Linking Biodiversity Data Using Evolutionary History

Emily Jane McTavish

University of California, Merced
ejmctavish@ucmerced.edu, Twitter:@snacktavish

"How can we do a better job of understanding the diversity of life?"
Paul Hebert

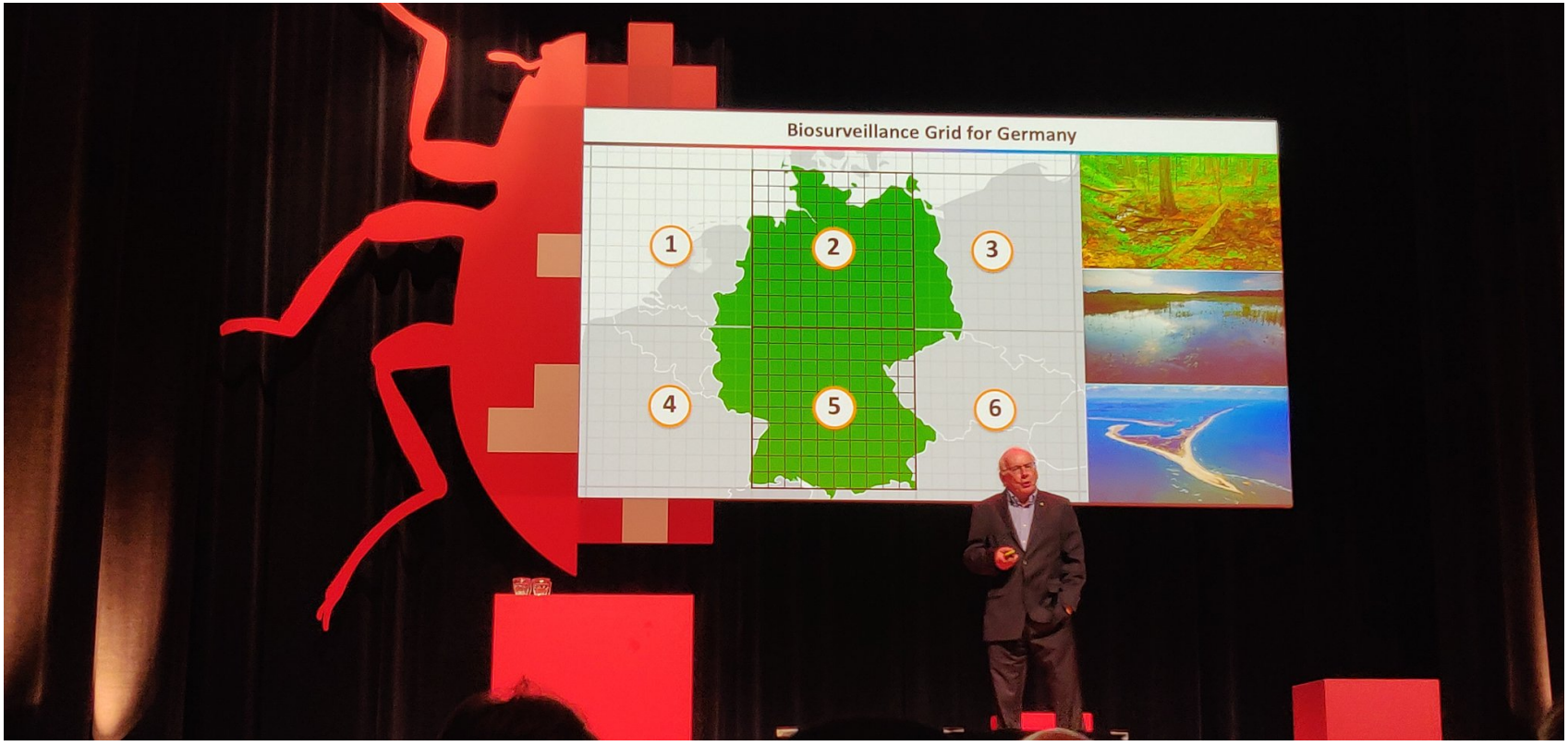
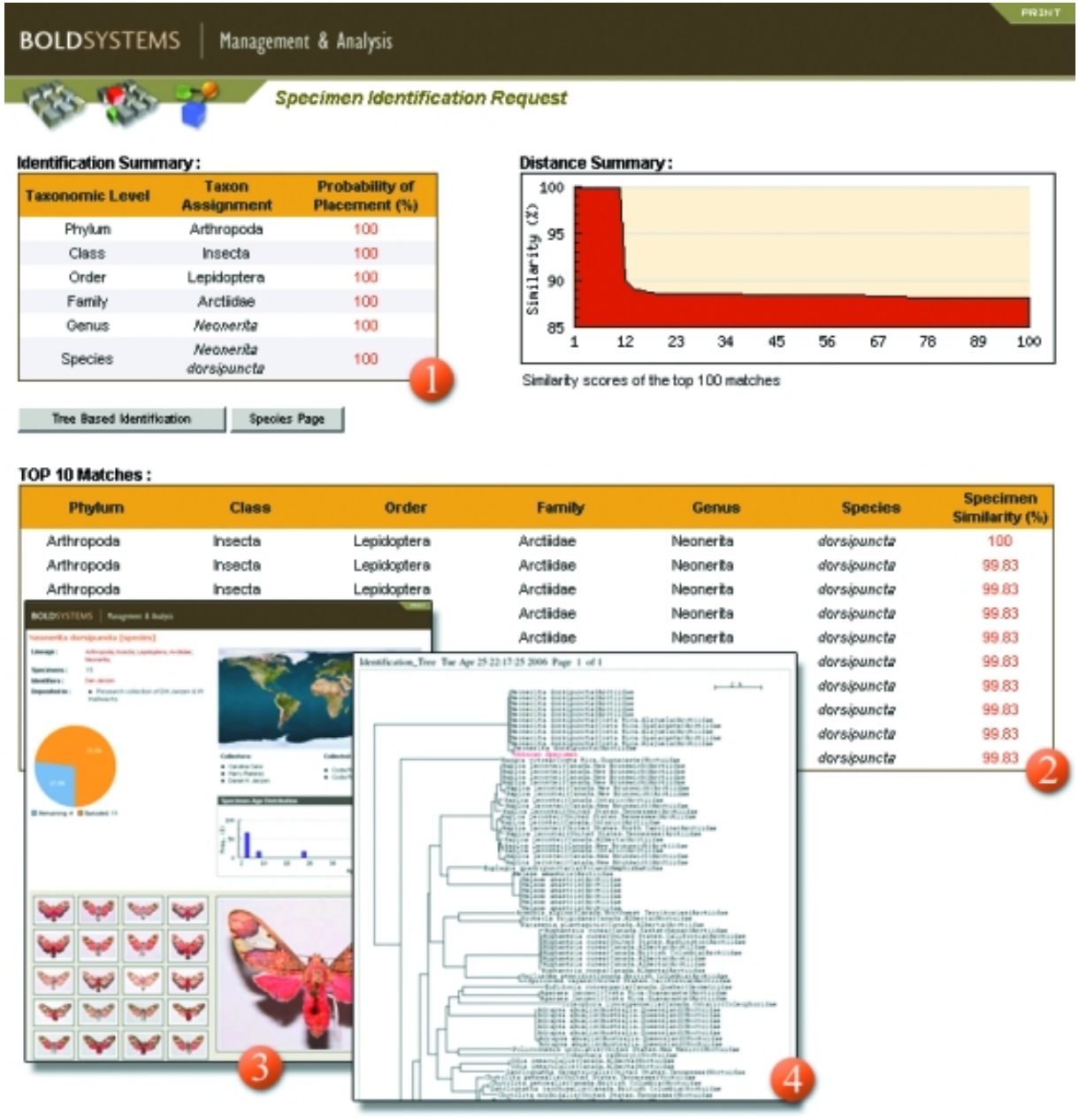
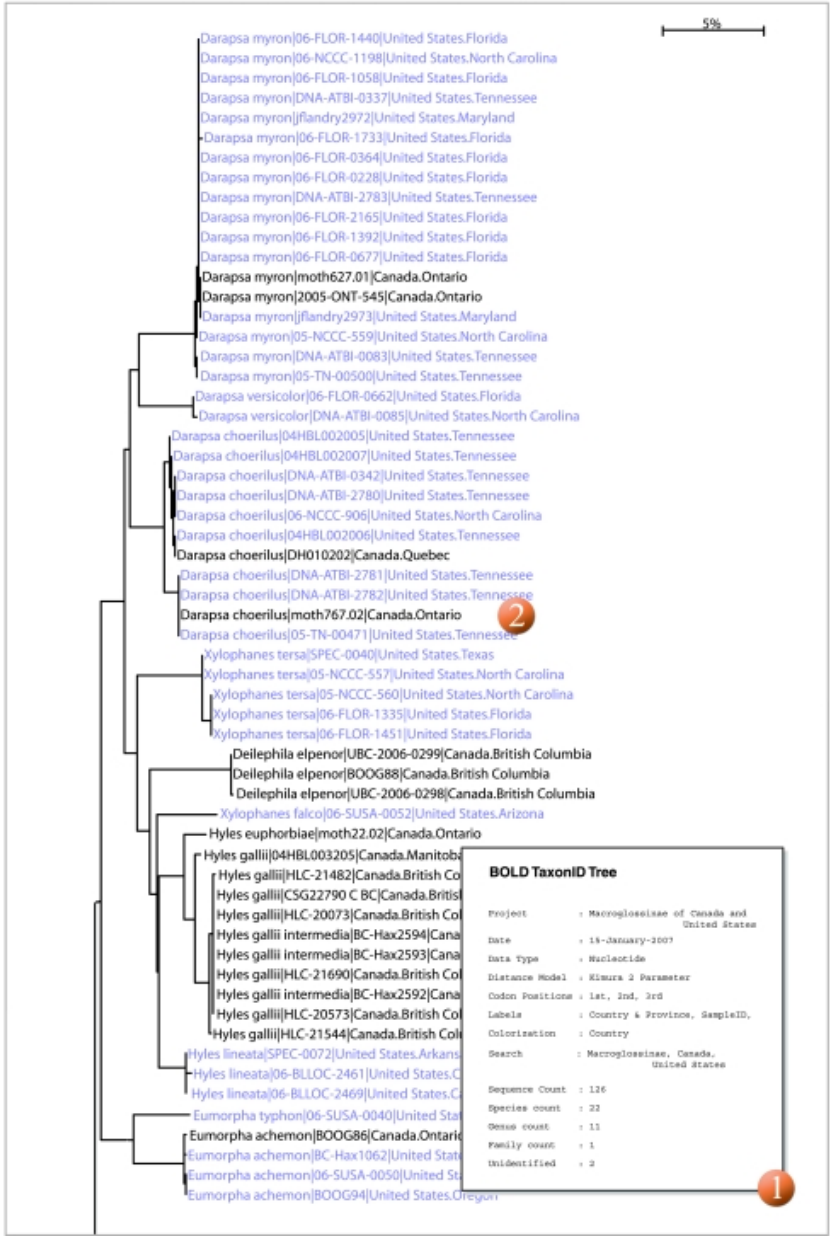


Image @yvan2935 Twitter



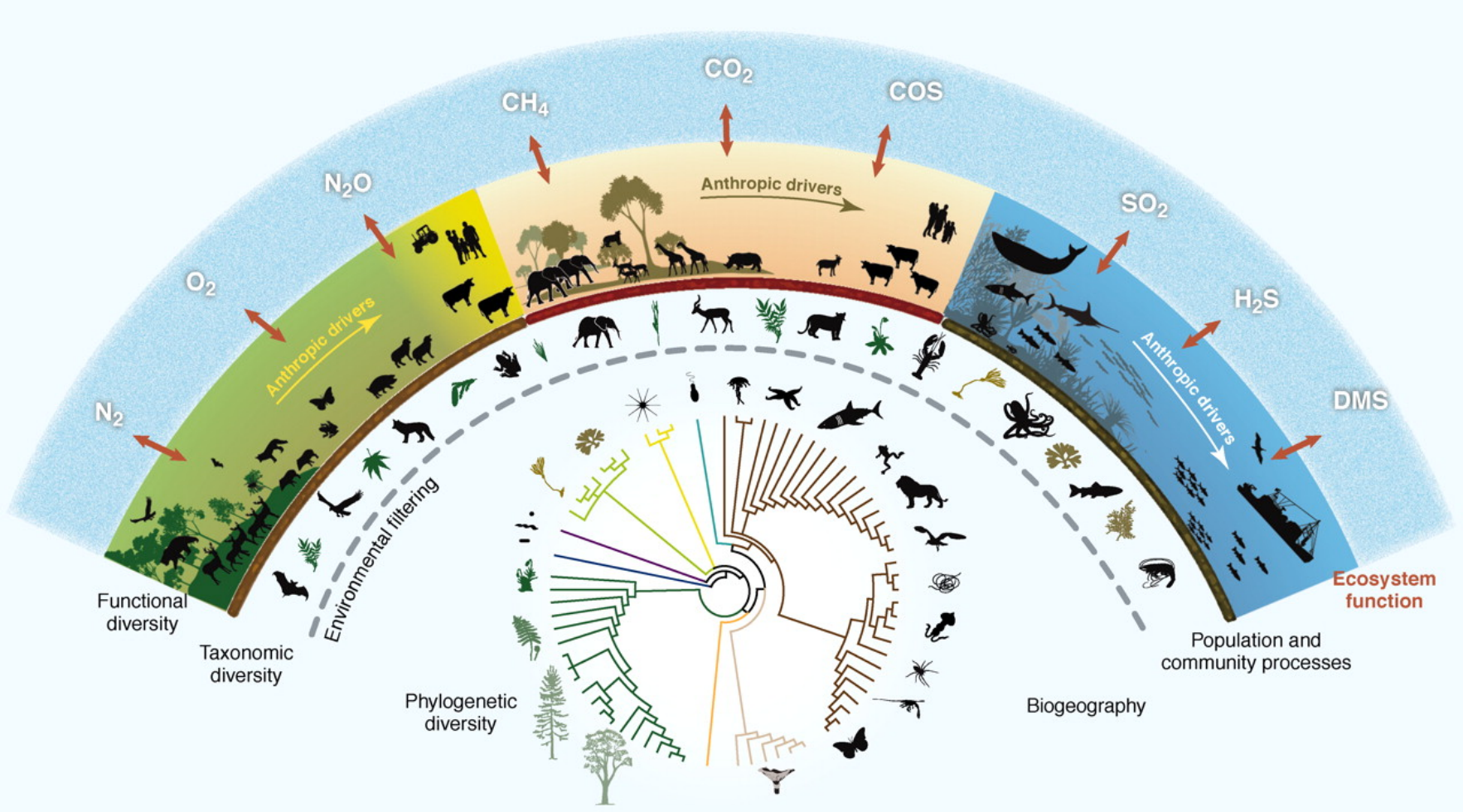
Ratnasingham and Hebert (2007)



Ratnasingham and Hebert (2007)

Barcodes allow us to identify, name, and place lineages in their evolutionary context

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

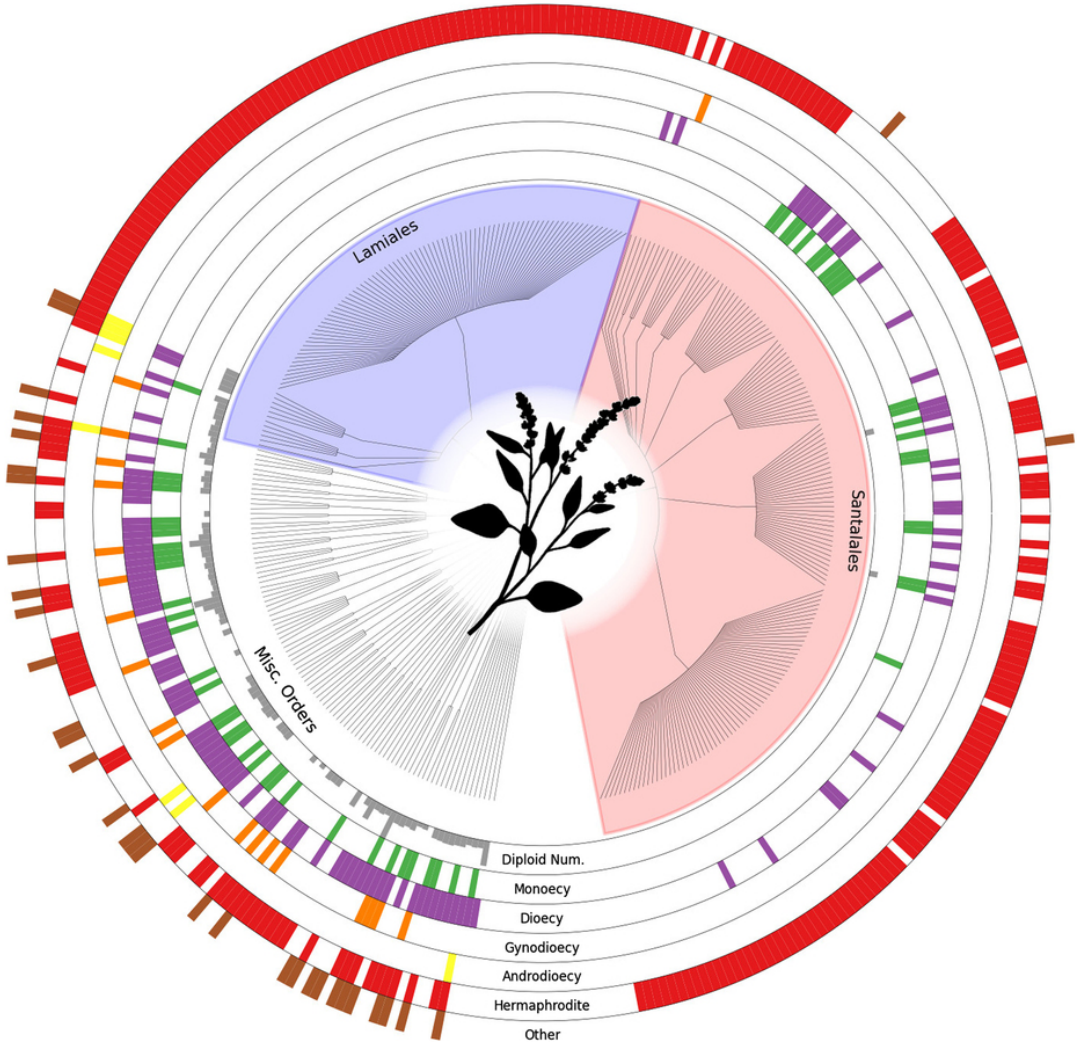


Naeem et al. (2012)

Taxonomy is often used as a proxy for shared evolutionary history

Carrie Seltzer, iNaturalist

Evolutionary transitions in plant mating



Tree of Sex Consortium (2014)

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 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.


What we have is artisanal
when we need industrial.
Paul Herbert

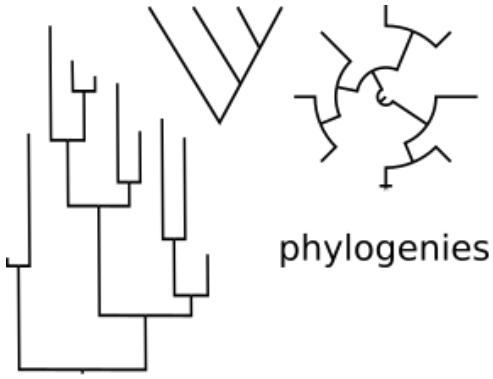

OPEN
Tree of Life

Goal: Build a tree of all life.

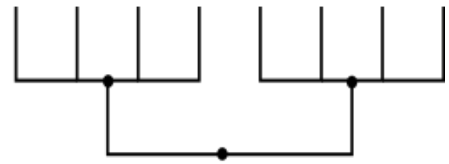
Goal: Build a tree of all life.

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

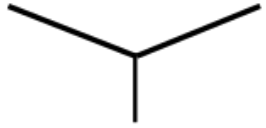
tree.opentreeoflife.org



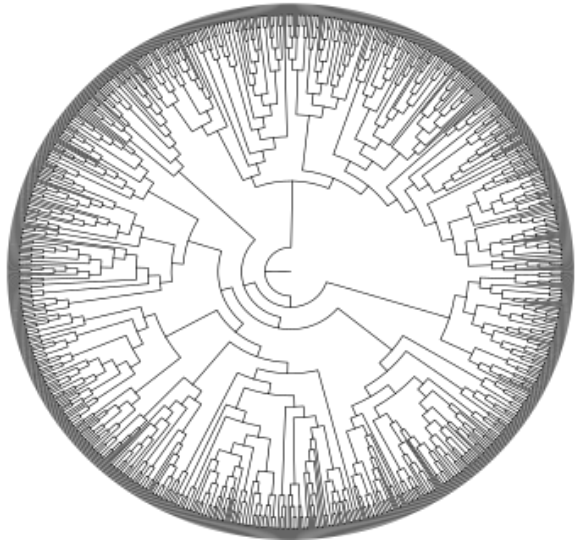
phylogenies

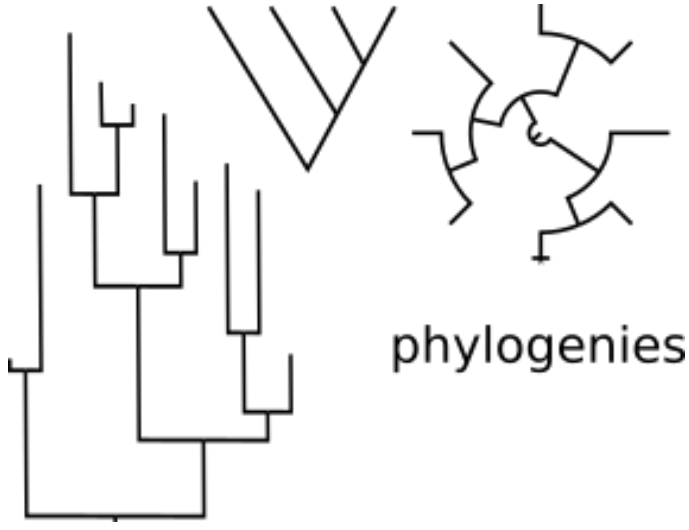


taxonomy

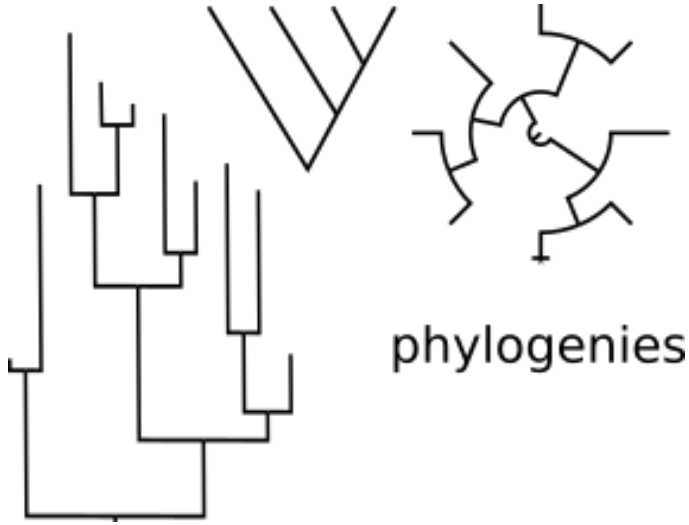


unified tree of life





phylogenies



Current synthetic tree

1,212 community curated representative phylogenies

87,682 tips from phylogeny

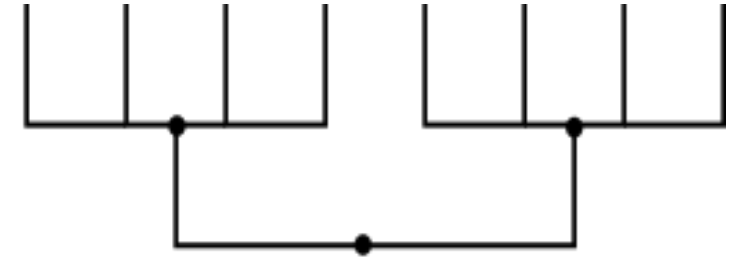
New draft posted October 9, 2019, will be updated with more input trees every few months

Redelings and Holder (2017)

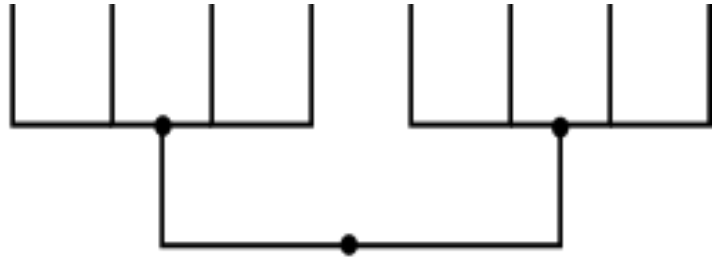
Input trees

- 253 individual curators of 4,431 uploaded studies
- Easy to use browser based interface
- Trees can be uploaded from any source, does not have to be own data
- 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)



taxonomy



taxonomy

2.7 million named taxa

Merges NCBI, Index fungorum, Silva, IRMNG, GBIF and other taxonomies

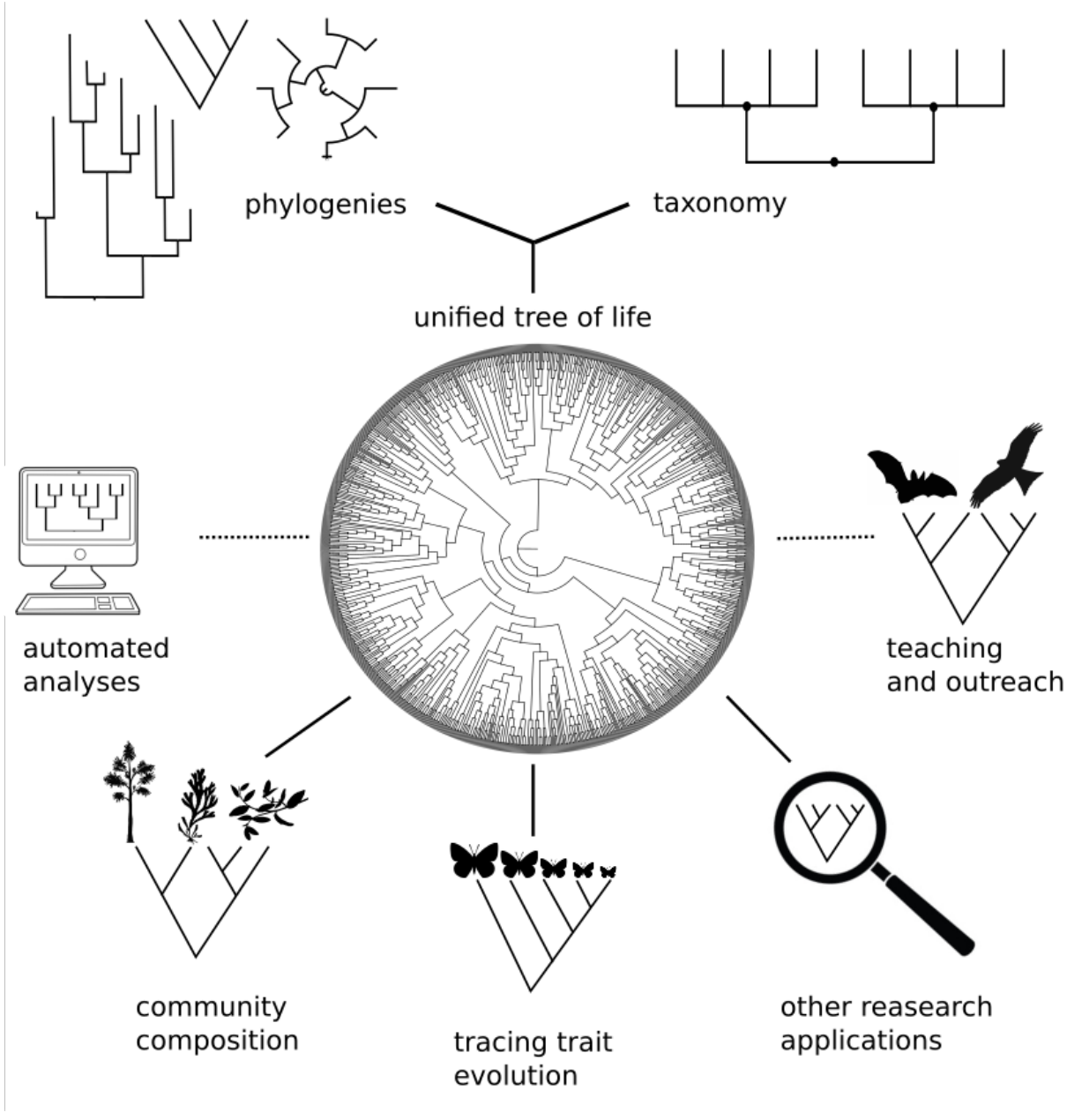
Scaffold for combining ranked phylogenetic estimates

New drafts released as inputs change

Rees and Cranston (2017)

Taxonomic versioning

- OpenTree Taxonomy (OTT) v.3.1 published October 14, 2019
- Includes newest NCBI updates
- Version including updated GBIF names forthcoming in November.



McTavish et al. (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) ROpenSci

CC0 license provides fully open access for downstream re-usability

CC0 license provides fully open access for downstream re-usability

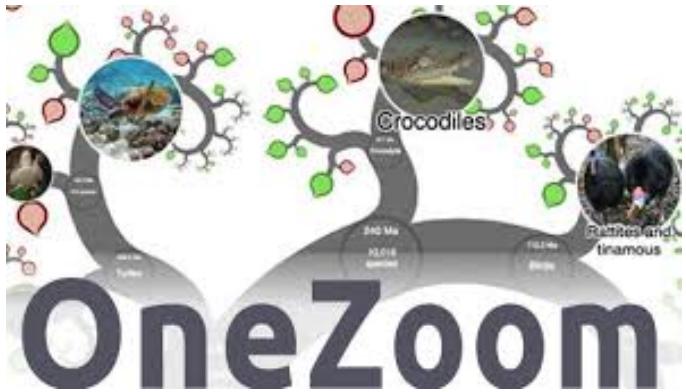
Open Tree provides the tree backend for:

PHYLOtastic

DATElife

Lifemap

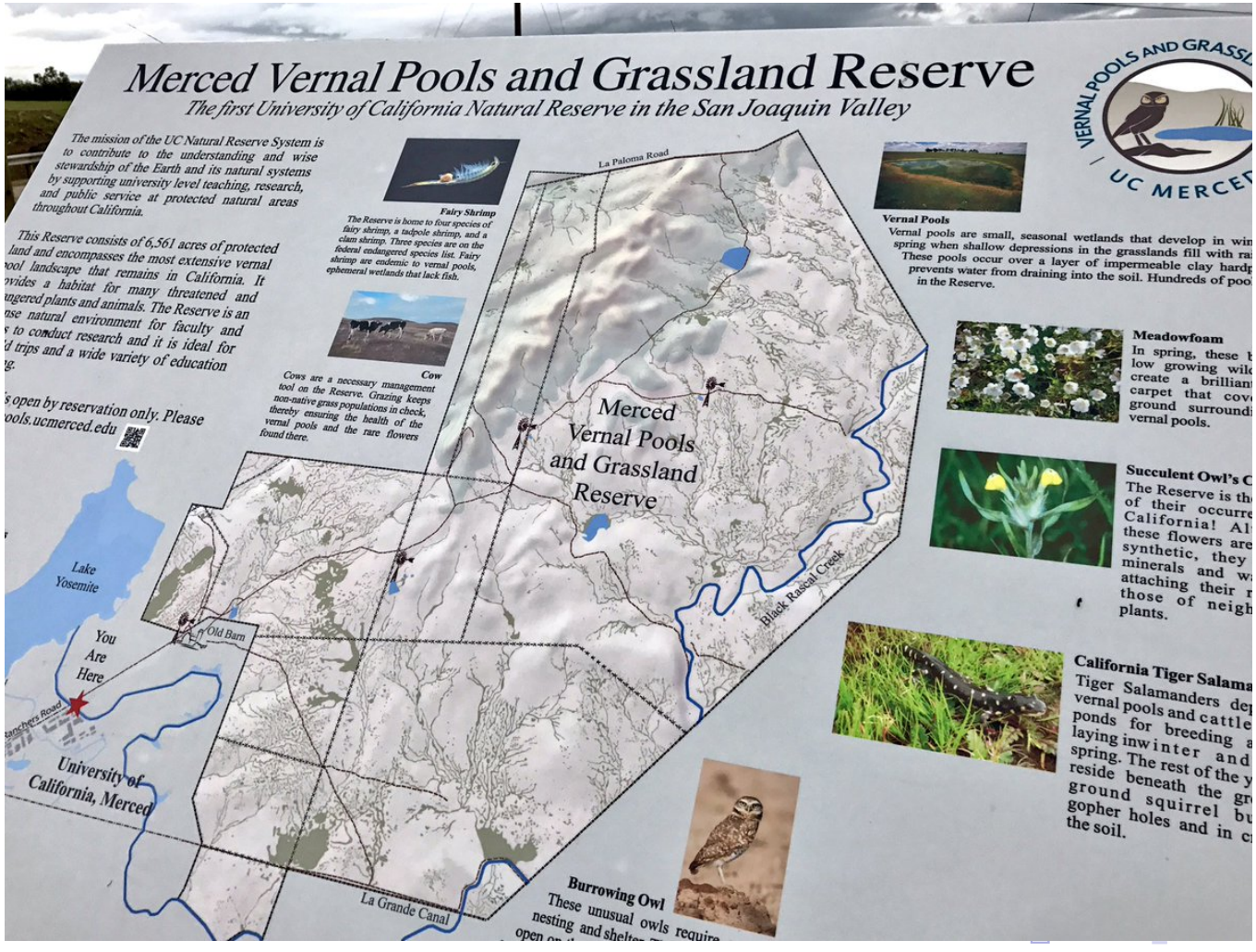
OPEN Tree of Life



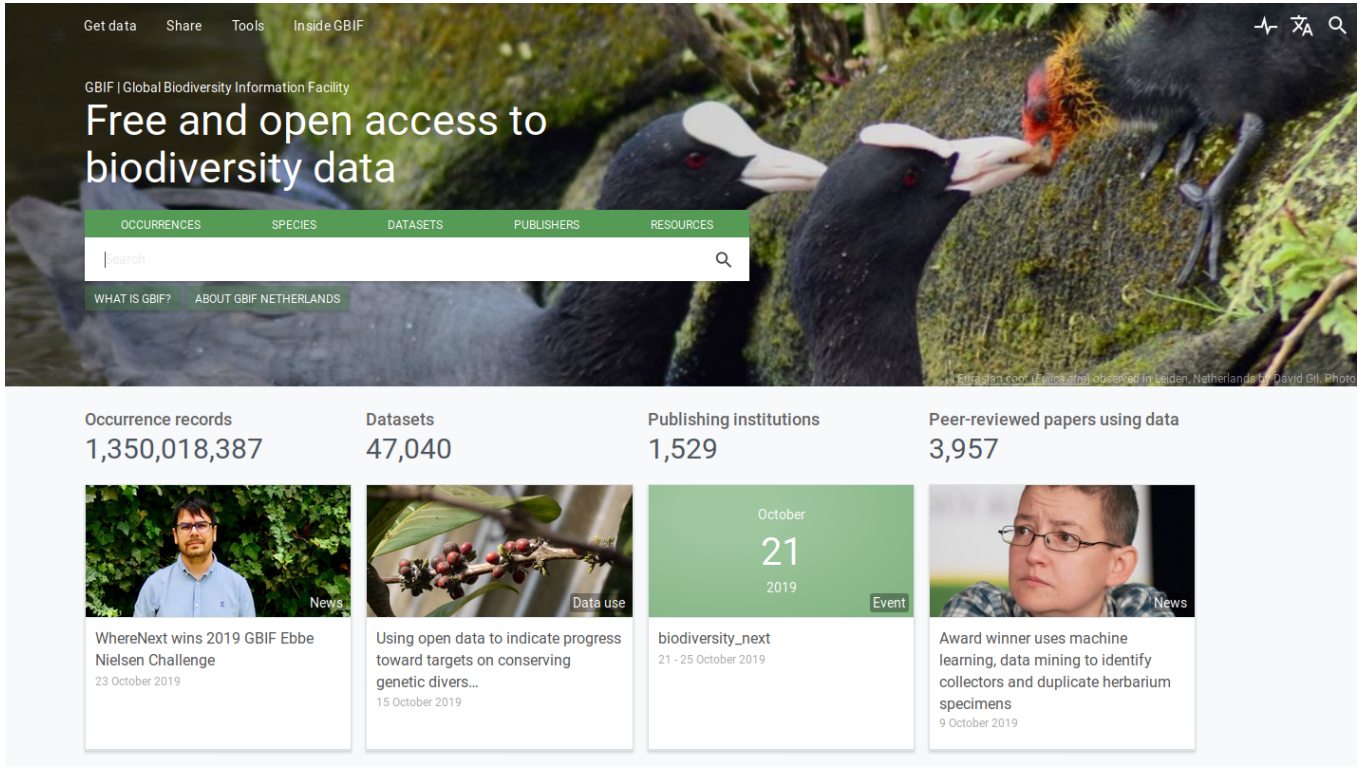
How can we integrate phylogenetic estimates into biodiversity research?

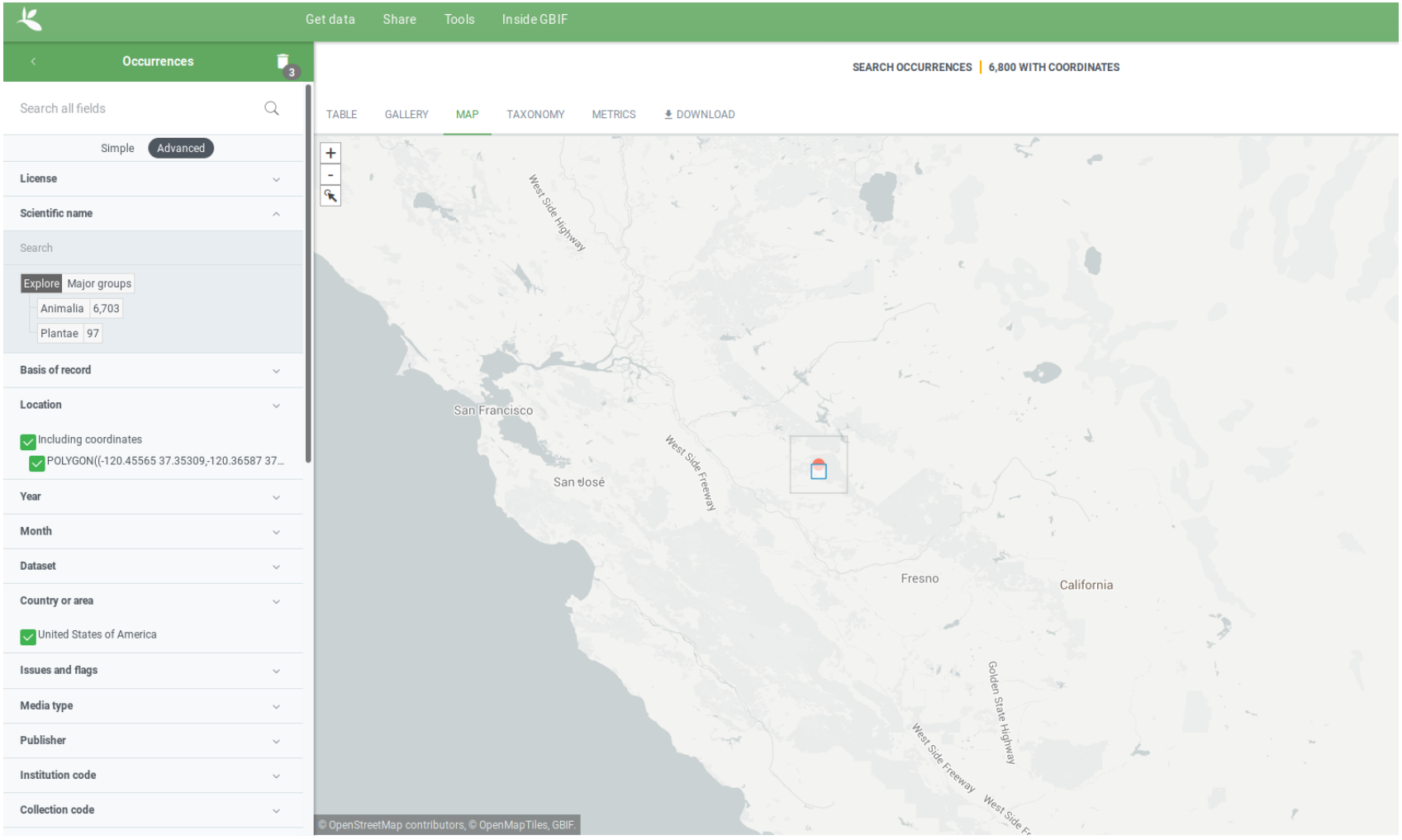
How can we integrate phylogenetic estimates into biodiversity research?

Case study: What is the evolutionary diversity of the species found at the UC Merced Vernal Pools Reserve

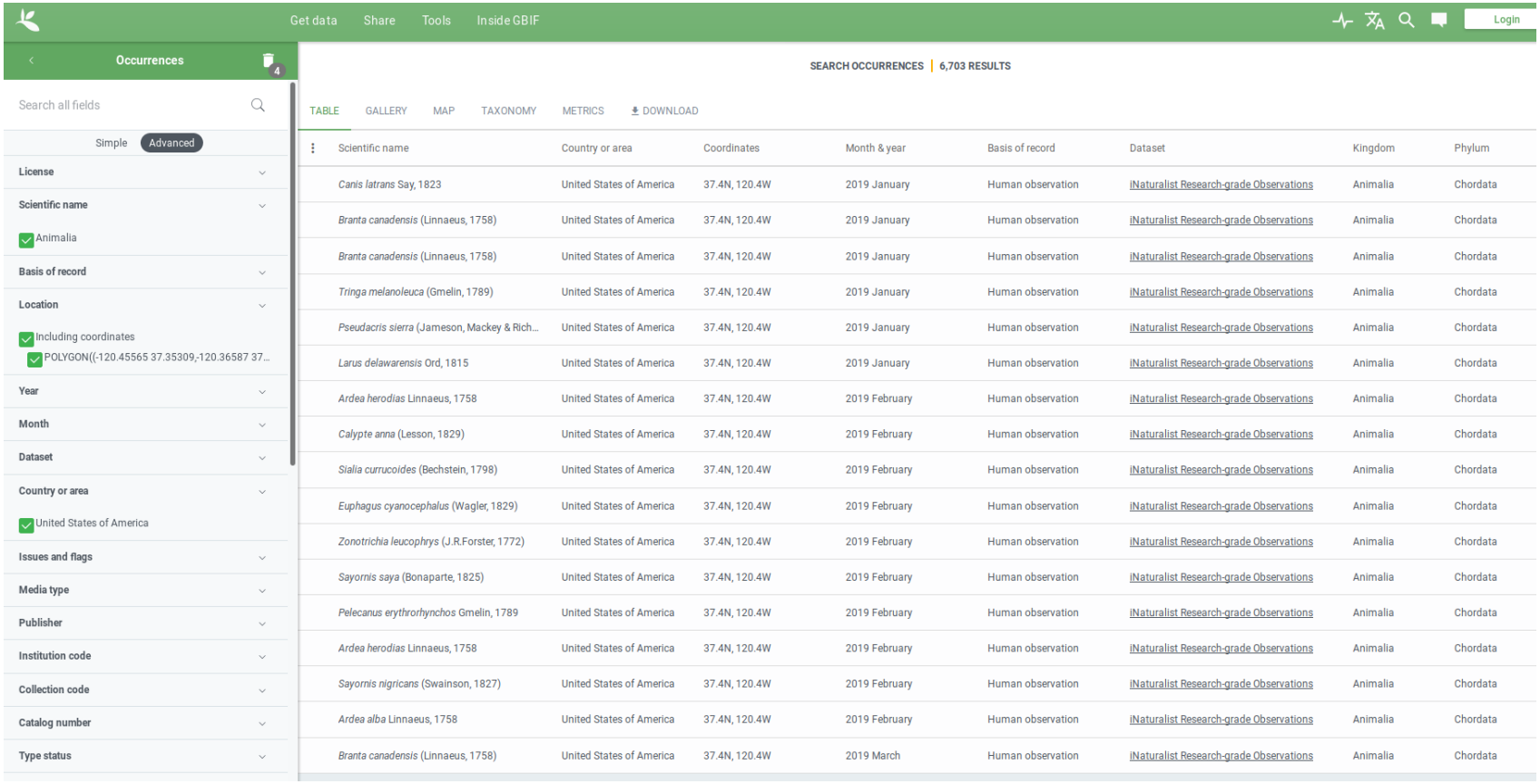


Get data from Global Biodiversity Information Facility




Get data
Share
Tools
Inside GBIF
Occurrences
3
Search all fields
Simple
Advanced
License
Scientific name
Search
Explore
Major groups
Animalia 6,703
Plantae 97
Basis of record
Location
Including coordinates
POLYGON((-120.45565 37.35309,-120.36587 37...
Year
Month
Dataset
Country or area
United States of America
Issues and flags
Media type
Publisher
Institution code
Collection code
SEARCH OCCURRENCES | 6,800 WITH COORDINATES
TABLE
GALLERY
MAP
TAXONOMY
METRICS
DOWNLOAD
West Side Highway
San Francisco
San José
West Side Freeway
Fresno
California
Golden State Highway
West Side Freeway
West Side Fr
© OpenStreetMap contributors, © OpenMapTiles, GBIF.

6,909 records in GBIF



GBIF.org (17 October 2019) GBIF Occurrence Download https://doi.org/10.15468/dl.9bigak #CiteTheDoi!

https://www.inaturalist.org/observations/32478397

iNaturalist Explore Community More

Black Saddlebags (*Tramea lacerata*) Research Grade

lunasare
19 observations

Observed:
Sep 10, 2019 · 8:39 AM PDT

Submitted:
Sep 10, 2019 · 10:27 AM PDT

University of California, Merced... Show Details

Activity

~~lunasare~~ ~~suggested an ID~~ ID Withdrawn 1mo

~~Saddlebag Gliders~~
~~Genus *Tramea*~~

joshualincoln suggested an ID Improving 1mo

Black Saddlebags
Tramea lacerata

Community Taxon

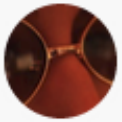
What's this?

Black Saddlebags (*Tramea lacerata*)
Cumulative IDs: 3 of 3

0 **2/3rds** 3

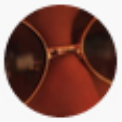
Agree About

Projects (1)



kathyclaypolebiggs suggested an ID 1mo

Black Saddlebags
Tramea lacerata

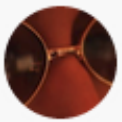


kathyclaypolebiggs commented 1mo

So glad for this submission as it upgrades a prior sight only record for the species in Merced County! Could you please also submit this to Odonata Central where we record all such records. https://www.odonatacentral.org

lunasare suggested an ID 1mo

Black Saddlebags
Tramea lacerata



kathyclaypolebiggs commented 1mo

September 10, 2019
Merced County
iNatauralist
University of California, Merced, Merced
!! Black Saddlebags Tramea lacerata –photograph updates prior sight only record CA Chart

name or id

Open Tree taxonomy: **Tramea lacerata**

The current taxonomy version is ott3.1 (click for more information). See the OTT documentation for an explanation of the taxon flags used below, e.g., `extinct`

Taxon details

species **Tramea lacerata** ncbi:126233 (gbif:1428475, irmng:10341497) `sibling_higher` (OTT id 340907)

View this taxon in the current synthetic tree

Lineage

life > cellular organisms > Eukaryota > Opisthokonta > Holozoa > Metazoa > Eumetazoa > Bilateria > Protostomia > Ecdysozoa > Panarthropoda > Arthropoda > Mandibulata > Pancrustacea > Hexapoda > Insecta > Dicondylia > Pterygota > Palaeoptera > Odonata > Epiprocta > Anisoptera > Cavilabiata > Libellulidae > Tramea

Can access taxon information directly from the Open Tree API's

```
~$ curl -X POST https://api.opentreeoflife.org/v3/taxonomy/taxon_info -H 'conte
nt-type:application/json' -d '{"source_id":"gbif:1428475"}'

{
 "flags": [
  "sibling_higher"
 ],
 "is_suppressed": false,
 "is_suppressed_from_synth": false,
 "name": "Tramea lacerata",
 "ott_id": 340907,
 "rank": "species",
 "source": "ott3.1draft2",
 "synonyms": [],
 "tax_sources": [
  "ncbi:126233",
  "gbif:1428475",
  "irmng:10341497"
 ],
 "unique_name": "Tramea lacerata"
```

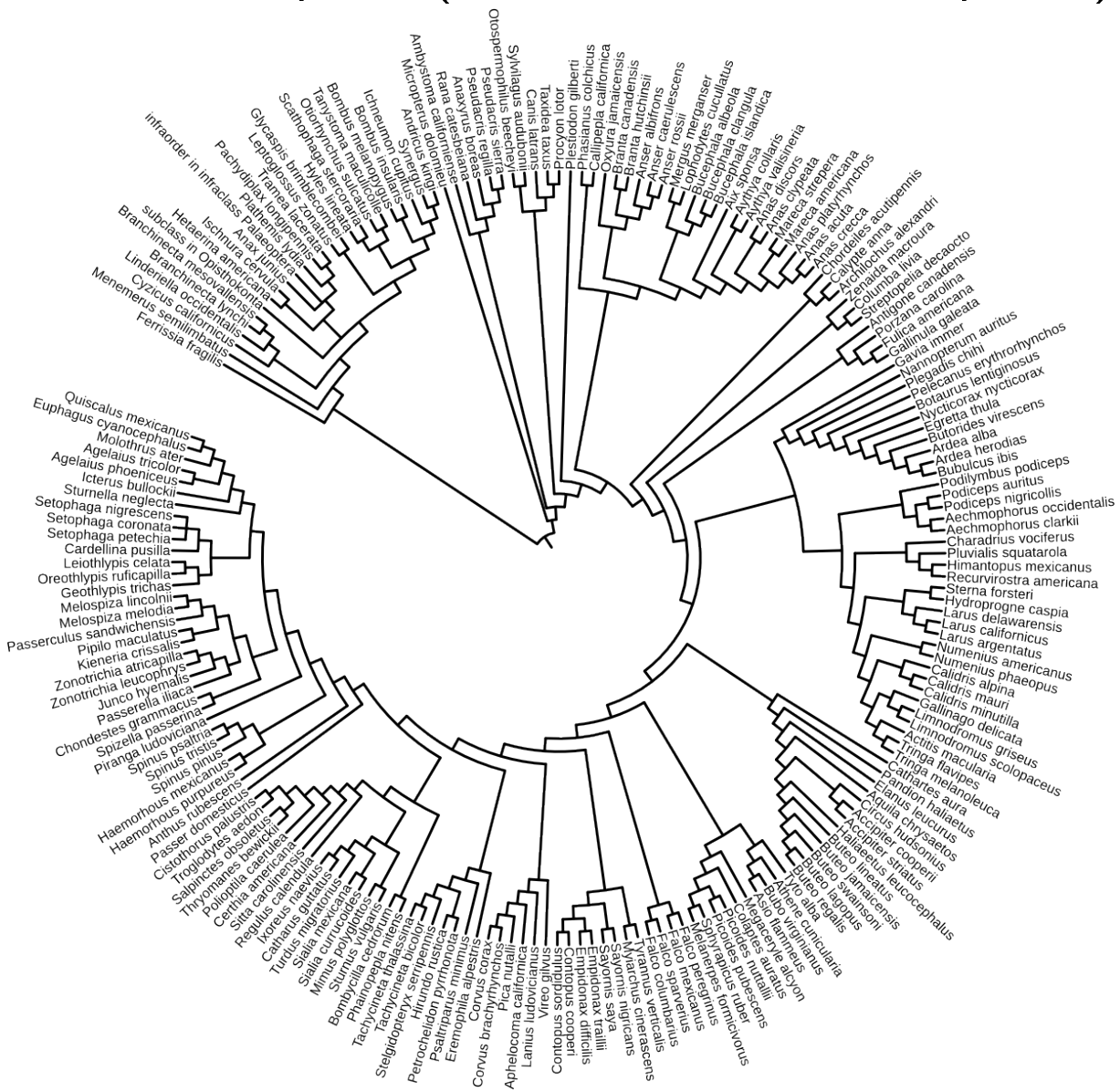
API documentation at `https://github.com/OpenTreeOfLife/germinator/wiki/Open-Tree-of-Life-Web-APIs`

Use the GBIF ids to get a synthetic tree for all species in the reserve

Python code on Github,

`https://github.com/McTavishLab/biodiversity_next`

Animals of the vernal pools (6,708 records of 205 species)



Tree vis via `https://itol.embl.de`, phylogenies from 34 studies

Automatic download of references for phylogenetic inferences used in induced subtree (34 studies went into the animal tree).

```
Cho, S., Zwick A., Regier J., Mitter C., Cummings M.P., Yao J., Du Z., Zhao H., Kawahara A.Y.,
Weller S.J., Davis D.R., Baixeras J., Brown J.W., & Parr C. 2011. Can deliberately incomplete
gene sample augmentation improve a phylogeny estimate for the advanced moths and butterflies (
Hexapoda: Lepidoptera)?. Systematic Biology 60 (6): 782-796.
http://dx.doi.org/10.1093/sysbio/syr079

Hedtke, S.M., Patiny S., & Danforth B. 2013. The bee tree of life: a supermatrix approach to
apoid phylogeny and biogeography. BMC Evolutionary Biology 13: 138.
http://dx.doi.org/10.1186/1471-2148-13-138

Meredith, R.W., Janecka J., Gatesy J., Ryder O.A., Fisher C., Teeling E., Goodbla A., Eizirik
E., Simao T., Stadler T., Rabosky D., Honeycutt R., Flynn J., Ingram C., Steiner C., Williams
T., Robinson T., Herrick A., Westerman M., Ayoub N., Springer M., & Murphy W. 2011. Impacts of
the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. Science 334
(6055): 521-524.
http://dx.doi.org/10.1126/science.1211028

Yuchi Zheng, John J. Wiens, 2016, 'Combining phylogenomic and supermatrix approaches, and a
time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162
species', Molecular Phylogenetics and Evolution, vol. 94, pp. 537-547
http://dx.doi.org/10.1016/j.ympev.2015.10.009

Wright, April M., Kathleen M. Lyons, Matthew C. Brandley, David M. Hillis. 2015. Which came
first: The lizard or the egg? Robustness in phylogenetic reconstruction of ancestral states.
Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 324 (6): 504-516
http://dx.doi.org/10.1002/jez.b.22642
```

Potential for biodiversity loss

SPECIES | ACCEPTED

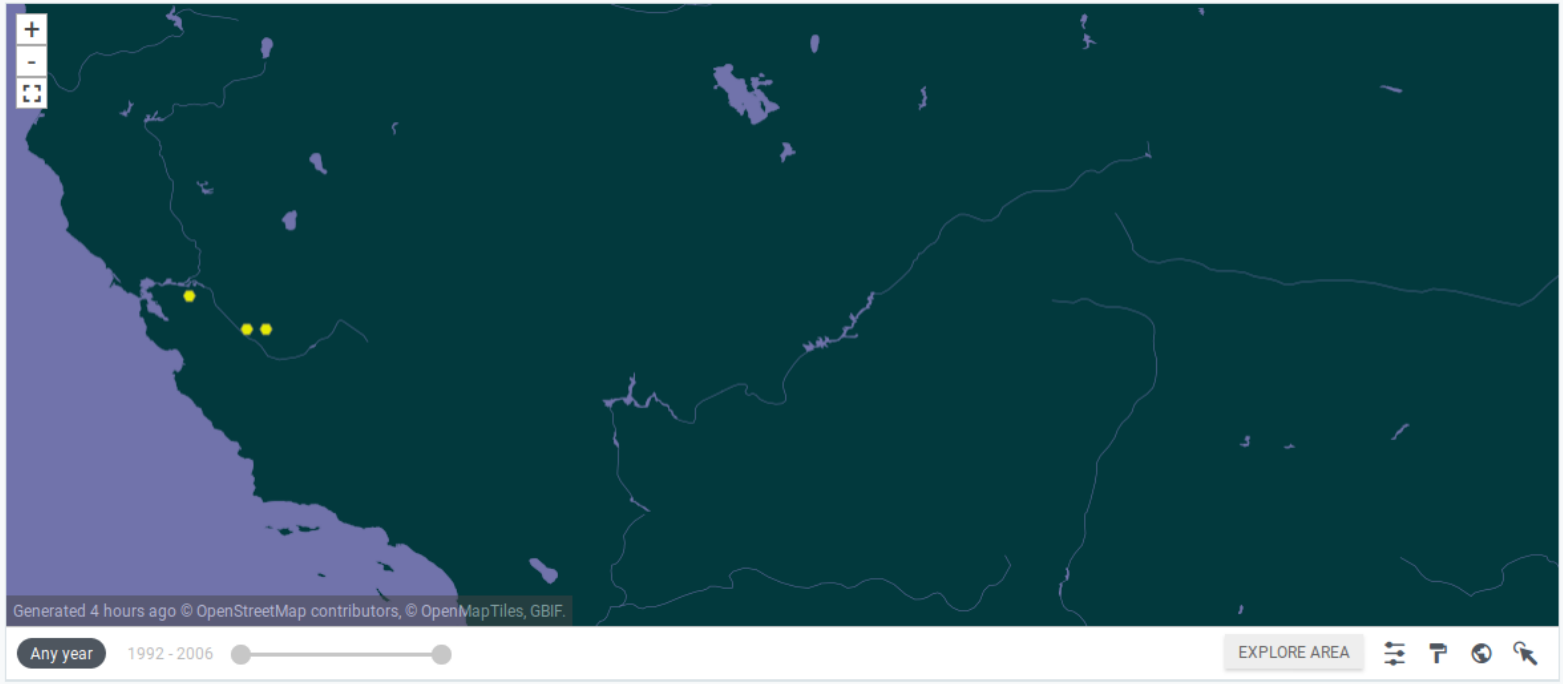
Branchinecta mesovallensis Belk & Fugate, 2000

Published in: Belk, D.; Fugate, M. (2000). Two New Branchinecta (Crustacea: Anostraca) from the Southwestern United States. *The Southwestern Naturalist*. 45(2): 111-117.
source: Catalogue of Life

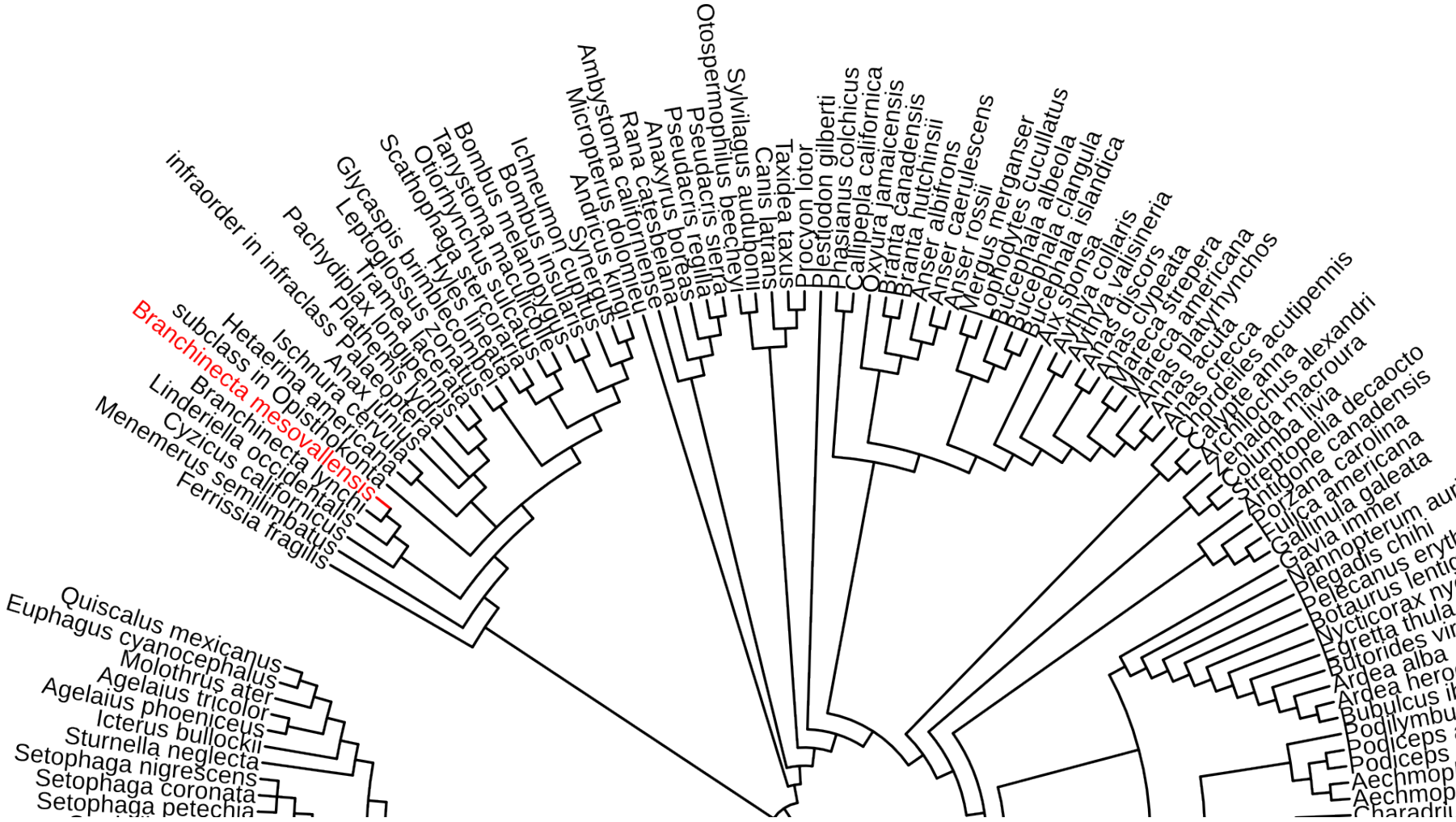
Mid-valley fairy shrimp In English

OVERVIEW METRICS REFERENCE TAXON 161 OCCURRENCES

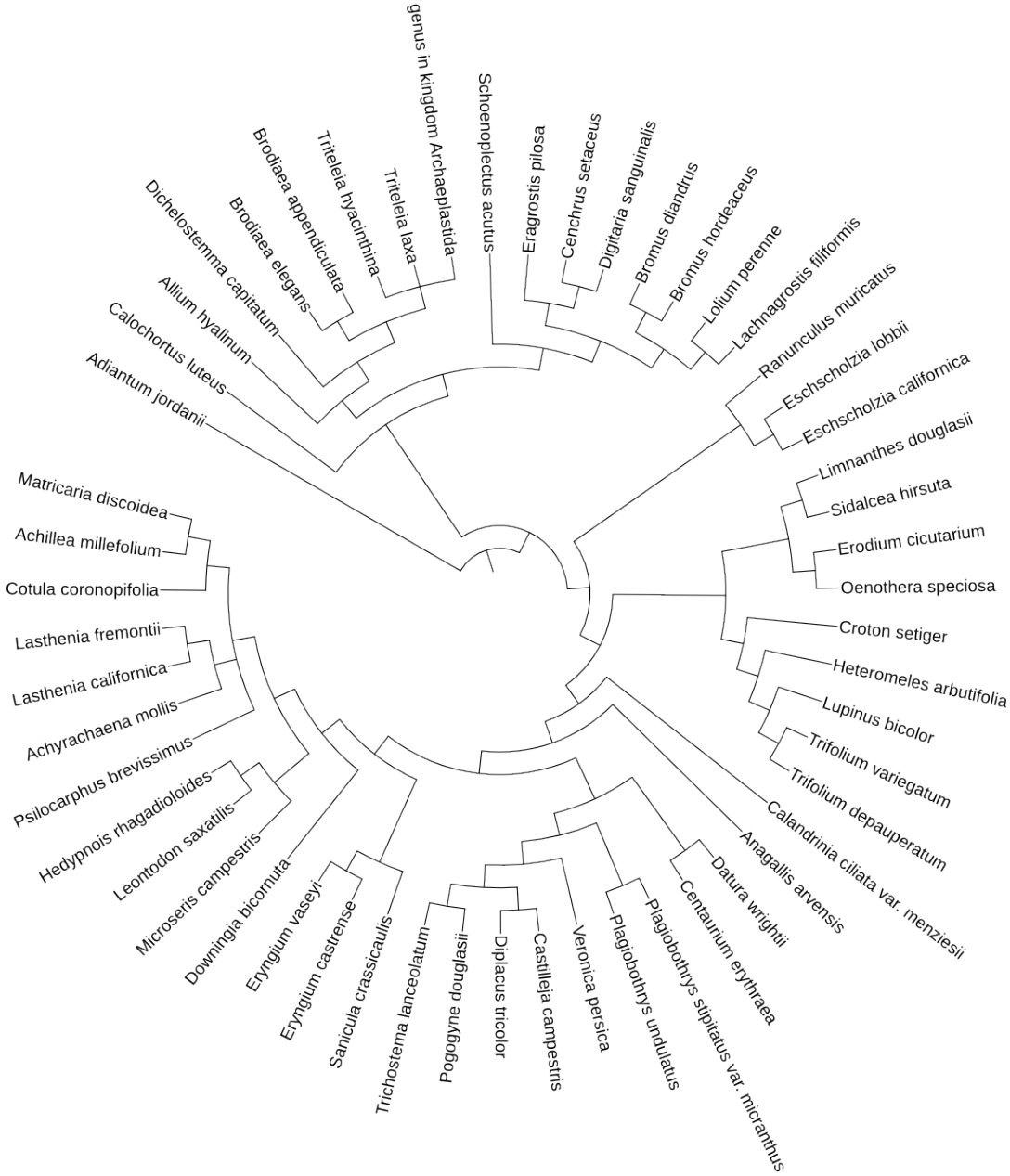
5 GEOREFERENCED RECORDS



IUCN STATUS


Quiscalus mexicanus
Euphagus cyanocephalus
Molothrus ater
Agelaius tricolor
Agelaius phoeniceus
Icterus bullockii
Sturnella neglecta
Setophaga nigrescens
Setophaga coronata
Setophaga petechia
Ferrissia fragilis
Menemerus semilimbatus
Cyzicus californicus
Linderiella occidentalis
Branchinecta lynchi
Branchinecta mesovallensis
subclass in Opisthokonta
Hetaerina americana
Ischnura cervula
Anax junius
Plathemis lydia
Pachydiplax longipennis
Tramea lacerata
Leptoglossus zonatus
Glycaspis brimblecombei
Scathophaga stercoraria
Hylemya lineata
Otiorhynchus sulcatus
Tanystoma maculicolle
Bombus melanopygus
Bombus insularis
Ichneumon cuprifer
infraorder in infraclass Palaeoptera
Synergus
Andricus kingi
Micropterus dolomieu
Ambystoma californiense
Rana catesbeiana
Anaxyrus boreas
Pseudacris regilla
Pseudacris sierra
Otospermophilus beecheyi
Sylvilagus audubonii
Canis latrans
Taxidea taxus
Procyon lotor
Plestiodon gilberti
Phasianus colchicus
Callipepla californica
Oxyura jamaicensis
Branta canadensis
Branta hutchinsii
Anser albifrons
Anser caerulescens
Anser rossii
Mergus merganser
Lophodytes cucullatus
Bucephala albeola
Bucephala clangula
Bucephala islandica
Aythya collaris
Aythya valisineria
Anas discors
Anas clypeata
Mareca strepera
Mareca americana
Anas platyrhynchos
Anas acuta
Anas crecca
Chordeiles acutipennis
Calypte anna
Archilochus alexandri
Zenaida macroura
Columba livia
Streptopelia decaocto
Antigone canadensis
Porzana carolina
Fulica americana
Gallinula galeata
Gavia immer
Nannopterum aur
Plegadis chihi
Pelecanus eryth
Botaurus lentig
Nycticorax ny
Egretta thula
Butorides vir
Ardea alba
Ardea hero
Bubulcus il
Podilymbu
Podiceps
Podiceps
Aechmop
Aechmop
Charadriu

Plants of the vernal pools (100 records of 61 species)



67 studies inform the relationships in the plant tree

Miao Sun, Douglas E. Soltis, Pamela S. Soltis, Xinyu Zhu, J. Gordon Burleigh, Zhiduan Chen, 2015, 'Deep phylogenetic incongruence in the angiosperm clade Rosidae', Molecular Phylogenetics and Evolution, vol. 83, pp. 156-166
http://dx.doi.org/10.1016/j.ympev.2014.11.003

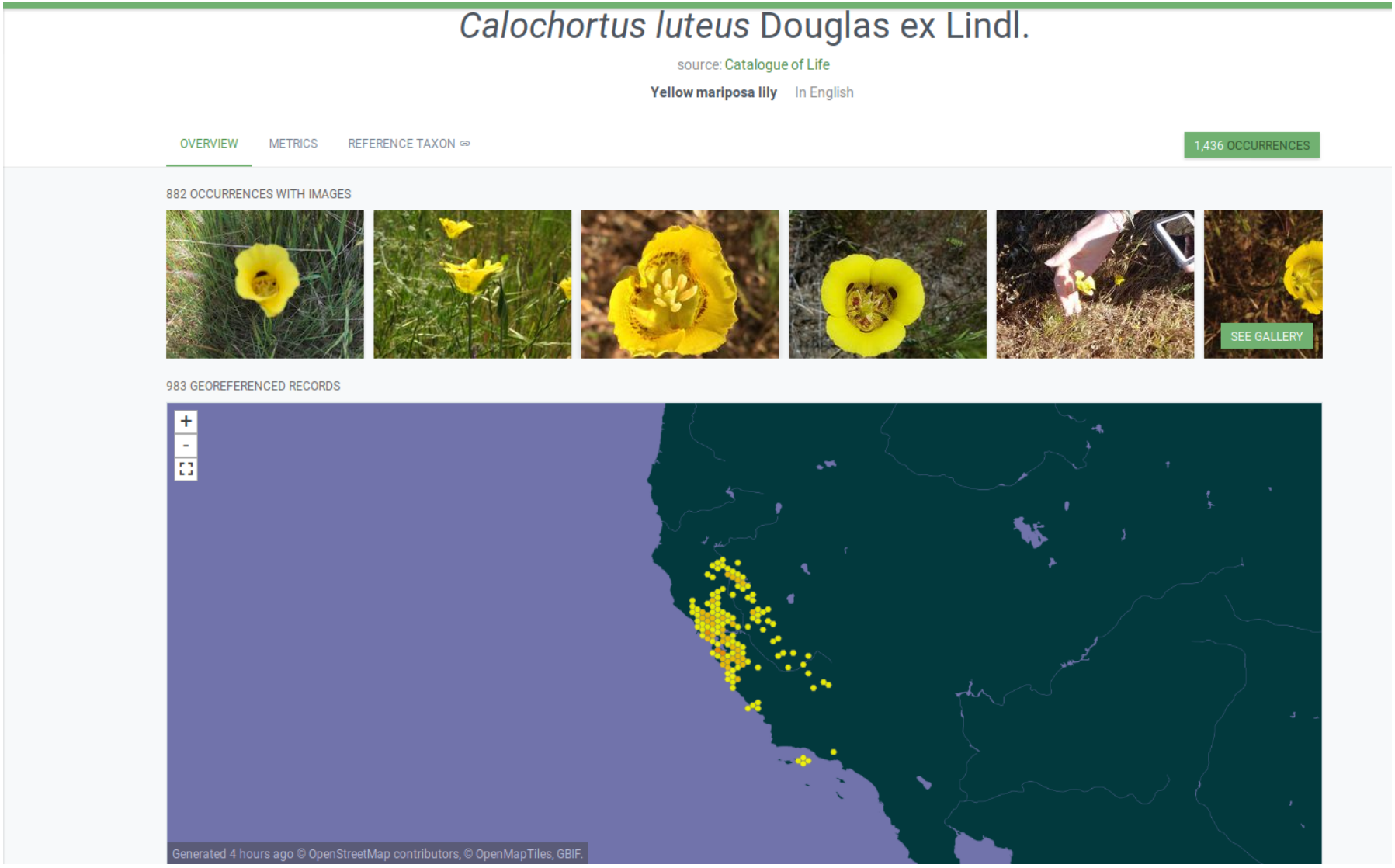
Magallon, Susana, Sandra Gomez-Acevedo, Luna L. Sanchez-Reyes, Tania Hernandez-Hernandez. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytologist 207 (2): 437-453
http://dx.doi.org/10.1111/nph.13264

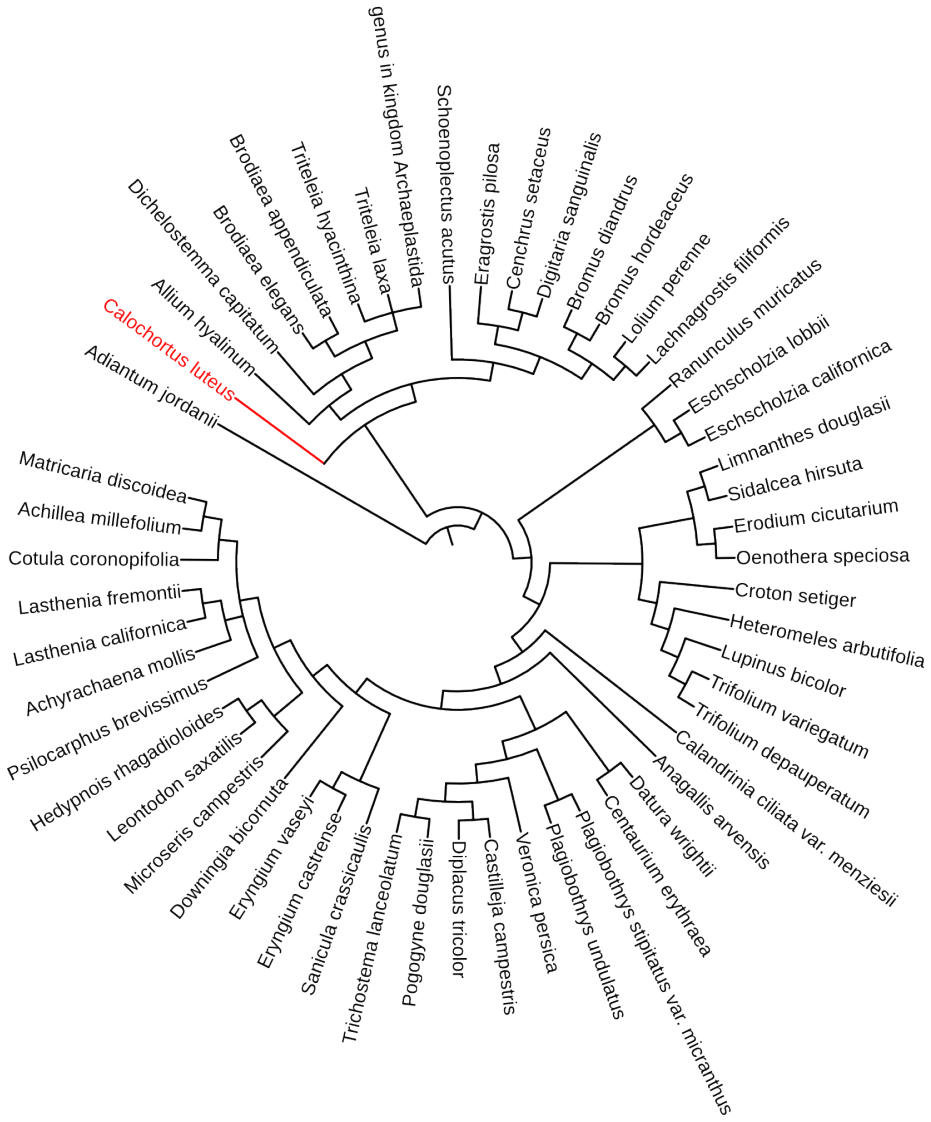
Sun, Yanxia, Michael J. Moore, Shoujun Zhang, Pamela S. Soltis, Douglas E. Soltis, Tingting Zhao, Aiping Meng, Xiaodong Li, Jianqiang Li, Hengchang Wang. 2016. Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. Molecular Phylogenetics and Evolution 96: 93-101
http://dx.doi.org/10.1016/j.ympev.2015.12.006

Zhang S., Soltis D., Li D., Yang Y., & Yi T. 2011. Multi-gene analysis provides a well-supported phylogeny of Rosales. Molecular Phylogenetics and Evolution, 60(1): 21-8.
http://dx.doi.org/10.1016/j.ympev.2011.04.008

Wang, Hengchang, Michael J. Moore, Pamela S. Soltis, Charles D. Bell, Samuel F. Brockington, Roolse Alexandre, Charles C. Davis, Maribeth Latvis, Steven R. Manchester, and Douglas E. Soltis. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. Proceedings of the National Academy of Sciences 106, no. 10 (March 10): 3853 -3858. doi:10.1073/pnas.0813376106.
http://dx.doi.org/10.1073/pnas.0813376106

Potential for biodiversity loss




genus in kingdom Archaeplastida
Schoenoplectus acutus
Eragrostis pilosa
Cenchrus setaceus
Digitaria sanguinalis
Bromus diandrus
Bromus hordeaceus
Lolium perenne
Lachnagrostis filiformis
Ranunculus muricatus
Eschscholzia lobbii
Eschscholzia californica
Limnanthes douglasii
Sidalcea hirsuta
Erodium cicutarium
Oenothera speciosa
Croton setiger
Heteromeles arbutifolia
Lupinus bicolor
Trifolium variegatum
Trifolium depauperatum
Calandrinia ciliata var. menziesii
Anagallis arvensis
Datura wrightii
Centaurium erythraea
Plagiobothrys stipitatus var. micranthus
Plagiobothrys undulatus
Veronica persica
Castilleja campestris
Diplacus tricolor
Pogogyne douglasii
Trichostema lanceolatum
Sanicula crassicaulis
Eryngium castrense
Eryngium vaseyi
Downingia bicornuta
Microseris campestris
Leontodon saxatilis
Hedypnois rhagadioloides
Psilocarphus brevissimus
Achyrachaena mollis
Lasthenia californica
Lasthenia fremontii
Cotula coronopifolia
Achillea millefolium
Matricaria discoidea
Adiantum jordanii
Calochortus luteus
Allium hyalinum
Dichelostemma capitatum
Brodiaea elegans
Brodiaea appendiculata
Triteleia hyacinthina
Triteleia laxa

Conclusions

Phylogeny not taxonomy, can and should be used to capture evolutionary relationships among taxa.

Conclusions

Phylogeny not taxonomy, can and should be used to capture evolutionary relationships among taxa.

Open Tree cross-links phylogenetic and taxonomic information across biodiversity informatics resources

Conclusions

Phylogeny not taxonomy, can and should be used to capture evolutionary relationships among taxa.

Open Tree cross-links phylogenetic and taxonomic information across biodiversity informatics resources

A variety of tools and approaches provides wide access to Open Tree resources

Coming soon!

- Automated updating with new genetic data
- Branch lengths / Node ages
- Private data stores
- Custom synthesis
- Links to wikidata

Holder, McTavish  ABI

Cranston

Contribute your knowledge!

`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

Lab group:
Martha Kandziora
Luna Luisa Sanchez Reyes
Lesly Lopez Fang
Jasper Toscani-Field
Joseline Velasquez

Citations

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.

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., 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.

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

Ratnasingham, S. and Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, 7(3):355.

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*, (5).

Tree of Sex Consortium (2014). Tree of Sex: A database of sexual systems. *Scientific Data*, 1:140015.