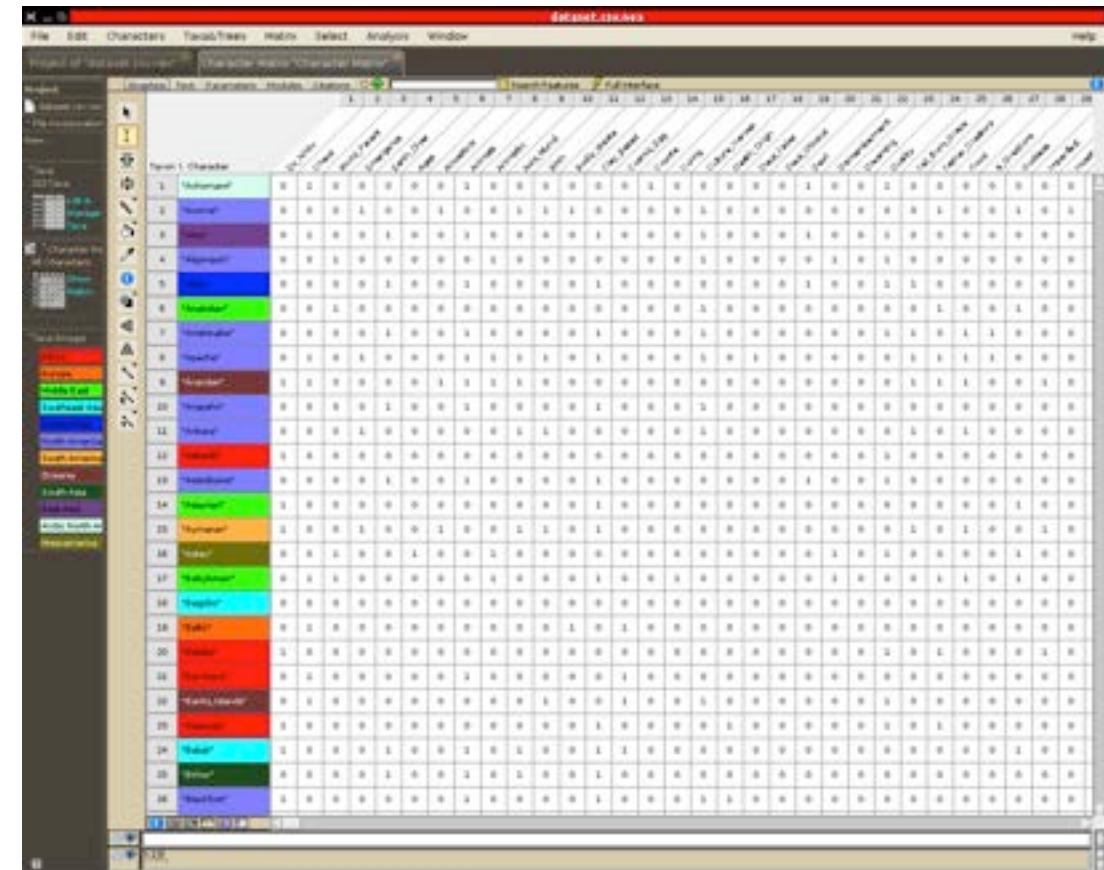- 40 loci from hybrid capture data

Different types of characters

Genome rearrangements

Morphological character matrices

Muller Element

A B C D E F

D. melanogaster
D. erecta
D. ananassae
D. pseudoobscura
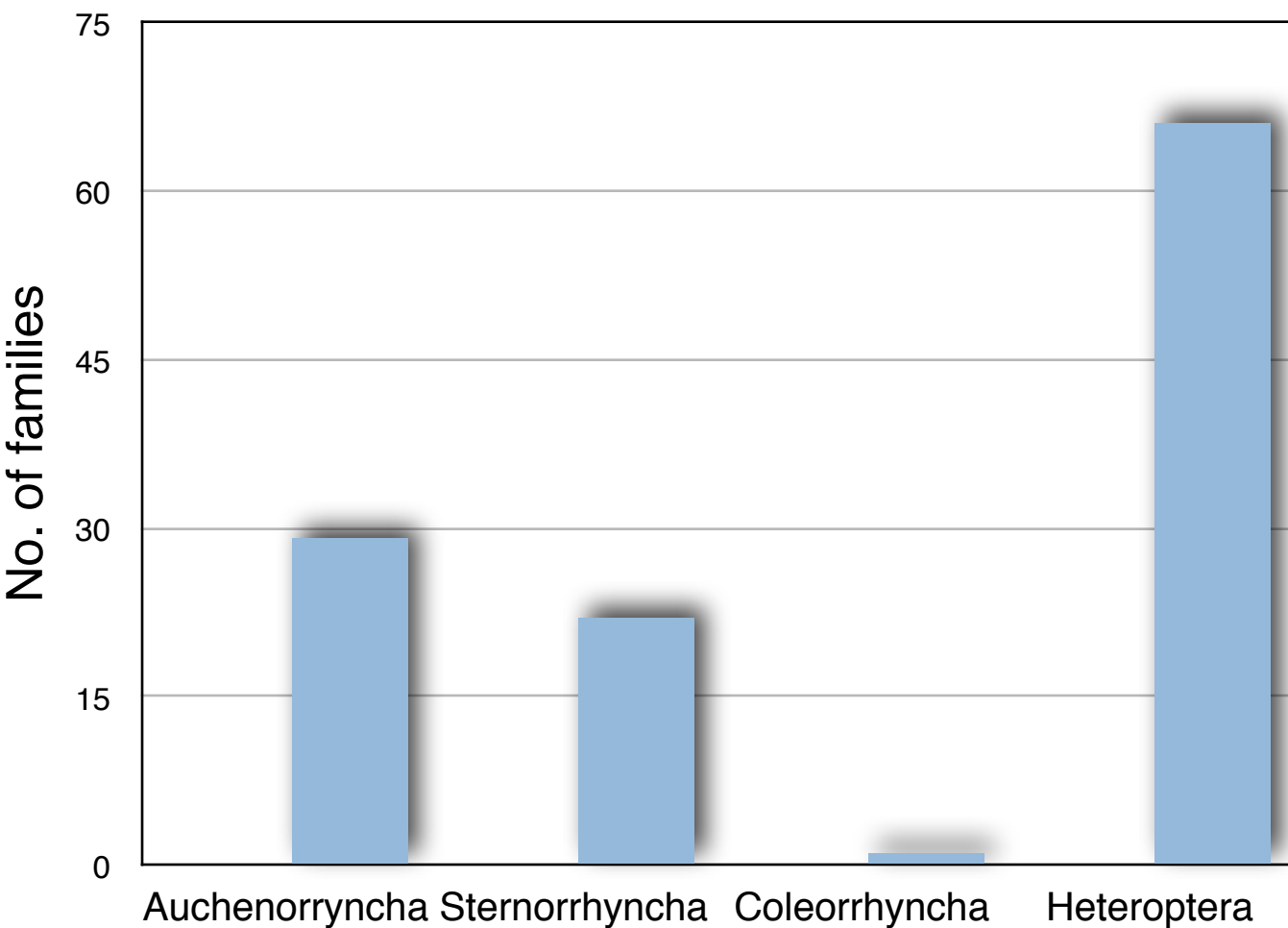D. willistoni
D. virilis
D. mojavensis
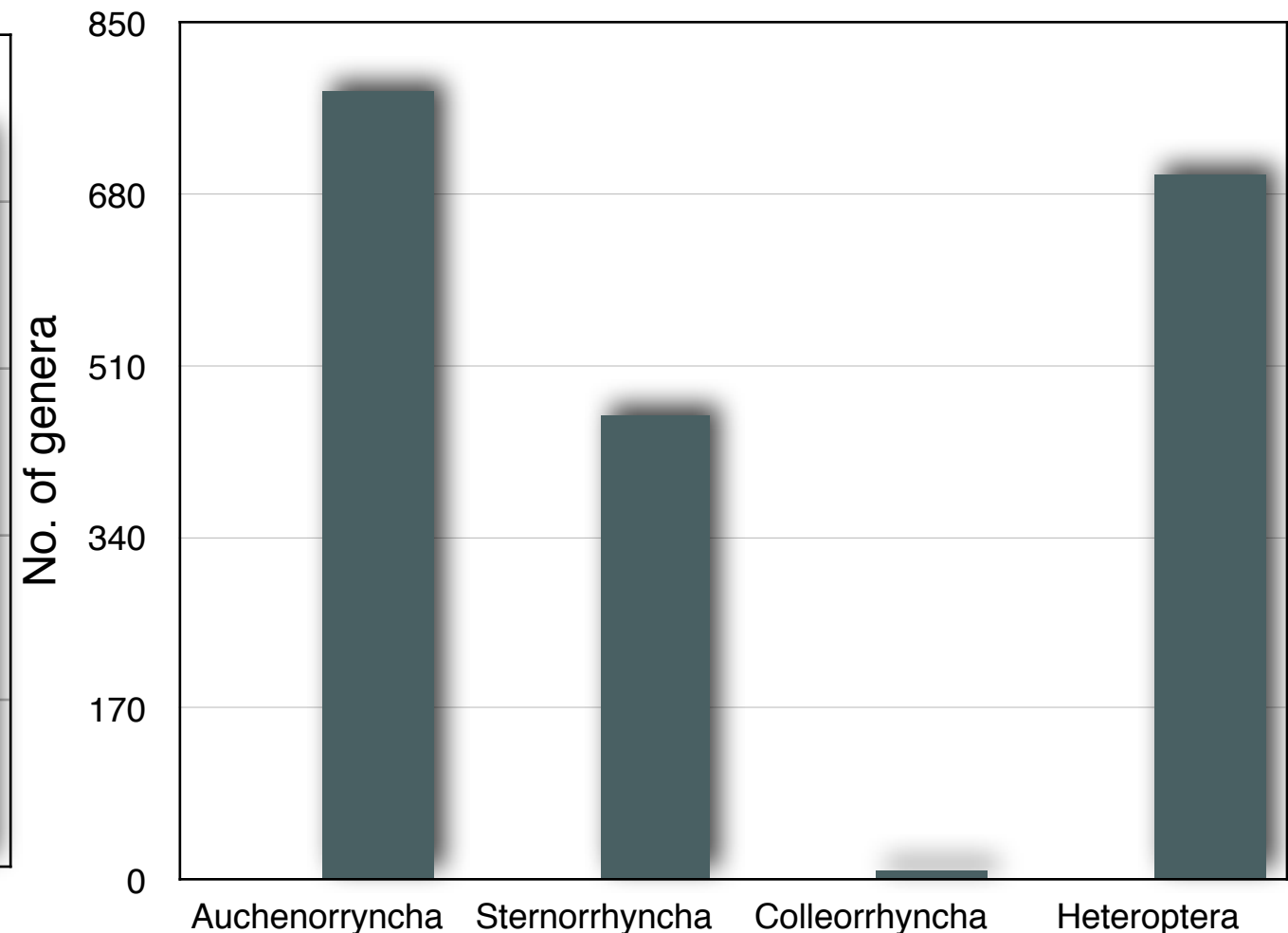D. grimshawi

Bhutkar et al. 2008

Taxonomic diversity of GenBank Sanger data

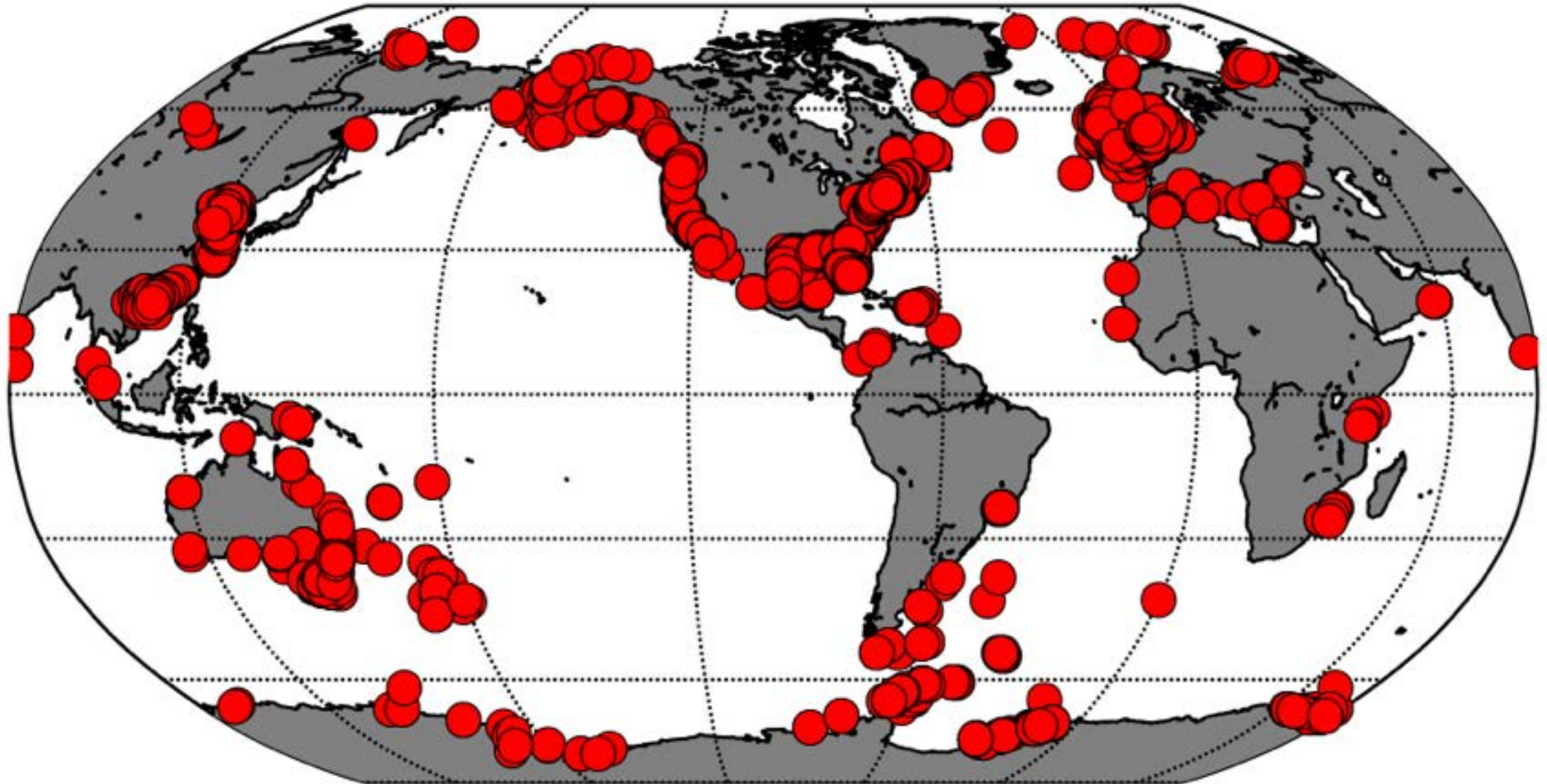Family diversity sampled within suborders

Generic diversity sampled within suborders

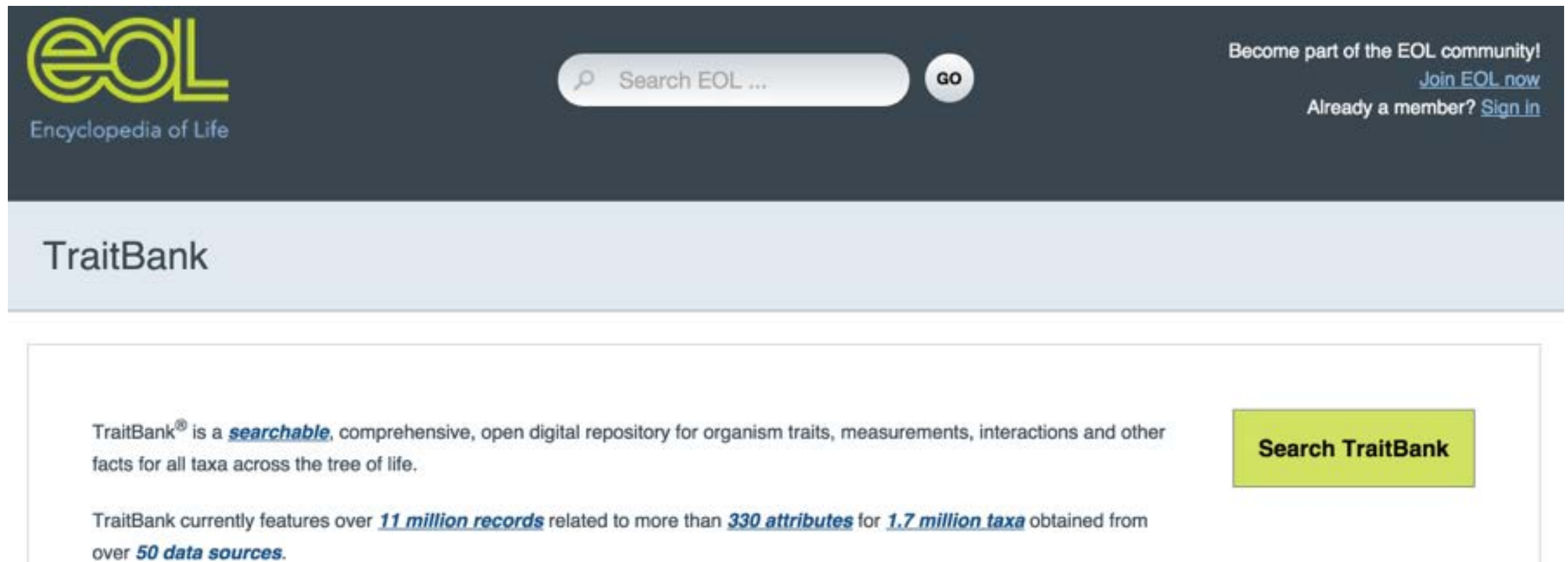## Open Tree Hemiptera Taxonomy Stats

- 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

- 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?