

Advancing Scientific Research with Digitized Data: *using and reusing biodiversity data*

Pamela S. Soltis

Florida Museum of Natural History

University of Florida



iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. All images used with permission or are free from copyright.



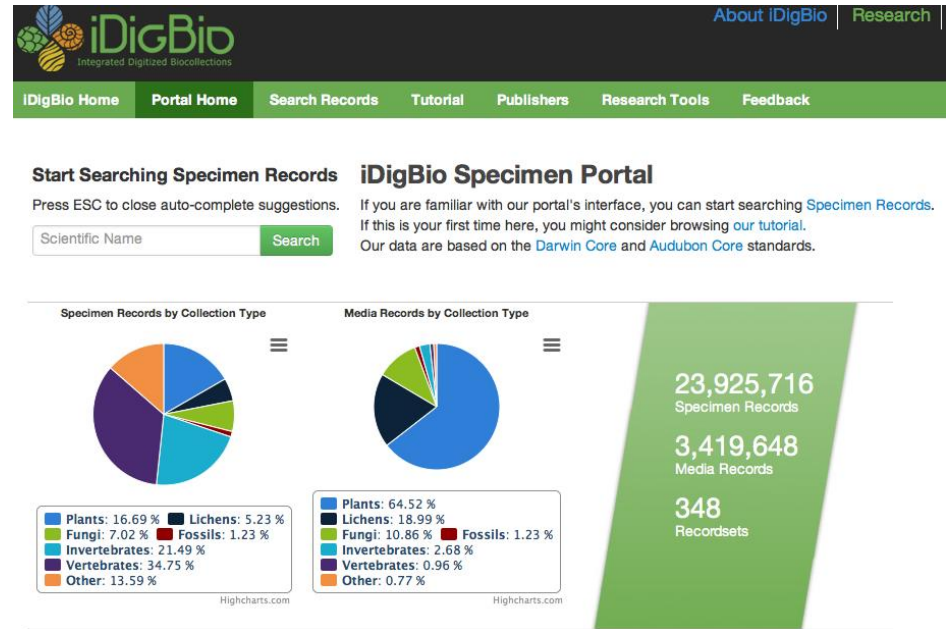
Collections: The Library of Life

>1600 natural history
collections
in the US alone
1-2 billion specimens
in the US
3-4 billion specimens
worldwide



Research @ iDigBio

- Data portal

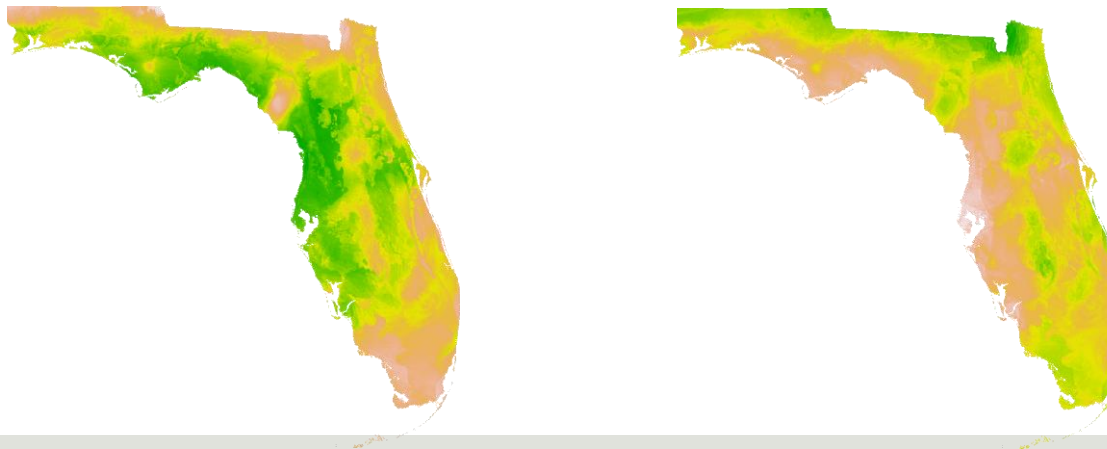


The screenshot shows the iDigBio Specimen Portal interface. At the top, there is a navigation bar with links for 'About iDigBio' and 'Research'. Below this is a green header with 'iDigBio Integrated Digitized Biocollections' and a menu with 'iDigBio Home', 'Portal Home', 'Search Records', 'Tutorial', 'Publishers', 'Research Tools', and 'Feedback'. The main content area is titled 'Start Searching Specimen Records' and 'iDigBio Specimen Portal'. It includes a search box for 'Scientific Name' and a 'Search' button. Below the search box are two pie charts: 'Specimen Records by Collection Type' and 'Media Records by Collection Type'. To the right of these charts is a green box displaying record counts: 23,925,716 Specimen Records, 3,419,648 Media Records, and 348 Recordsets.

Collection Type	Percentage
Plants	16.69 %
Fungi	7.02 %
Invertebrates	21.49 %
Vertebrates	34.75 %
Other	13.59 %
Lichens	5.23 %
Fossils	1.23 %

Collection Type	Percentage
Plants	64.52 %
Fungi	10.86 %
Invertebrates	2.68 %
Vertebrates	0.96 %
Other	0.77 %
Lichens	18.99 %
Fossils	1.23 %

- Computational workflows – research use



Linking Collections to...

- Phylogenies
- Ecology
- Genomics
- Paleontology
- Living Collections
- Other Repositories

neon
National Ecological Observatory Network, Inc.
Paleobiology Database

Quick search

UTEX

Raphitoma, Dufong, Eulytoceras, Diptera, Mammals, Africa, Mireosaurus, Diptera, Robustirynchia, Vallinacera, Aganicia, Leptacanthus, Mammals, Africa

Welcome to Morphbank
User: Guest [\[click to login\]](#)

TreeBASE
A Database of Phylogenetic Knowledge

About NCBI
National Center for Biotechnology Information

About NCBI	NCBI at a Glance	A Science Primer	Databases and Tools
Human Genome Resources	Model Organisms Guide	Outreach and Education	News

NCBI at a Glance
A Science Primer
Databases and Tools
Human Genome Resources
Model Organisms Guide
Outreach and Education
News



Synthetic Research and Complex Data Analysis

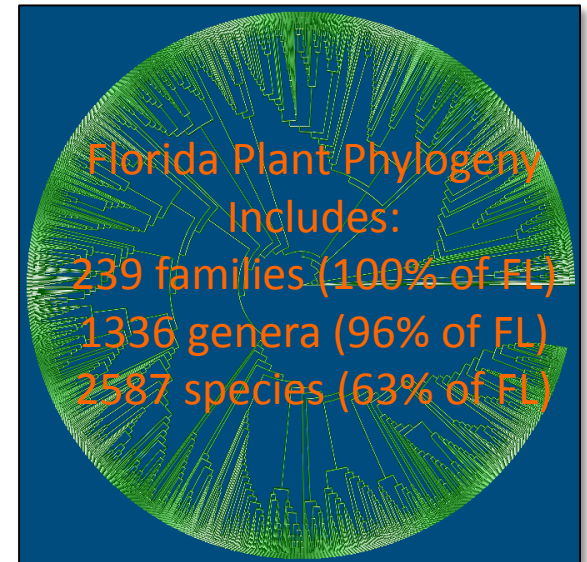
- Large data sets
- Heterogeneous data
- Multiple software packages
- Multi-step analytical workflows
- Training

Synthetic Research: Connections

- Connecting with ecological data
 - Ecological niche modeling
 - Climate change
 - Polyploidy
- Connecting with phylogeny
 - PhyloJive
 - Open Tree of Life
- Connecting with software and tools

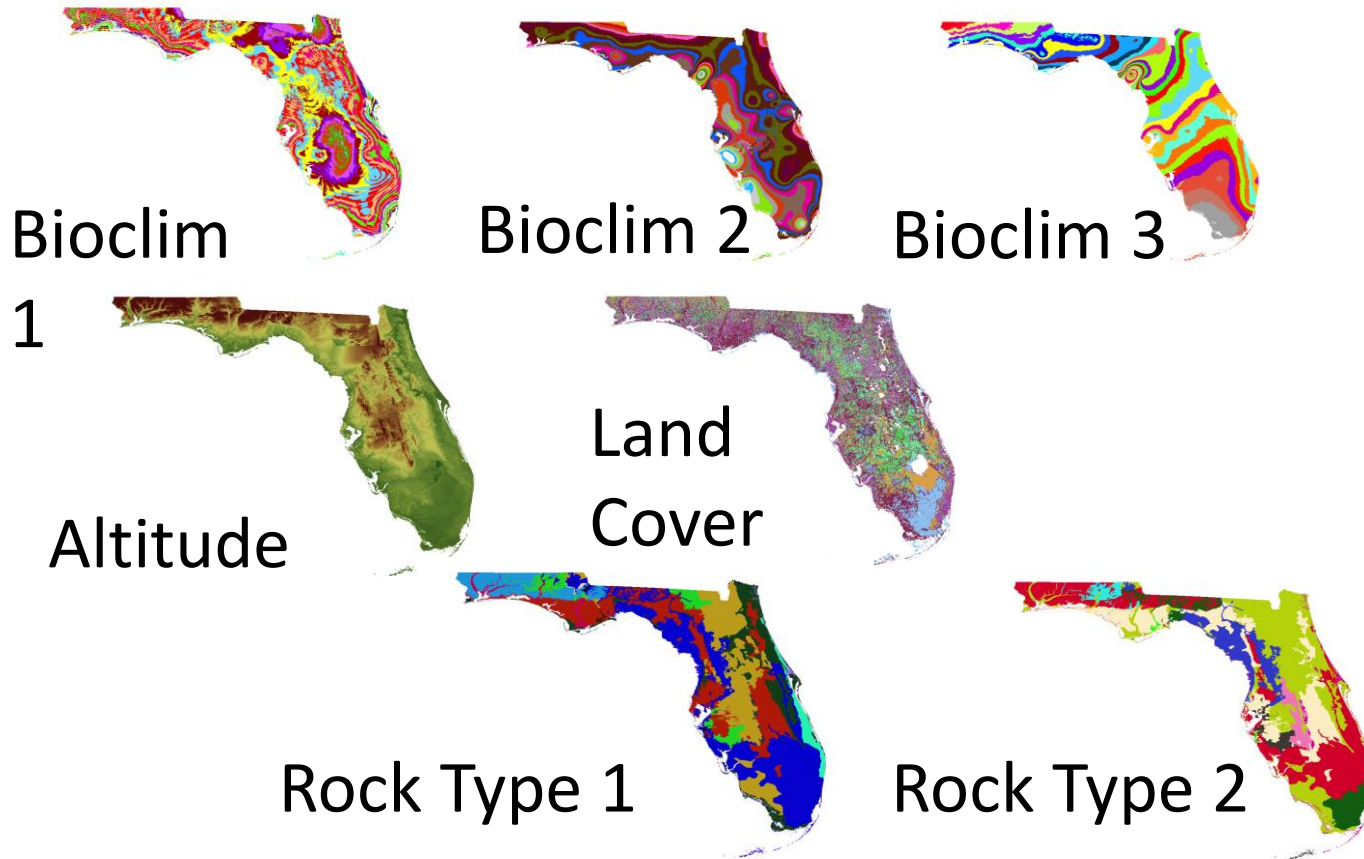
Florida Plant Diversity in a Changing Climate

Integrating herbarium specimen data, ENM, climate change models, and phylogeny



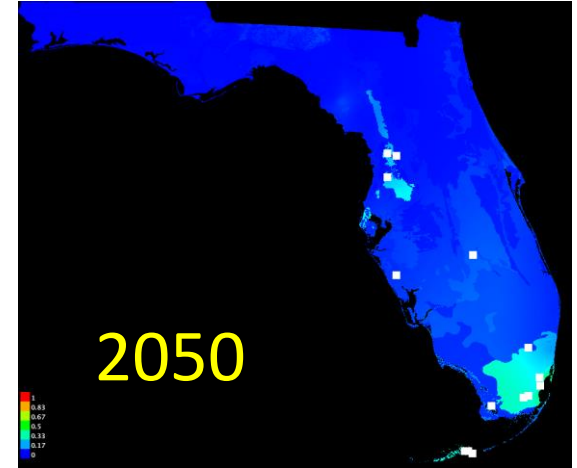
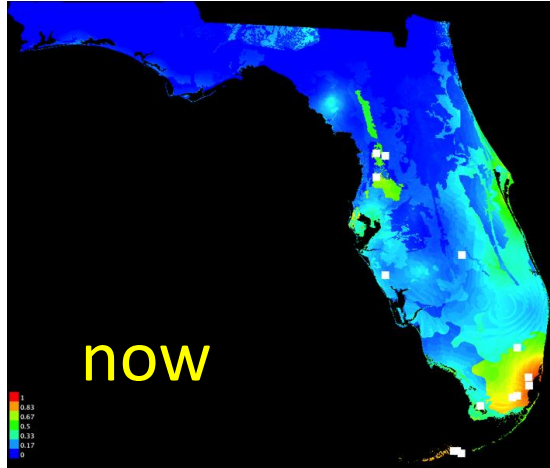
Charlotte Germain-Aubrey

Ecological Niche Modeling: Environmental Layers

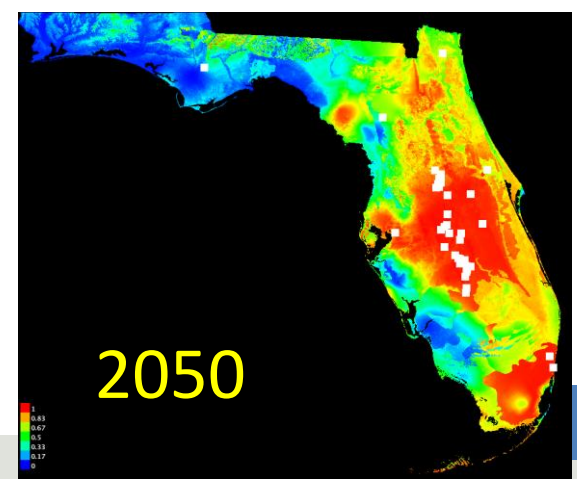
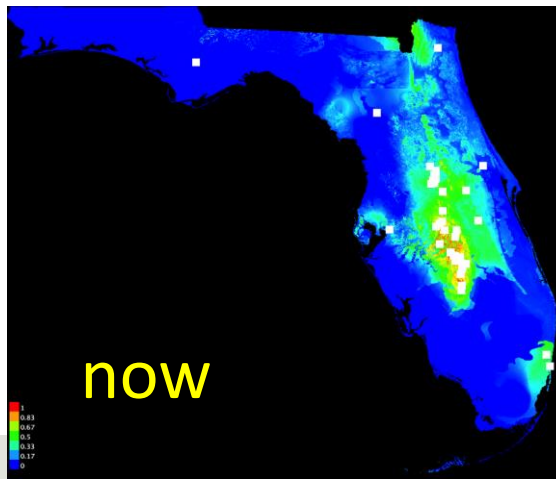


Responses to Climate Change: Winners & Losers

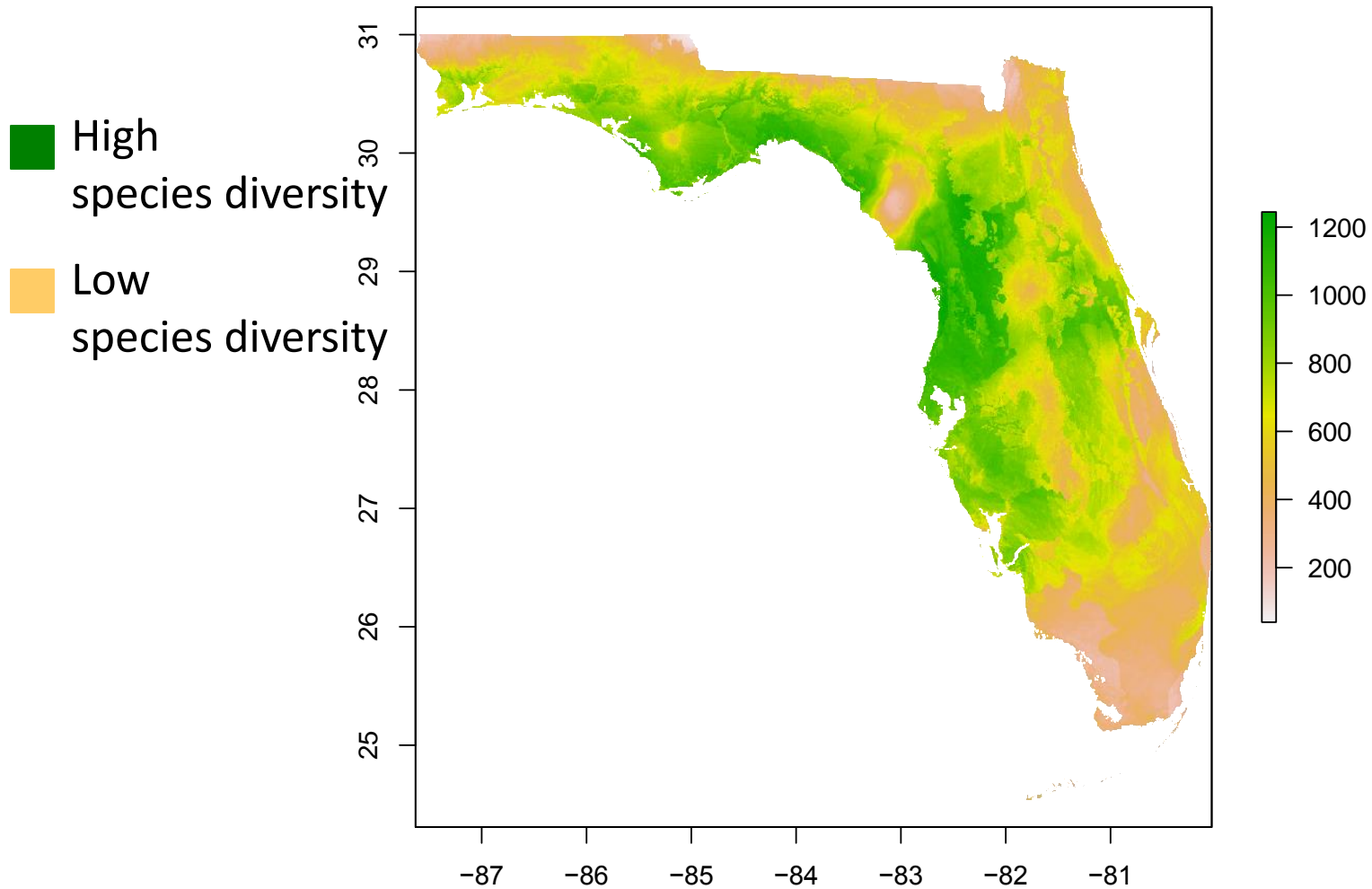
Abildgaardia ovata (flatspike sedge)



Prunus geniculata (scrub plum)

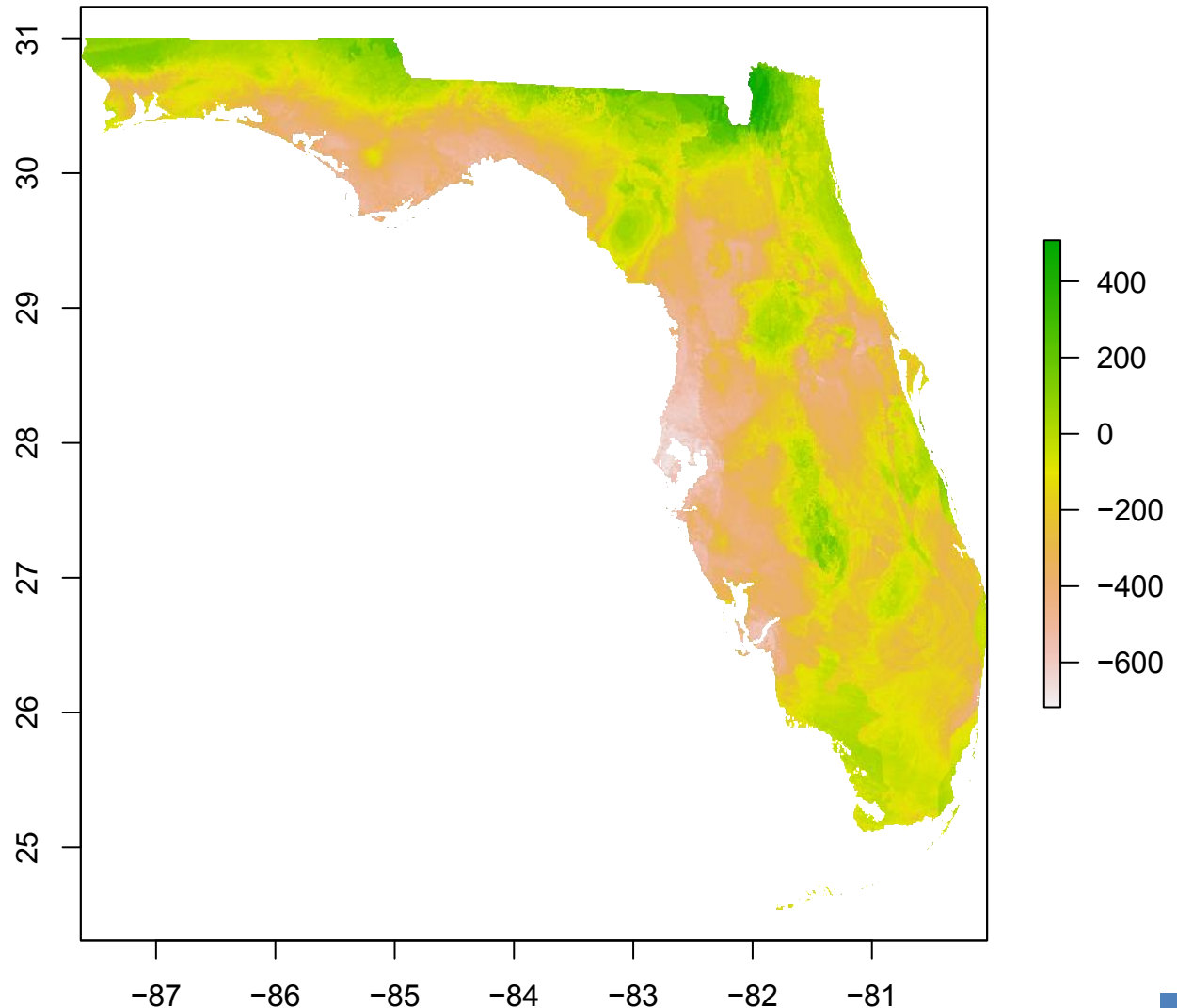


Florida plant diversity heat map: now



Between now and 2050...

- Panhandle species moving NORTH!
- Peninsula species moving SOUTH!



spp 2050 - #spp now

R package coming soon...

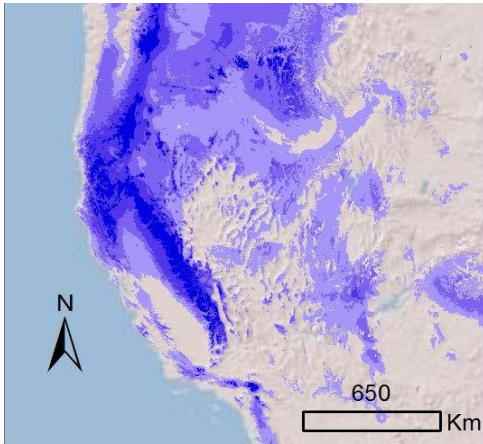
- Data cleaning
 - Lat/long – round decimal points
 - Remove duplicates
 - Bring points back to the coast
 - Removing outliers using a shape file of the desired area
- Extracting climate data from year of collection
 - Prism format
 - Bioclim format
 - If no year given, then default average
- Output formatted for Maxent and Biomod2



C. Germain-Aubrey



Distributions of Polyploids and their Diploid Progenitors



- Do polyploids have broader ‘ecological amplitude’ and/or ranges than diploid parents?
- ENM for polyploids and parents
- Predicted distributions under models of climate change
 - Will polyploids and diploids respond in similar ways?

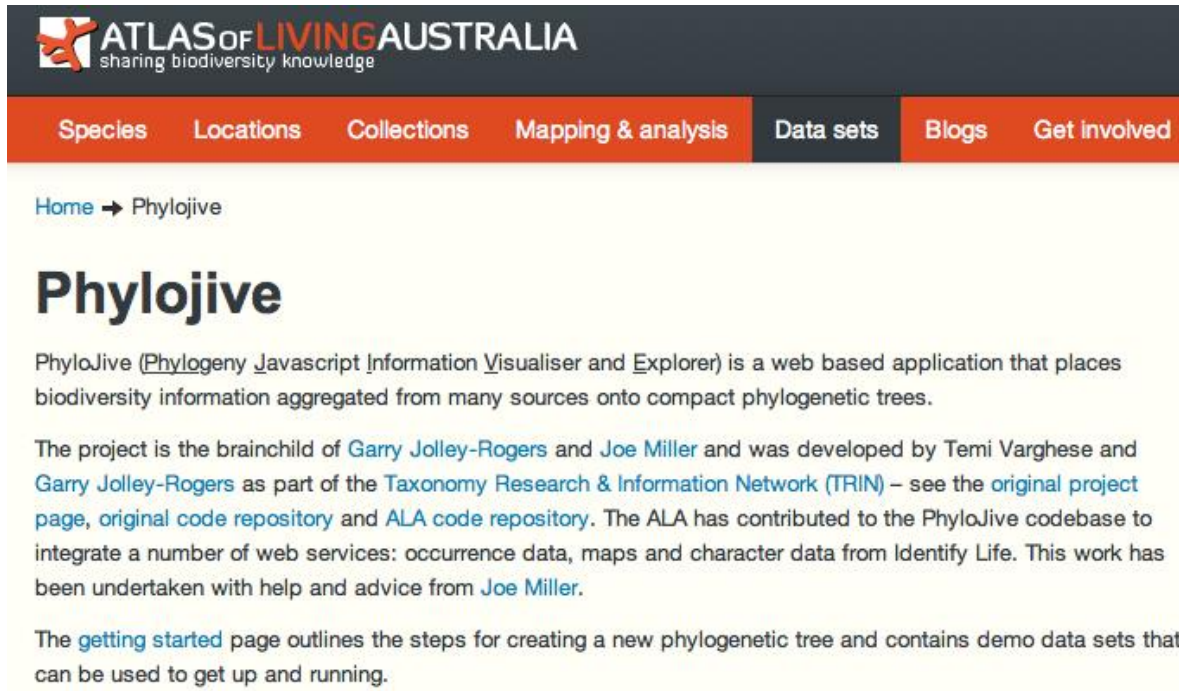
Blaine Marchant

Linking to Phylogeny

The screenshot shows the OneZoom website interface. At the top is a navigation bar with the following links: OneZoom, Home, Embed, Software, Impacts, News, About, Gallery, and Future. The main content area features a stylized phylogenetic tree of plants. The tree is rooted in a thick green stem labeled "Seed Plants". From this stem, three main branches emerge: a brown branch on the left labeled "Gymnosperms", a central green branch labeled "Flowering Plants", and a brown branch on the right labeled "Mesangiosperms". The "Flowering Plants" branch further divides into two sub-branches: a green one labeled "Water-lilies, Water-shields and more" and a brown one labeled "Star Anise, Lemon Wood and more". The "Mesangiosperms" branch also divides into two sub-branches: a green one labeled "Water-lilies, Water-shields and more" and a brown one labeled "Star Anise, Lemon Wood and more". At the bottom of the page, a dark grey banner contains the text "Click to see how OneZoom works".

PhyloJIVE

Links biodiversity data to trees
Joe Miller & Garry Jolley-Rogers
phylojive.ala.org.au/



ATLAS of LIVING AUSTRALIA
sharing biodiversity knowledge

Species Locations Collections Mapping & analysis **Data sets** Blogs Get Involved

[Home](#) → [Phylojive](#)

Phylojive

PhyloJive ([Phy](#)logeny [J](#)avascript [I](#)nformation [V](#)isualiser and [E](#)xplorer) is a web based application that places biodiversity information aggregated from many sources onto compact phylogenetic trees.

The project is the brainchild of [Garry Jolley-Rogers](#) and [Joe Miller](#) and was developed by Temi Varghese and [Garry Jolley-Rogers](#) as part of the [Taxonomy Research & Information Network \(TRIN\)](#) – see the [original project page](#), [original code repository](#) and [ALA code repository](#). The ALA has contributed to the PhyloJive codebase to integrate a number of web services: occurrence data, maps and character data from Identify Life. This work has been undertaken with help and advice from [Joe Miller](#).

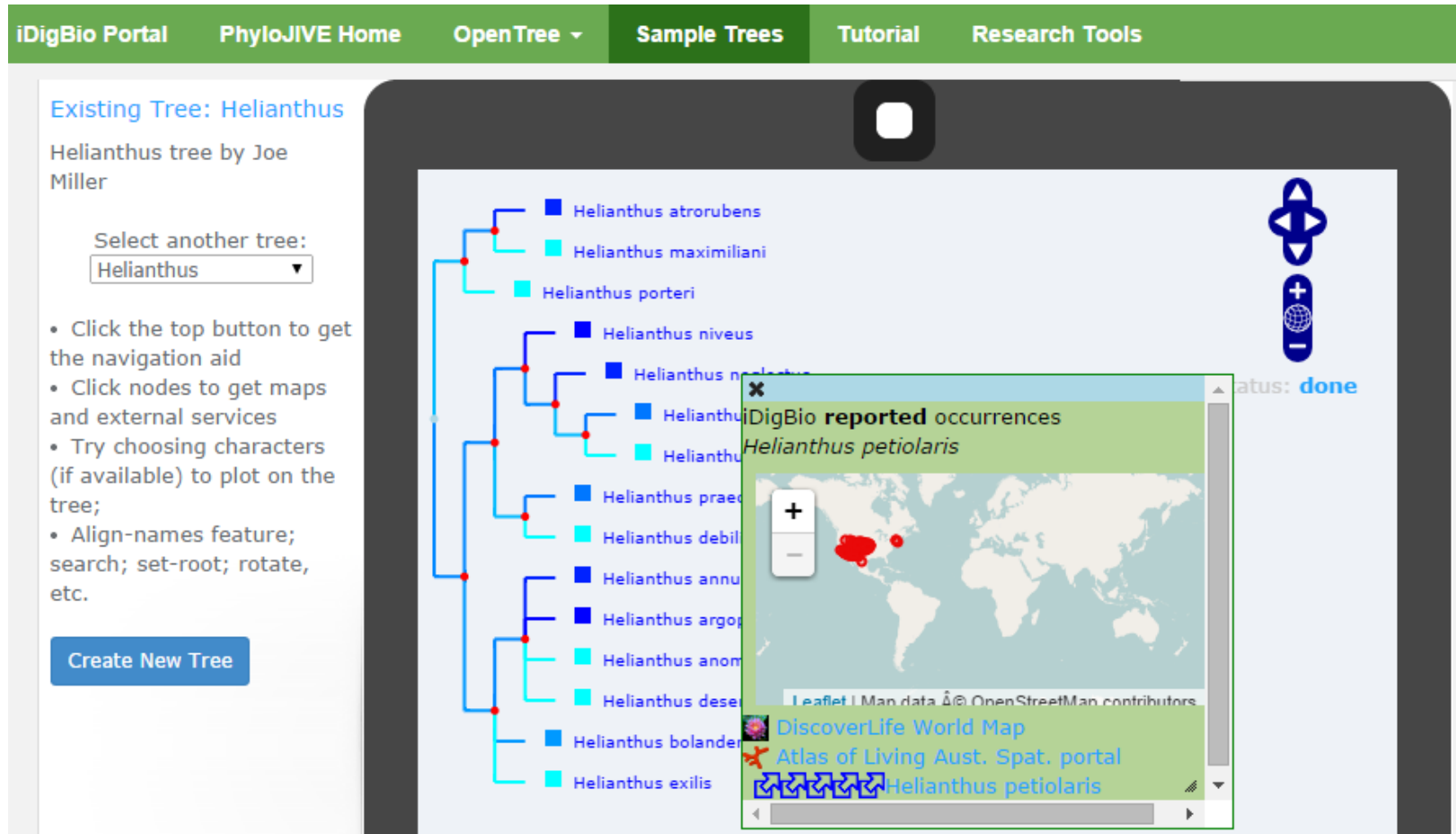
The [getting started](#) page outlines the steps for creating a new phylogenetic tree and contains demo data sets that can be used to get up and running.



A. buxifolia

Source: Australian Plant
Image Index Image by: Macd

PhyloJIVE instance in iDigBio



The screenshot displays the iDigBio Portal interface for the PhyloJIVE instance. The navigation bar includes links for iDigBio Portal, PhyloJIVE Home, OpenTree, Sample Trees, Tutorial, and Research Tools. The main content area shows an existing tree titled "Existing Tree: Helianthus" by Joe Miller. A dropdown menu allows selecting another tree from the "Helianthus" genus. A list of species is shown on the right, including Helianthus atrorubens, Helianthus maximiliani, Helianthus porteri, Helianthus niveus, Helianthus mollis, Helianthus scaberrimus, Helianthus debilis, Helianthus annuus, Helianthus argenteus, Helianthus anomus, Helianthus desertorum, Helianthus bolanderi, and Helianthus exilis. A map window is open, showing "iDigBio reported occurrences" for Helianthus petiolaris, with red dots indicating locations in the western United States. The map includes a search bar, zoom controls, and a status bar showing "status: done".

- Developed by Garry Jolley-Rogers, Joe Miller, and Temi Varghese
- Integrates biodiversity data with phylogeny
- <http://phylojive.acis.ufl.edu/>

A. Matsunaga

PhyloJIVE instance in iDigBio

iDigBio Portal PhyloJIVE Home OpenTree ▾ Sample Trees Tutorial Research Tools

Existing Tree

Helianthus tre
Miller

Select and
Helianthus

- Click the top
- Click nodes
- Try choosing
- Align-names
- search; set-ro
- etc.

Create New T

OPEN
Tree of Life
A community-assembled ToL

is: done

- Developed by Garry Jolley-Rogers, Joe Miller, and Temi Varghese
- Integrates biodiversity data with phylogeny
- <http://phylojive.acis.ufl.edu/>

A. Matsunaga

PhyloJIVE instance in iDigBio

The screenshot shows the iDigBio Portal interface for PhyloJIVE. The navigation bar includes links for iDigBio Portal, PhyloJIVE Home, OpenTree, Sample Trees, Tutorial, and Research Tools. On the left, there is a sidebar with the heading 'Existing Tree' and a list of trees, including 'Helianthus tre Miller'. Below this is a search box containing 'Helianthus' and a list of instructions: 'Click the top of the navigation', 'Click nodes and external s', 'Try choosing (if available) t tree;', and 'Align-names search; set-ro etc.'. A 'Create New T' button is visible at the bottom of the sidebar. The main content area features the 'OPEN Tree of Life' logo, which includes a stylized leaf and the text 'A community-assembled ToL'. On the right side of the main area, there is a vertical toolbar with navigation icons (up, down, left, right, zoom in, zoom out) and a 'done' status indicator.

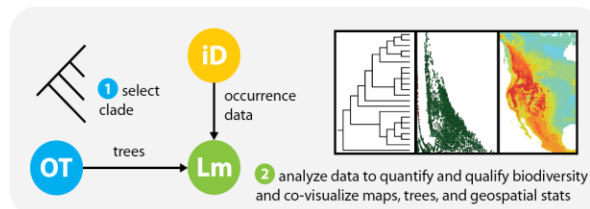
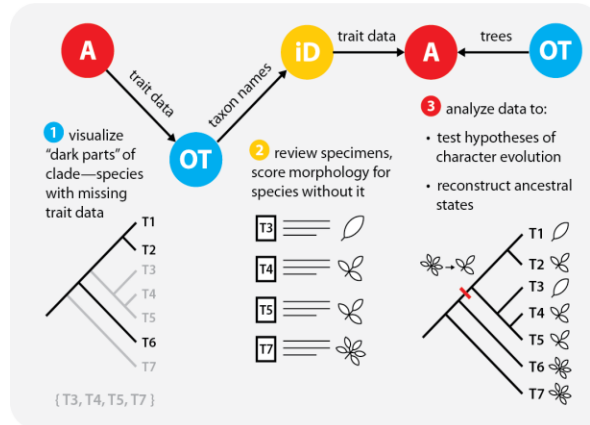
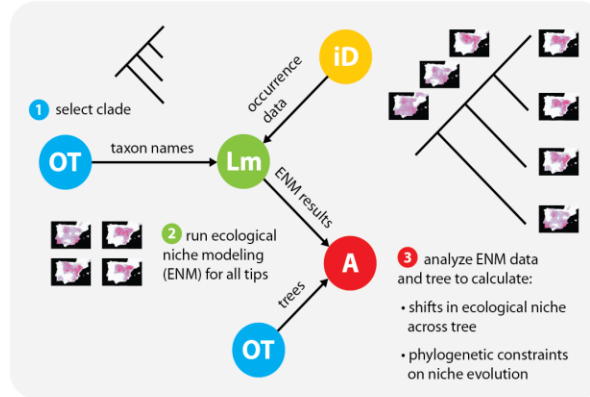
1.8 million described species

- Developed by Garry Jolley-Rogers, Joe Miller, and Temi Varghese
- Integrates biodiversity data with phylogeny
- <http://phylojive.acis.ufl.edu/>

A. Matsunaga

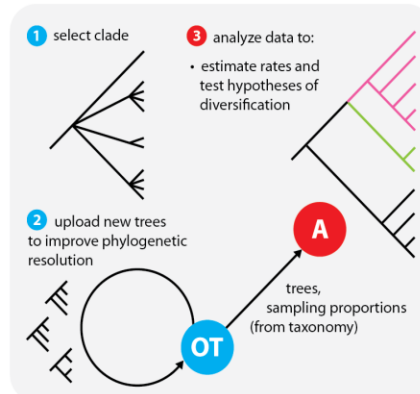
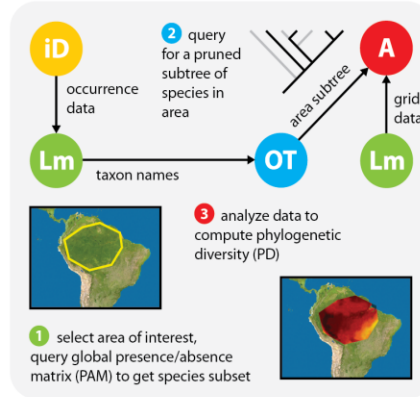
Connecting Trees, Specimens, Tools

EXAMPLE WORKFLOWS:



RESOURCES:

- Lm** Lifemapper
 - ecological niche modeling
 - biodiversity and range analysis
 - visualization
- A** Arbor
 - evolutionary models
 - comparative methods
 - visualization
- OT** Open Tree of Life
 - phylogenies
 - taxonomy / names
 - visualization
- iD** iDigBio
 - trait data
 - specimen data / images
 - fossil data / images

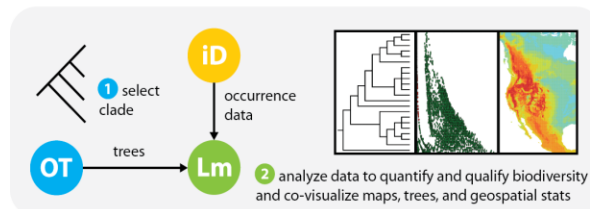
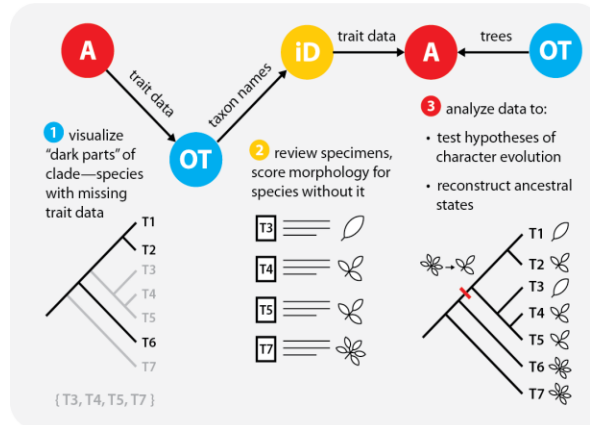
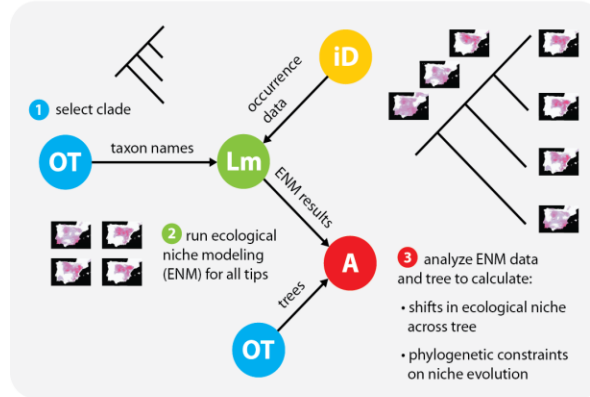


Connecting Trees, Specimens, Tools



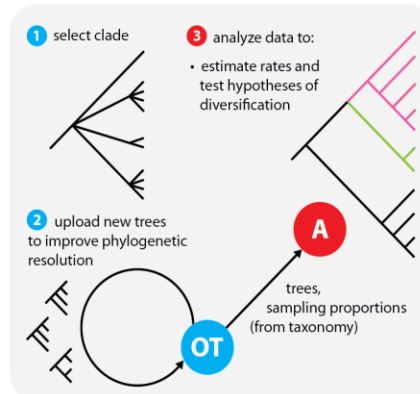
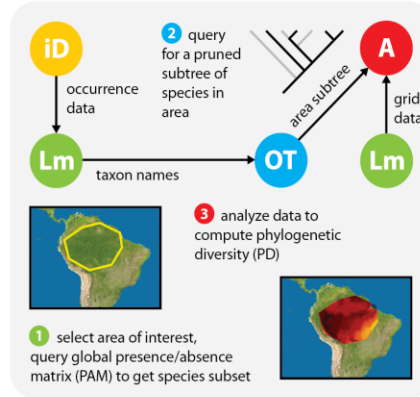
Connecting Trees, Specimens, Tools

EXAMPLE WORKFLOWS:



RESOURCES:

- Lm** Lifemapper
 - ecological niche modeling
 - biodiversity and range analysis
 - visualization
- A** Arbor
 - evolutionary models
 - comparative methods
 - visualization
- OT** Open Tree of Life
 - phylogenies
 - taxonomy / names
 - visualization
- iD** iDigBio
 - trait data
 - specimen data / images
 - fossil data / images



Connecting Trees, Specimens, Tools

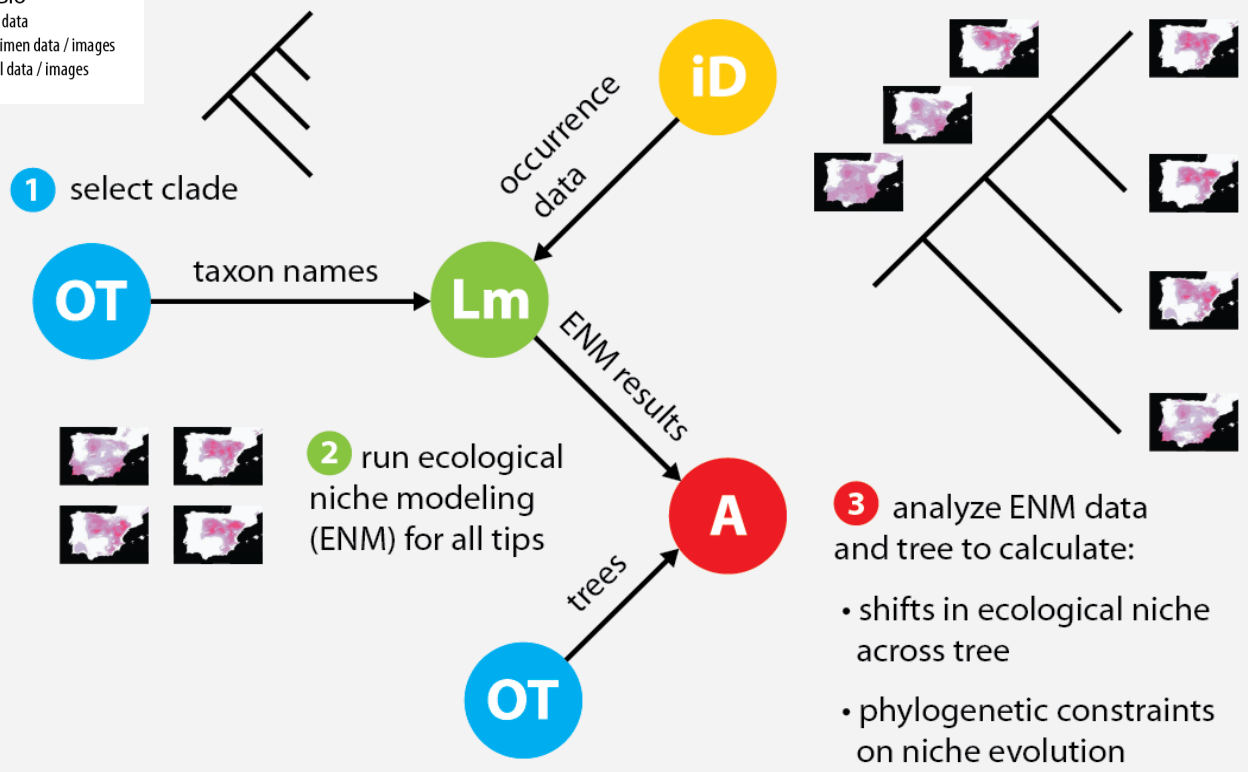
RESOURCES:

Lm Lifemapper
 • ecological niche modeling
 • biodiversity and range analysis
 • visualization

A Arbor
 • evolutionary models
 • comparative methods
 • visualization

OT Open Tree of Life
 • phylogenies
 • taxonomy / names
 • visualization

iD iDigBio
 • trait data
 • specimen data / images
 • fossil data / images



Connecting Trees, Specimens, Tools



ABI Innovation: Connecting resources to enable large-scale biodiversity analyses

D. Soltis, P. Soltis, J. Fortes, A. Matsunaga,
J. Beach, J. Soberon, S. Smith

RESOURCES:



Lm Lifemapper

- ecological niche modeling
- biodiversity and range analysis
- visualization



A Arbor

- evolutionary models
- comparative methods
- visualization



OT Open Tree of Life

- phylogenies
- taxonomy / names
- visualization

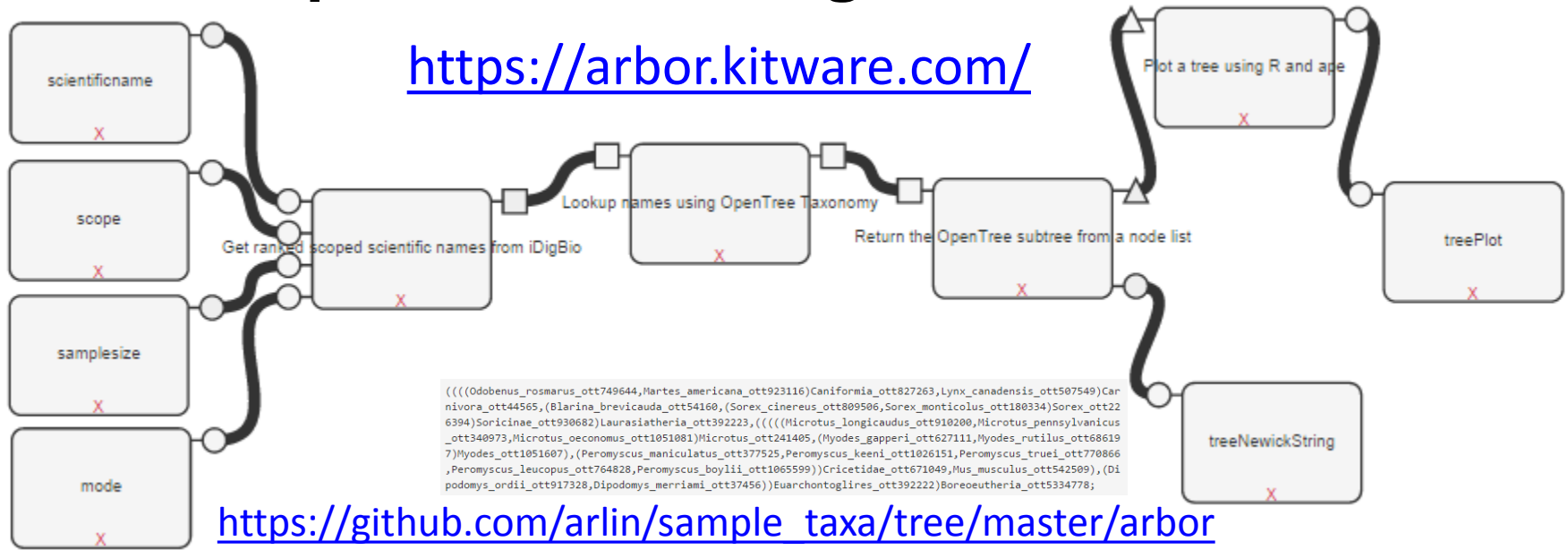


iD iDigBio

- trait data
- specimen data / images
- fossil data / images

Arbor, OpenTree, and iDigBio

<https://arbor.kitware.com/>



https://github.com/arlin/sample_taxa/tree/master/arbor

Workflow to get an induced tree from a configurable iDigBio query

scientificname:

scope:

samplesize:

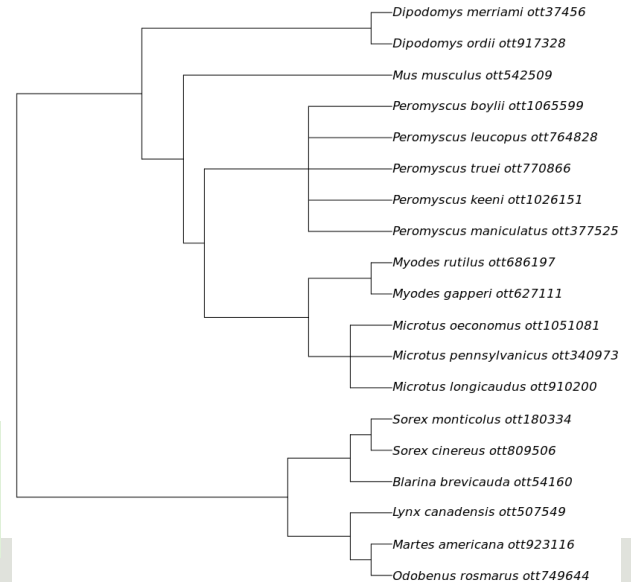
mode:

```

1 {
2   "query": {
3     "query_string": {
4       "default_field": "order",
5       "query": "rodentia"
6     }
7   },
8   "aggregations": {
9     "my_agg": {
10      "terms": {
11        "field": "scientificname",
12        "size": 100
13      }
14    }
15  }
16 }
  
```

Success! Produced the following outputs:

- Workflow to get an induced tree from a configurable iDigBio query treeNewickString [string]
- Workflow to get an induced tree from a configurable iDigBio query treePlot [image]



Specimen Records AND Images



GEOLocate  29.65, -82.32

number,dwc:preparations,dwc:identificationVerificationStatus,idigbio:subfamily,idigbio:preparationCount,fcc:pickedBy,dwc:eventRemarks,dwc:VerbatimEventDate,dwc:associatedReferences,idigbio:endangeredStatus,dwc:locationAccordingTo,dwc:georeferenceSources,dwc:associatedSequences,dwc:formation,dwc:higherClassification,dwc:catalogNumber,dwc:verbatimSRS,dwc:higherGeography,dwc:individualCount,dwc:decimalLongitude,dwc:datasetName,dwc:month,dwc:georeferencedBy,dwc:eventTime,dwc:identificationQualifier,idigbio:

Systematics and Taxonomy



Linnea (twinflower)

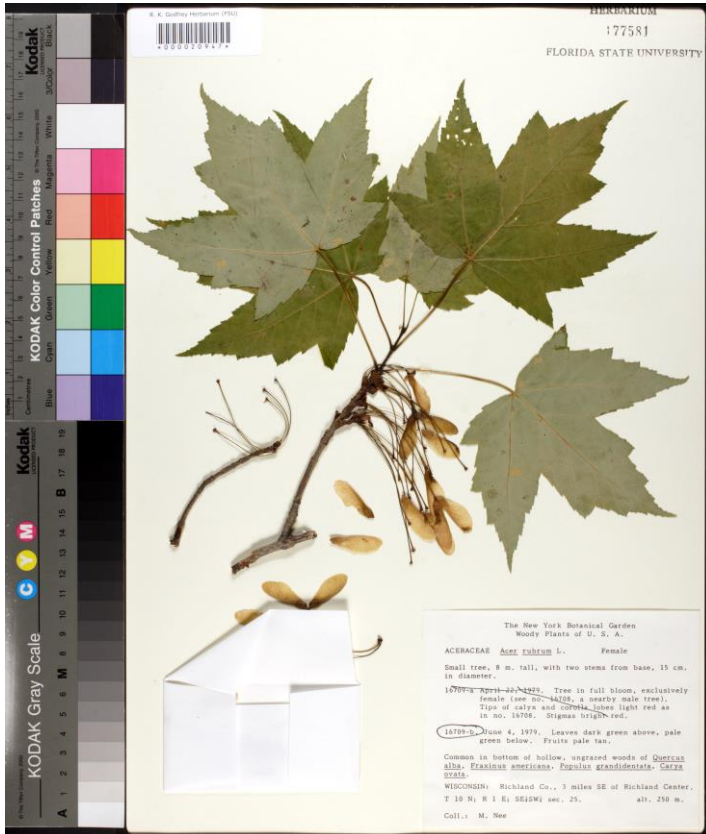
Carl Linné, aka Carolus Linnaeus

TRY

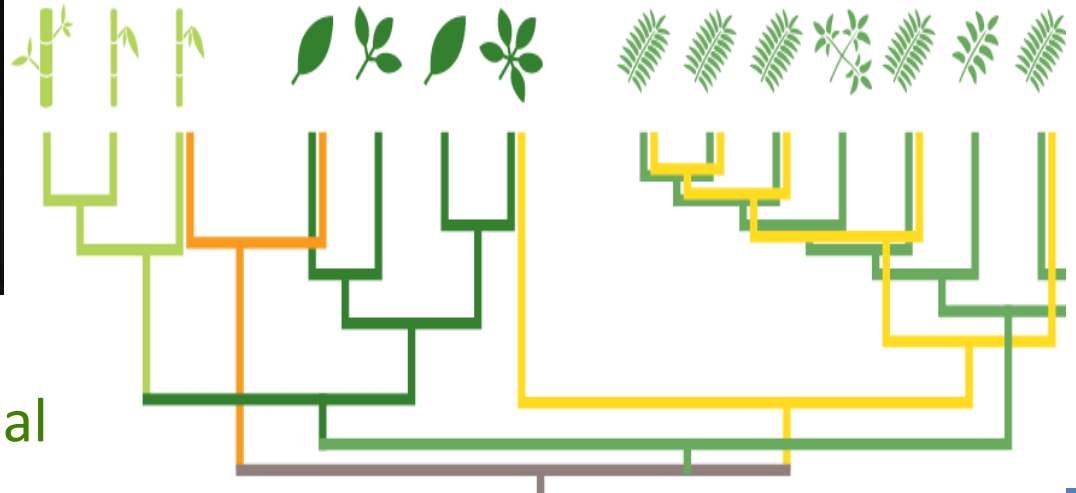
Plant Trait Database

PhotosyntheticPathway
Respiration LeafArea NfixationCapacity
SLA RegenerationCapacity PlantLifespan
WoodDensity GrowthForm
PhenologyType LeafN
LeafP LeafLongevity PhotosyntheticCapacity
MaxPlantHeight SeedMass

Using Images to Infer Functional Traits

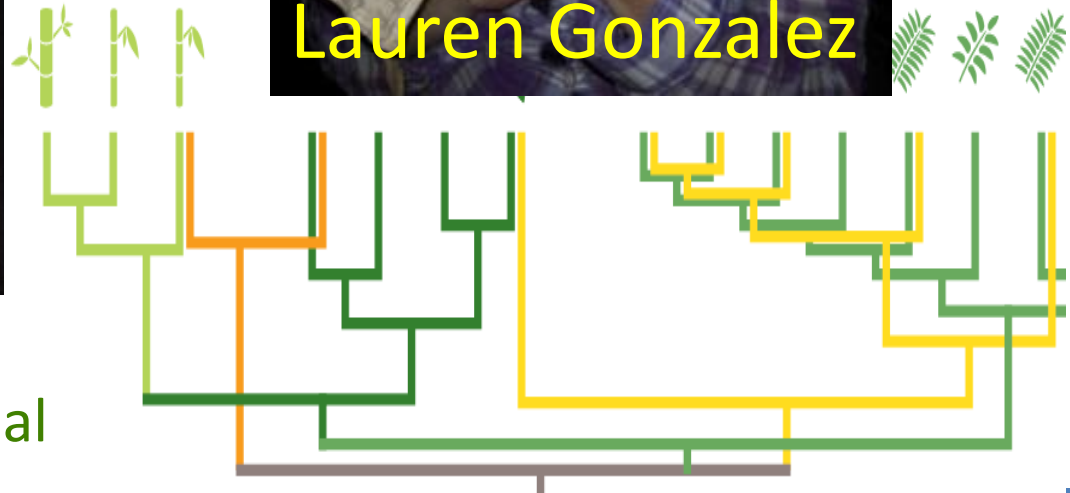
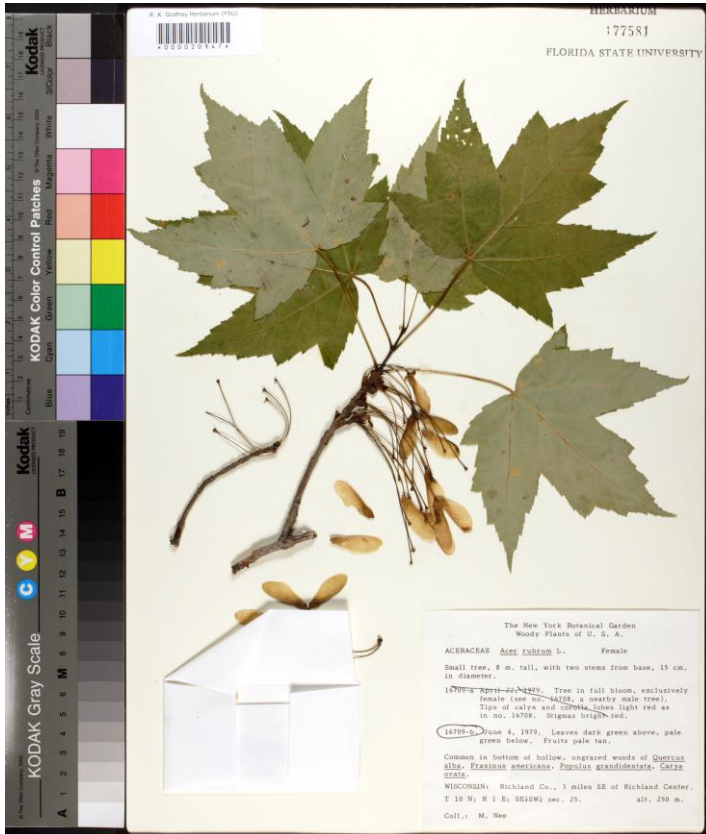


Photosynthetic Pathway
 Respiration Leaf Area Nfixation Capacity
 SLA Regeneration Capacity Plant Lifespan
 Wood Density Growth Form
 Phenology Type Leaf N
 Leaf P Leaf Longevity Photosynthetic Capacity
 Max Plant Height Seed Mass



Connect to ecology
 Evolution of plant functional
 traits

Using Images to Infer Functional Traits



Connect to ecology
Evolution of plant functional
traits

Many Research Uses for Specimen Data

- Connections to other resources, e.g. GenBank
- Ecological Niche Modeling
- Integration with phylogeny, e.g. PhyloJIVE
- Complex integration of phylogeny, specimens, ENM, other heterogeneous data
- Images as sources of traits for ecological studies
- Others???

- iDigBio Research Applications Working Group

Thank you!

iDigBio colleagues: J. Fortes, A. Matsunaga
J. Miller, D. Soltis, S. Smith, J. Beach, J. Soberon, L. Harmon
C. Germain-Aubrey, B. Marchant, L. Gonzalez



www.idigbio.org

psoltis@flmnh.ufl.edu



facebook.com/iDigBio



twitter.com/iDigBio



vimeo.com/idigbio



idigbio.org/rss-feed.xml



webcal://www.idigbio.org/events-calendar/export.ics

Linking to GenBank



Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [NCBI News](#)

- **Examples of the /specimen_voucher information:**

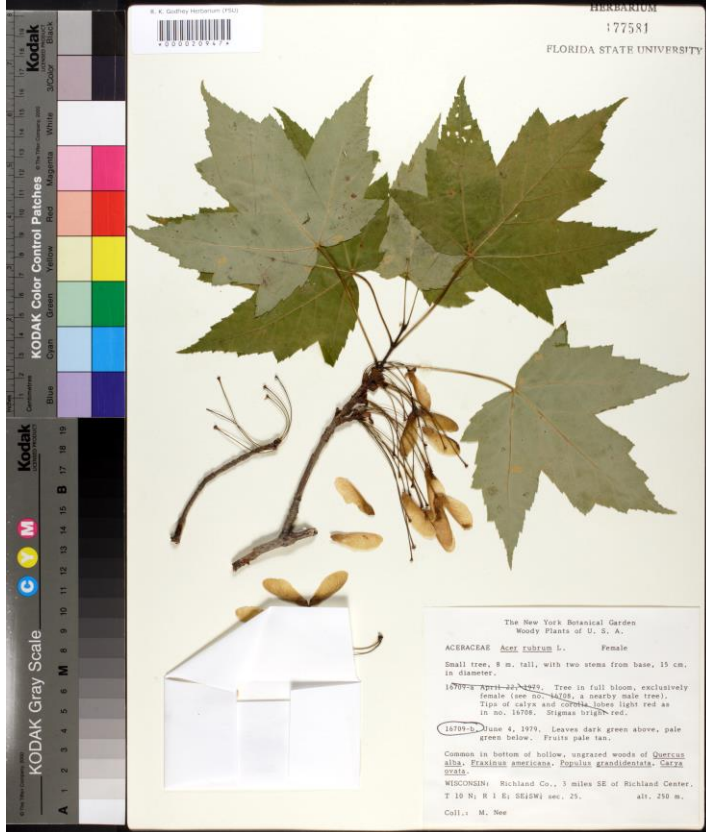
```
/specimen_voucher="UAM:Mamm:52179"
```

```
/specimen_voucher="AMCC:101706"
```

```
/specimen_voucher="USNM:field series 8798"
```

```
/specimen_voucher="personal:Dan Janzen:99-SRNP-2003"
```

```
/specimen_voucher="99-SRNP-2003"
```



Connecting ecology to
specimens
Correlation of plant functional
traits

