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?

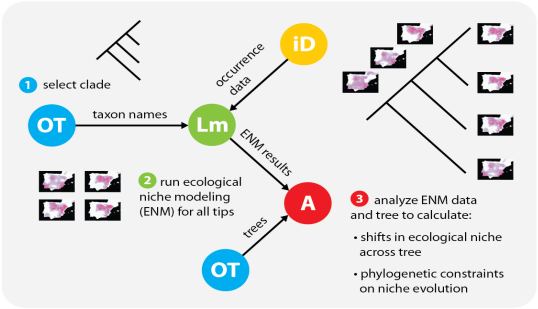
- Monitor shifts in biodiversity through time
- Track invasive species
- Ecological Niche/Species Distribution Modeling
- Track phenological shifts
- Integrate with evolutionary history
- Past movements and climate change
- Landscape genetics




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

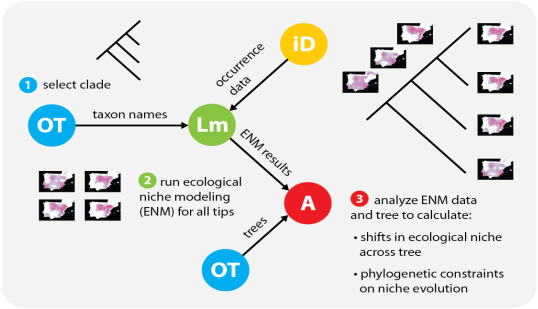


- Facilitating research
  - Developing methods
  - Aggregating tools
  - ‘workbench’ model

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



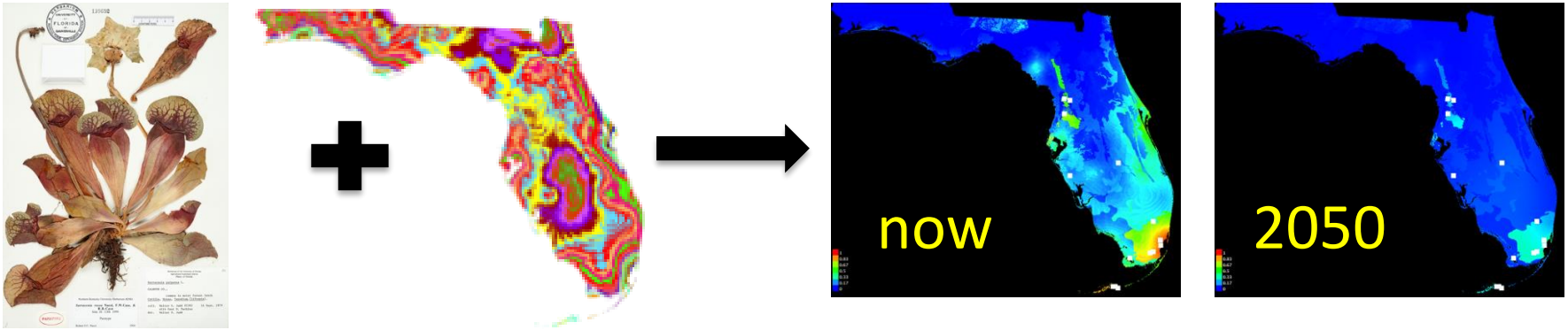
- ➢ Facilitating research
  - Developing methods
  - Aggregating tools
  - 'workbench' model

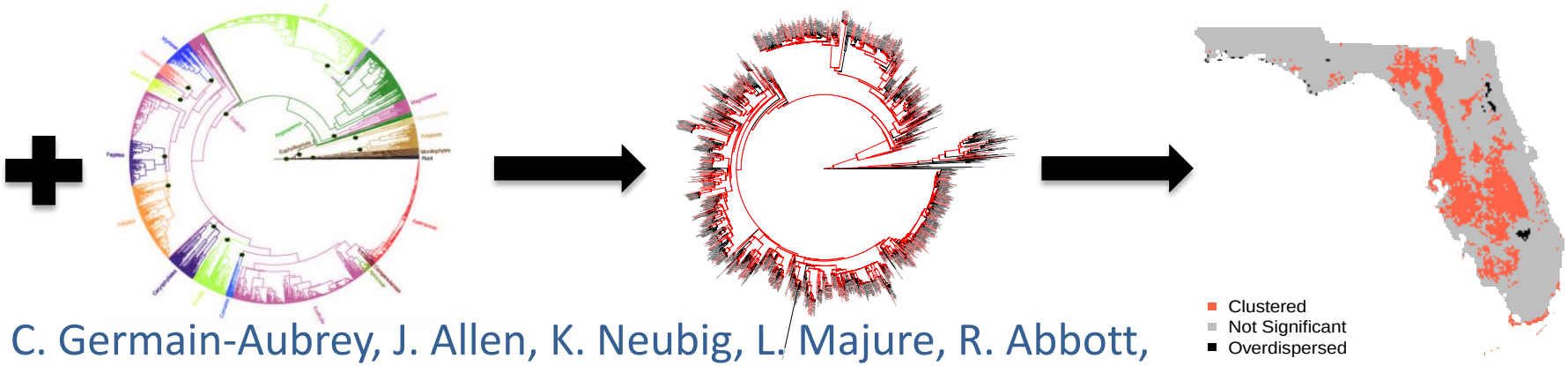
Facilitating Research

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

Facilitating Research

- **Further development of methods**





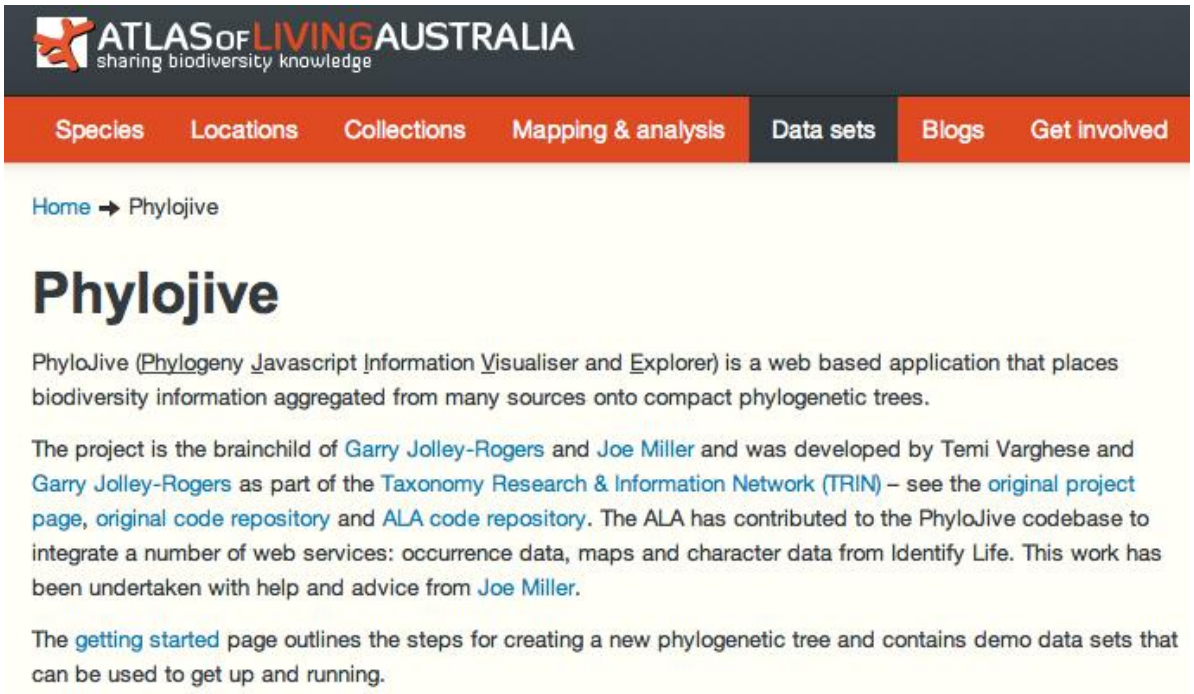
C. Germain-Aubrey, J. Allen, K. Neubig, L. Majure, R. Abbott, M. Whitten, J. M. Ponciano, B. Mishler, S. Laffan, T. Lemay, R. Guralnick, D. Soltis

Facilitating Research

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

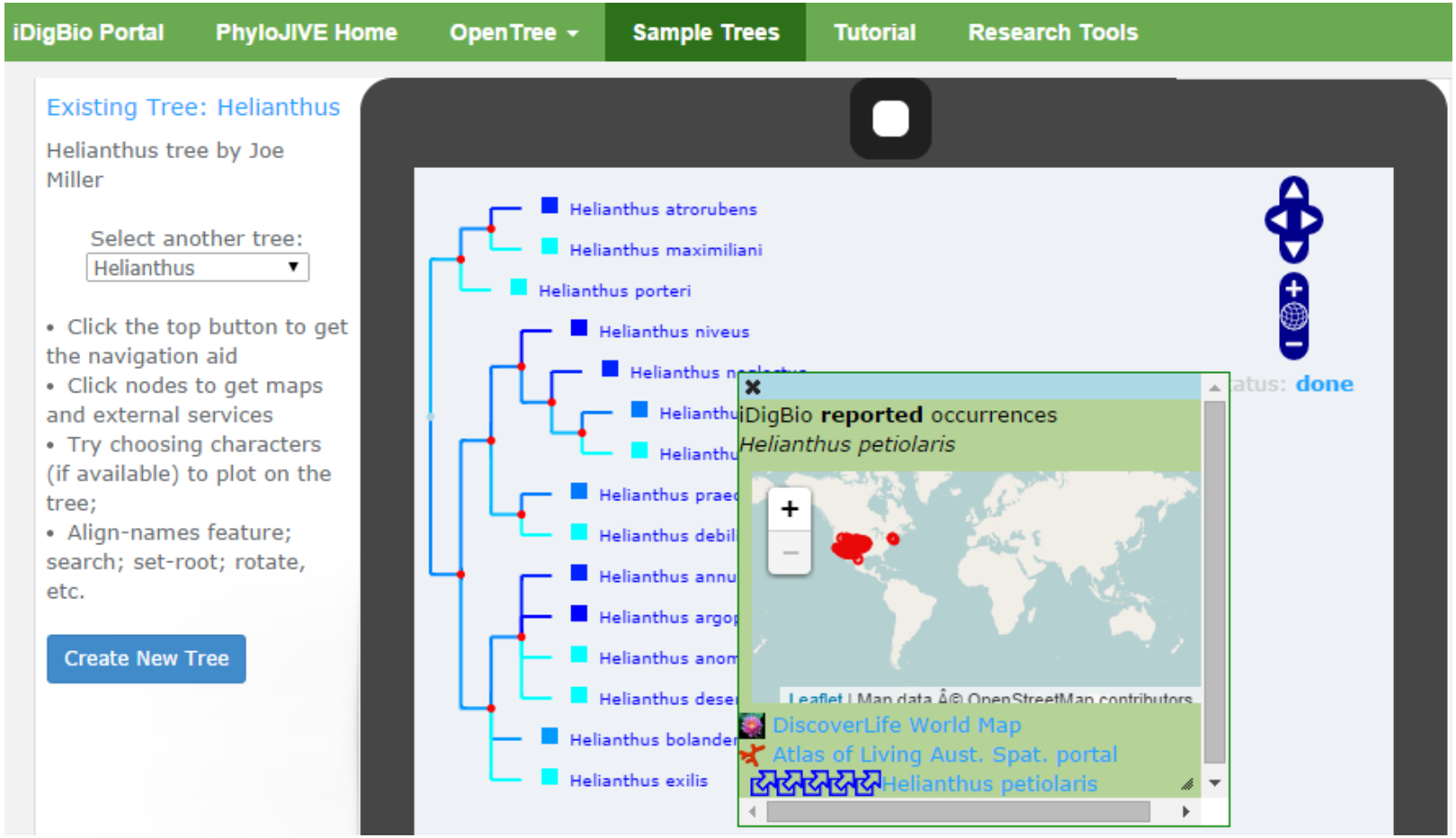
PhyloJIVE

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



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

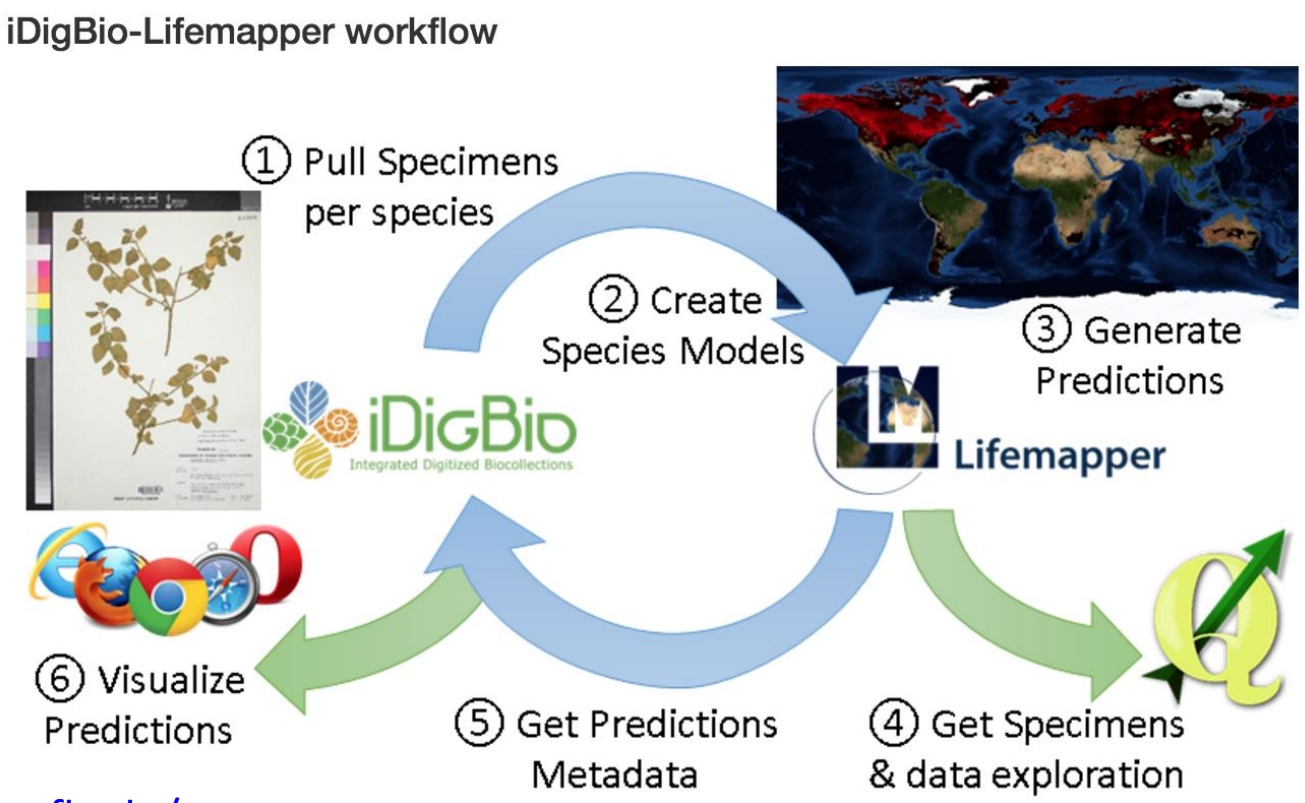
PhyloJIVE instance in iDigBio



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

A. Matsunaga

iDigBio – Lifemapper Instance



A. Matsunaga
A. Stewart

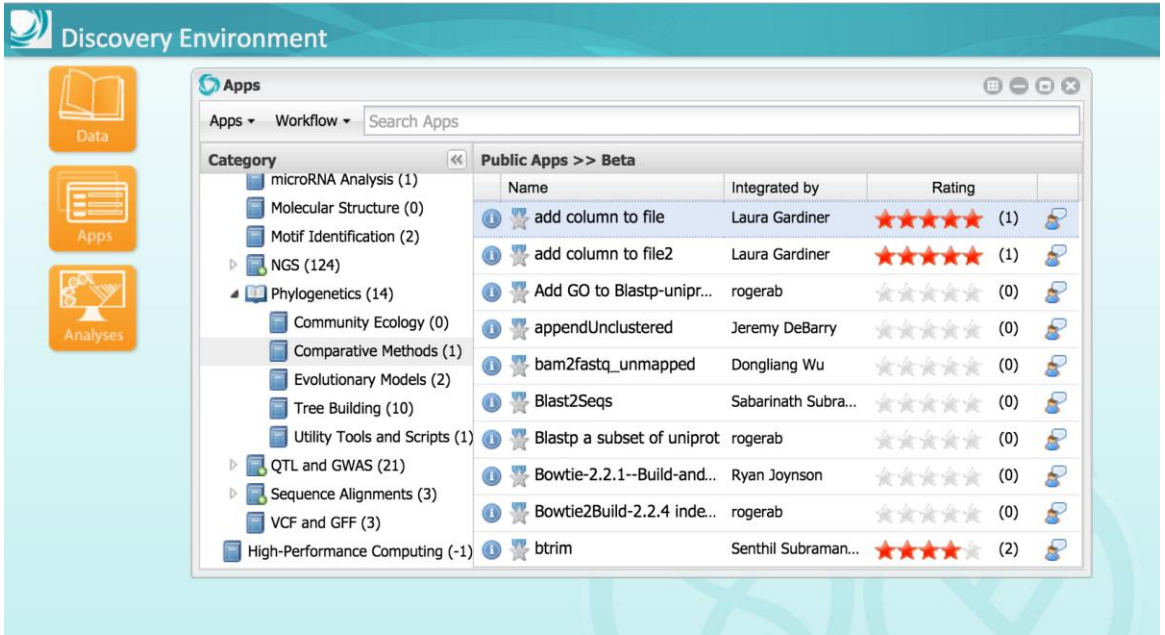
http://lifemapper.acis.ufl.edu/

Facilitating Research

- Further development of methods

- Aggregating tools

iPlant Collaborative™

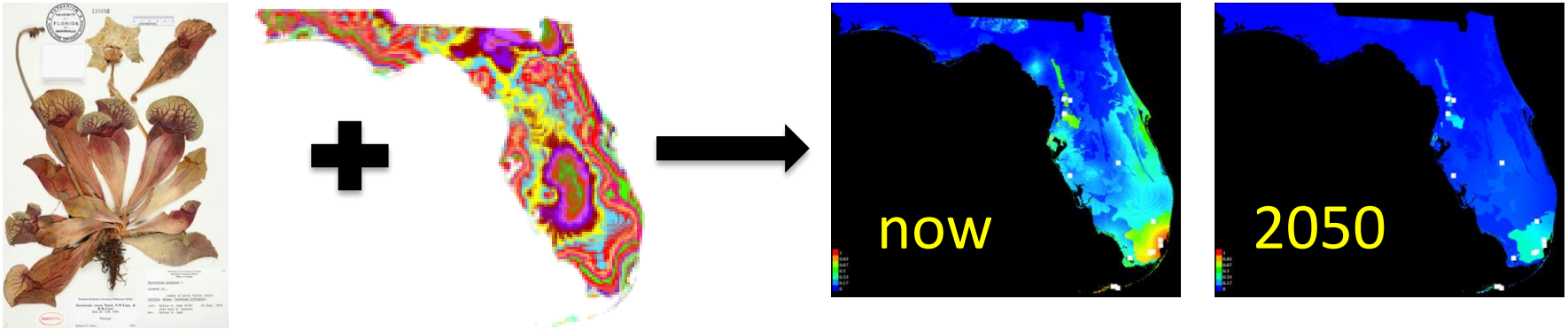


Facilitating Research

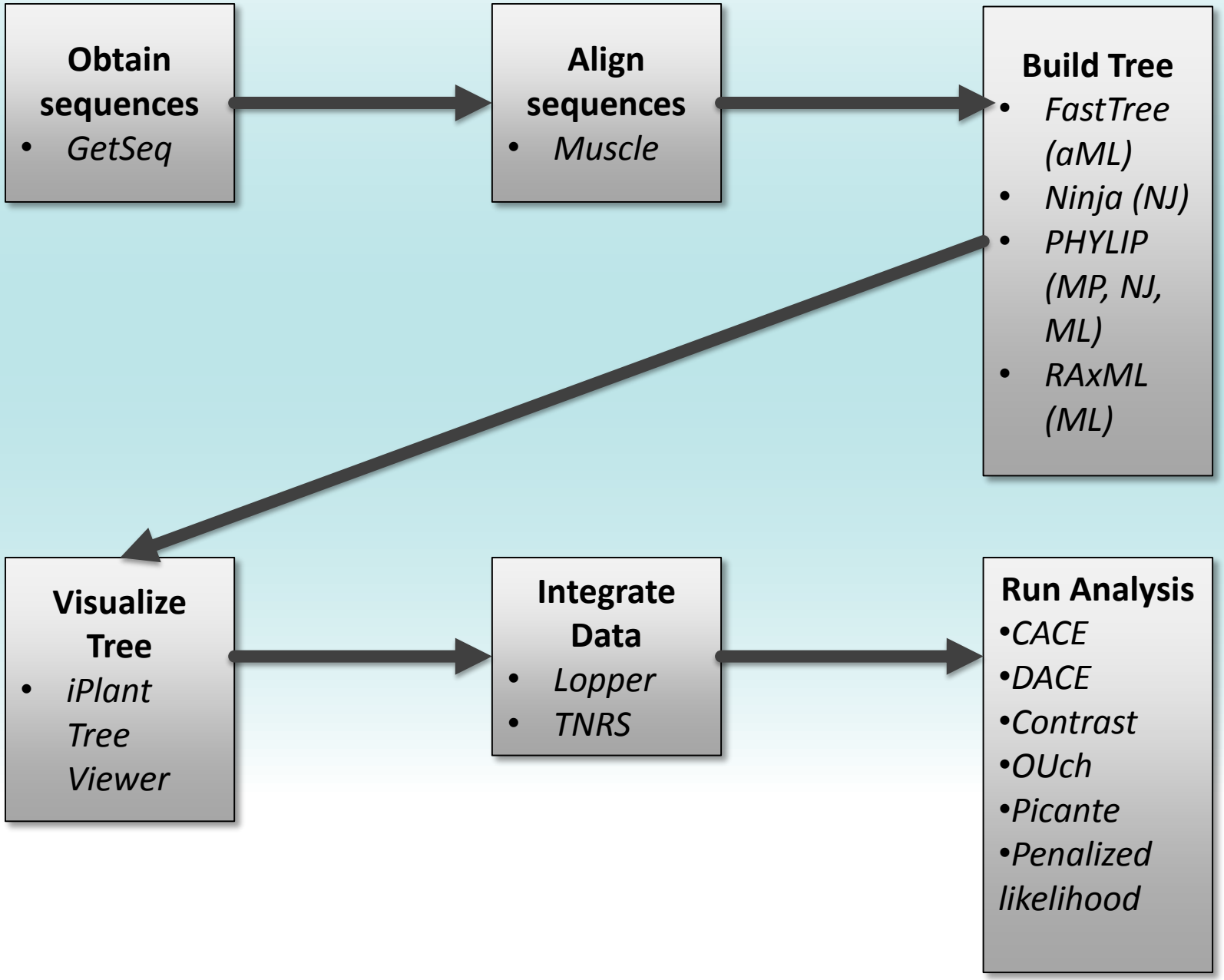
- Further development of methods
- Aggregating tools
- **'Workbench' – enhanced workflows**
- Research applications workshops

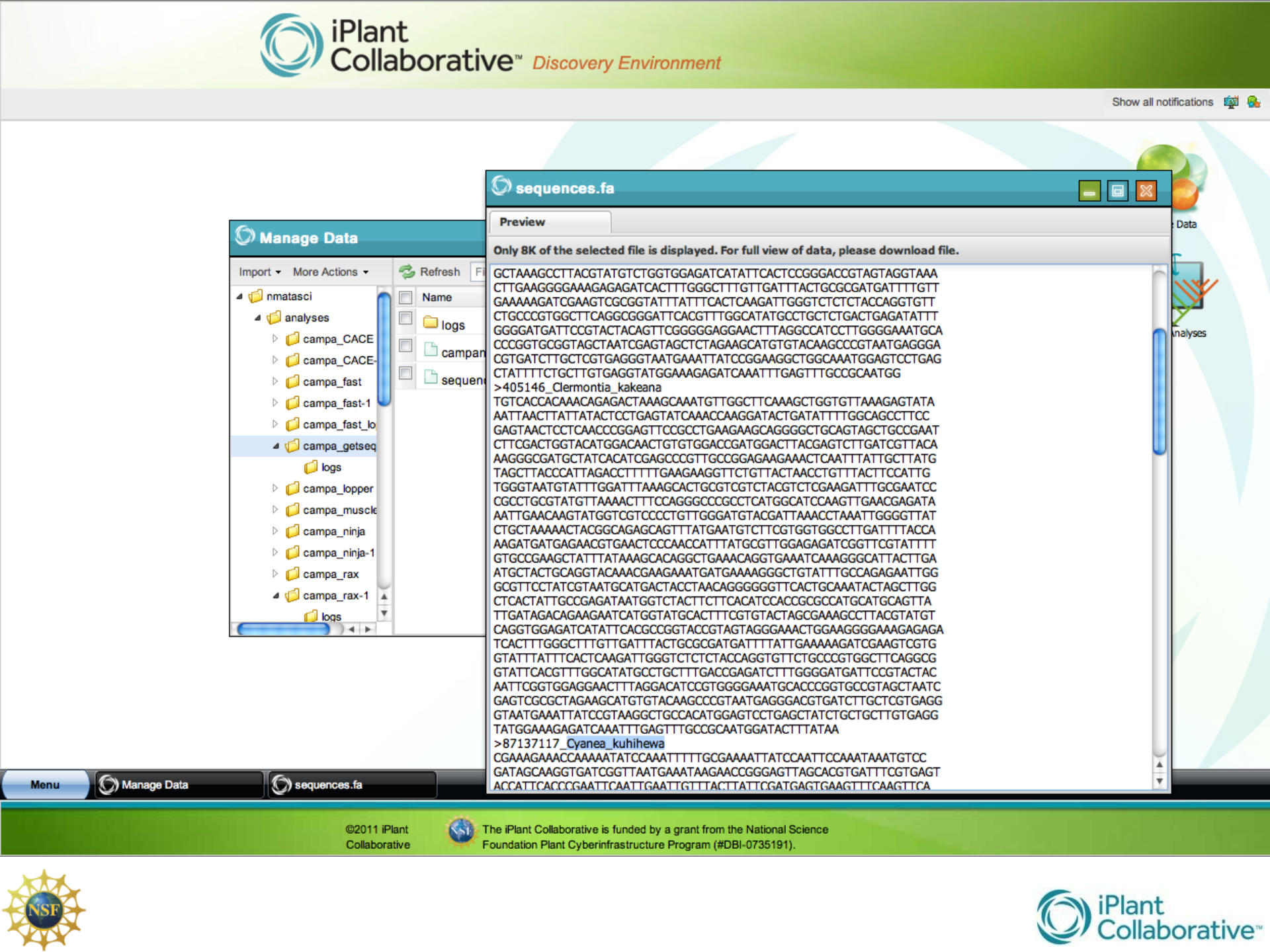
Facilitating Research

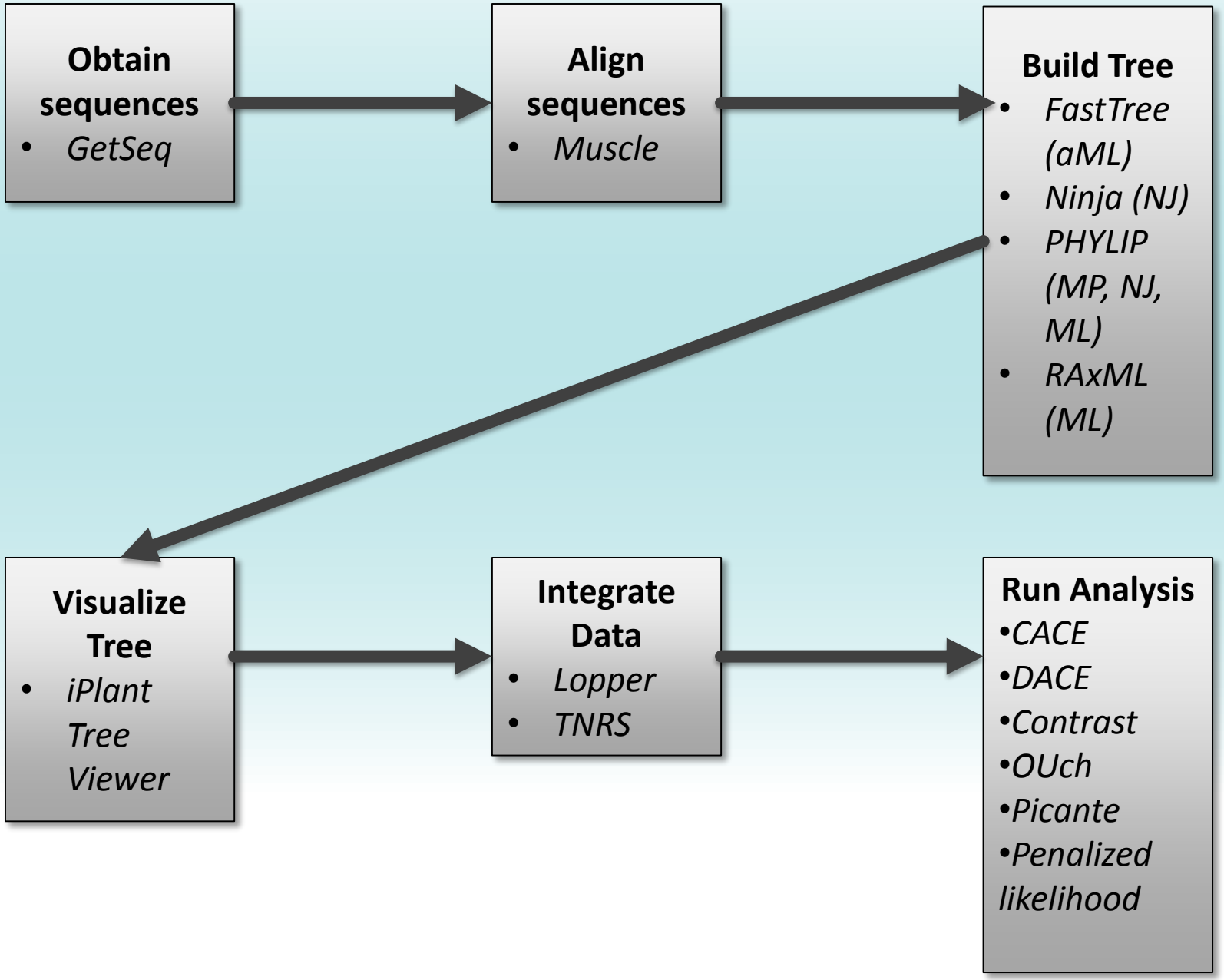
- **Toward a Biodiversity Workbench**



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


Obtain sequences
• GetSeq
Align sequences
• Muscle
Build Tree
• FastTree (aML)
• Ninja (NJ)
• PHYLIP (MP, NJ, ML)
• RAxML (ML)
Visualize Tree
• iPlant Tree Viewer
Integrate Data
• Lopper
• TNRS
Run Analysis
•CACE
•DACE
•Contrast
•OUch
•Picante
•Penalized likelihood


iPlant Collaborative™ Discovery Environment
Show all notifications
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
Name
logs
campan
sequen
sequences.fa
Preview
Only 8K of the selected file is displayed. For full view of data, please download file.
GCTAAAGCCTTACGTATGTCTGGTGGAGATCATATTCACTCCGGGACCGTAGTAGGTAAA
CTTGAAGGGGAAAGAGAGATCACTTTGGGCTTTGTTGATTTACTGCGCGATGATTTTGTT
GAAAAAGATCGAAGTCGCGGTATTTATTTCACTCAAGATTGGGTCTCTCTACCAGGTGTT
CTGCCCGTGGCTTCAGGCGGGATTCACGTTTGGCATATGCCTGCTCTGACTGAGATATTT
GGGGATGATTCCGTACTACAGTTCGGGGGAGGAACTTTAGGCCATCCTTGGGGAAATGCA
CCCGGTGCGGTAGCTAATCGAGTAGCTCTAGAAGCATGTGTACAAGCCCGTAATGAGGGA
CGTGATCTTGCTCGTGAGGGTAATGAAATTATCCGGAAGGCTGGCAAATGGAGTCCTGAG
CTATTTTCTGCTTGTGAGGTATGGAAAGAGATCAAATTTGAGTTTGCCGCAATGG
>405146_Clermontia_kakeana
TGTCACCACAAACAGAGACTAAAGCAAATGTTGGCTTCAAAGCTGGTGTTAAAGAGTATA
AATTAACTTATTATACTCCTGAGTATCAAACCAAGGATACTGATATTTTGGCAGCCTTCC
GAGTAACTCCTCAACCCGGAGTTCCGCCTGAAGAAGCAGGGGCTGCAGTAGCTGCCGAAT
CTTCGACTGGTACATGGACAACTGTGTGGACCGATGGACTTACGAGTCTTGATCGTTACA
AAGGGCGATGCTATCACATCGAGCCCGTTGCCGGAGAAGAAACTCAATTTATTGCTTATG
TAGCTTACCCATTAGACCTTTTTGAAGAAGGTTCTGTTACTAACCTGTTTACTTCCATTG
TGGGTAATGTATTTGGATTTAAAGCACTGCGTCGTCTACGTCTCGAAGATTTGCGAATCC
CGCCTGCGTATGTTAAAACTTTCCAGGGCCCGCCTCATGGCATCCAAGTTGAACGAGATA
AATTGAACAAGTATGGTCGTCCCCTGTTGGGATGTACGATTAAACCTAAATTGGGGTTAT
CTGCTAAAAACTACGGCAGAGCAGTTTATGAATGTCTTCGTGGTGGCCTTGATTTTACCA
AAGATGATGAGAACGTGAACTCCCAACCATTTATGCGTTGGAGAGATCGGTTCGTATTTT
GTGCCGAAGCTATTTATAAAGCACAGGCTGAAACAGGTGAAATCAAAGGGCATTACTTGA
ATGCTACTGCAGGTACAAACGAAGAAATGATGAAAAGGGCTGTATTTGCCAGAGAATTGG
GCGTTCCTATCGTAATGCATGACTACCTAACAGGGGGGTTCACTGCAAATACTAGCTTGG
CTCACTATTGCCGAGATAATGGTCTACTTCTTCACATCCACCGCGCCATGCATGCAGTTA
TTGATAGACAGAAGAATCATGGTATGCACTTTCGTGTACTAGCGAAAGCCTTACGTATGT
CAGGTGGAGATCATATTCACGCCGGTACCGTAGTAGGGAAACTGGAAGGGGAAAGAGAGA
TCACTTTGGGCTTTGTTGATTTACTGCGCGATGATTTTATTGAAAAAGATCGAAGTCGTG
GTATTTATTTCACTCAAGATTGGGTCTCTCTACCAGGTGTTCTGCCCGTGGCTTCAGGCG
GTATTCACGTTTGGCATATGCCTGCTTTGACCGAGATCTTTGGGGATGATTCCGTACTAC
AATTCGGTGGAGGAACTTTAGGACATCCGTGGGGAAATGCACCCGGTGCCGTAGCTAATC
GAGTCGCGCTAGAAGCATGTGTACAAGCCCGTAATGAGGGACGTGATCTTGCTCGTGAGG
GTAATGAAATTATCCGTAAGGCTGCCACATGGAGTCCTGAGCTATCTGCTGCTTGTGAGG
TATGGAAAGAGATCAAATTTGAGTTTGCCGCAATGGATACTTTATAA
>87137117_Cyanea_kuhihewa
CGAAAGAAACCAAAAATATCCAAATTTTTGCGAAAATTATCCAATTCCAAATAAATGTCC
GATAGCAAGGTGATCGGTTAATGAAATAAGAACCGGGAGTTAGCACGTGATTTCGTGAGT
ACCATTCACCCGAATTCAATTGAATTGTTTACTTATTCGATGAGTGAAGTTTCAAGTTCA
Menu
Manage Data
sequences.fa
©2011 iPlant Collaborative
The iPlant Collaborative is funded by a grant from the National Science Foundation Plant Cyberinfrastructure Program (#DBI-0735191).
NSF
iPlant Collaborative™


Obtain sequences
• GetSeq
Align sequences
• Muscle
Build Tree
• FastTree (aML)
• Ninja (NJ)
• PHYLIP (MP, NJ, ML)
• RAxML (ML)
Visualize Tree
• iPlant Tree Viewer
Integrate Data
• Lopper
• TNRS
Run Analysis
•CACE
•DACE
•Contrast
•OUch
•Picante
•Penalized likelihood

Connecting Trees, Specimens, Tools

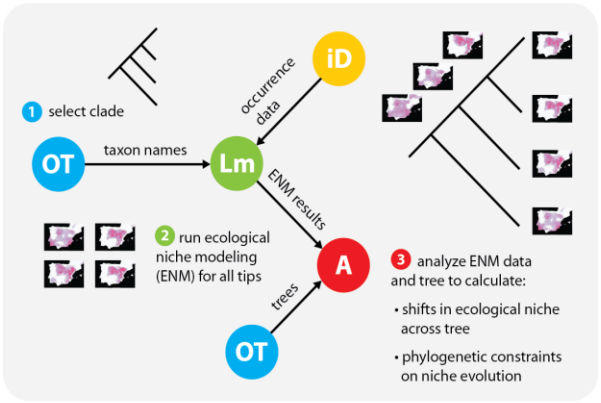
iDigBio
Integrated Digitized Biocollections

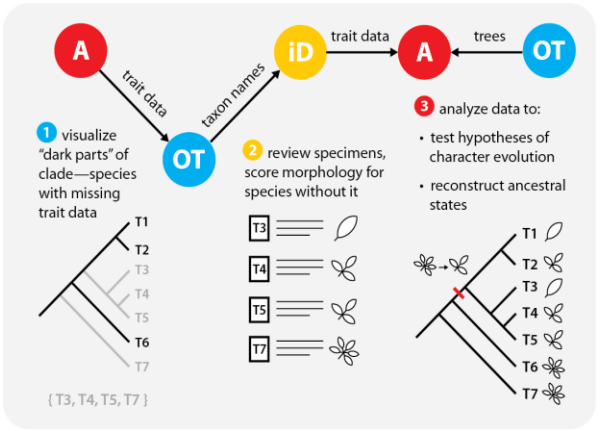
OPEN
Tree of Life
A community-assembled ToL

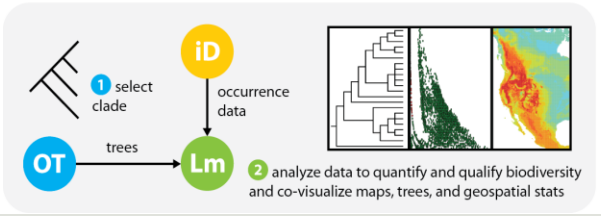
Lifemapper

ARBOR
revolutionary workflows

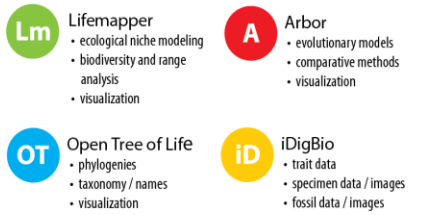
EXAMPLE WORKFLOWS:

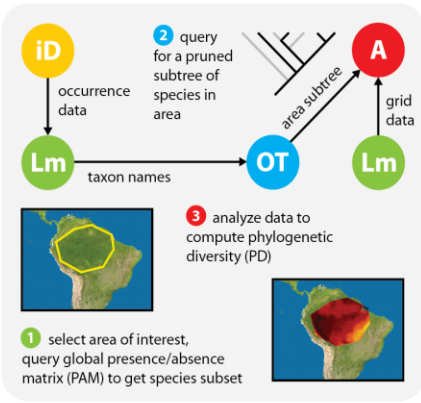


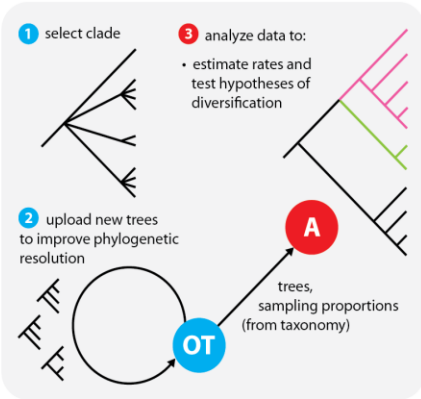




RESOURCES:







Connecting Trees, Specimens, Tools

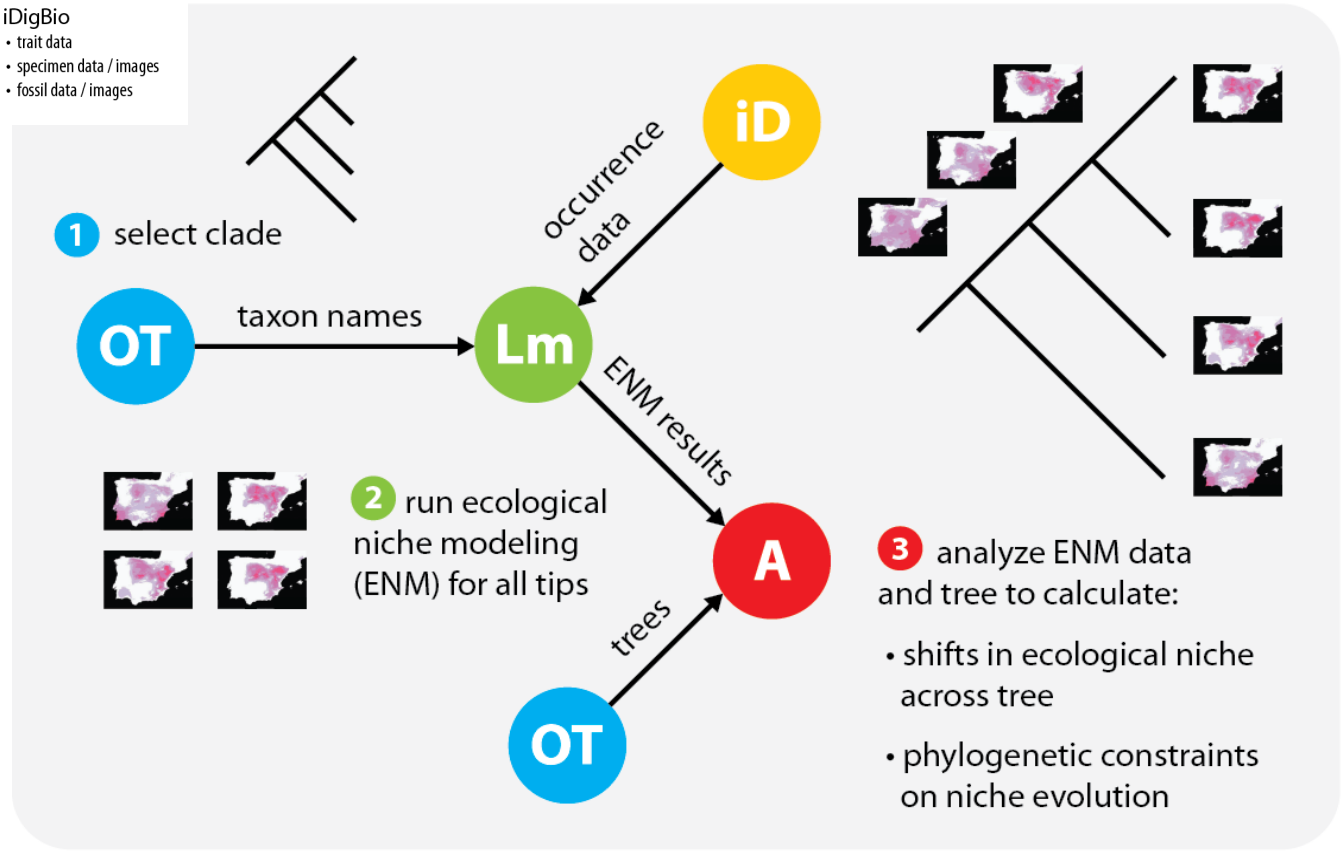
RESOURCES:

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:

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

Images as Sources of Trait Data

Images as Sources of Trait Data

TRY Plant Trait Database

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

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*

iDigBio
Integrated Digitized Biocollections

Thank you!

Blaine Marchant

Lauren Gonzalez

Charlotte Germain-Aubrey

Greg Stull

Andre Naranjo

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.

iDigBio
Integrated Digitized Biocollections

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

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.