

Using Digitized Data in Scientific Research *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.



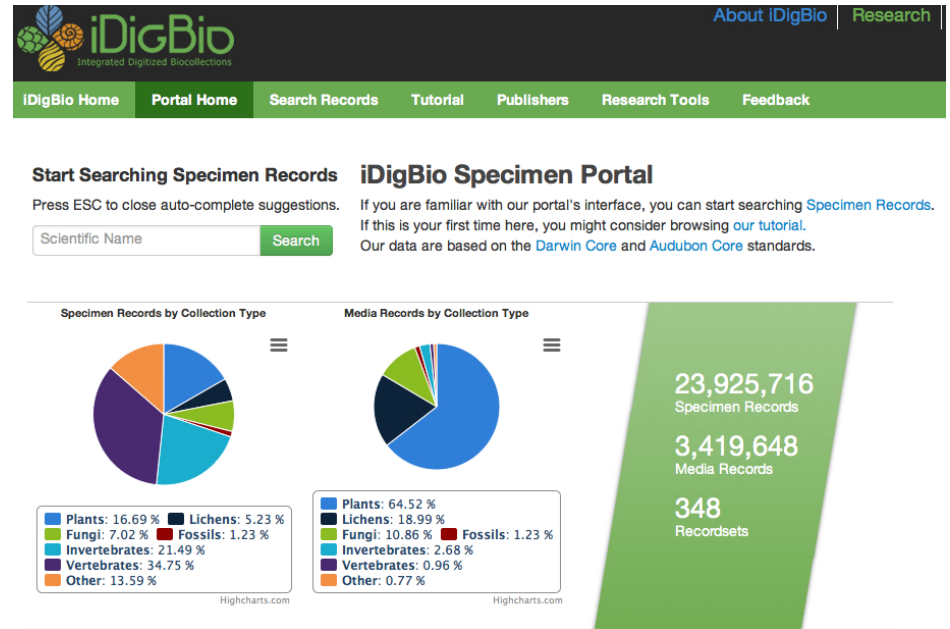
Collections: The Library of Life

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



Research @ iDigBio

- Data portal

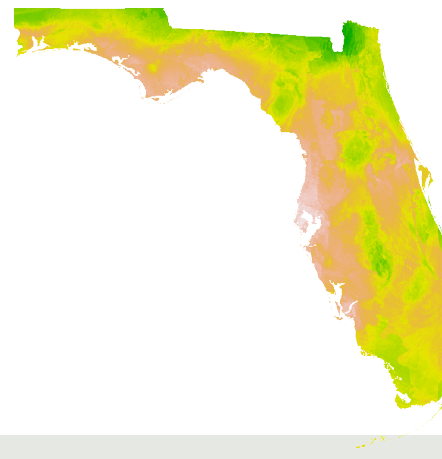
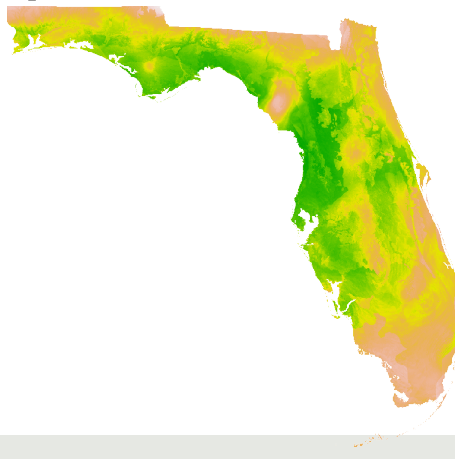


The screenshot shows the iDigBio Specimen Portal interface. At the top is the iDigBio logo and navigation links: About iDigBio, Research, iDigBio Home, Portal Home, Search Records, Tutorial, Publishers, Research Tools, and Feedback. Below the navigation is a search bar with the text "Scientific Name" and a "Search" button. To the right of the search bar is a section titled "iDigBio Specimen Portal" with a sub-heading "Start Searching Specimen Records". It contains text: "Press ESC to close auto-complete suggestions." and "If you are familiar with our portal's interface, you can start searching [Specimen Records](#). If this is your first time here, you might consider browsing [our tutorial](#). Our data are based on the [Darwin Core](#) and [Audubon Core](#) standards." Below the search bar are two pie charts: "Specimen Records by Collection Type" and "Media Records by Collection Type". To the right of the charts is a green vertical bar with the following statistics: 23,925,716 Specimen Records, 3,419,648 Media Records, and 348 Recordsets.

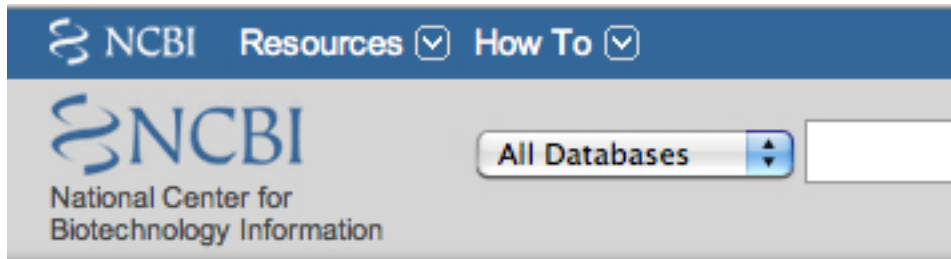
| Collection Type | Percentage |
|-----------------|------------|
| Plants | 16.69 % |
| Fungi | 7.02 % |
| Invertebrates | 21.49 % |
| Vertebrates | 34.75 % |
| Other | 13.59 % |
| Lichens | 5.23 % |
| Fossils | 1.23 % |

| 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 and environment



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"
```

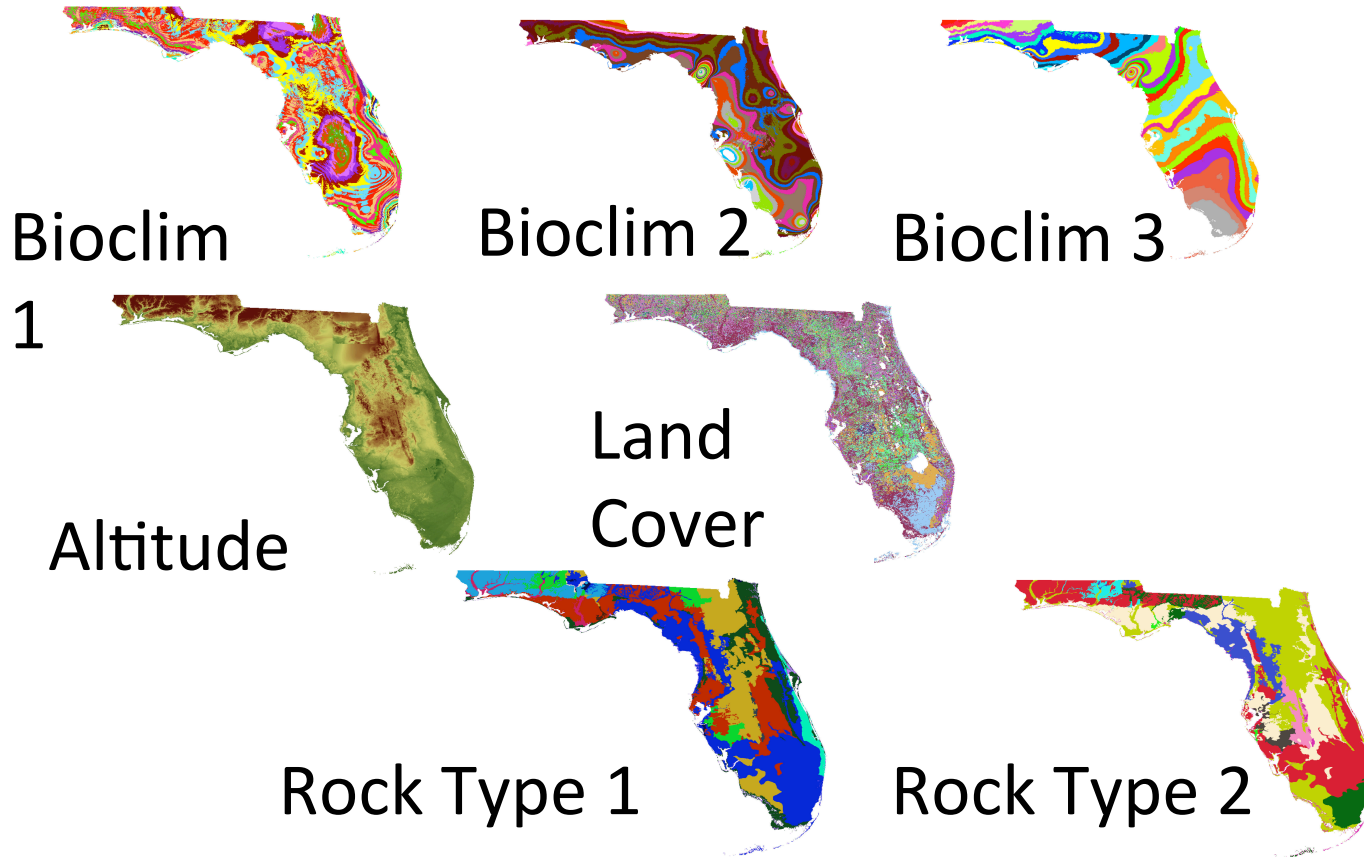

Ecological Niche Modeling: locality information



GEOLocate  29.65, -82.32

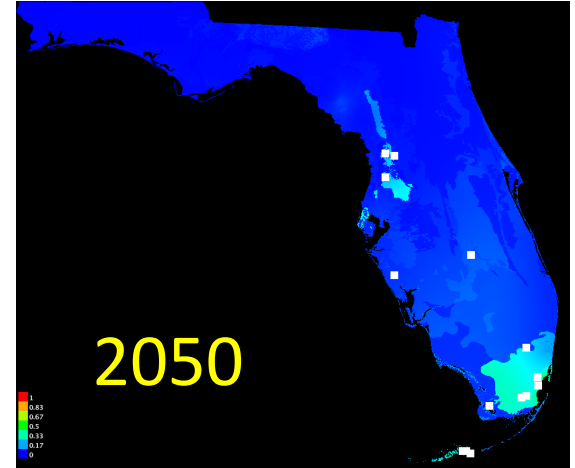
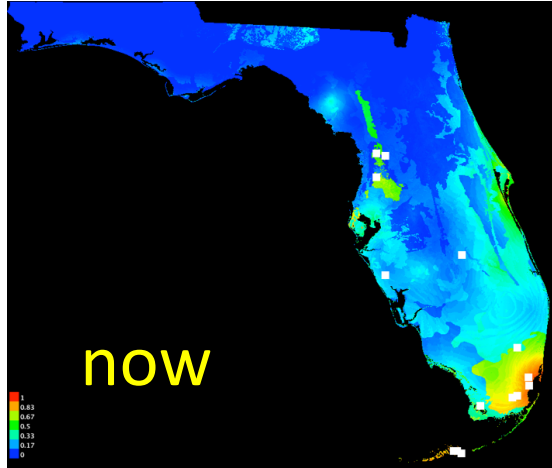
number,dwc:preparations,dwc:identificationVerificationStatus,idigbio:subfamily,idigbio:preparationCount,fcc:pickedBy,dwc:eventRemarks,dwc:VerbatimEventDate,dwc:associatedReferences,idigbio:endangeredStatus,dwc:locationAccordingTo,dwc:georeferenceSources,dwc:associatedSequences,dwc:formation,dwc:higherClassification,dwc:catalogNumber,dwc:verbatimSRS,dwc:higherGeo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

Ecological Niche Modeling: Environmental Layers

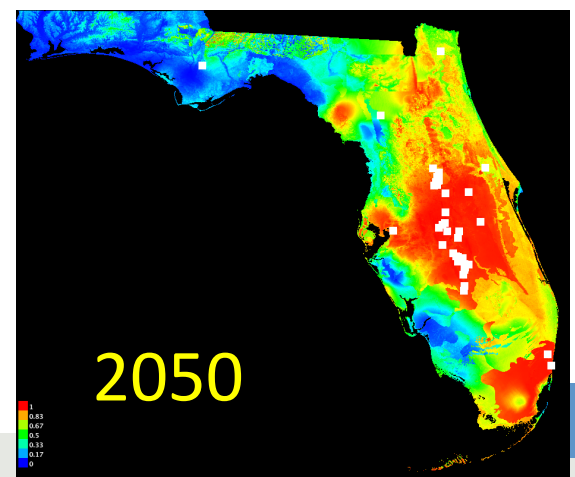
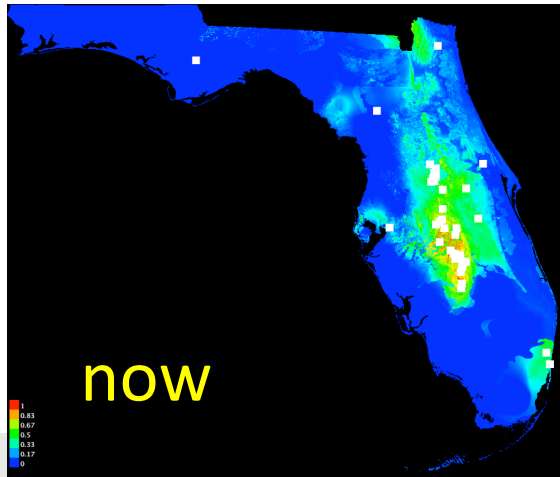


Responses to Climate Change: Winners & Losers

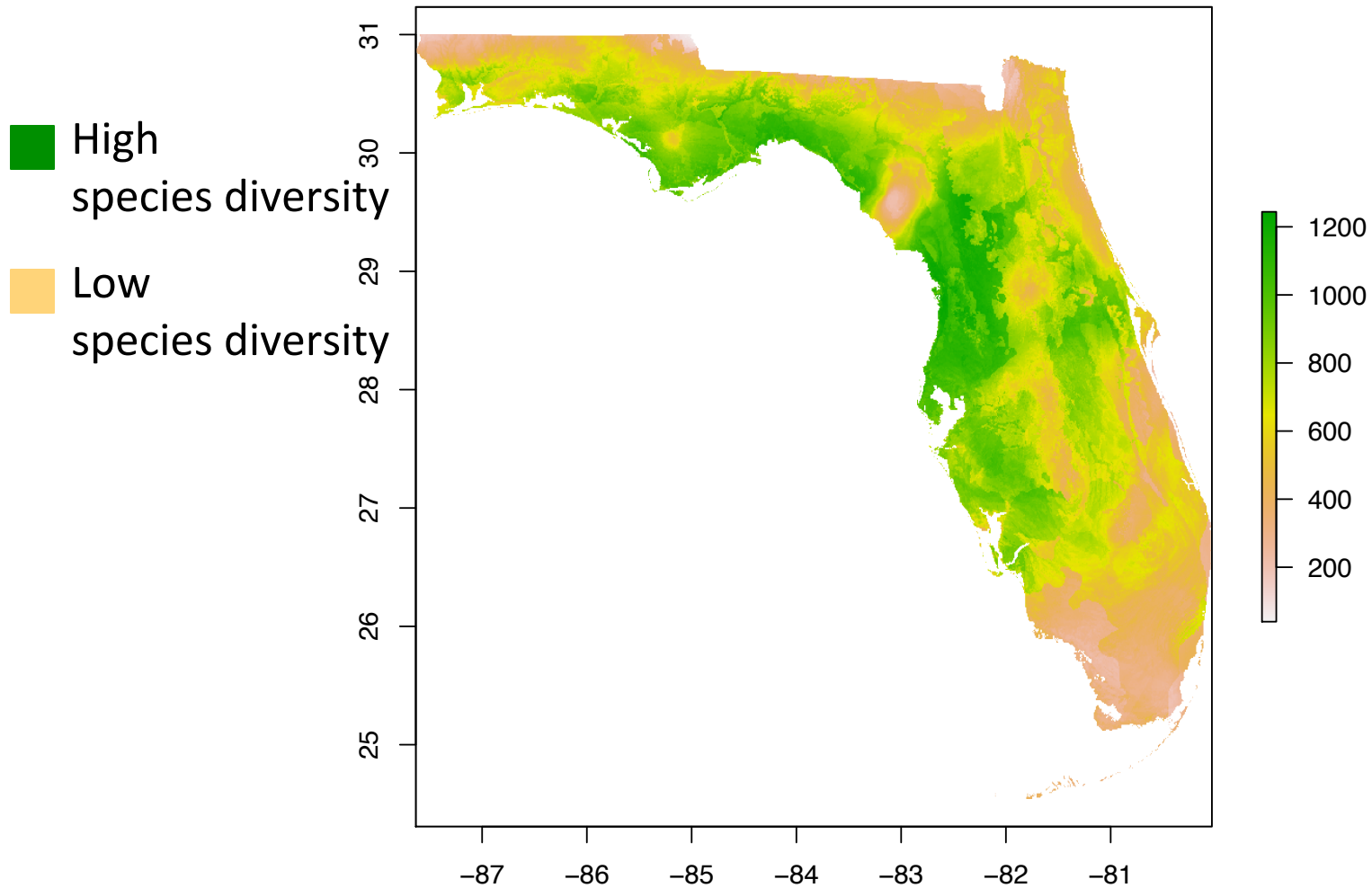
Abildgaardia ovata (flatspike sedge)



Prunus geniculata (scrub plum)

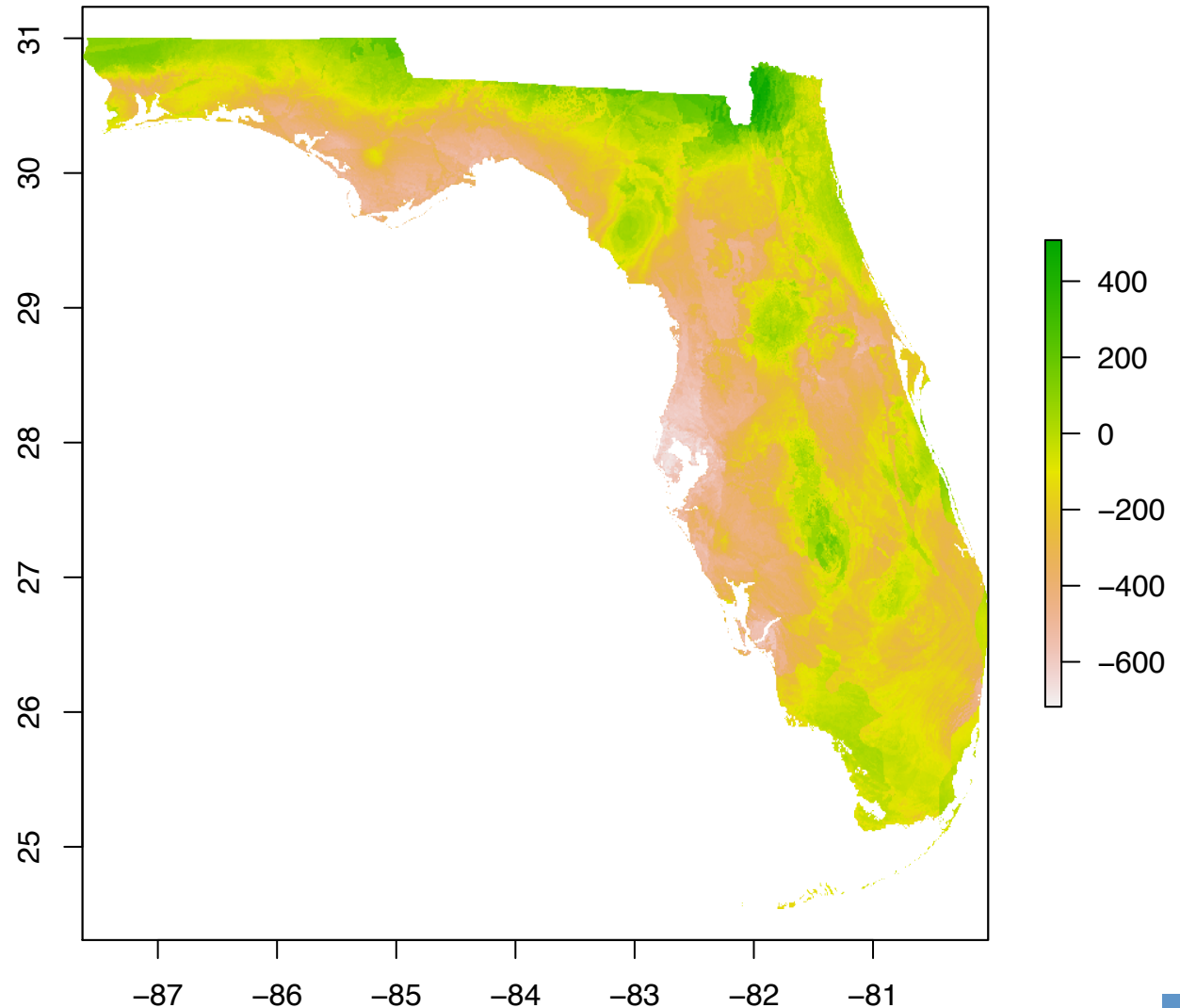


Florida plant diversity heat map: now



Between now and 2050...

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



spp 2050 - #spp now

Linking to Phylogeny

The screenshot shows the OneZoom website interface. At the top, a dark navigation bar contains the OneZoom logo and menu items: 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". Each of these main branches further divides into smaller, more detailed sub-branches. Three circular callouts are overlaid on the tree, each containing text and a small tree icon: "Water-lilies, Water-shields and more" is positioned above the central "Flowering Plants" branch; "Star Anise, Lemon Wood and more" is positioned above the right "Mesangiosperms" branch; and "Gymnosperms" is positioned above the left "Gymnosperms" branch. At the bottom of the screenshot, a dark grey bar contains the text "Click to see how OneZoom works".

OneZoom Home Embed Software Impacts News About Gallery Future

Water-lilies, Water-shields and more

Star Anise, Lemon Wood and more

Gymnosperms

Flowering Plants

Mesangiosperms

Seed Plants

Click to see how OneZoom works

PhyloJIVE

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



[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](#).

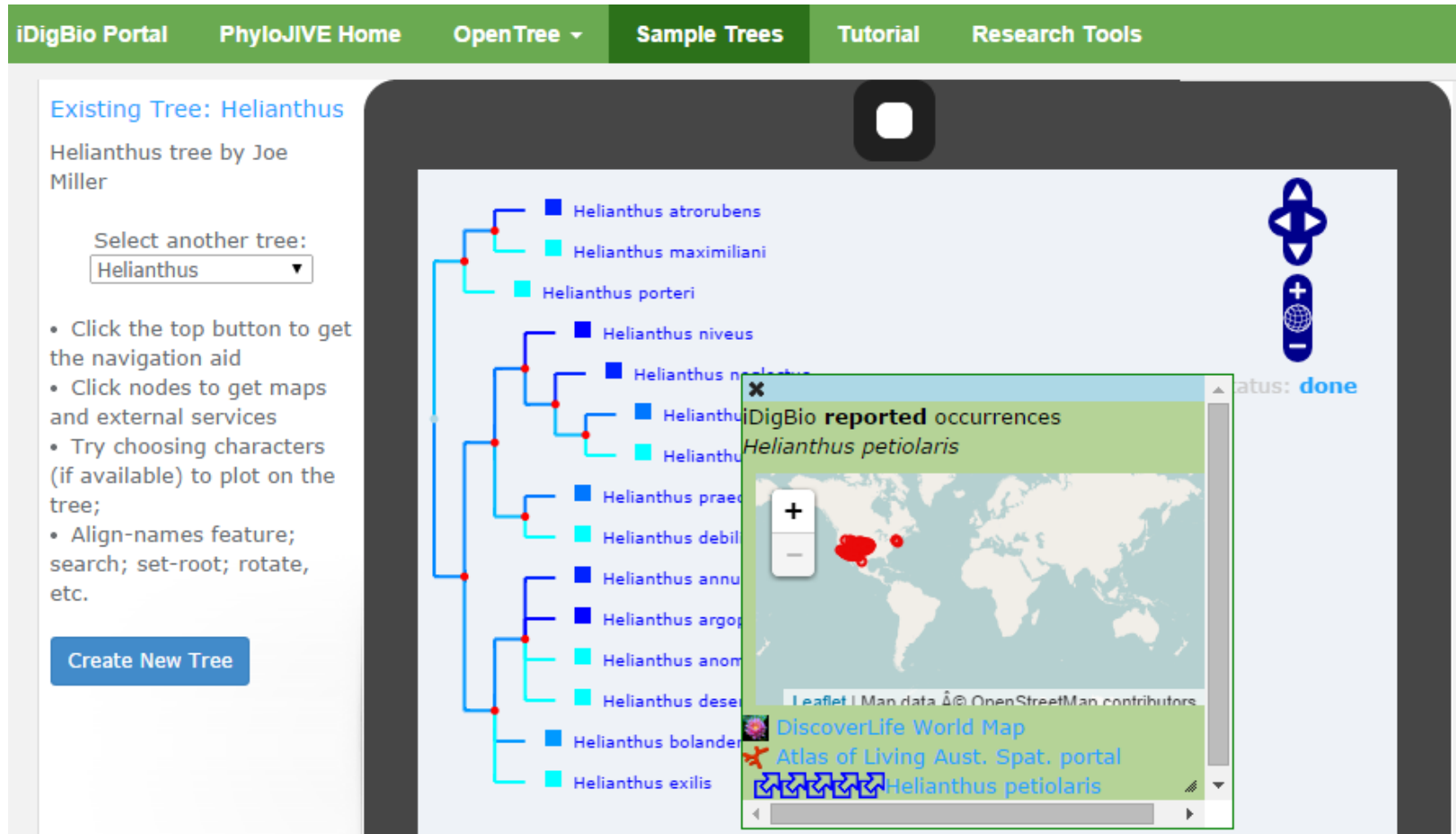
The [getting started](#) page outlines the steps for creating a new phylogenetic tree and contains demo data sets that can be used to get up and running.



A. buxifolia

Source: Australian Plant
Image Index
Image by: Macdonald, C.

PhyloJIVE instance in iDigBio



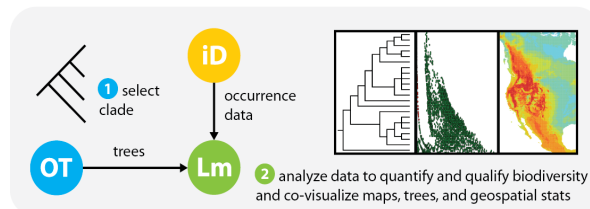
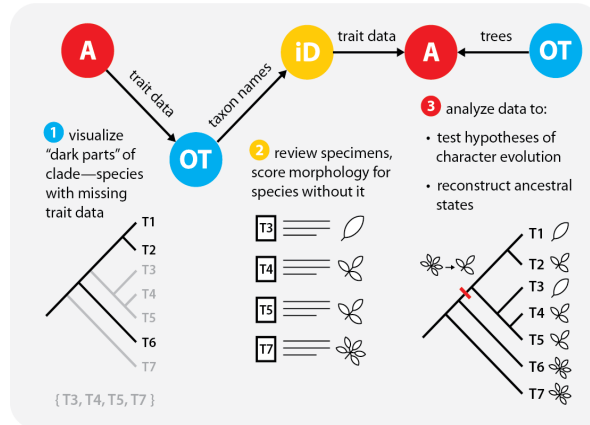
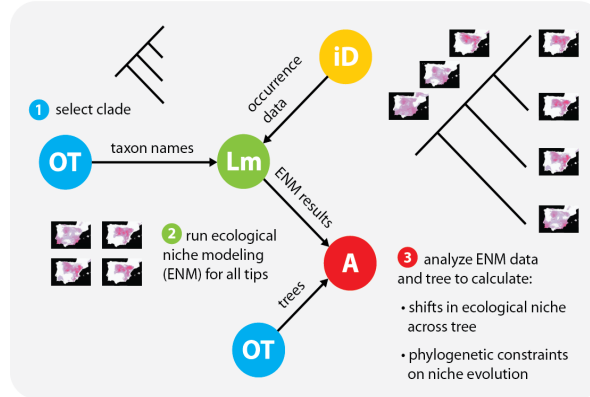
The screenshot displays the iDigBio Portal interface for the PhyloJIVE instance. The navigation bar includes links for iDigBio Portal, PhyloJIVE Home, OpenTree, Sample Trees, Tutorial, and Research Tools. The main content area shows an existing tree titled "Existing Tree: Helianthus" by Joe Miller. A 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 praecox, Helianthus debilis, Helianthus annuus, Helianthus argophyllus, 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 zoom control and a status bar at the bottom with links to DiscoverLife World Map, Atlas of Living Aust. Spat. portal, and Helianthus petiolaris.

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

A. Matsunaga

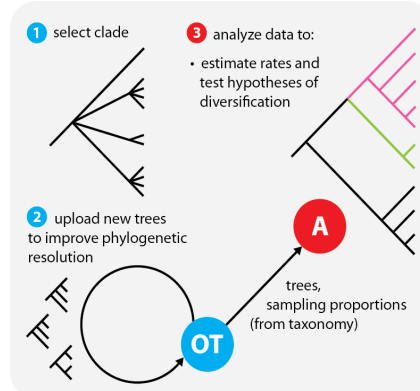
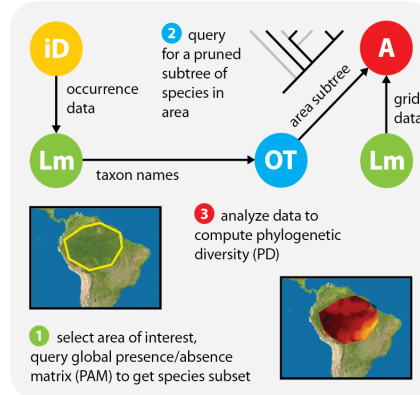
Connecting Trees, Specimens, Tools

EXAMPLE WORKFLOWS:

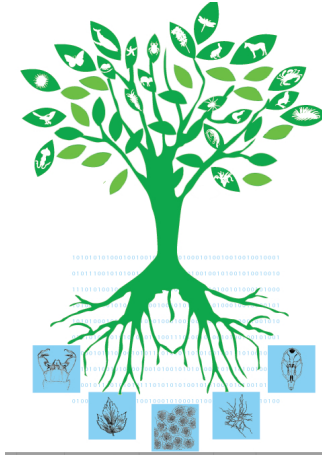


RESOURCES:

- 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

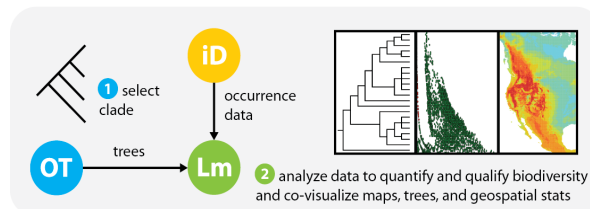
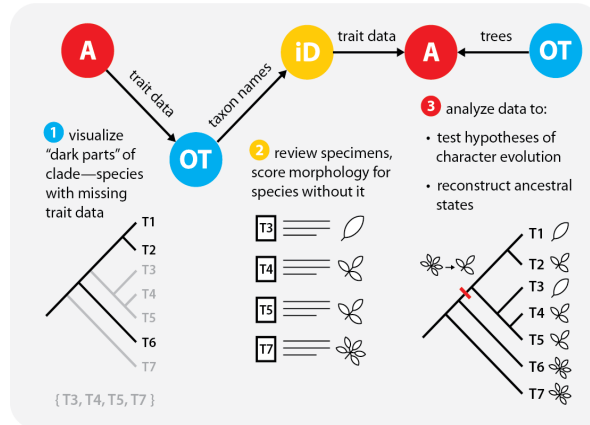
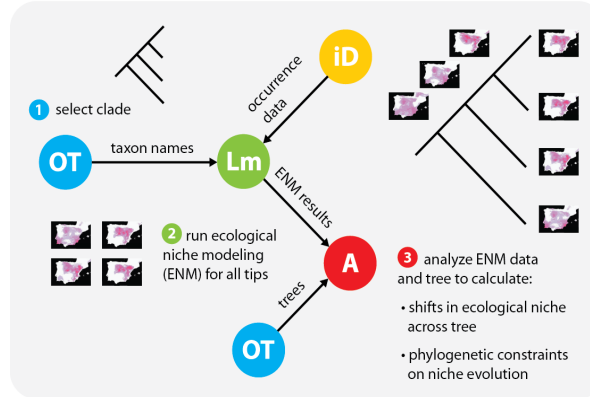


Connecting Trees, Specimens, Tools



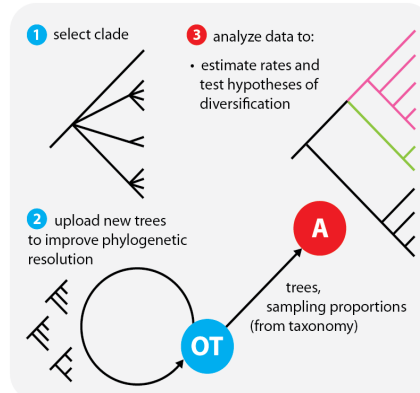
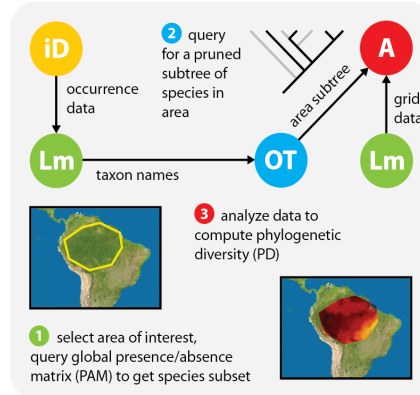
Connecting Trees, Specimens, Tools

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

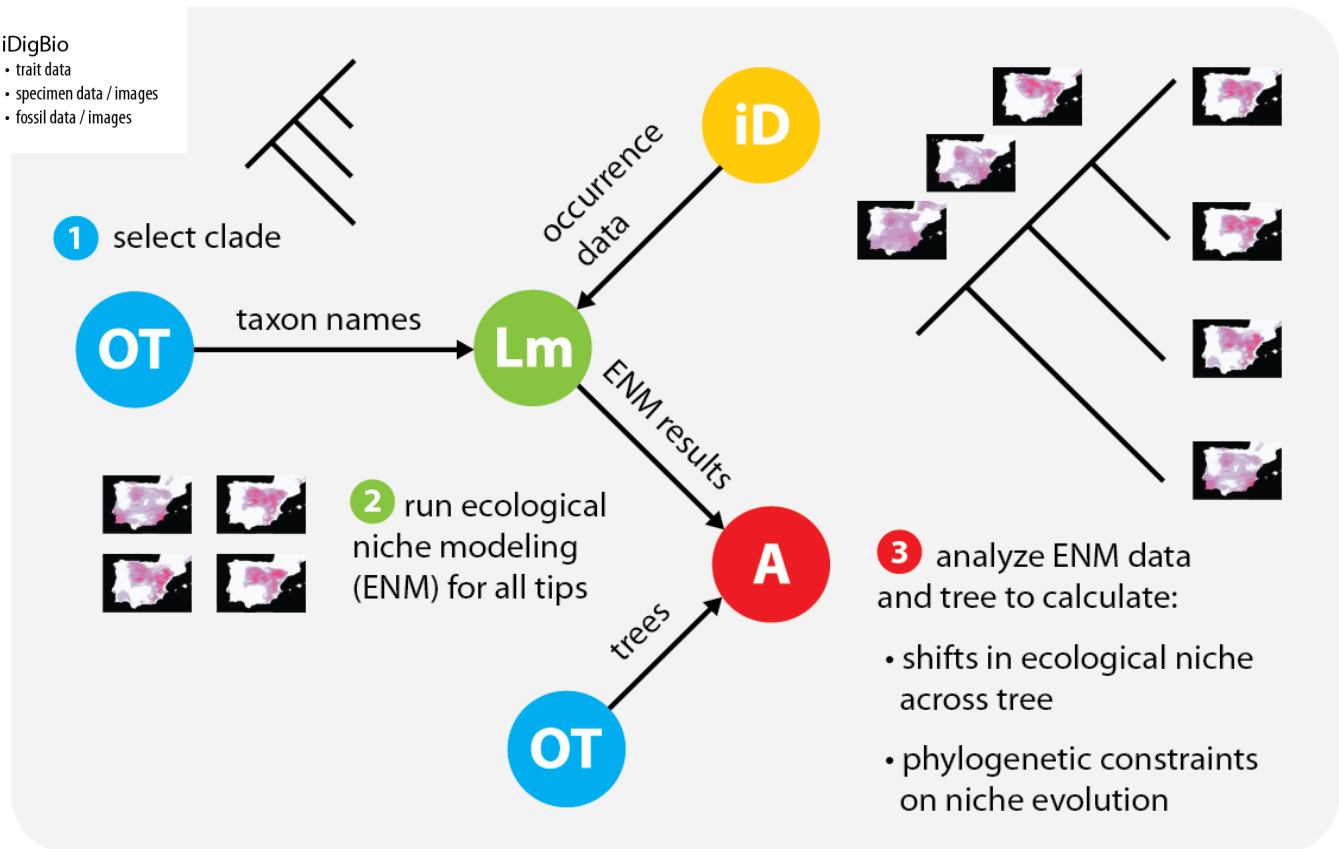
RESOURCES:

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

A Arbor
 • evolutionary models
 • comparative methods
 • visualization

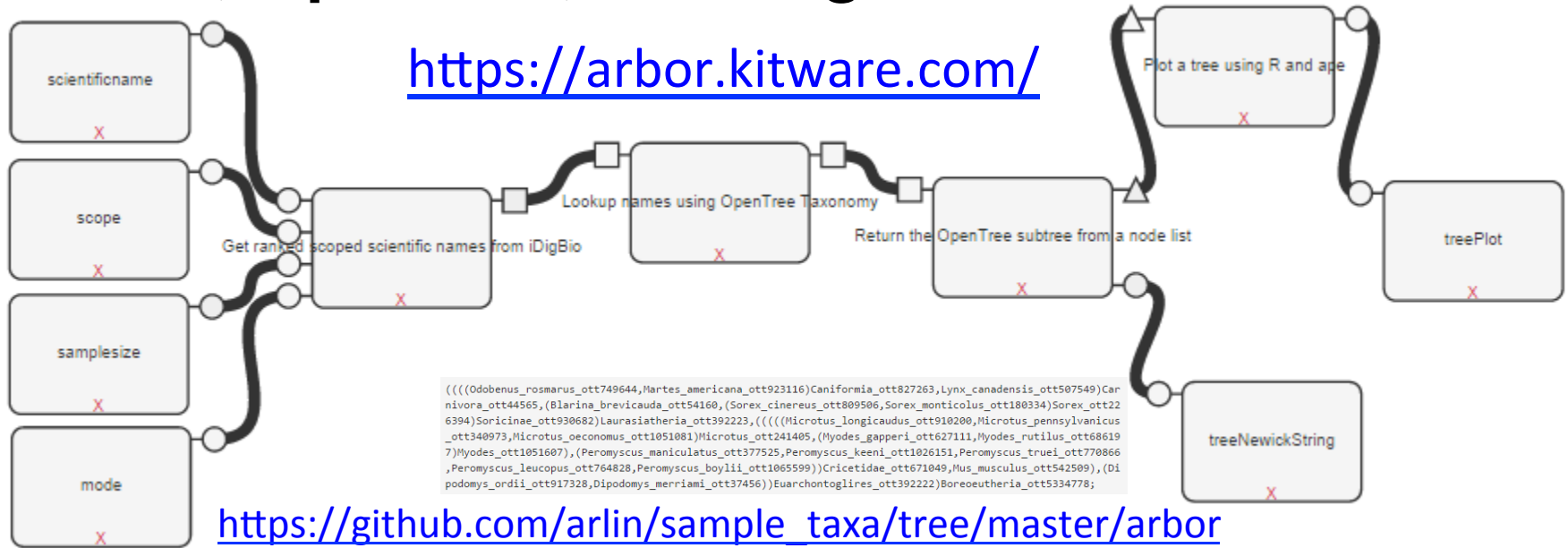
OT Open Tree of Life
 • phylogenies
 • taxonomy / names
 • visualization

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



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:

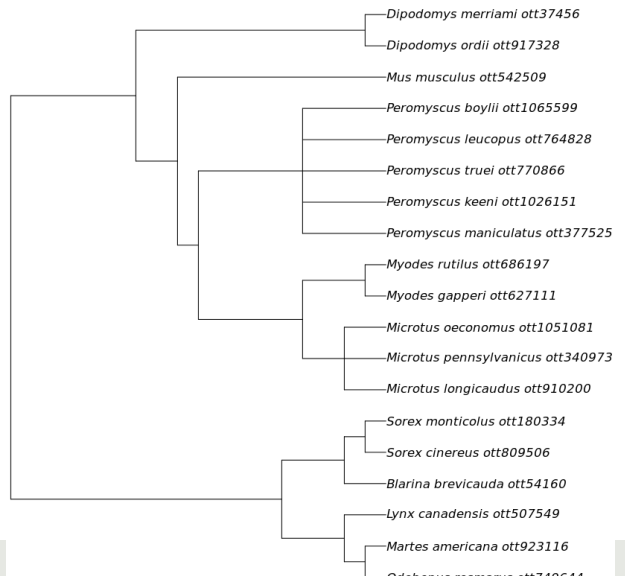
mode:

```

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:

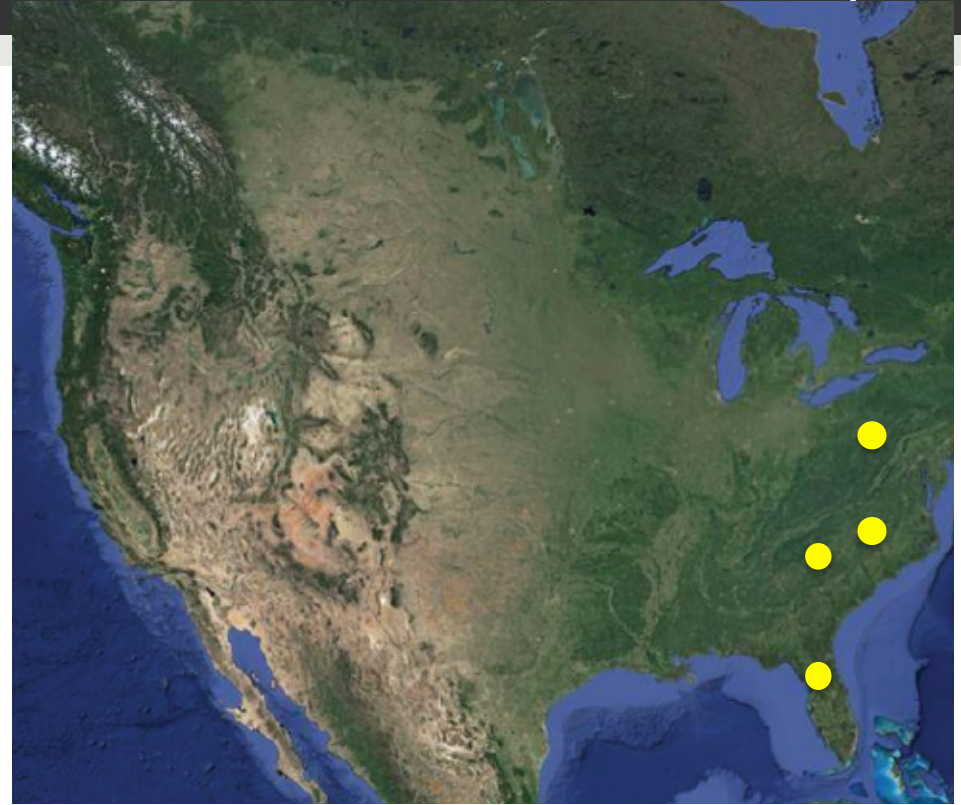
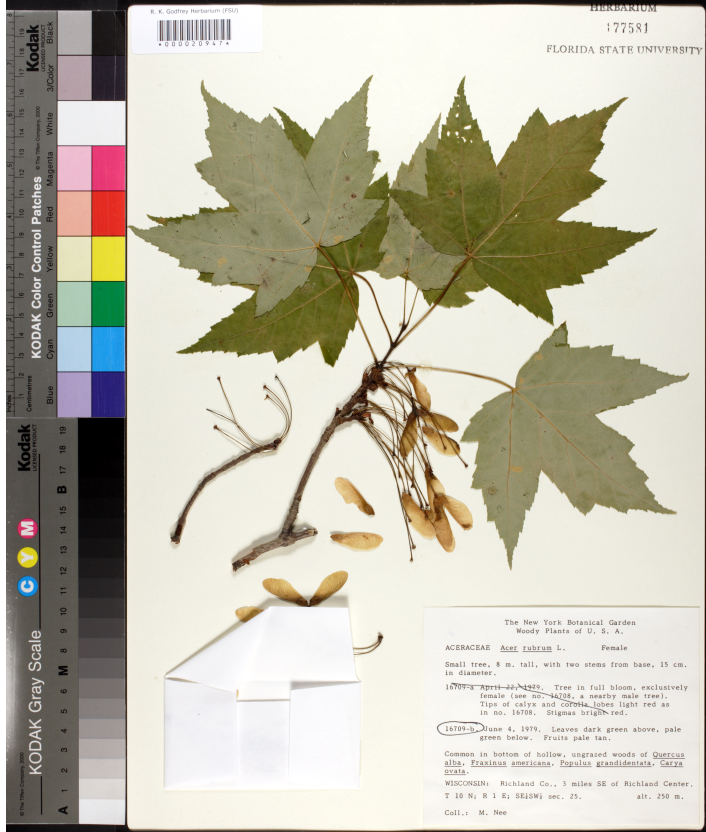
- Workflow to get an induced tree from a configurable iDigBio query treeNewickString [string]
- Workflow to get an induced tree from a configurable iDigBio query treePlot [image]



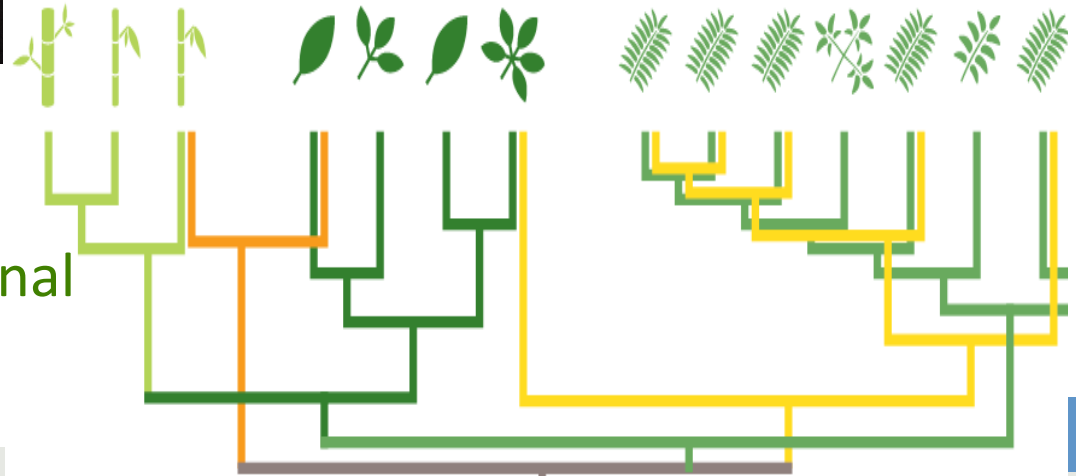
TRY

Plant Trait Database

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



Connecting ecology to specimens
Correlation of plant functional traits



Many 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???
-
- Research Applications Working Group

Thank you!



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>