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

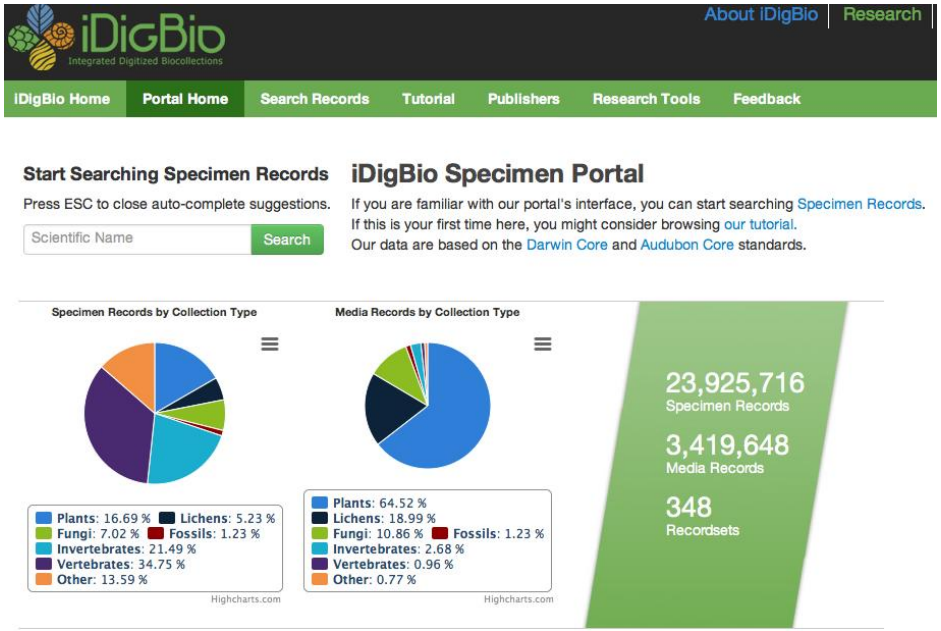
>1600 natural history collections in the US alone

1-2 billion specimens in the US

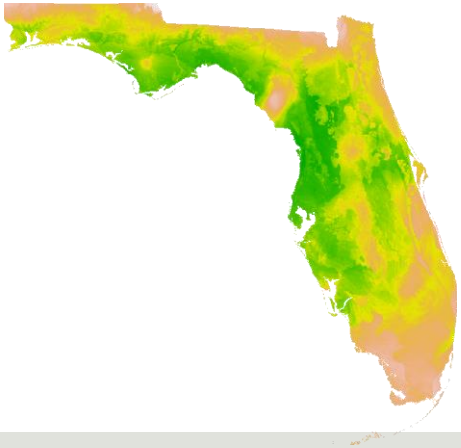
3-4 billion specimens worldwide

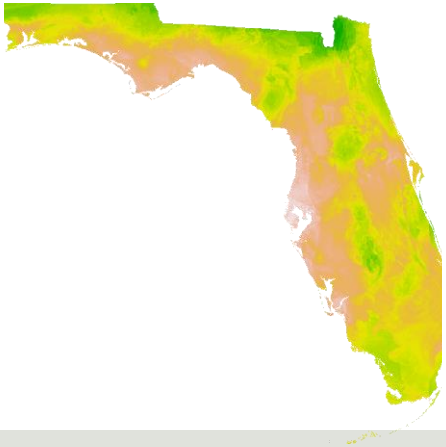
Research @ iDigBio

- Data portal



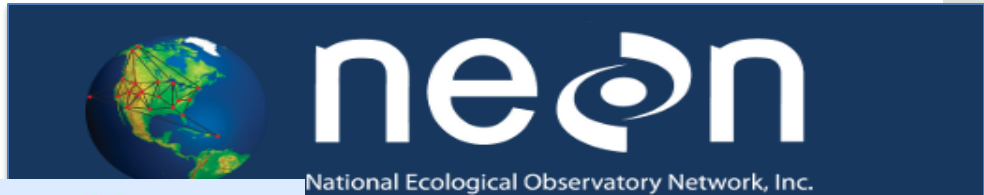
- Computational workflows – research use

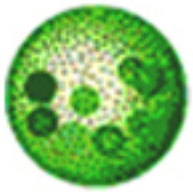




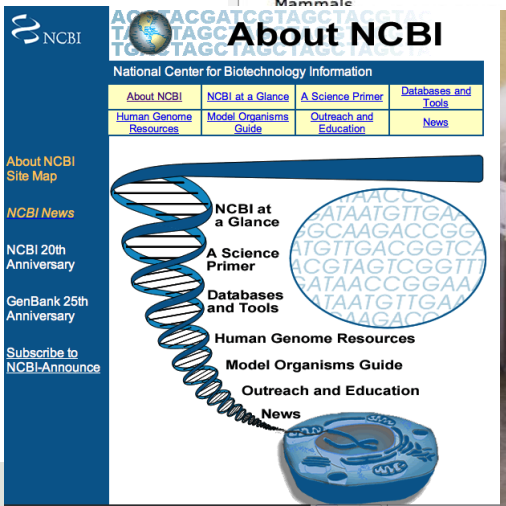
Linking Collections to…

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





UTEX



Synthetic Research and Complex Data Analysis

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

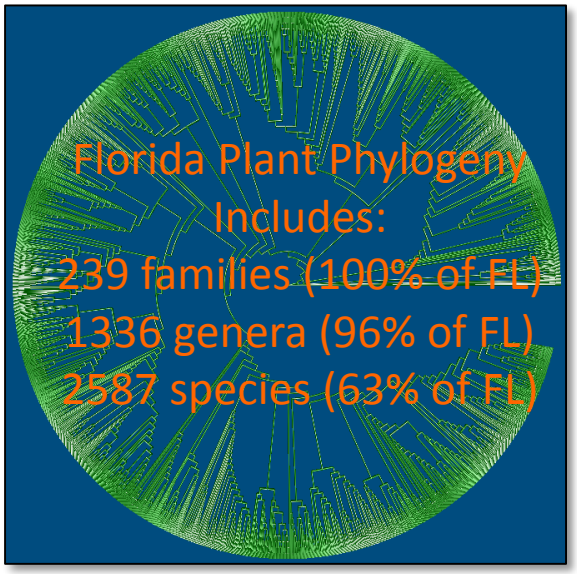
Synthetic Research: Connections

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

iDigBio
Integrated Digitized Biocollections

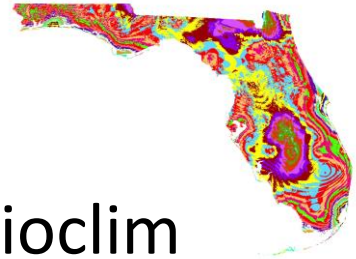
Florida Plant Diversity in a Changing Climate

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

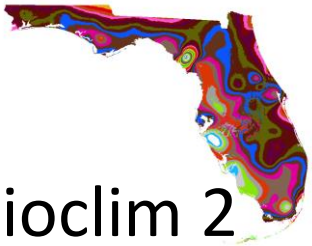


Charlotte Germain-Aubrey

Ecological Niche Modeling: Environmental Layers

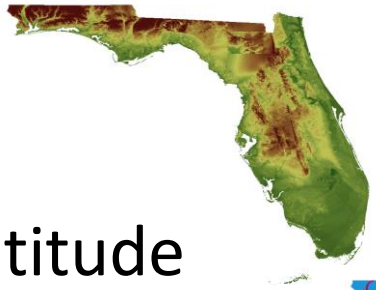


Bioclim 1



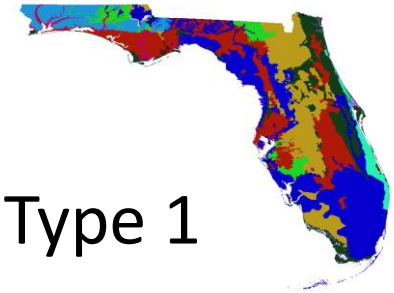
Bioclim 2

Bioclim 3

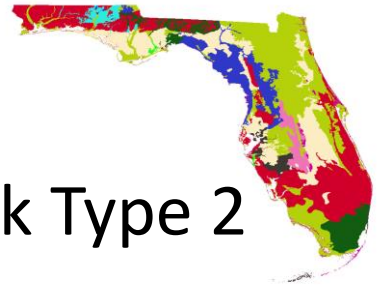


Altitude

Land Cover



Rock Type 1

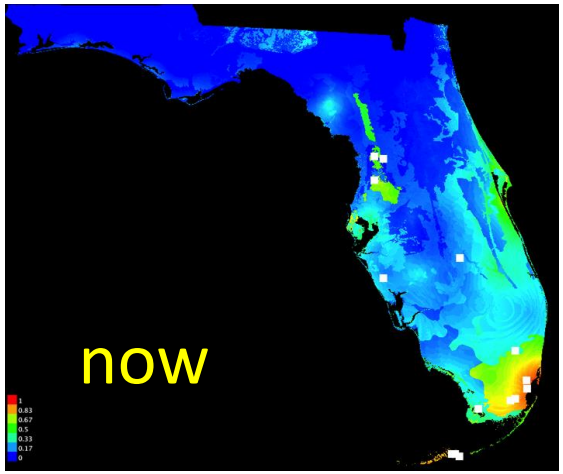


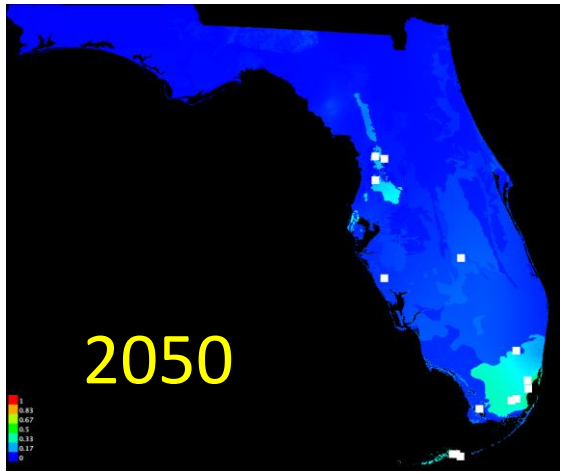
Rock Type 2

C. Germain-Aubrey et al.

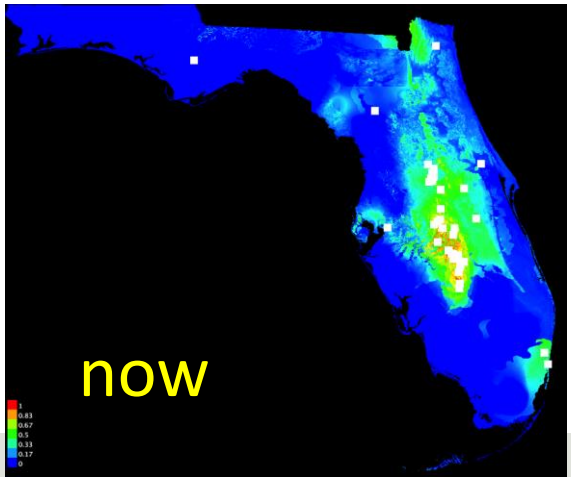
Responses to Climate Change: Winners & Losers

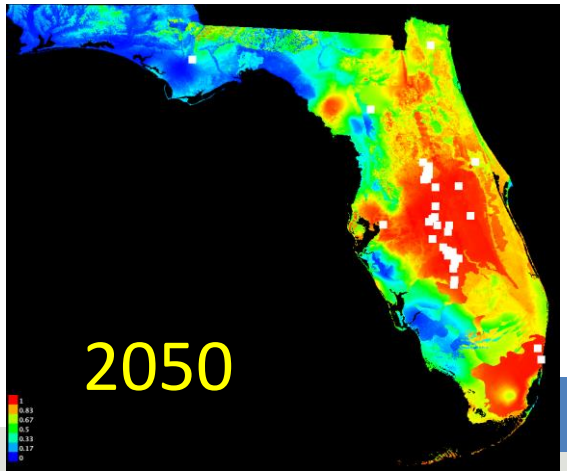
Abildgaardia ovata (flatspike sedge)





Prunus geniculata (scrub plum)

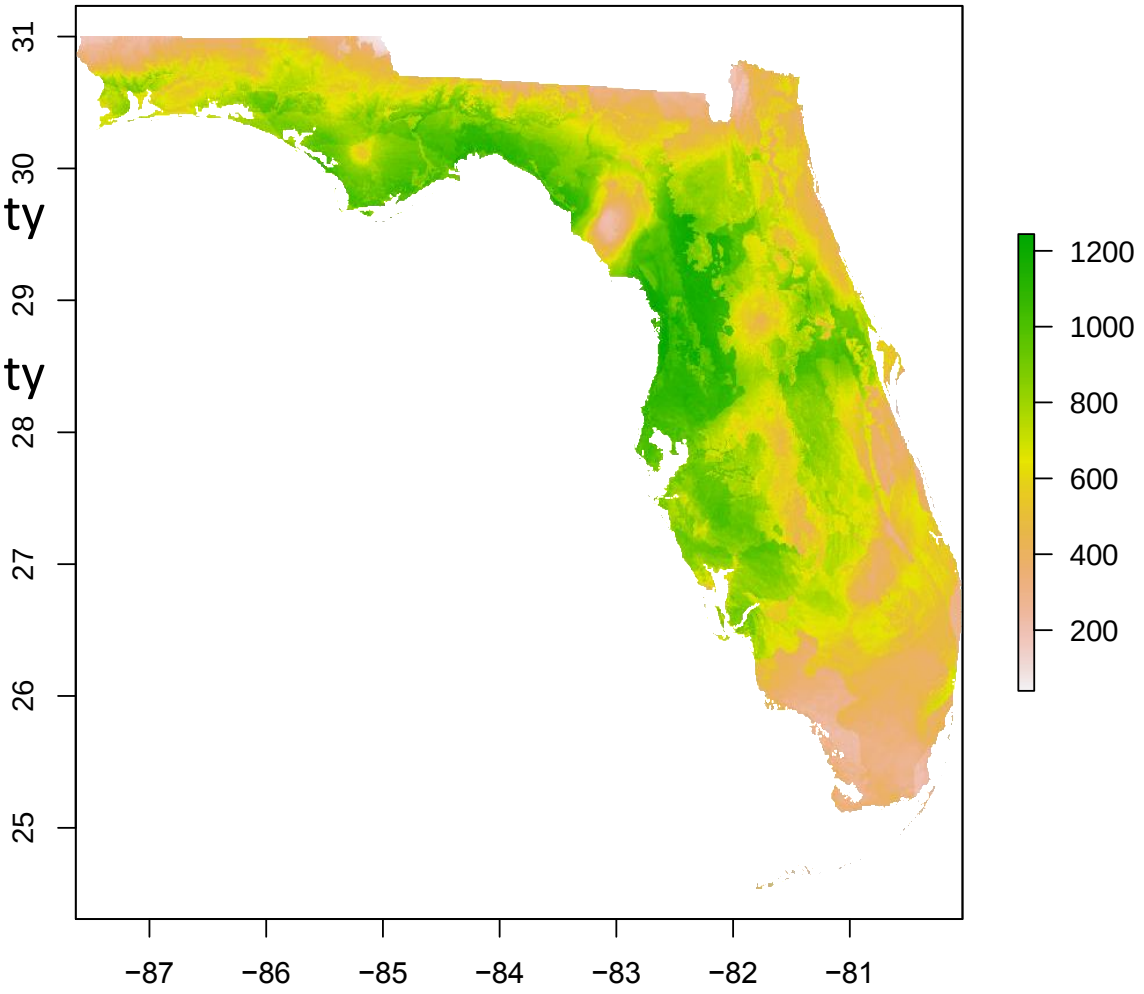




Florida plant diversity heat map: now

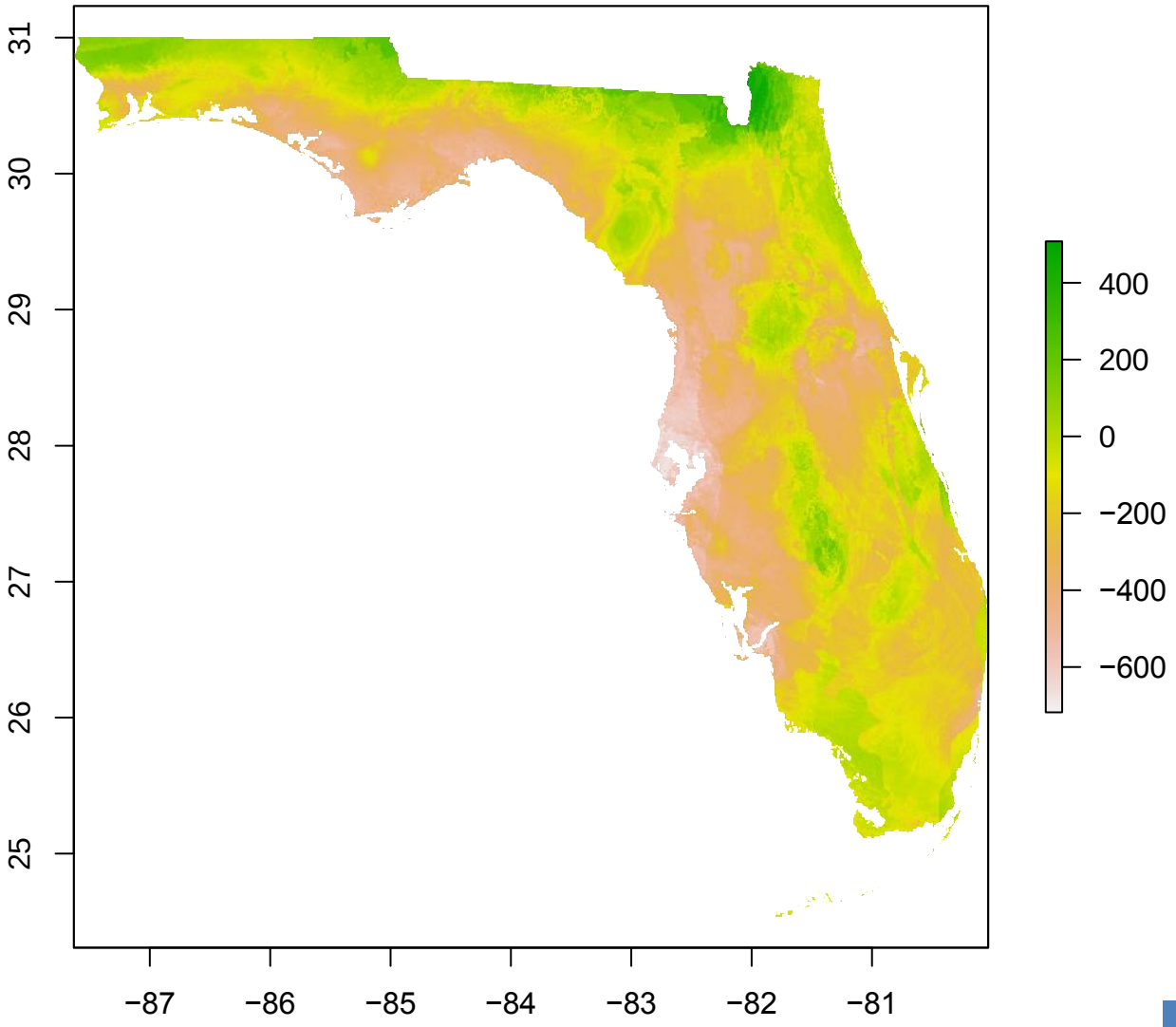
High species diversity

Low species diversity



Between now and 2050…

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



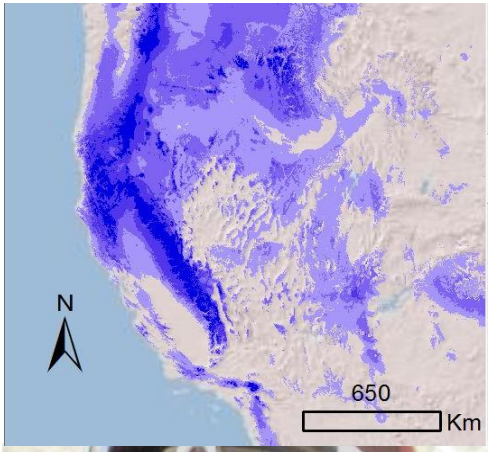
spp 2050 - #spp now

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

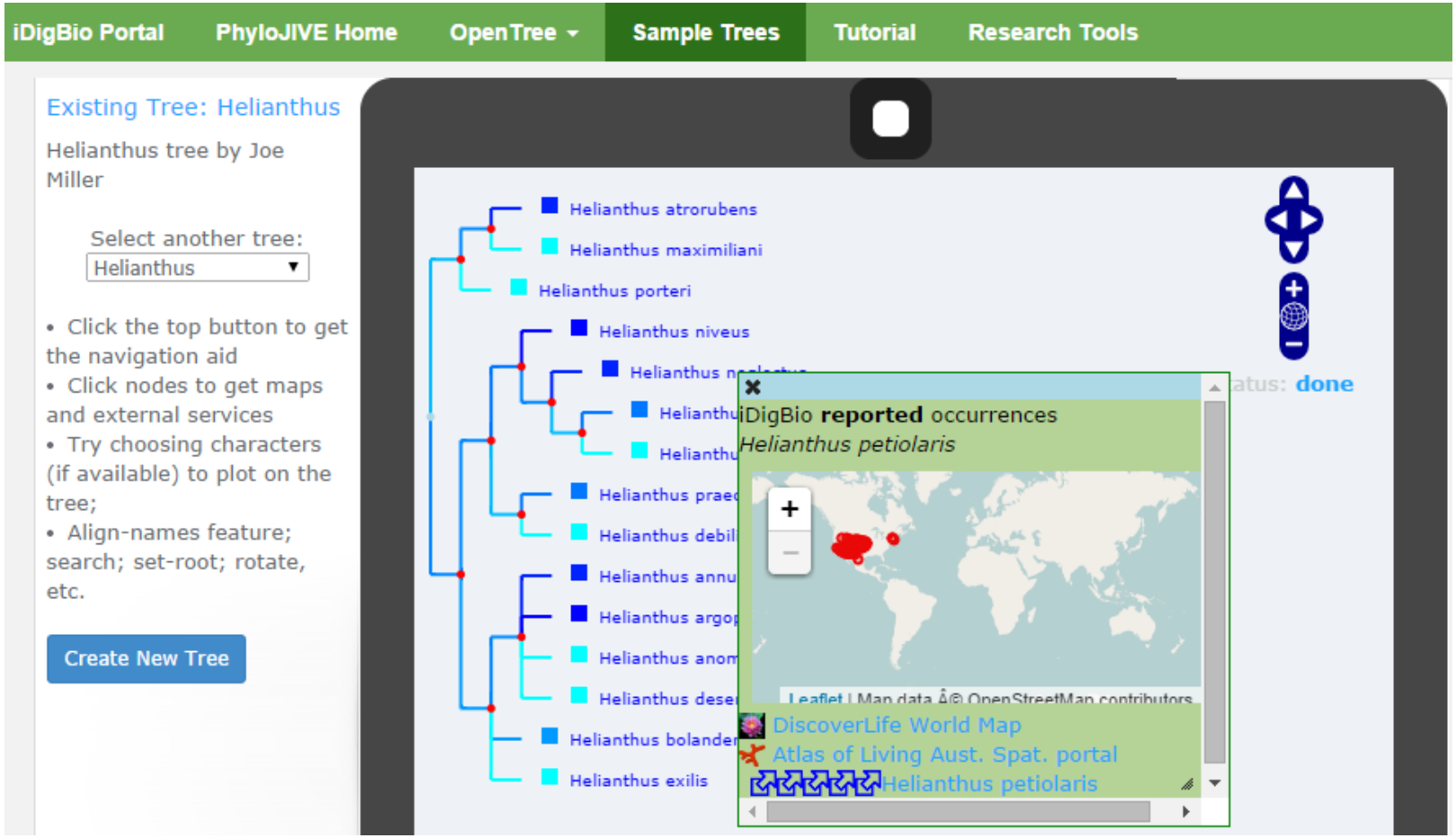
Linking to Phylogeny

PhyloJIVE

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

A. buxifolia
Source: Australian Plant Image Index Image by: Macd

PhyloJIVE instance in iDigBio



- 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

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

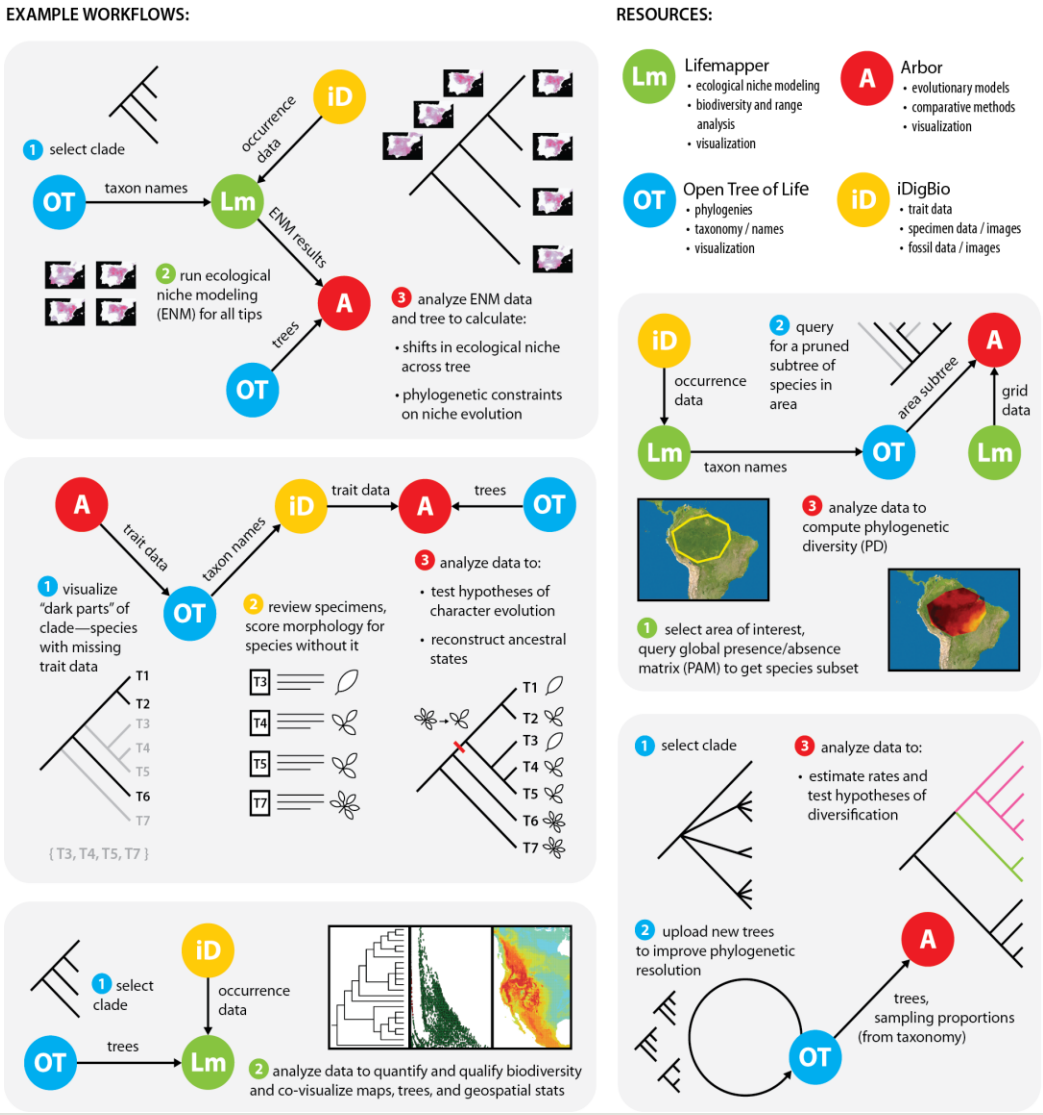
A. Matsunaga

PhyloJIVE instance in iDigBio

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

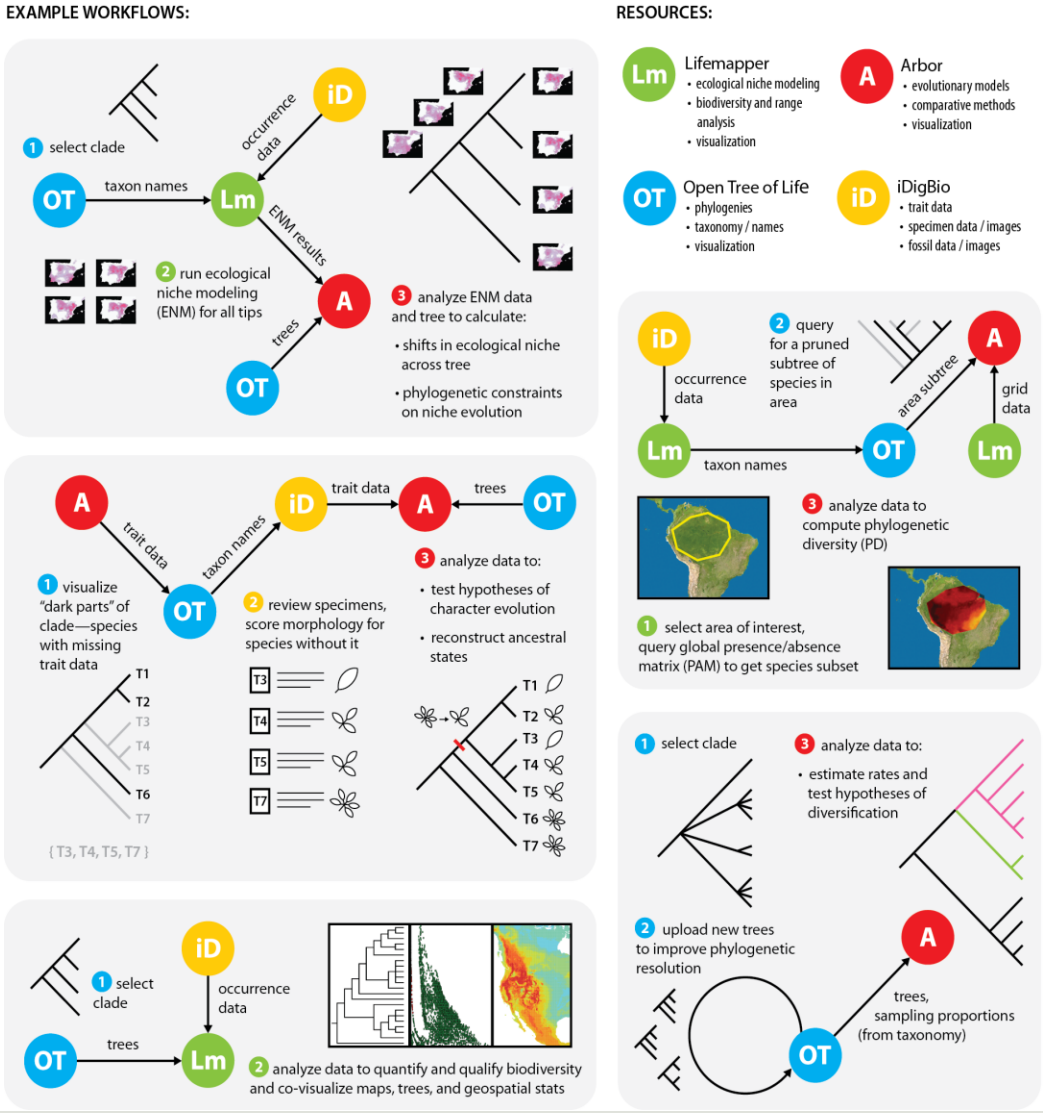
A. Matsunaga

Connecting Trees, Specimens, Tools



Connecting Trees, Specimens, Tools

Connecting Trees, Specimens, Tools



Connecting Trees, Specimens, Tools

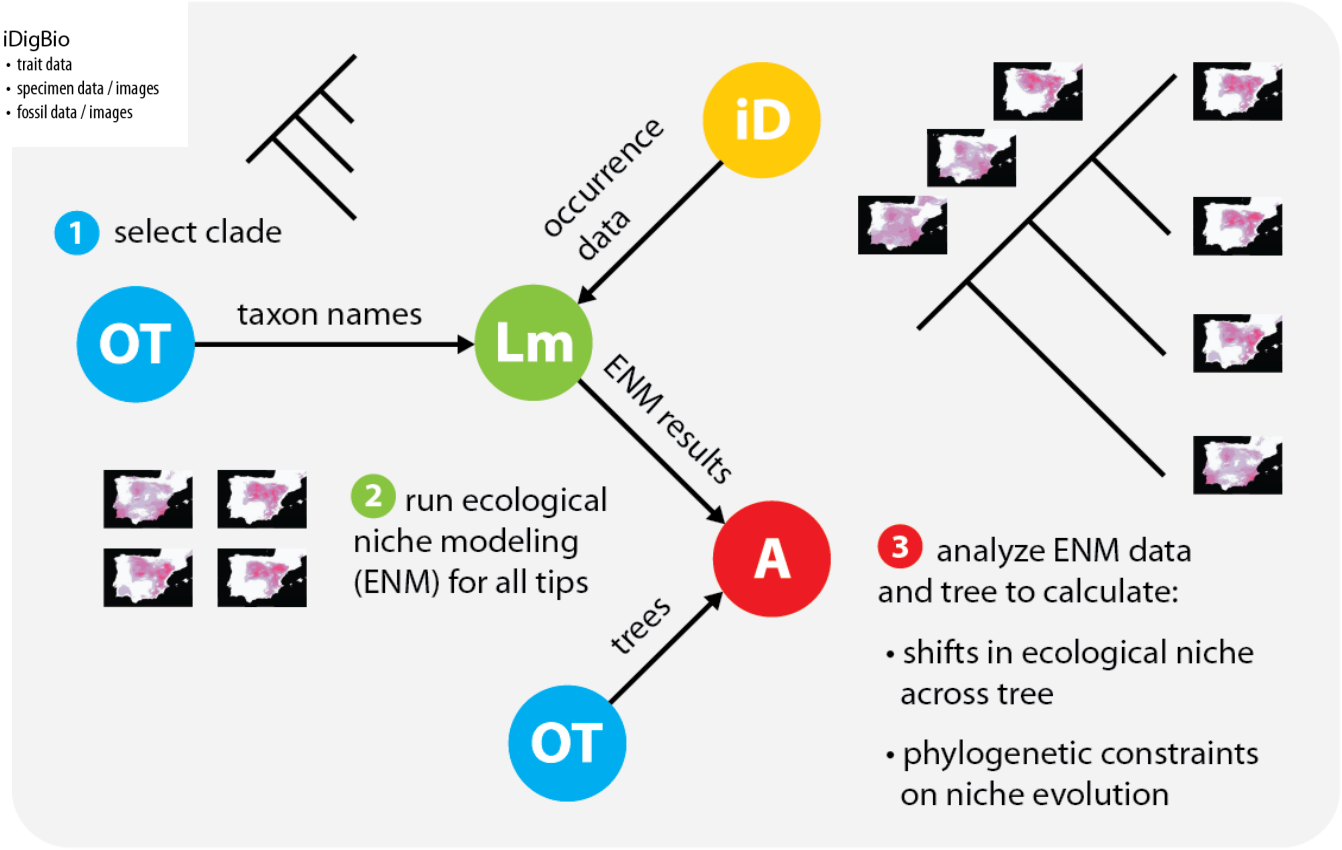
RESOURCES:

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

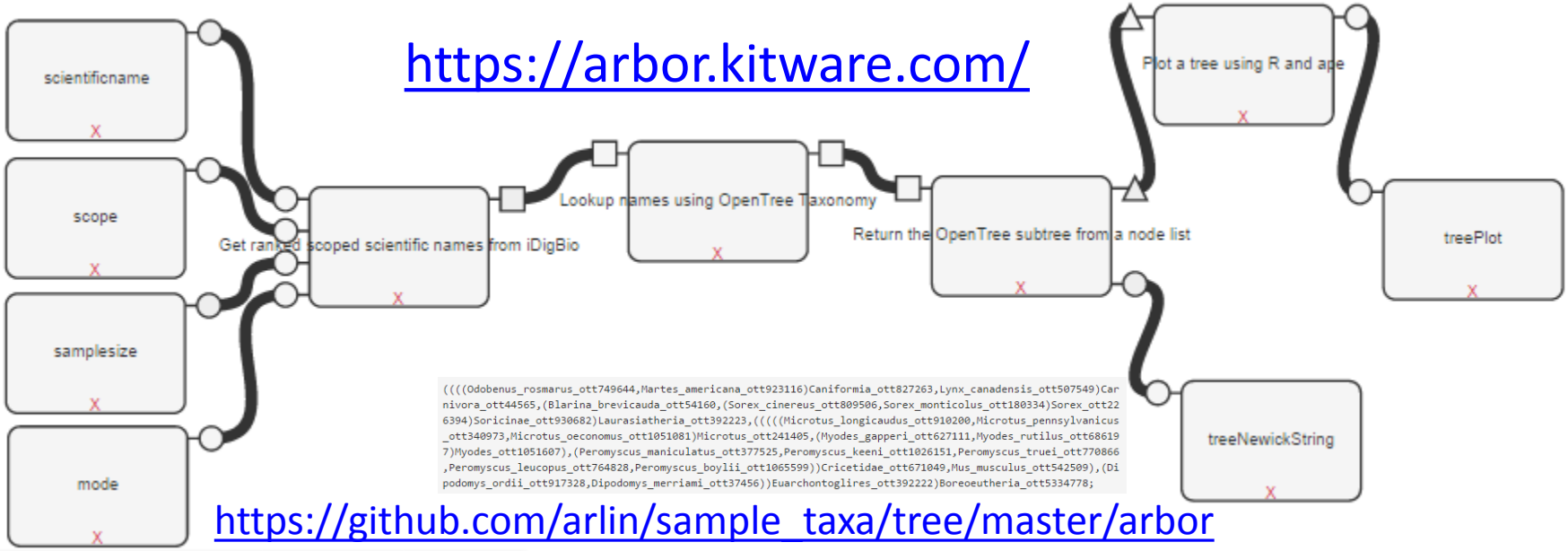
A Arbor
- evolutionary models
- comparative methods
- visualization

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

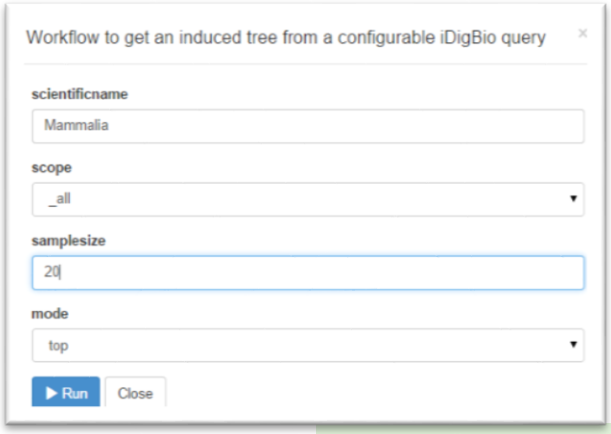
iD iDigBio
- trait data
- specimen data / images
- fossil data / images

Arbor, OpenTree, and iDigBio

https://arbor.kitware.com/



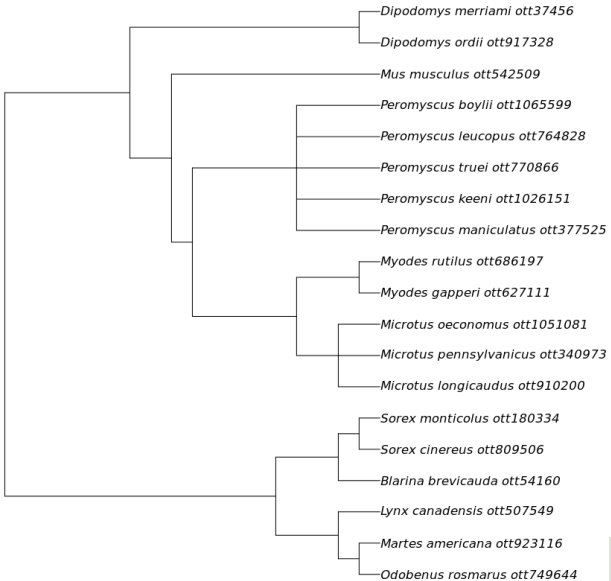
https://github.com/arlin/sample_taxa/tree/master/arbor




```
{
    "query": {
        "query_string" : {
            "default_field" : "order",
            "query" : "rodentia"
        }
    },
    "aggregations": {
        "my_agg": {
            "terms": {
                "field": "scientificname",
                "size": 100
            }
        }
    }
}
```

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]



A. Matsunaga

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:

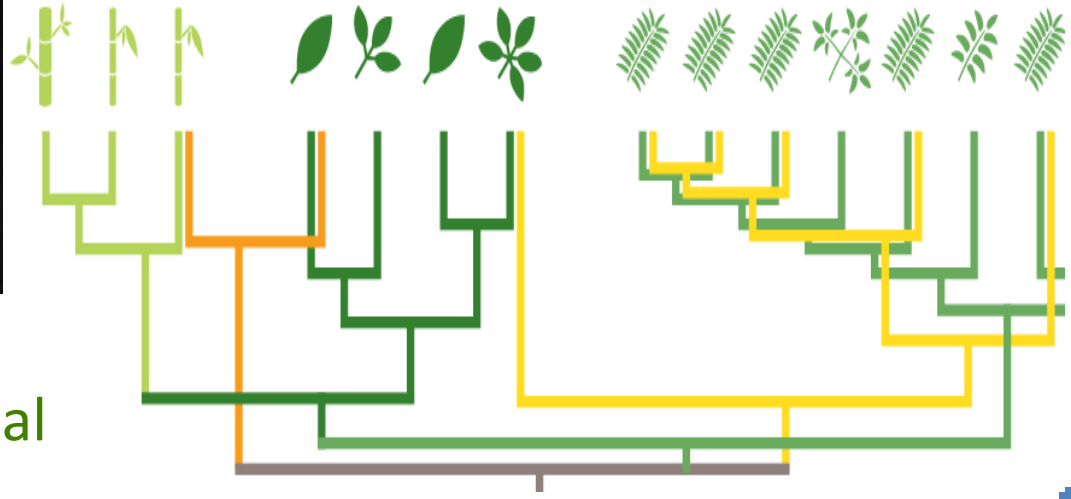
Systematics and Taxonomy

Carl Linné, aka Carolus Linnaeus

Linnea (twinflower)

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

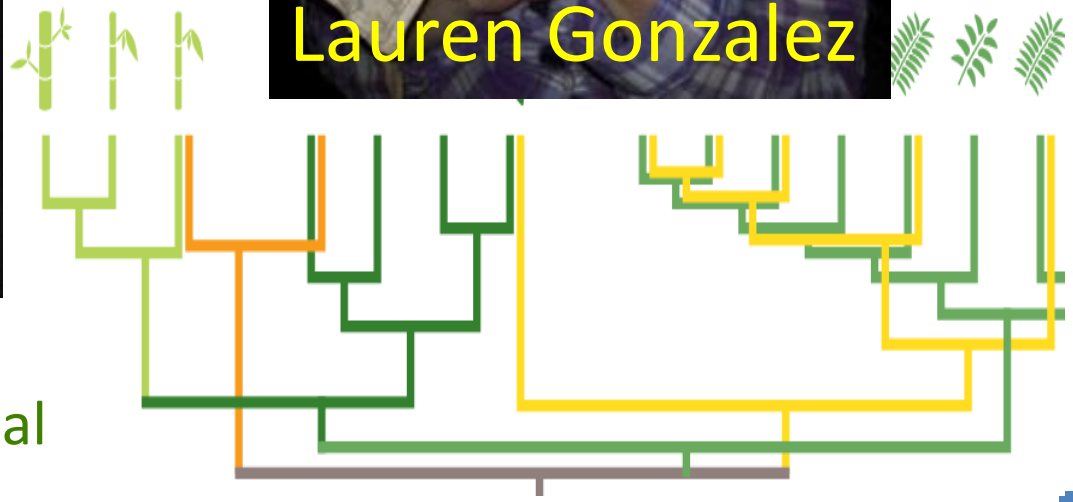
Using Images to Infer Functional Traits



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

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.

Linking to GenBank

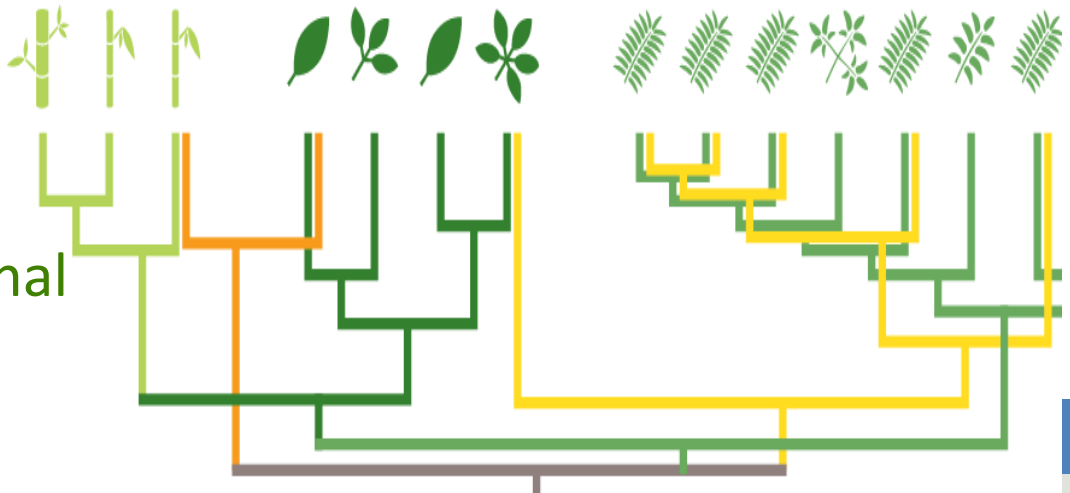
Welcome to NCBI

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