

Digitized Data in Biodiversity Research *Summit 2015*

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



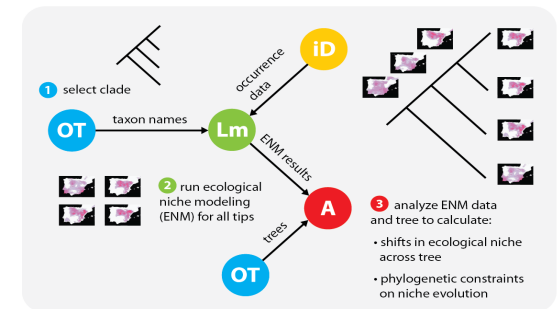
What Can We Do with Specimen Data?

- Monitor shifts in biodiversity through time
- Track invasive species
- Ecological Niche/Species Distribution Modeling
- Track phenological shifts
- Integrate with evolutionary history
- Past movements and climate change
- Landscape genetics



Enabling Research Using Specimen Data

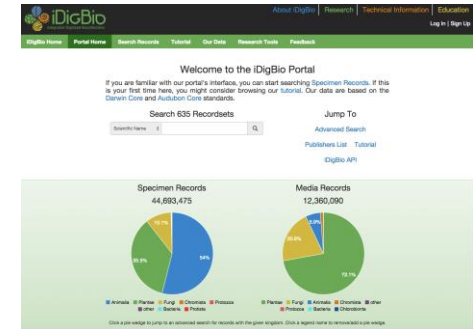
- Making data available through the portal
 - >46 M specimen records
 - Augmented records
 - Diverse means of searching data
- Developing examples and research workflows
 - Florida plant diversity, polyploidy, evolution of breeding systems, paleoclimate and glacial refugia
 - Conceptual workflows
 - Scripts
 - R packages
- Facilitating research
 - Developing methods
 - Aggregating tools
 - ‘workbench’ model



Enabling Research Using Specimen Data

✓ Making data available through the portal

- >46 M specimen records
- Augmented records
- Diverse means of searching data

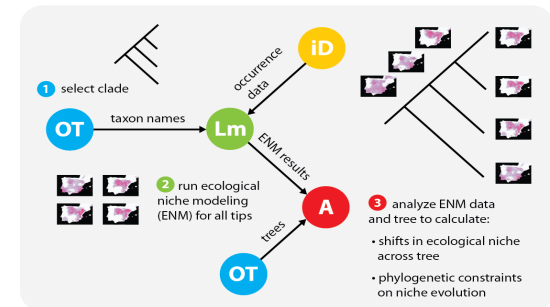


✓ Developing examples and research workflows

- Florida plant diversity, polyploidy, evolution of breeding systems, paleoclimate and glacial refugia
- Conceptual workflows
- Scripts
- R packages

➤ Facilitating research

- Developing methods
- Aggregating tools
- ‘workbench’ model

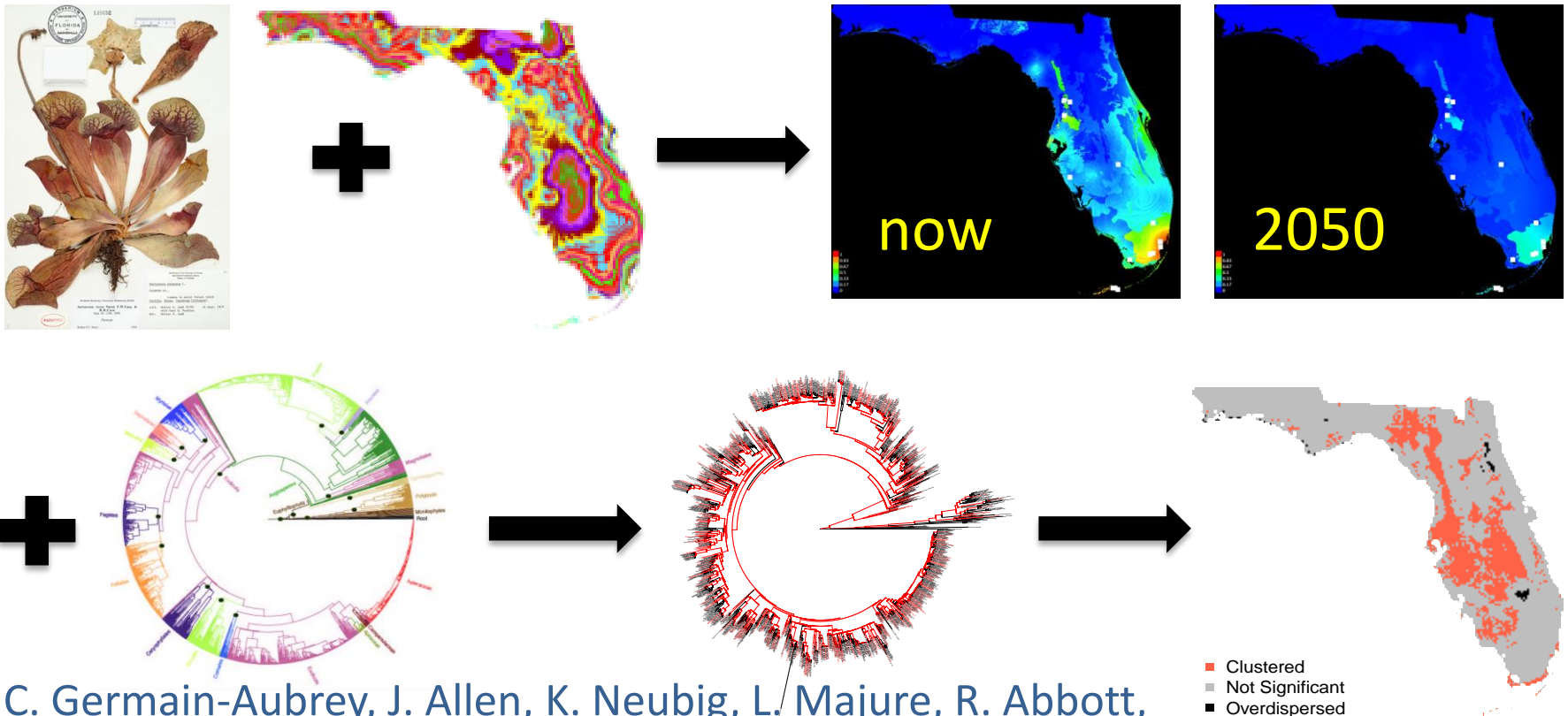


Facilitating Research

- Further development of methods
- Aggregating tools
- ‘Workbench’ – enhanced workflows
- Research applications workshops

Facilitating Research

- Further development of methods



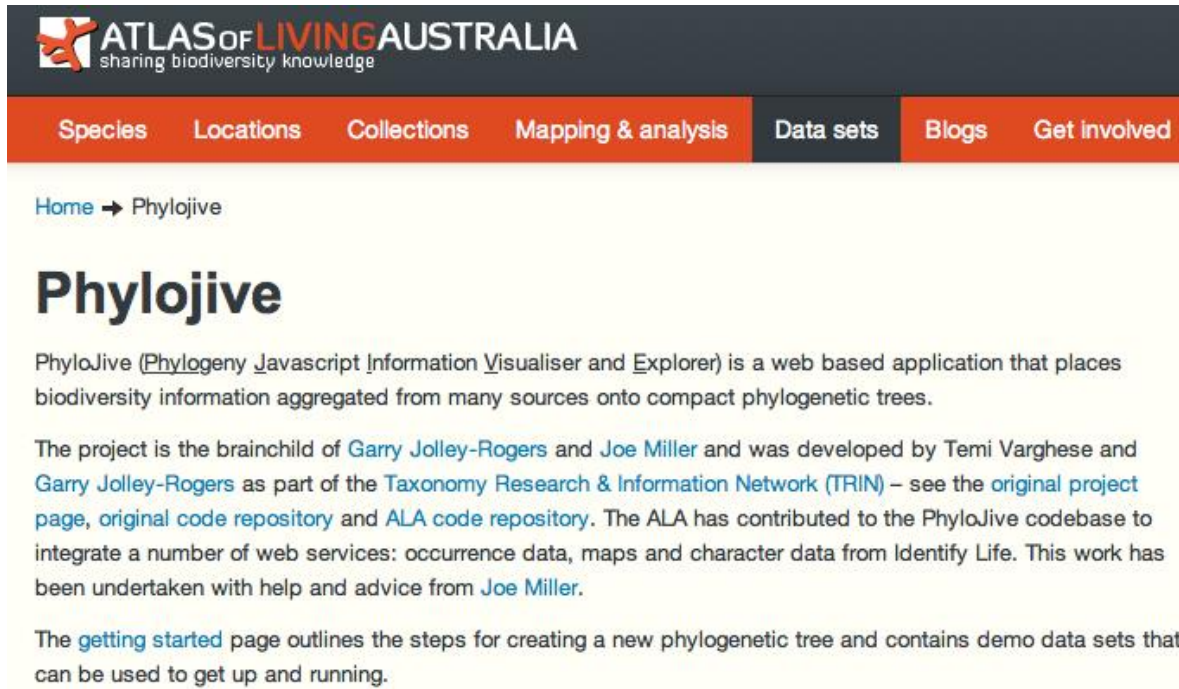
C. Germain-Aubrey, J. Allen, K. Neubig, L. Majure, R. Abbott,
M. Whitten, J. M. Ponciano, B. Mishler, S. Laffan, T. Lemay, R. Guralnick, D. Soltis

Facilitating Research

- Further development of methods
- **Aggregating tools**
- ‘Workbench’ – enhanced workflows
- Research applications workshops

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](#).

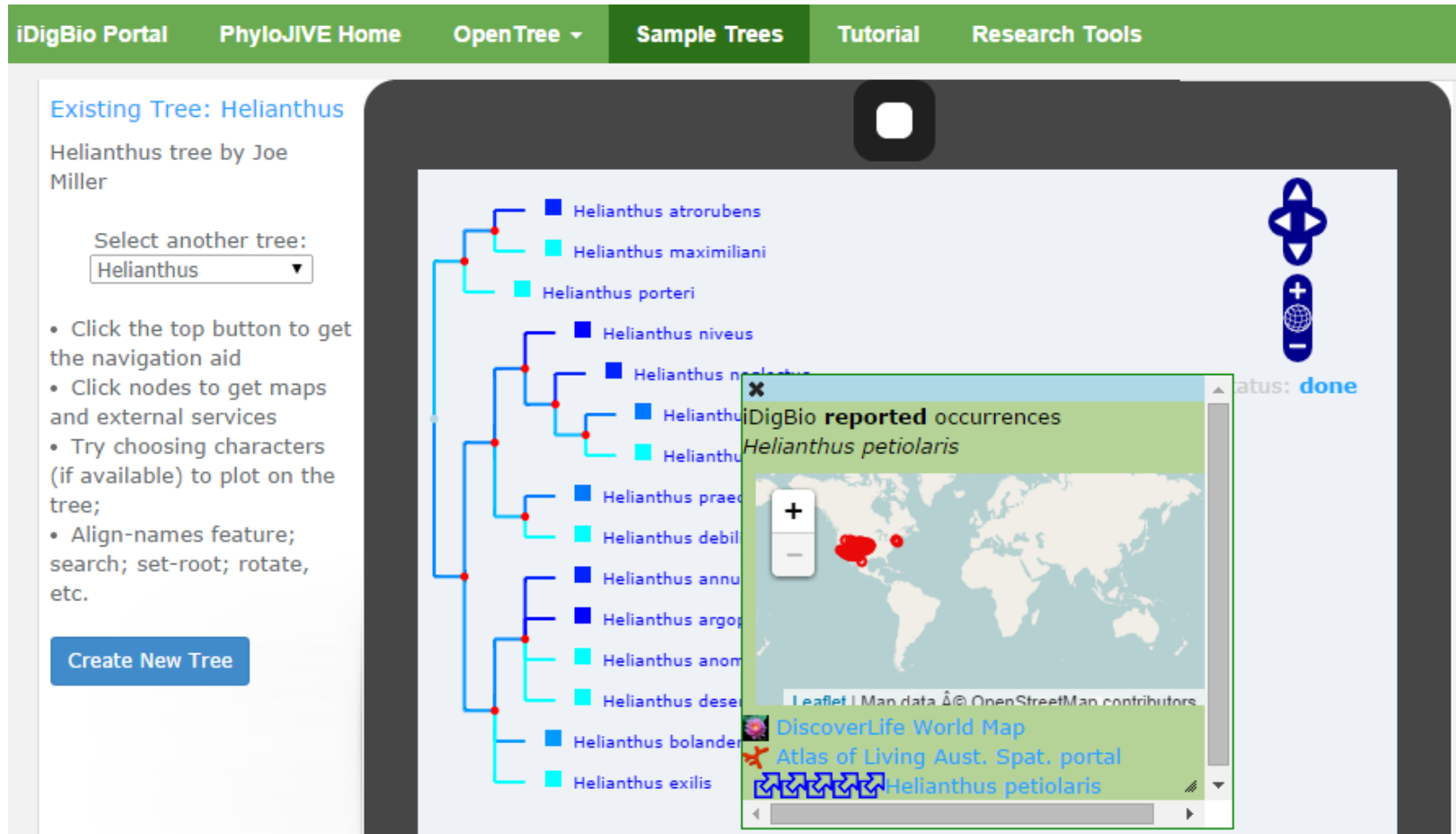
The [getting started](#) page outlines the steps for creating a new phylogenetic tree and contains demo data sets that can be used to get up and running.



A. buxifolia

Source: Australian Plant
Image Index Image by: Macd

PhyloJIVE instance in iDigBio

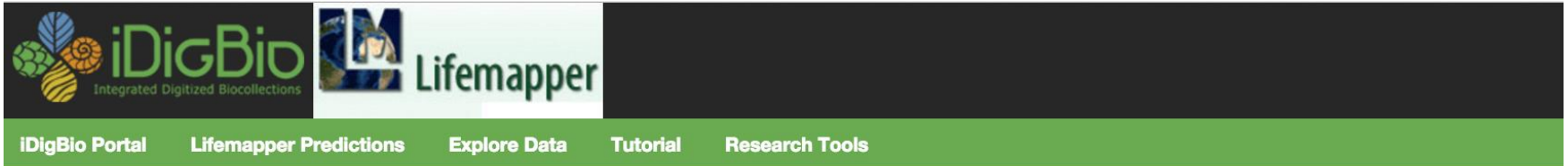


The screenshot shows the iDigBio Portal with the PhyloJIVE Home page. The navigation bar includes: iDigBio Portal, PhyloJIVE Home, OpenTree, Sample Trees, Tutorial, and Research Tools. The main content area is titled "Existing Tree: Helianthus" and shows a tree by Joe Miller. A dropdown menu allows selecting another tree from the "Helianthus" genus. A list of species is displayed on the right, including Helianthus atrorubens, Helianthus maximiliani, Helianthus porteri, Helianthus niveus, Helianthus mollis, Helianthus scaberrimus, 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.

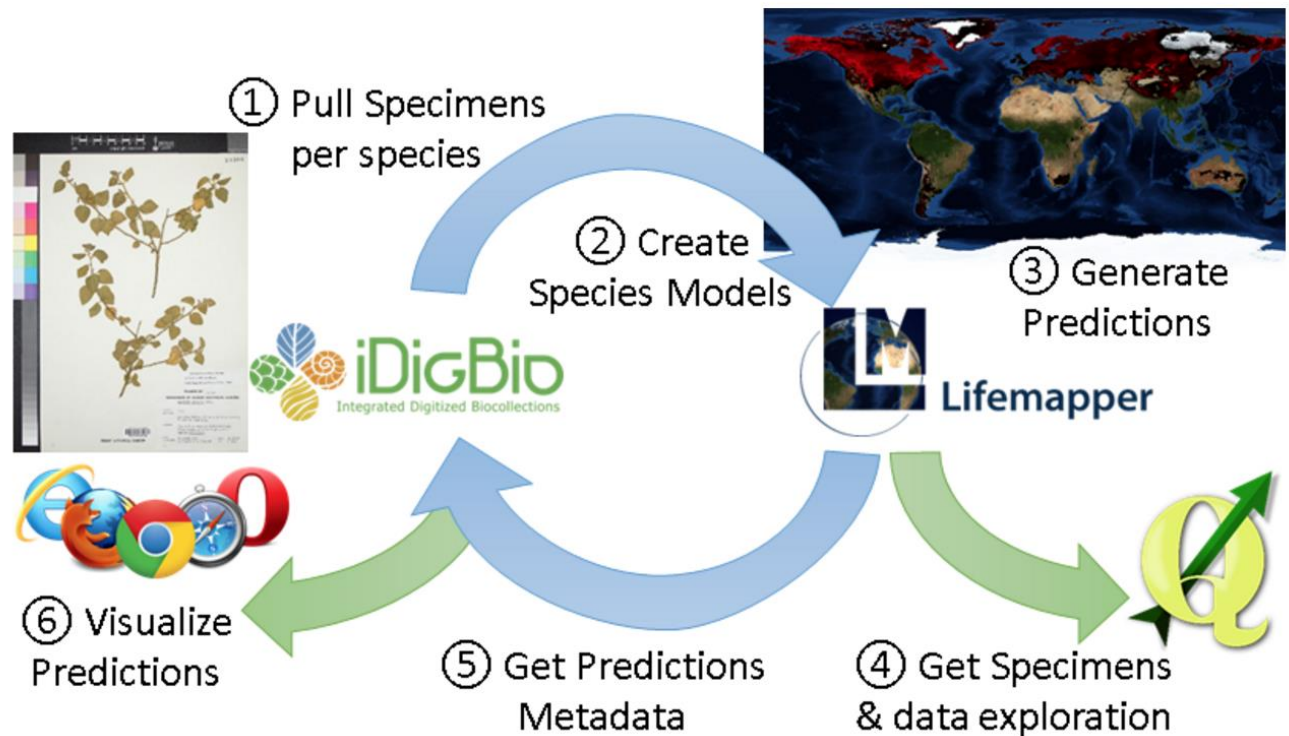
- Developed by Garry Jolley-Rogers, Joe Miller, and Temi Varghese
- Integrates biodiversity data with phylogeny
- <http://phylojive.acis.ufl.edu/>

A. Matsunaga

iDigBio – Lifemapper Instance



iDigBio-Lifemapper workflow



A. Matsunaga

A. Stewart

<http://lifemapper.acis.ufl.edu/>

Facilitating Research

- Further development of methods
- Aggregating tools



Discovery Environment

Apps

Apps Workflow Search Apps

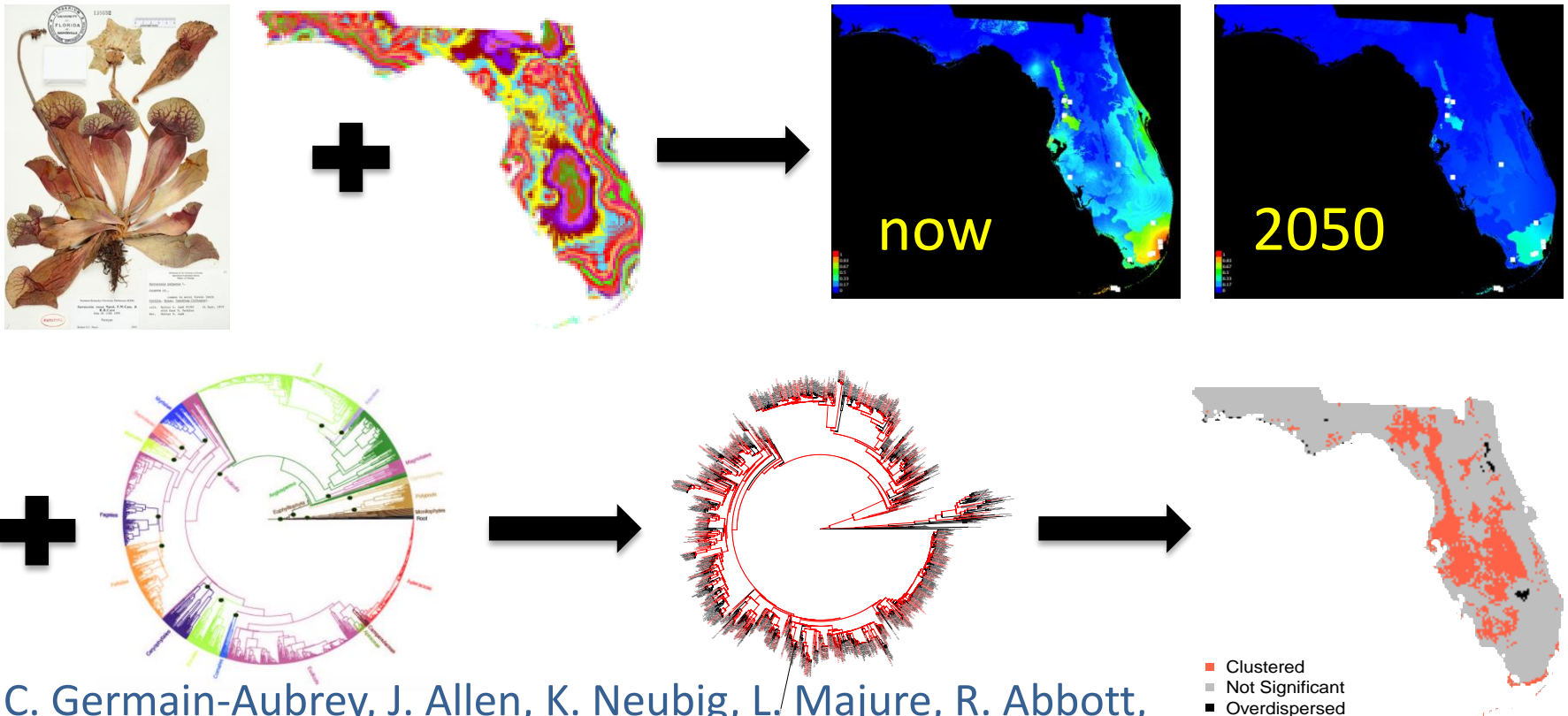
Category	Name	Integrated by	Rating
Public Apps >> Beta			
microRNA Analysis (1)	add column to file	Laura Gardiner	★★★★★ (1)
Molecular Structure (0)	add column to file2	Laura Gardiner	★★★★★ (1)
Motif Identification (2)	Add GO to Blastp-unipr...	rogerab	☆☆☆☆☆ (0)
NGS (124)	appendUnclustered	Jeremy DeBarry	☆☆☆☆☆ (0)
Phylogenetics (14)	bam2fastq_unmapped	Dongliang Wu	☆☆☆☆☆ (0)
Community Ecology (0)	Blast2Seqs	Sabarinath Subra...	☆☆☆☆☆ (0)
Comparative Methods (1)	Blastp a subset of uniprot	rogerab	☆☆☆☆☆ (0)
Evolutionary Models (2)	Bowtie-2.2.1--Build-and...	Ryan Joynson	☆☆☆☆☆ (0)
Tree Building (10)	Bowtie2Build-2.2.4 inde...	rogerab	☆☆☆☆☆ (0)
Utility Tools and Scripts (1)	btrim	Senthil Subraman...	★★★★★ (2)
QTL and GWAS (21)			
Sequence Alignments (3)			
VCF and GFF (3)			
High-Performance Computing (-1)			

Facilitating Research

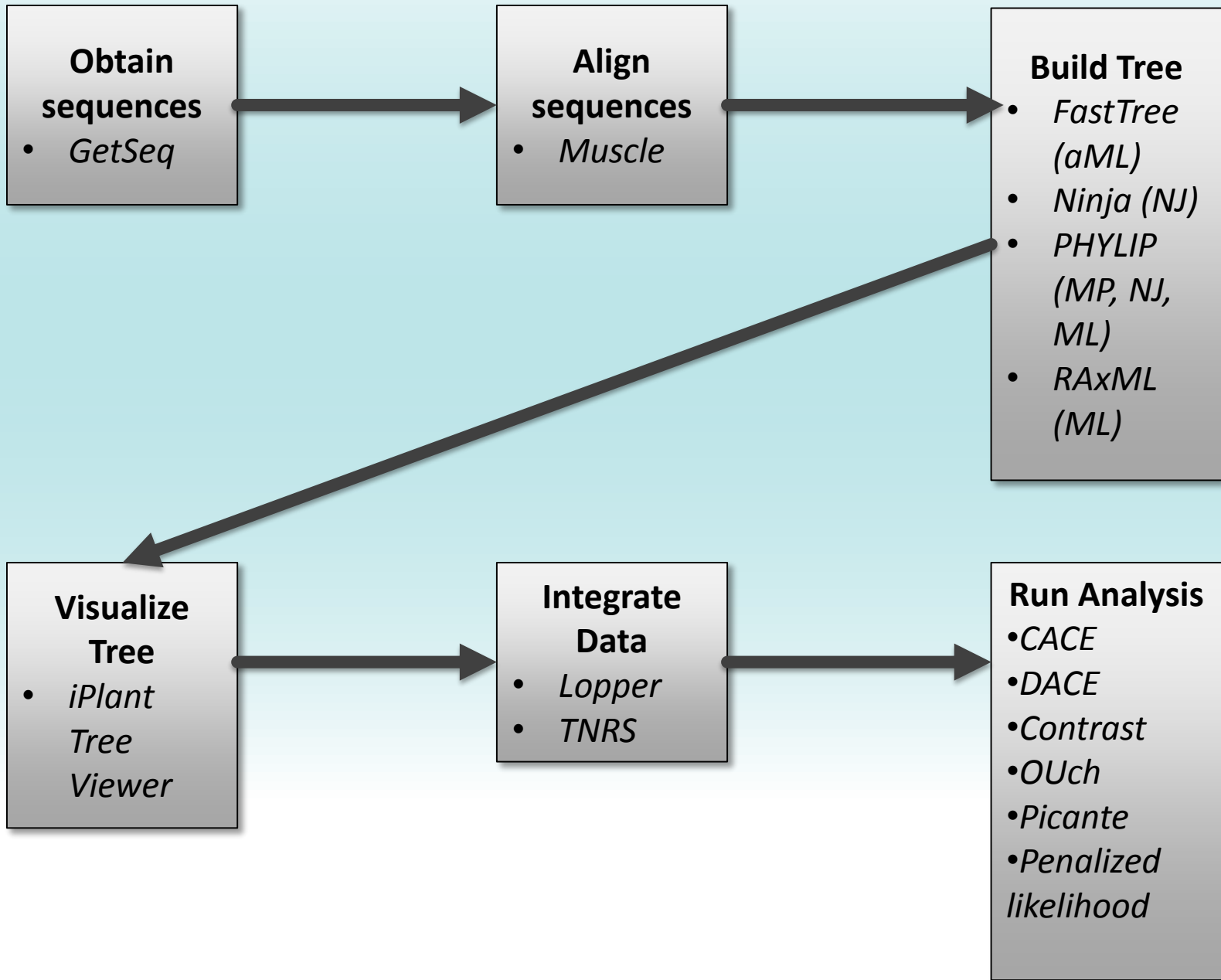
- Further development of methods
- Aggregating tools
- **‘Workbench’ – enhanced workflows**
- Research applications workshops

Facilitating Research

• Toward a Biodiversity Workbench



C. Germain-Aubrey, J. Allen, K. Neubig, L. Majure, R. Abbott,
M. Whitten, J. M. Ponciano, B. Mishler, S. Laffan, T. Lemay, R. Guralnick, D. Soltis



Manage Data

Import More Actions Refresh

- nmatasci
 - analyses
 - campa_CACE
 - campa_CACE-
 - campa_fast
 - campa_fast-1
 - campa_fast_lo
 - campa_getseq**
 - logs
 - campa_lopper
 - campa_muscle
 - campa_ninja
 - campa_ninja-1
 - campa_rax
 - campa_rax-1
 - logs

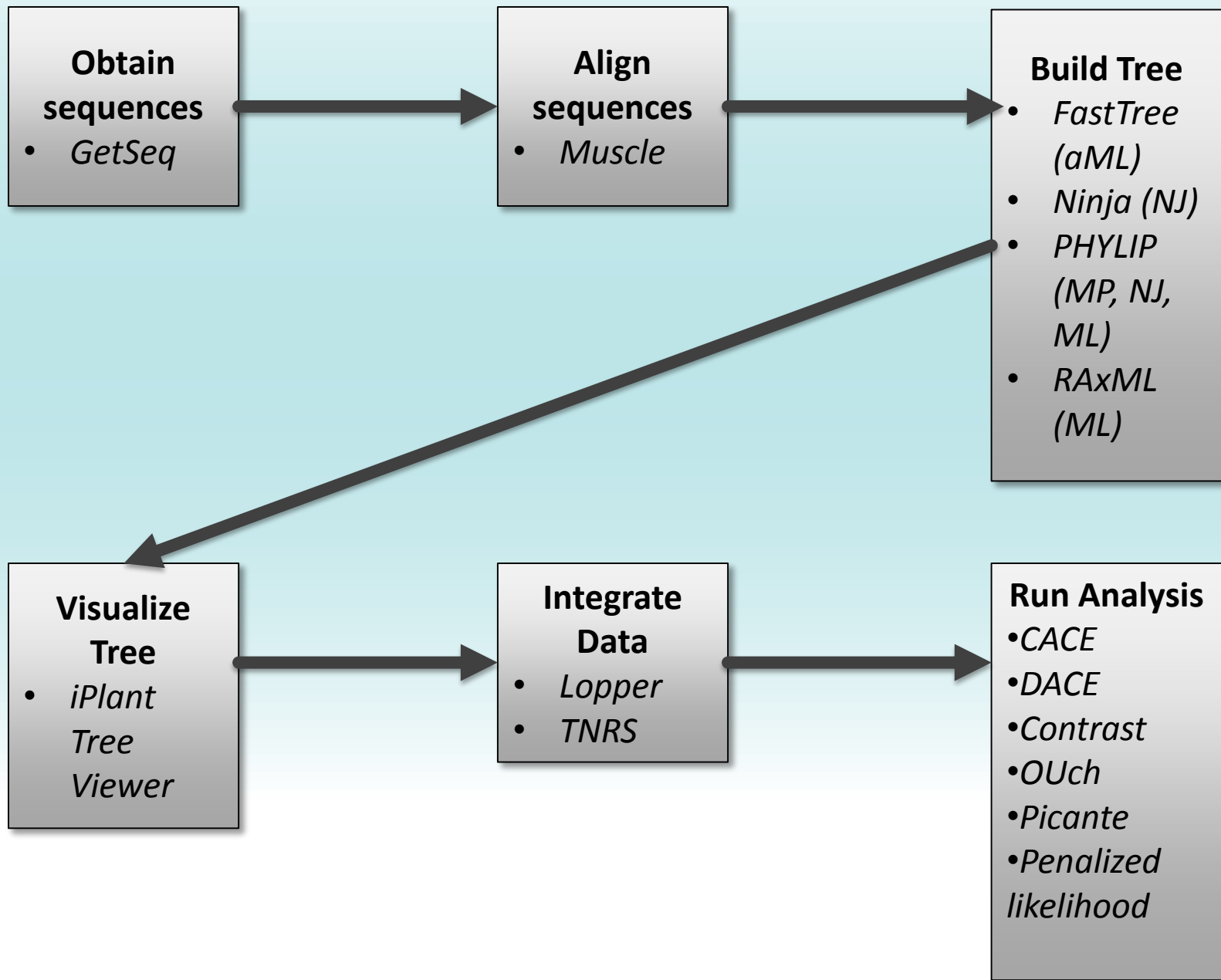
sequences.fa

Preview

Only 8K of the selected file is displayed. For full view of data, please download file.

```
GCTAAAGCCTTACGTATGCTGGTGGAGATCATATTCACCTCCGGGACCGTAGTAGGTA  
CTTGAAGGGAAAGAGAGATCACTTTGGGCTTTGTTGATTTACTGCGCGATGATTTTGT  
GAAAAAGATCGAAGTCGCGGATTTATTTCACTCAAGATTGGGTCTCTACCAGGTGT  
CTGCCGTGGCTT CAGGCGGATT CACGTTTGGCATATGCGCTGCTGACTGAGATTT  
GGGGATGATTCGCTACTACAGTTTCGGGGGAGGAACCTTAGGCCATCCTTGGGAAATGCA  
CCCGTGCGGTAGCTAATCGAGTAGCTCTAGAAGCATGTGTACAAGCCGTAATGAGGGA  
CGTGATCTTCTGCTGAGGGTAATGAAATATCGGAAGGCTGGCAAATGGAGCTCTGAG  
CTATTTCTGCTTGTGAGGTATGGAAAGAGATCAAATTTGAGTTTGCCGCAATGG  
>405146_Clermontia_akeana  
TGTCACCACAACAGAGACTAAGCAAATGTTGGCTTCAAAGCTGGTGTAAAGAGTATA  
AATTAAGCTTATTACTCCTGAGTATCAAACCAAGGATGACTGATTTTGGCAGCTTCC  
GAGTAACCTCCTCAACCCGGAGTTCCGCCTGAAGAAGCAGGGGCTGCAGTAGCTGCCGAAT  
CTTCGACTGGTACATGGACAACGTGTGGACCGATGGACTTACGAGTCTTGATCGTTACA  
AAGGGCGATGCTATCACATCGAGCCCGTTGCCGGAAGAAGAACTCAATTTATTGCTTATG  
TAGCTTACCCATTAGACCTTTTGAAGAAGGTTCTGTTACTAACCTGTTACTTCCATTG  
TGGGTAATGTATTTGGATTTAAGCACTGCGTCTACGCTCGAAGATTTGCGAATCC  
CGCTTCCGATGTTAAAACCTTTCCAGGGCCCGCTCATGGCATCCAAGTTGAACGAGATA  
AATTGAACAAGTATGGTGGTCCCCTGTTGGGATGTACGATTAACCTAAATTTGGGGTTAT  
CTGCTAAAACTACGGCAGAGCAGTTTATGAATGCTCTTGGTGGCCTTGATTTTACCA  
AAGATGATGAGAAGCTGAACTCCCAACCTTTATGCGTTGGAGAGATCGGTTGATTTTT  
GTGCCGAAGCTATTTAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA  
ATGCTACTGCGAGTACAACGAAGAAATGATGAAAAGGGCTGATTTGCCAGAGAATTGG  
GCGTTCCATCGTAATGCATGACTACCTAACAGGGGGTTCACTGCAAACTAGCTTGG  
CTCACTATGCCGAGATAATGGTCTACTTCTCACATCCACCCGCGCATGCATGCAGTTA  
TTGATAGACAGAAGAATCATGGTATGCACTTTGCTGTACTAGCGAAAGCCTTACGATGT  
CAGGTGGAGATCATATTCACGCCCGTACCGTAGTAGGAAACTGGAAGGGGAAAGAGAGA  
TCACTTTGGGCTTTGTTGATTTACTGCGCGATGATTTATTGAAAAAGATCGAAGTCGTG  
GTATTTATTTCACTCAAGATTGGGTCTCTACCAGGTGTTCTGCCCGTGGCTT CAGGCG  
GTATTCACGTTTGGCATATGCGCTGCTTACCCGAGATCTTTGGGGATGATTCGCTACTAC  
AATTCGGTGGAGGAACCTTAGGACATCCGTGGGAAATGCACCCGGTCCGCTAGCTAATC  
GAGTCGCGCTAGAAGCATGTGTACAAGCCGTAATGAGGGACGTGATCTTCTGCTGAGG  
GTAATGAAATTAACGTAAGGCTGCCACATGGAGTCTGAGCTATCTGCTGTTGTGAGG  
TATGGAAAGAGATCAAATTTGAGTTTGCCGCAATGGATACCTTTATAA  
>87137117_Cyanea_kuhihewa  
CGAAAGAAACCAAAATATCAAATTTTTCGCAAAATTAACAATTCAAAATAAATGTCC  
GATAGCAAGGTGATCGGTTAATGAAATAAGAACCAGGAGTAGCACGTGATTTCTGAGT  
ACCATTCAACCGAATCAATTTGAAATGTTTACTTATTTCGATGAGTGAAGTTTCAAGTTCA
```

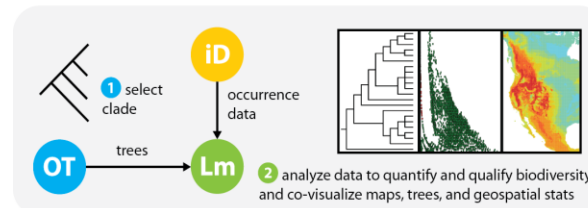
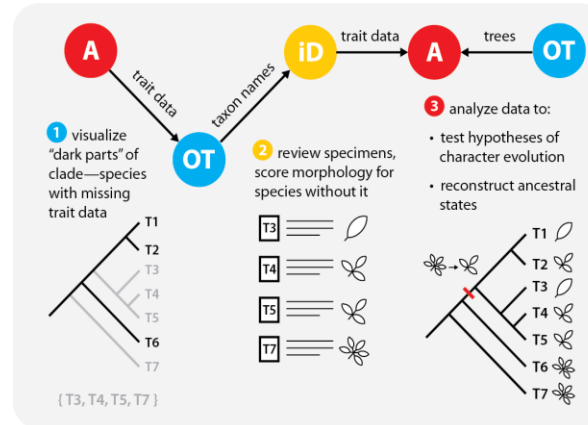
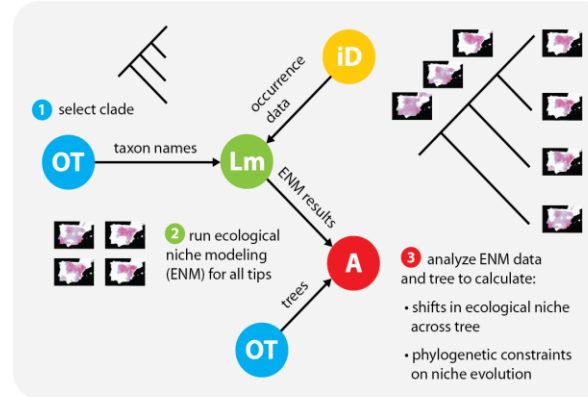




Connecting Trees, Specimens, Tools

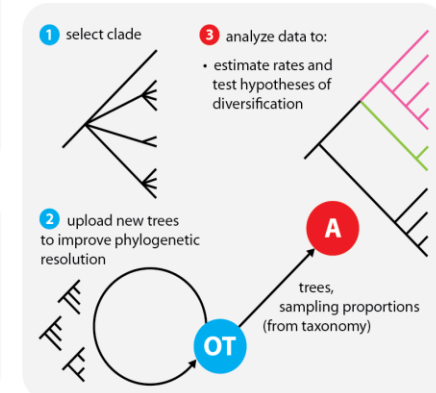
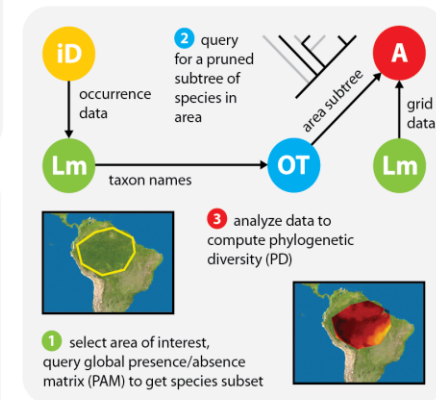


EXAMPLE WORKFLOWS:



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

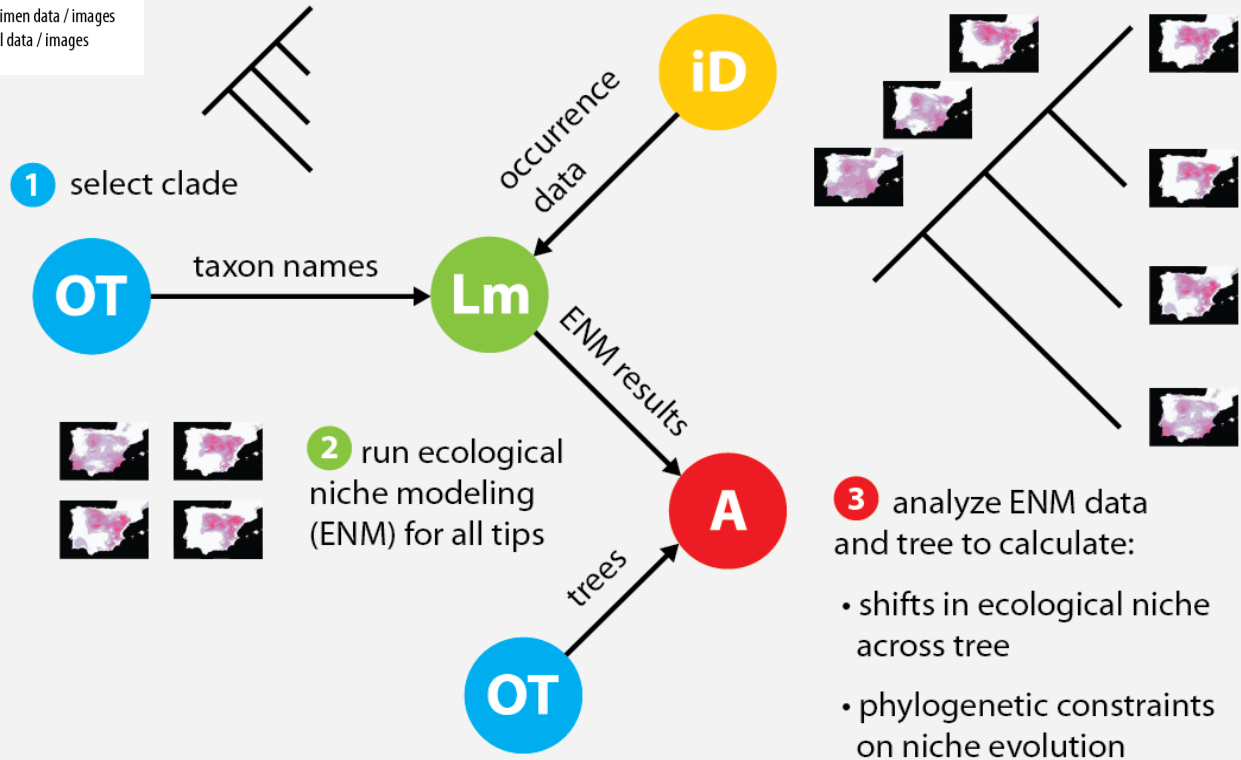
RESOURCES:

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: BiotaPhy Project
Connecting resources
to enable large-scale biodiversity analyses

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

RESOURCES:



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



Arbor
• evolutionary models
• comparative methods
• visualization



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



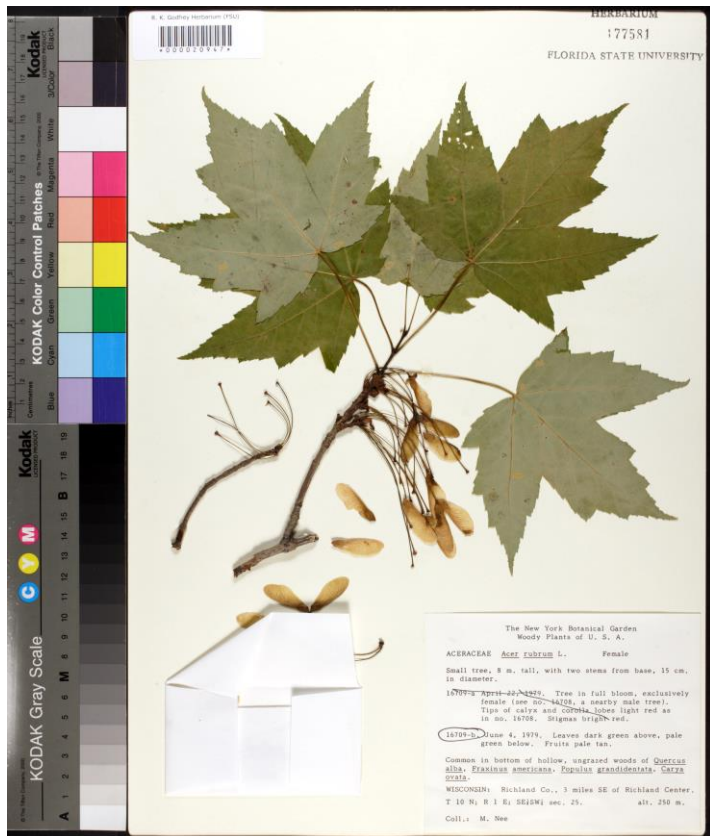
iDigBio
• trait data
• specimen data / images
• fossil data / images

Images as Sources of Trait Data



Photosynthetic Pathway
 Respiration Leaf Area N fixation 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

Images as Sources of Trait Data



Photosynthetic Pathway
 Respiration Leaf Area N fixation 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

Computer Science/Image Analysis

Facilitating Research

- Further development of methods
- Aggregating tools
- ‘Workbench’ – enhanced workflows
- **Research applications workshops**

Community input: topics, volunteers

Many Uses for Specimen Data in Research

- Connections to other resources, e.g. GenBank
- Ecological Niche/Species Distribution 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???

- *Discussions: Research Tool Development*
Data Use in Research and Education
- *Research Applications Working Group*

Thank you!



Charlotte Germain-Aubrey



Blaine Marchant



Lauren Gonzalez



Greg Stull



Andre Naranjo

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>