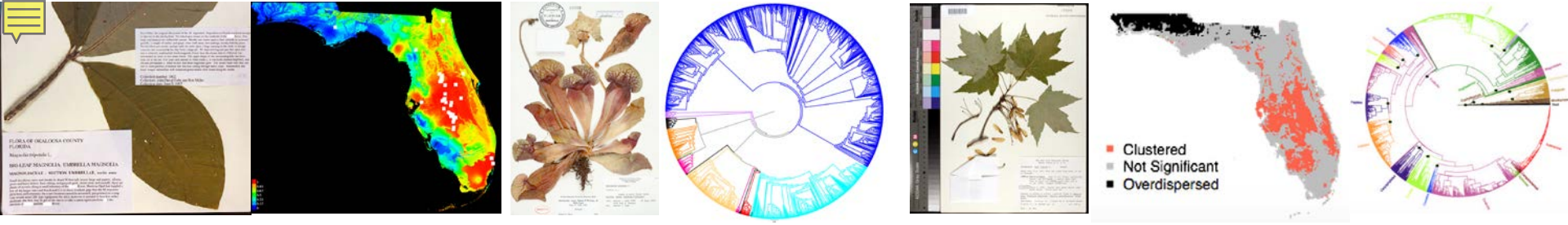


Linking Heterogeneous Data in Biodiversity Research

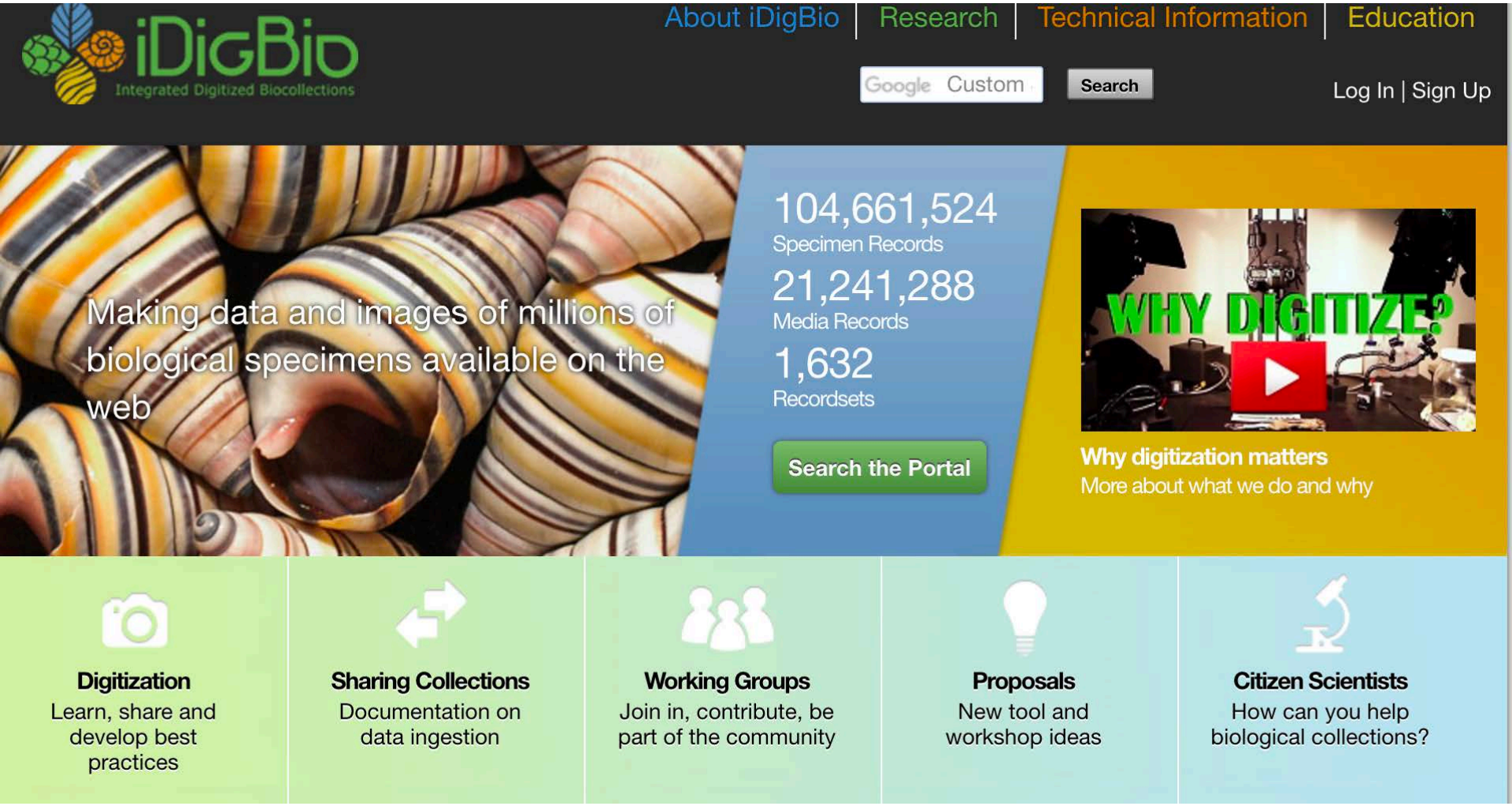
Pamela S. Soltis
University of Florida

FLORIDA MUSEUM OF NATURAL HISTORY

iDigBio
Integrated Digitized Biocollections

UF UNIVERSITY of FLORIDA

Biodiversity Research

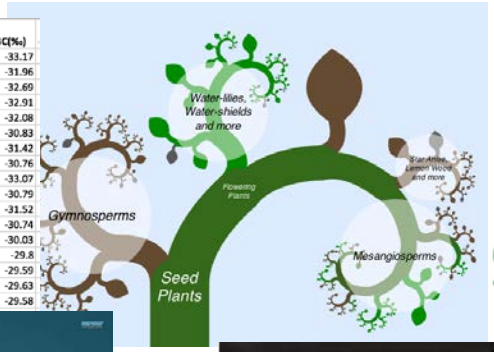


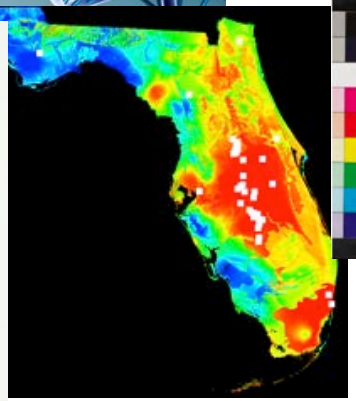
Biodiversity Research

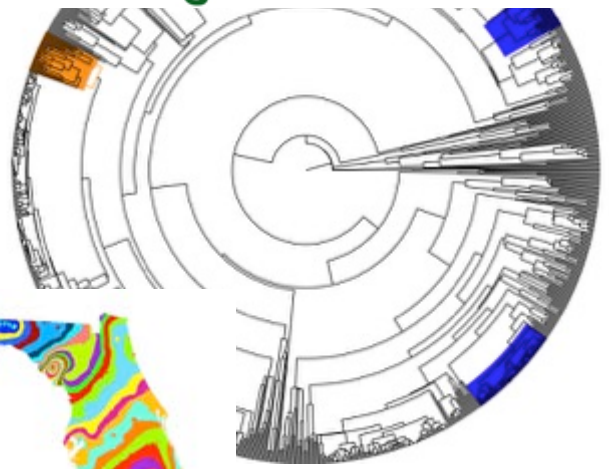
- Heterogeneous data
- Challenges
- Examples
- Solutions
- Summary

Heterogeneous Data

Genus	Species	Scientific name	Sample date	SLA dry mass (g)	leaf_area_cm2	SLA_cm2.g	LMA_g.m2	d13C(‰)
Chionanthu	virginicus	Chionanthus virginic	5/6/15	1.6	489.778	306.1	32.7	-33.17
Chionanthu	virginicus	Chionanthus virginic	5/6/15	2.36	462.018	195.8	51.1	-31.96
Chionanthu	virginicus	Chionanthus virginic	5/6/15	1.94	383.707	197.8	50.6	-32.69
Castanea	pumila	Castanea pumila	5/6/15	0.84				-32.91
Castanea	pumila	Castanea pumila	5/7/15	1.843	399.395	216.7	46.1	-32.08
Castanea	pumila	Castanea pumila	5/7/15	1.676	368.592	219.9	45.5	-30.83
Castanea	pumila	Castanea pumila	5/7/15	1.452	326.529	224.9	44.5	-31.42
Castanea	pumila	Castanea pumila	5/7/15	1.249	304.058	243.4	41.1	-30.76
Castanea	pumila	Castanea pumila	5/7/15	1.433	383.797	267.8	37.3	-33.07
Cartrema	americana	Cartrema americana	5/7/15	2.11	339.297	160.8	62.2	-30.79
Cartrema	americana	Cartrema americana	5/7/15	2.68	378.94	141.4	70.7	-31.52
Cartrema	americana	Cartrema americana	5/7/15	3.4	379.952	111.8	89.5	-30.74
Chionanthu	virginicus	Chionanthus virginic	5/7/15	2.866	393.473	137.3	72.8	-30.03
Diospyros	virginiana	Diospyros virginiana	5/7/15	2.085	383.958	184.2	54.3	-29.8
Quercus	laevis	Quercus laevis	5/7/15	1.445	134.457	93.0	107.5	-29.59
Cornus	florida	Cornus florida	5/8/15	1.546	229.312	148.3	67.4	-29.63
Cornus	florida	Cornus florida	5/8/15	2.136	282.992	132.5	75.5	-29.58







Challenges in Linking Heterogeneous Data

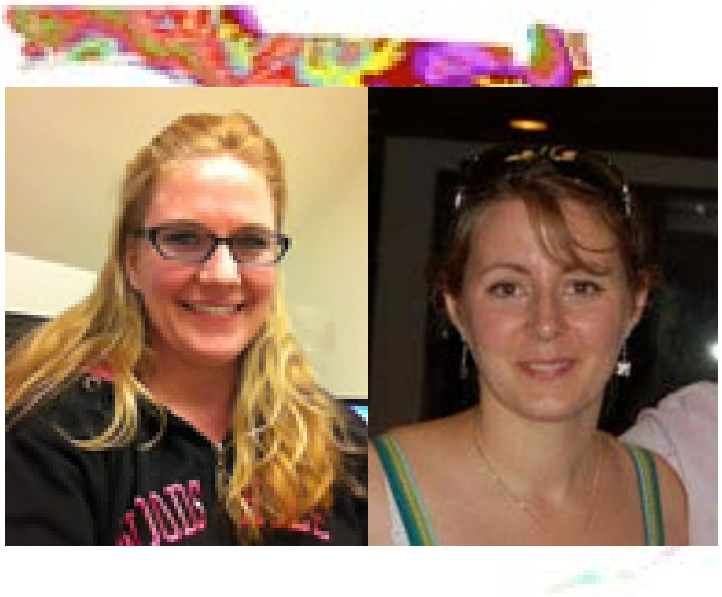
- Assembling data
- Data management and sharing
- Taxonomic names
- Patchy data
- Issues of scale: resolution, analysis
- Data integration

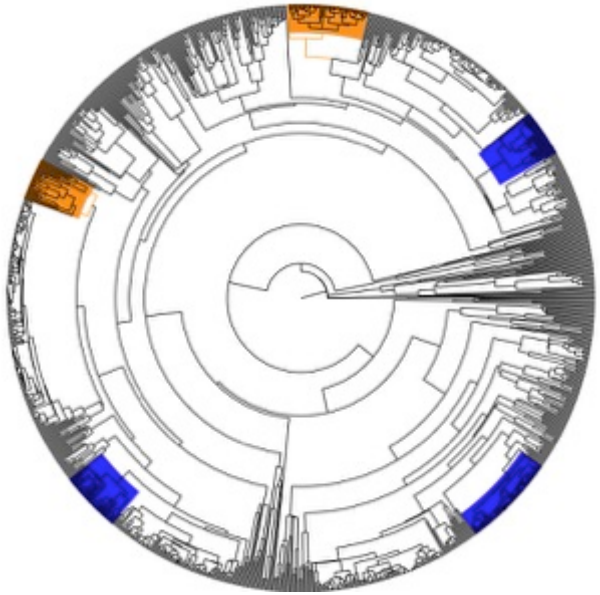
Examples

- Florida phylogenetic diversity
- Niche evolution in polyploids
- Ancient hybridization
- Phenology
- Traits from labels and images
- Spatial distribution of genome sizes

Florida Phylogenetic Diversity

Integrating herbarium specimen data, ENM, and phylogeny





Julie Allen, Charlotte Germain-Aubrey,

K. Neubig, L. Majure, R. Abbott, M. Whitten, N. Barve, H. Owens, J. M. Ponciano, B. Mishler, S. Laffan, R. Guralnick, D. Soltis

Florida Phylogenetic Diversity

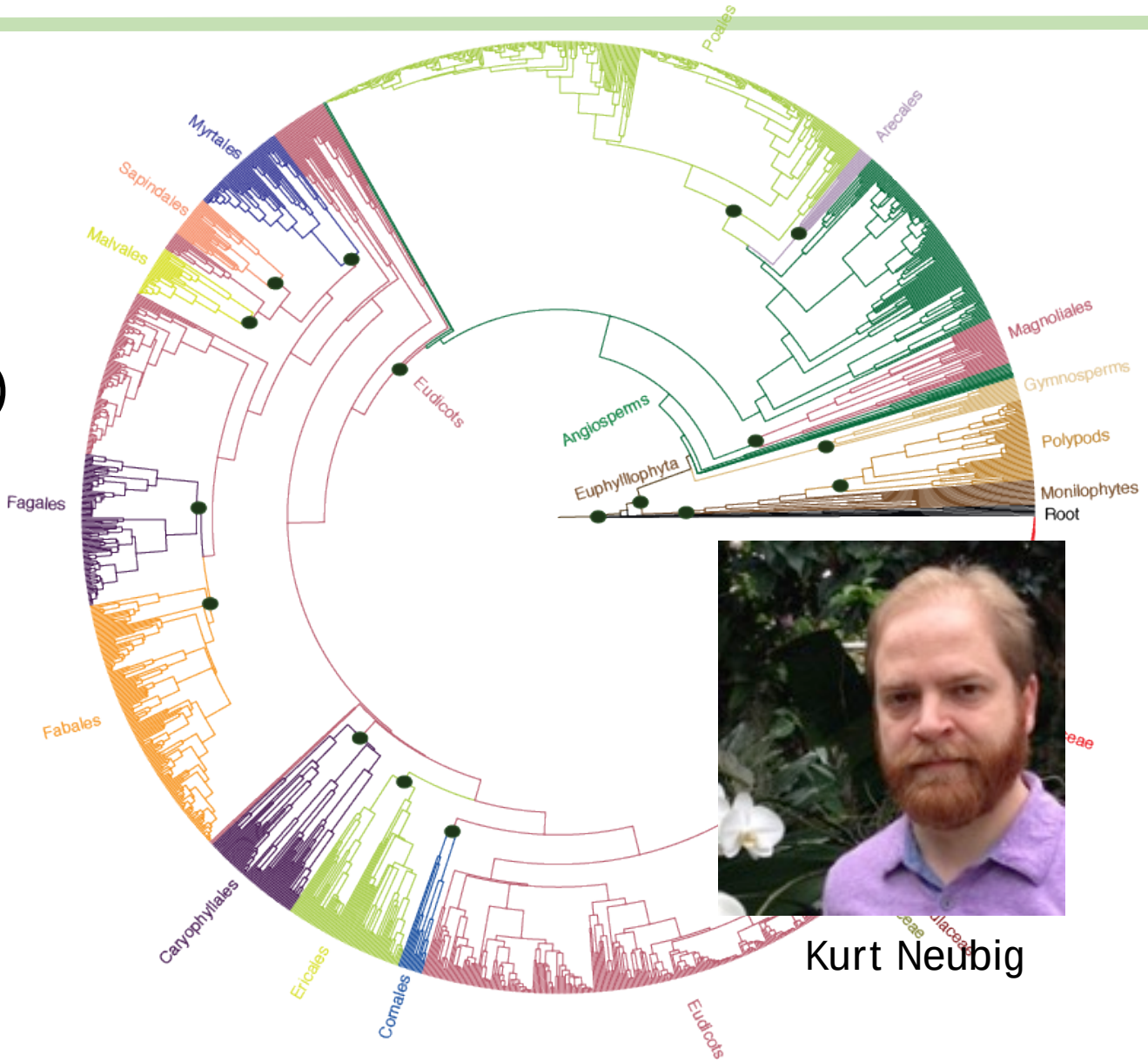
Modeling the Distribution of Species

- Location information and environmental data
- Maxent to model the range of each species
- For Florida plants:
  - 1,490 plant species (of 4100 species)
  - >511,000 georeferenced points (GPS)
  - Environmental features: temperature, precipitation, soil, etc.

Florida Phylogenetic Diversity

Florida Plant Phylogeny

1,490 species (37%)
685 genera (44%)
185 families (78%)
rbcL, *matK*
GenBank & new
RAxML
Dated with r8s

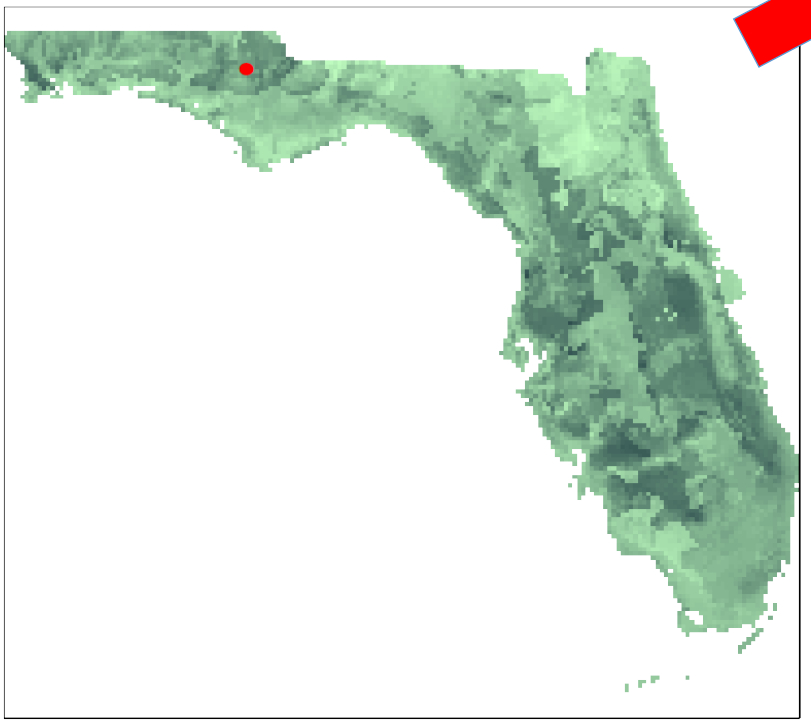


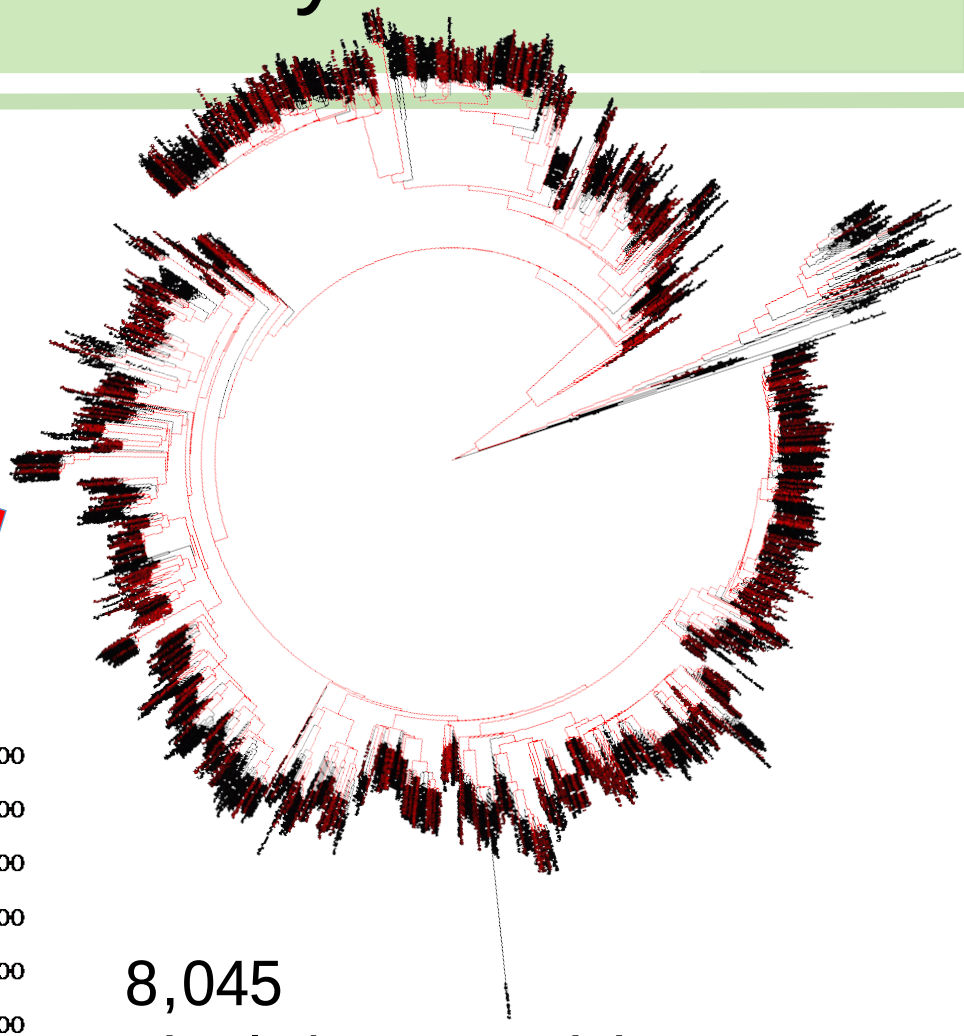
Kurt Neubig

Florida Phylogenetic Diversity

Phylogenetic Diversity:
≈ sum of branch lengths

Species list at each pixel
Generated from ENMs

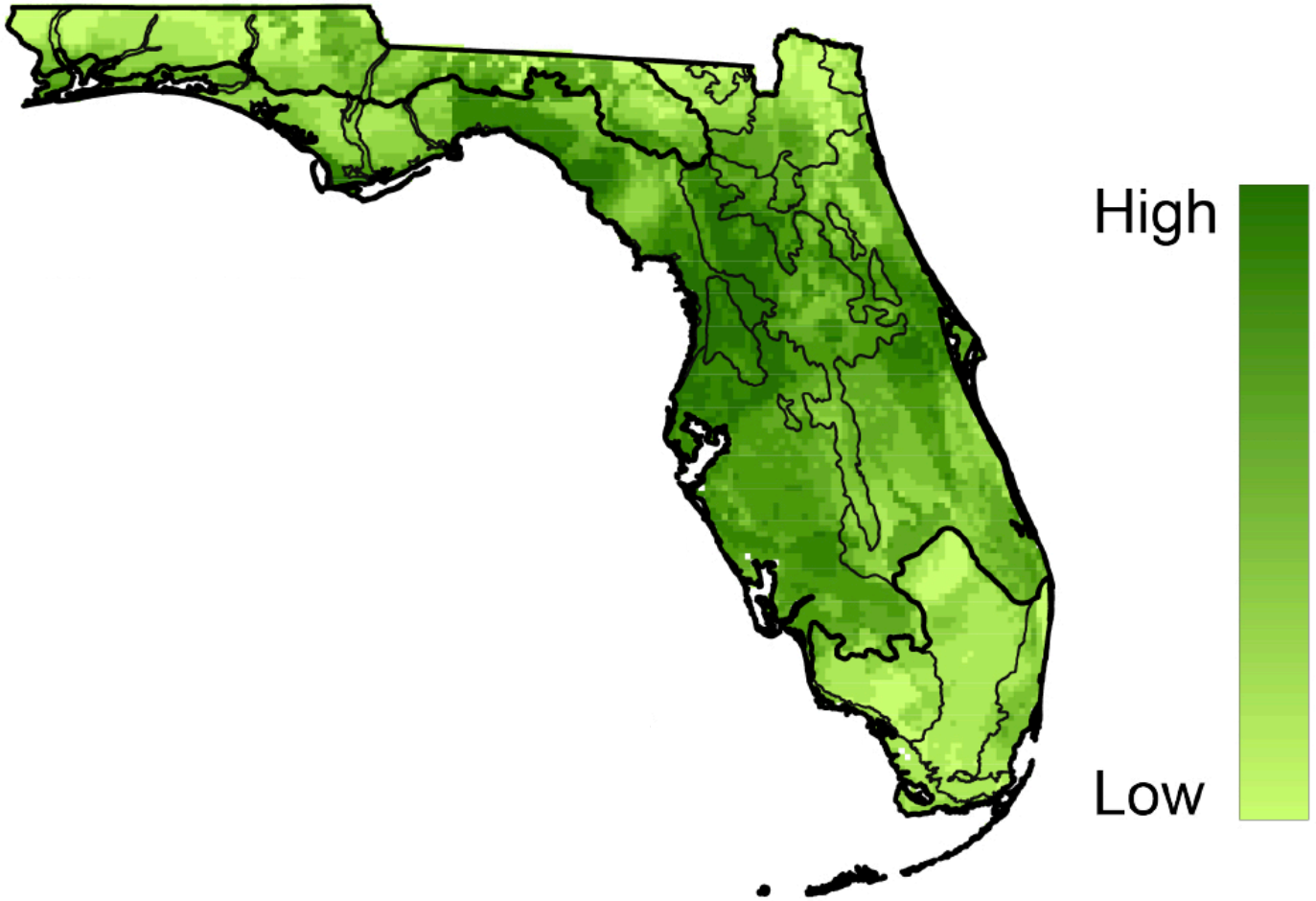




8,045
pixels/communities

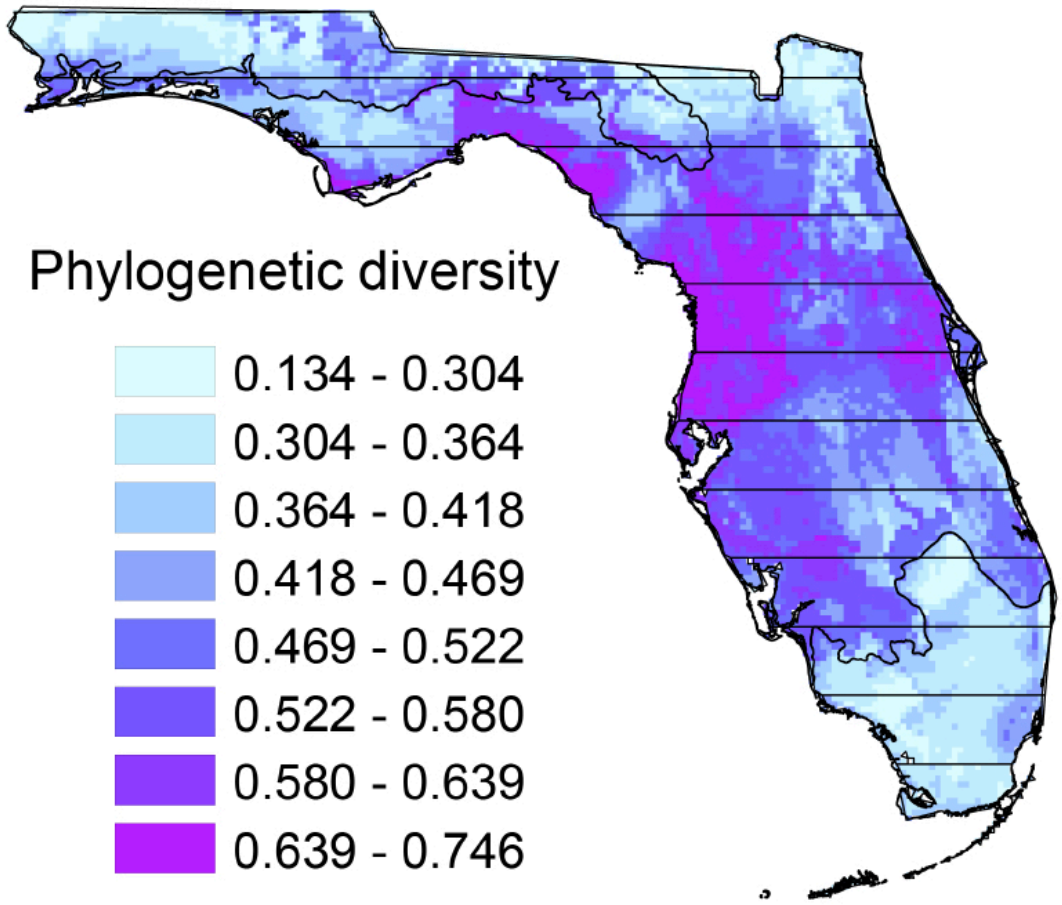
16 km^2 per pixel

Florida Phylogenetic Diversity



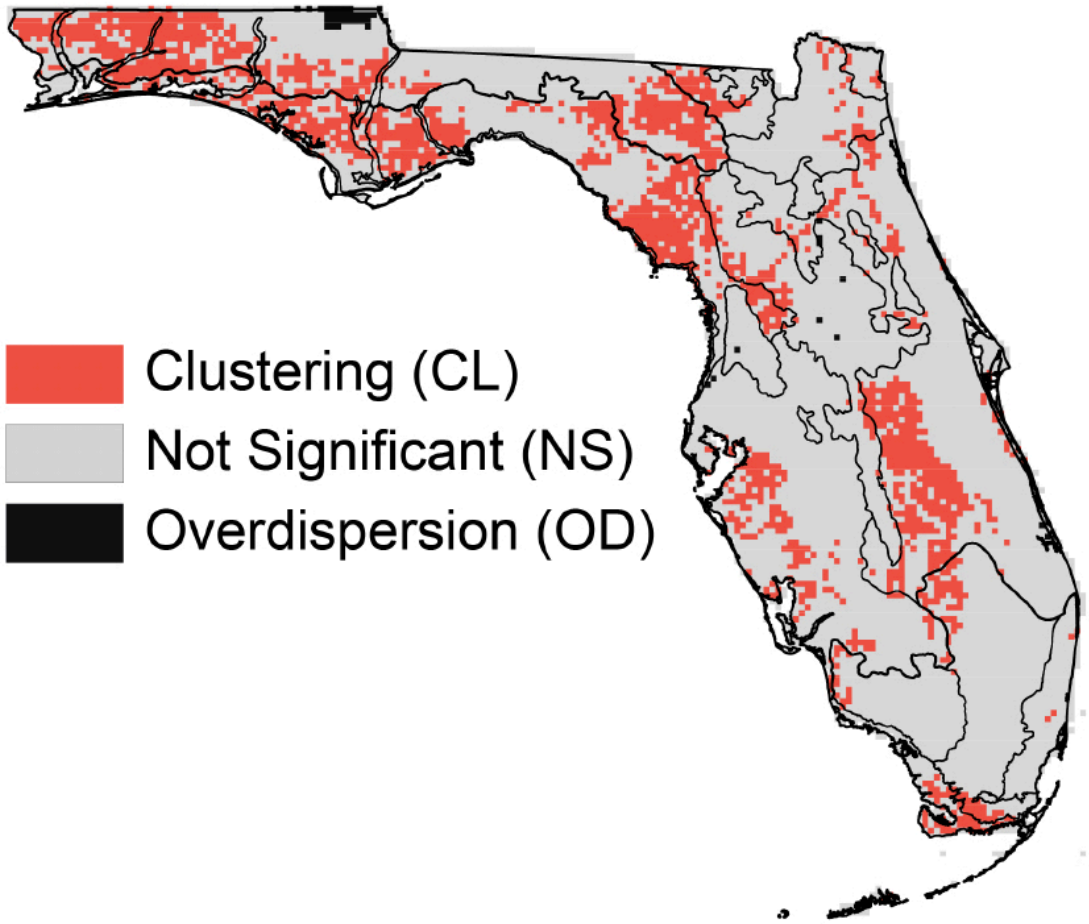
Florida Phylogenetic Diversity

Latitudinal patterns



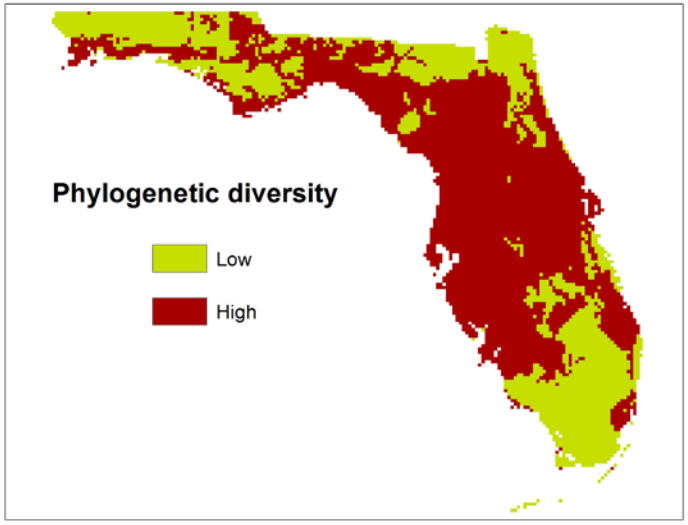
Florida Phylogenetic Diversity

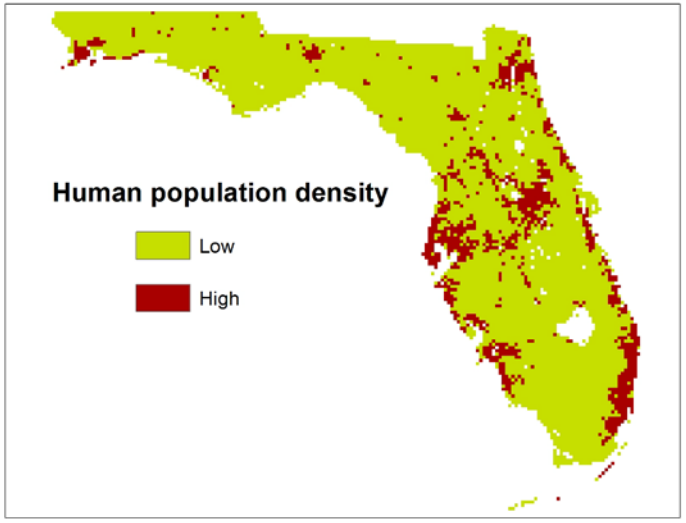
Clustering vs. overdispersion

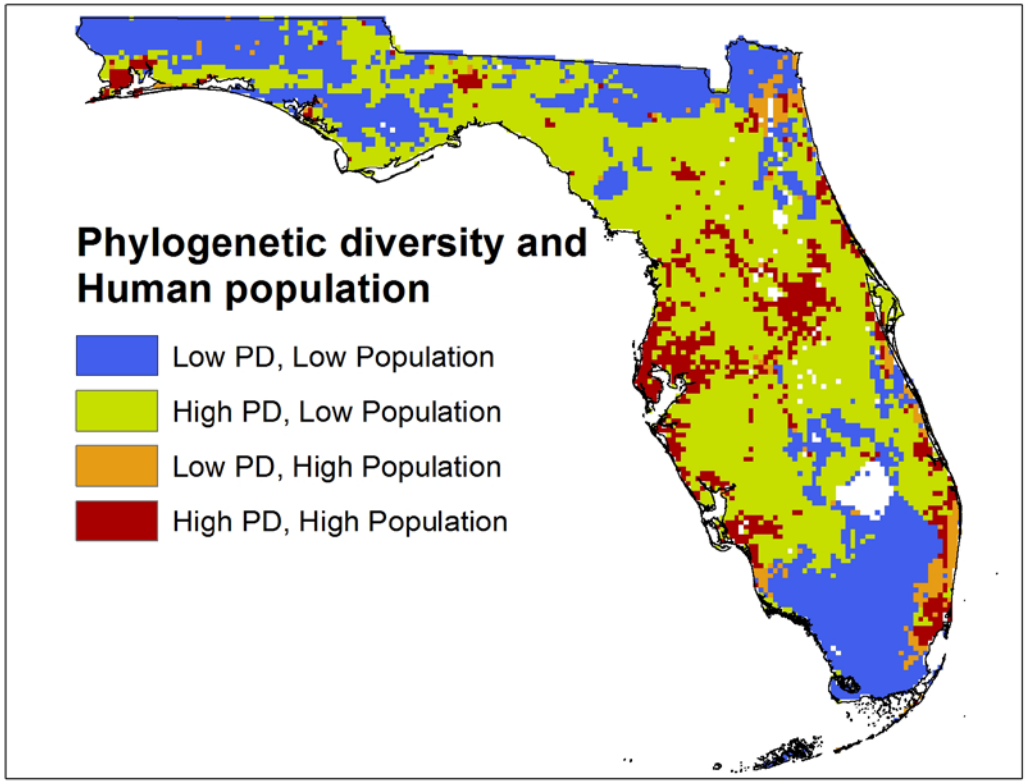


Florida Phylogenetic Diversity

Human population density



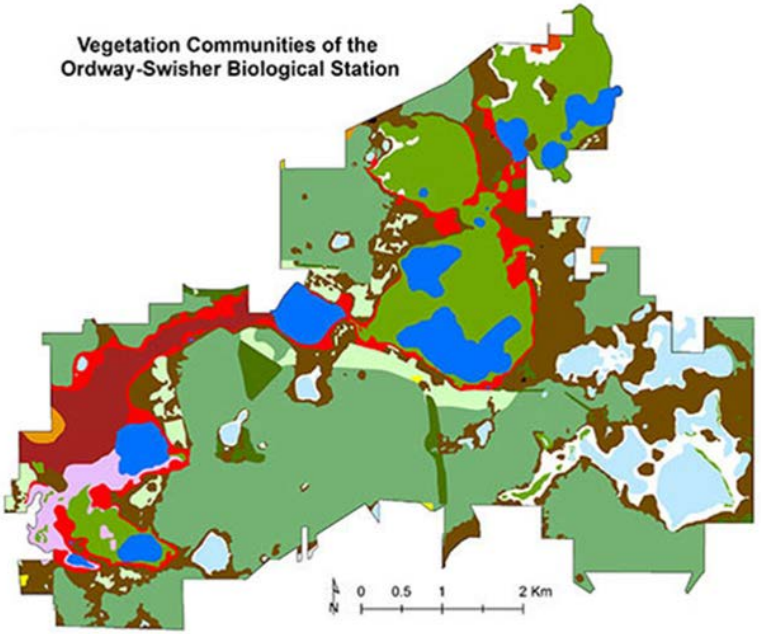


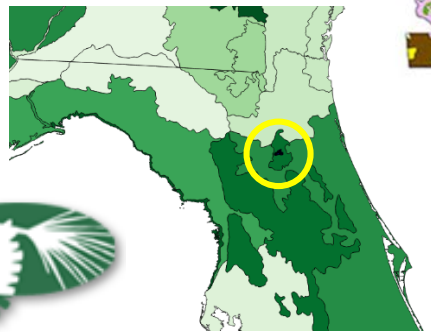


Florida Phylogenetic Diversity: Communities

Johanna Jantzen

Do levels of phylogenetic diversity vary among communities?





ORDWAY-SWISHER
BIOLOGICAL STATION

Florida Phylogenetic Diversity: Communities

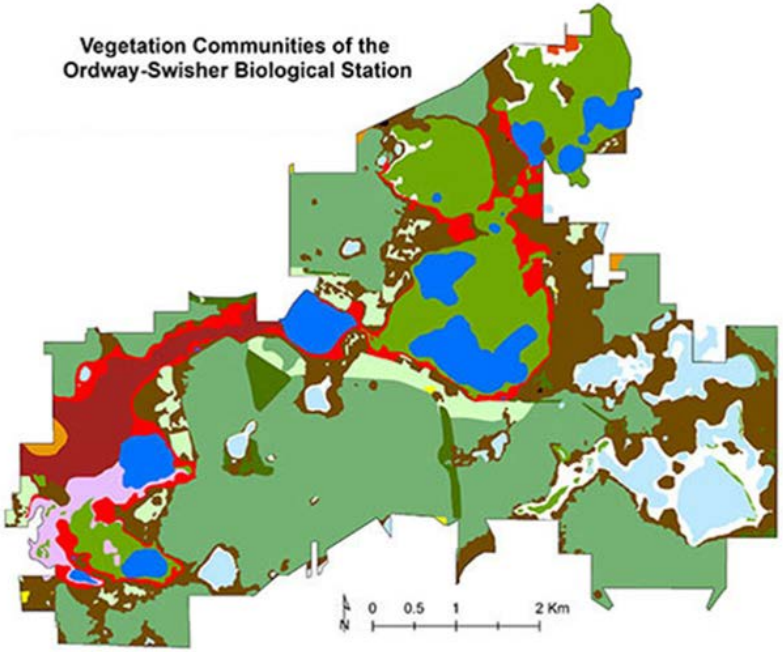
Johanna Jantzen

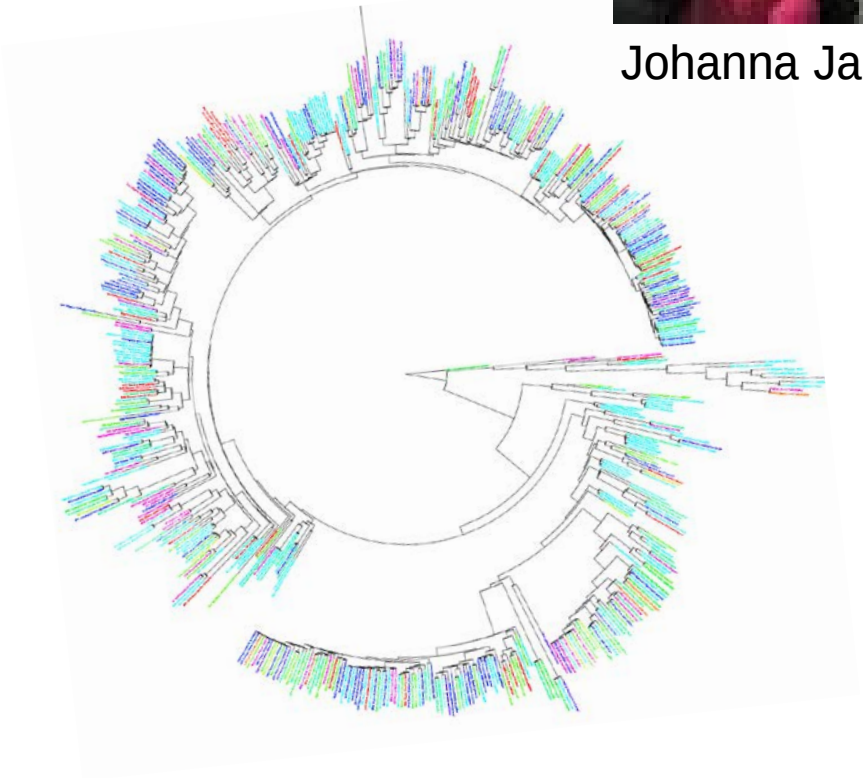
572 plant taxa

matK and *rbcL*

ML phylogeny reconstruction (RAxML)

PD calculations for 14 communities (Biodiverse)



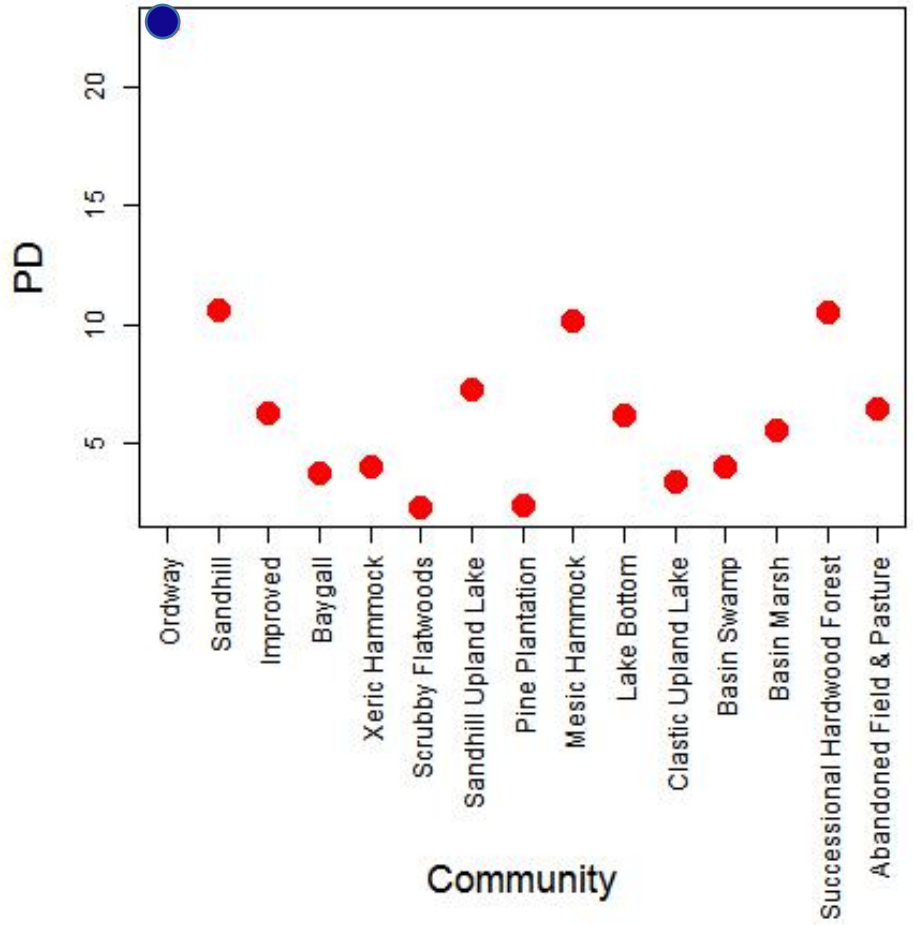


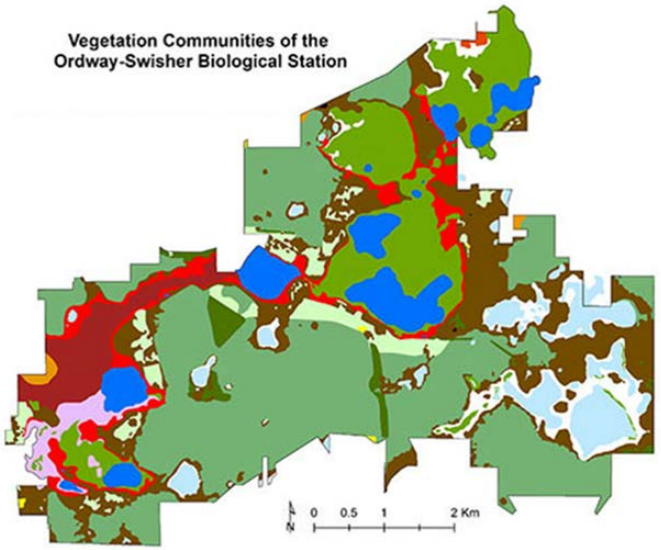
ORDWAY-SWISHER BIOLOGICAL STATION

Florida Phylogenetic Diversity: Communities

Johanna Jantzen

Phylogenetic diversity for 14 communities at OSBS: PD varies among communities



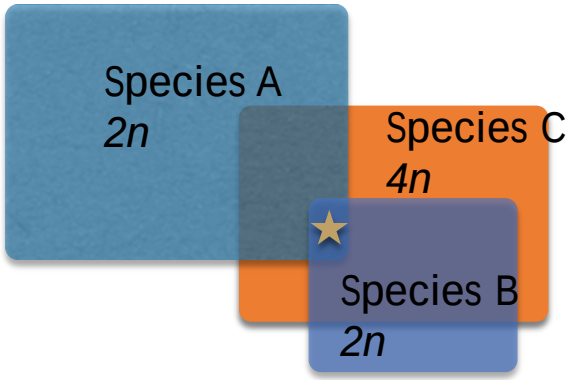


ORDWAY-SWISHER
BIOLOGICAL STATION

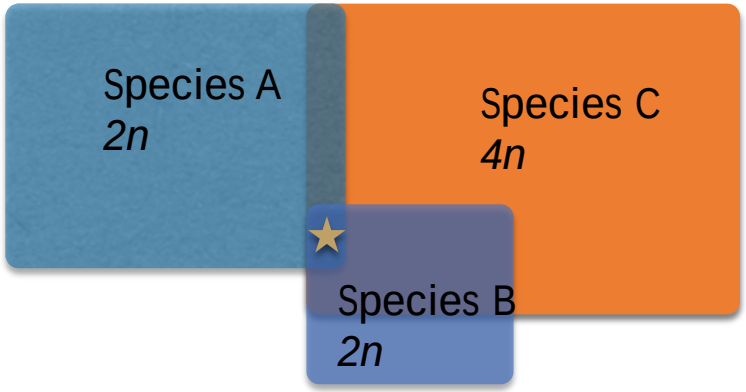
Niche Evolution in Allopolyploids

Blaine Marchant

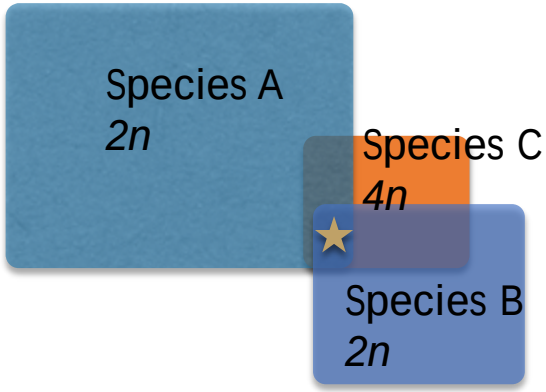
Niche Intermediacy



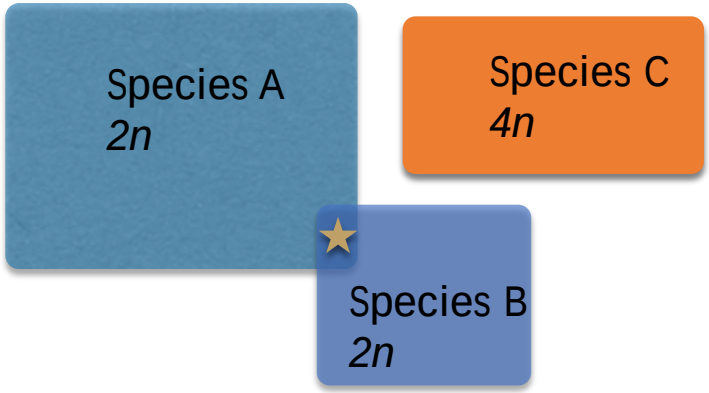
Niche Expansion



Niche Contraction



Niche Novelty



Marchant et al. 2016

iDigBio
Integrated Digitized Biocollections

About iDigBio | Research | Technical Information | Education

Google Custom Search

Log In | Sign Up

Making data and images of millions of biological specimens available on the web

104,661,524
Specimen Records

21,241,288
Media Records

1,632
Recordsets

Search the Portal

Why digitization matters
More about what we do and why

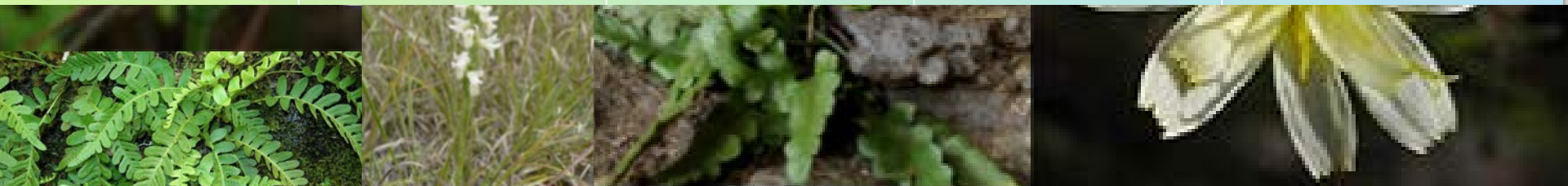
Digitization
Learn, share and develop best practices

Sharing Collections
Documentation on data ingestion

Working Groups
Join in, contribute, be part of the community

Proposals
New tool and workshop ideas

Citizen Scientists
How can you help biological collections?



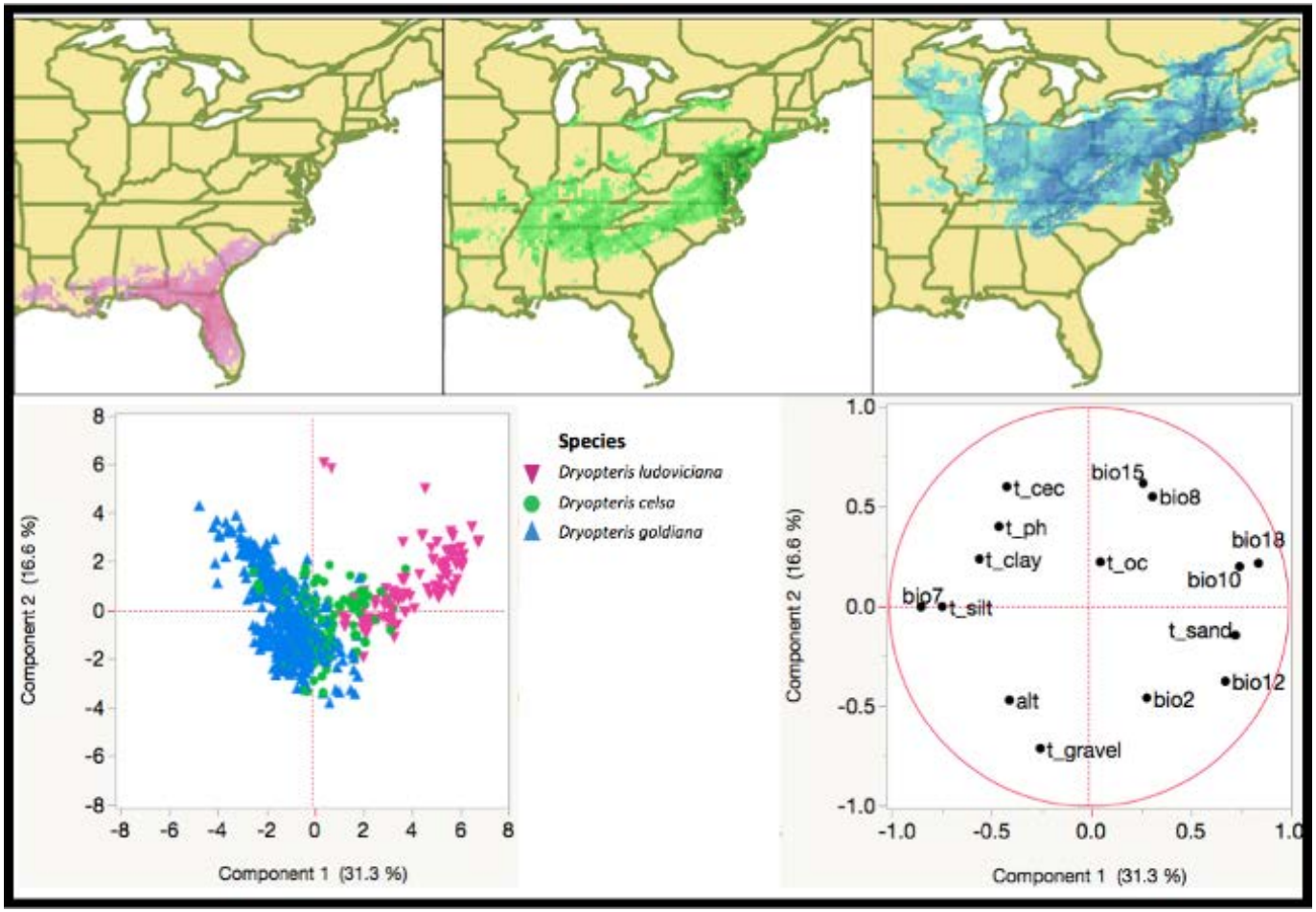
Niche Evolution in Allopolyploids

Blaine Marchant

Niche Intermediacy

Breadth: Parent < Polyploid < Parent

Overlap: Polyploid > 0.3



Dryopteris celsa

Nickrent, D.L. et al. 2006 onwards. *PhytoImages*. http://www.phytoimages.siu.edu

Marchant et al. 2016

Niche Evolution in Allopolyploids

Blaine Marchant

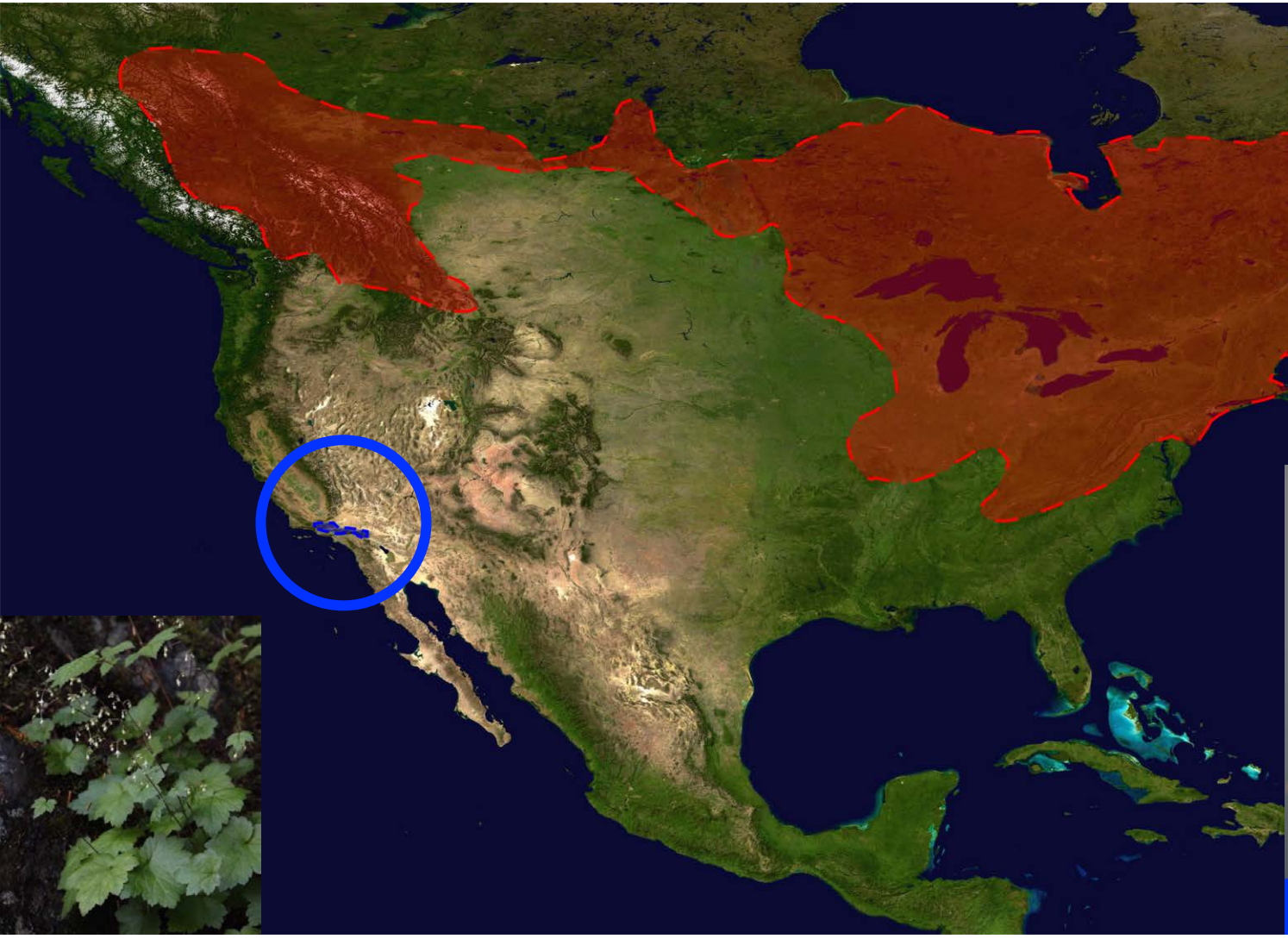
- 13 allopolyploids & parents
- Niche intermediacy: 8
- Niche contraction: 2
- Niche expansion: 2
- Niche novelty: 1

- More cases are needed!

Marchant et al. 2016

Ancient Hybridization: *Heuchera*

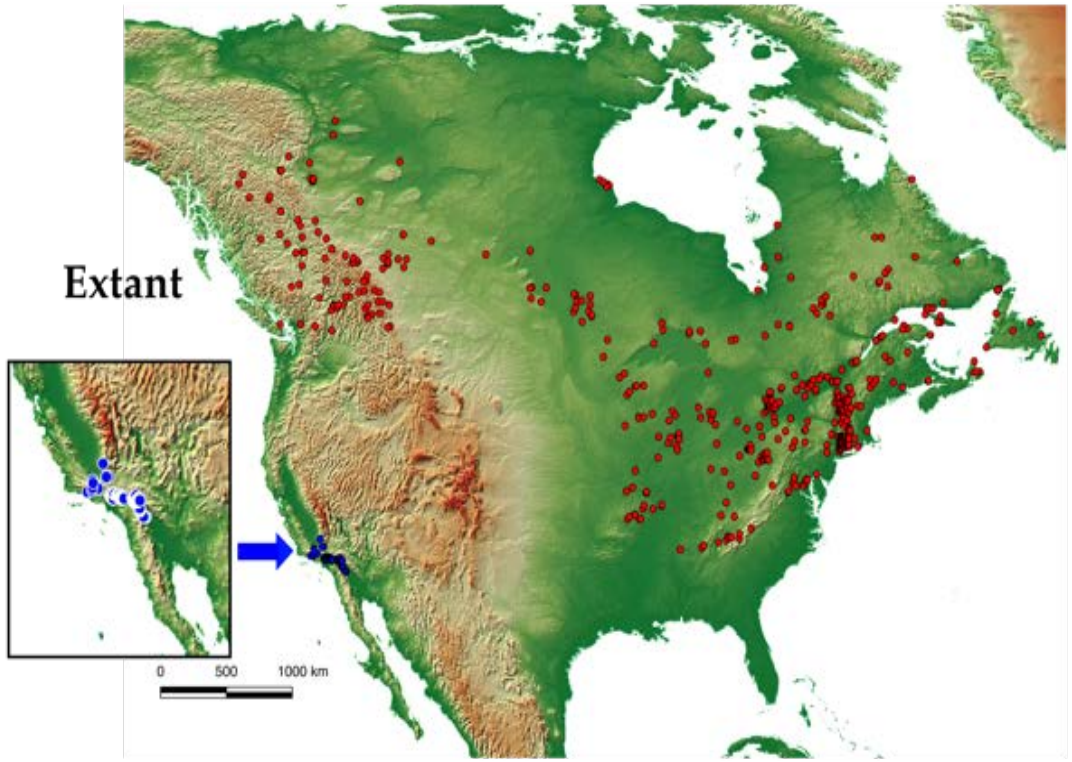
Ryan Folk



Ancient Hybridization: *Heuchera*

Ryan Folk

- Ancient chloroplast transfer: *Mitella* to *Heuchera*
- Species groups are currently allopatric
- How was hybridization/cp transfer accomplished?

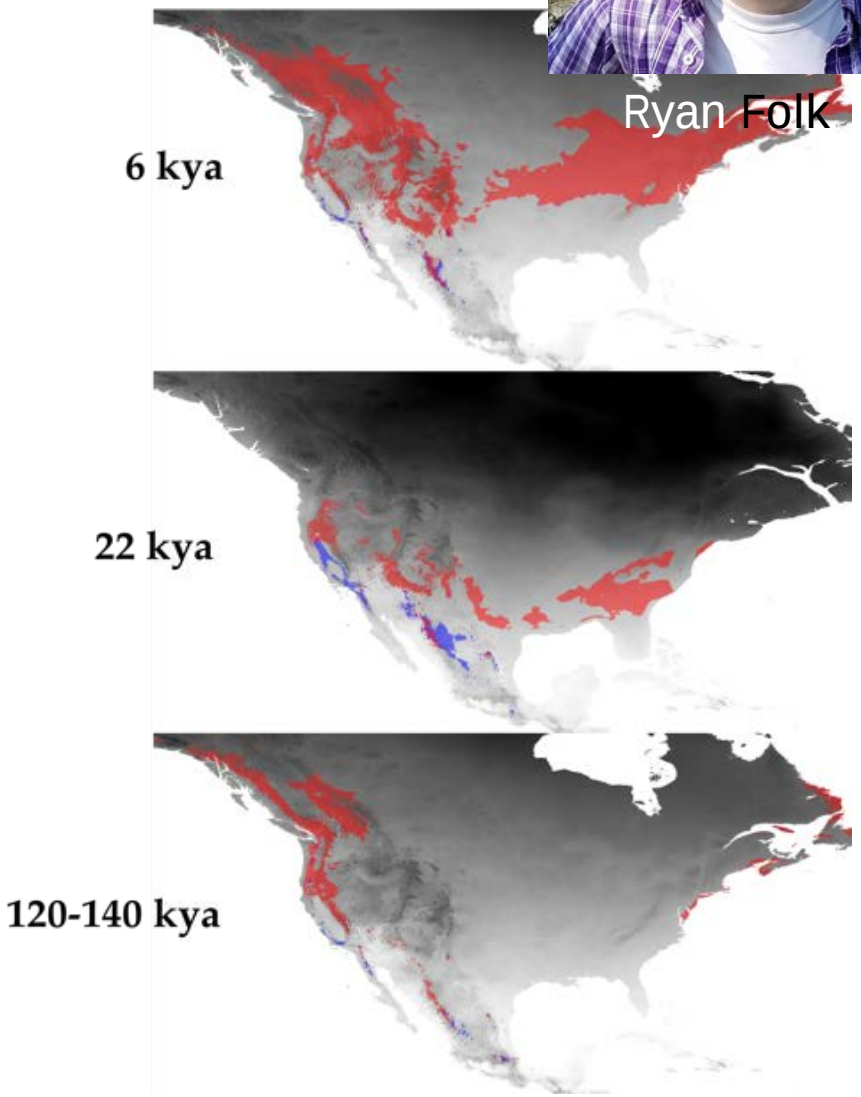


Ancient Hybridization: *Heuchera*


Ryan Folk

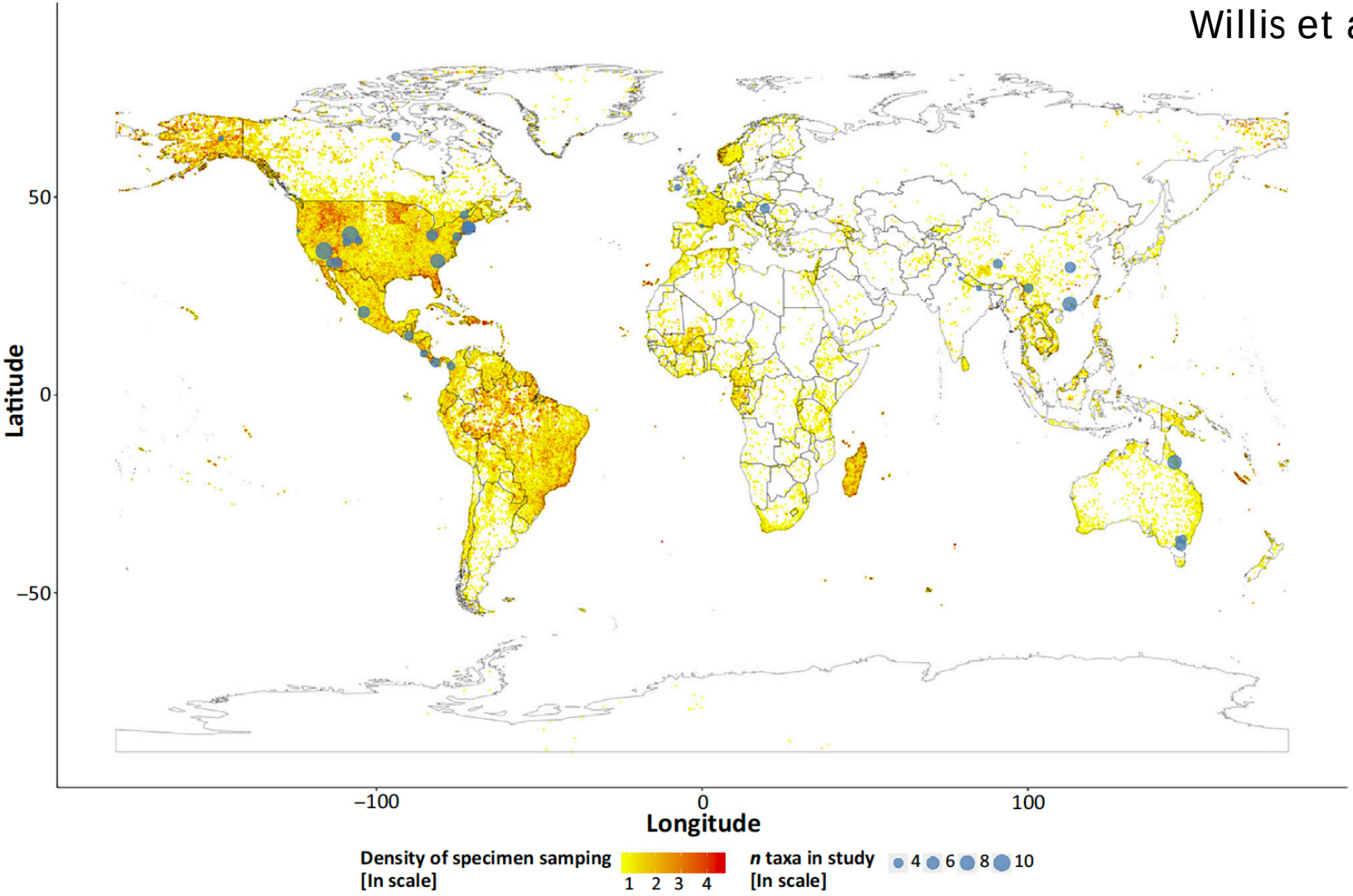
- Abiotic niches:
  - 12 climatic & environmental variables
- Ancestral niche reconstructions
- Projected niches into the past
- Biogeographic analyses

- Northern California the most likely region of overlap
- Pleistocene, LGM



Phenology: Flowering, Fruiting, Bud Burst

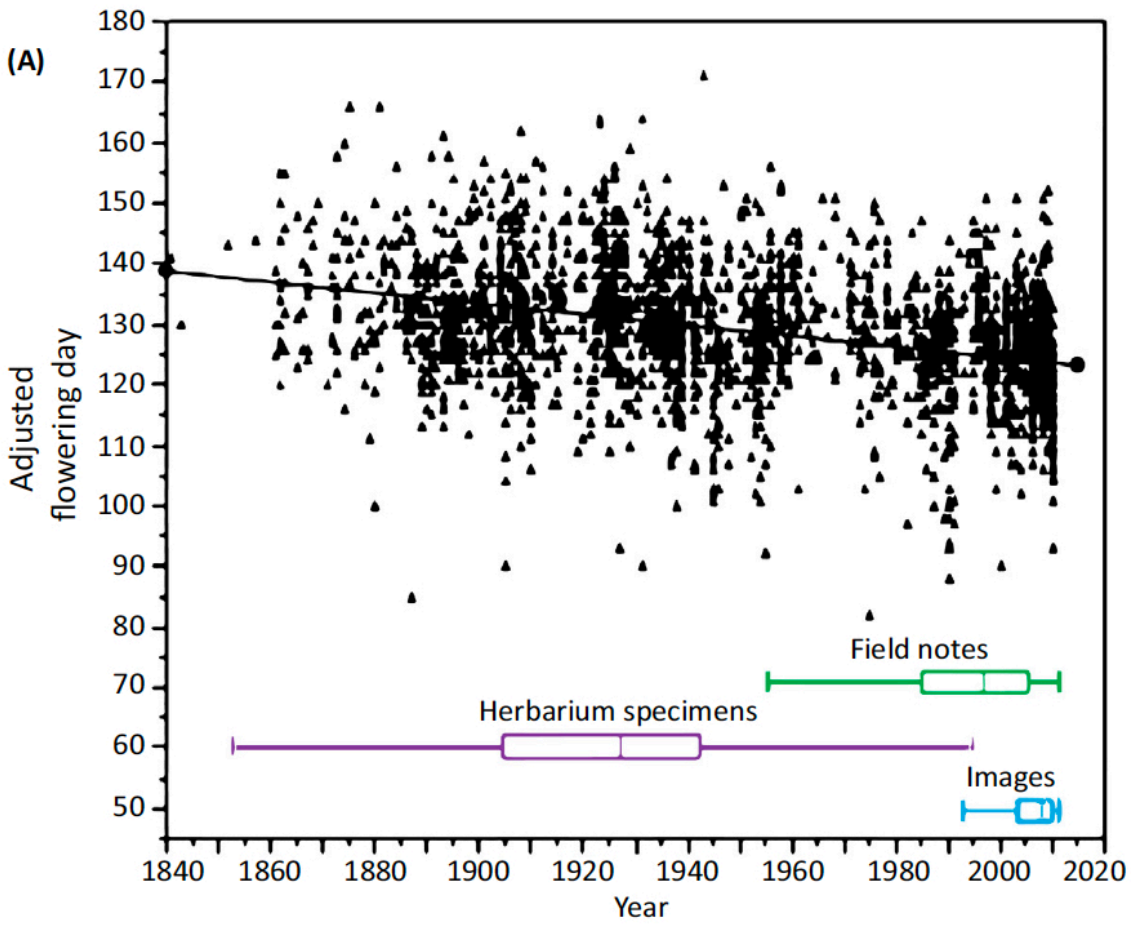
Willis et al. 2017



Trends in Ecology & Evolution

Phenology

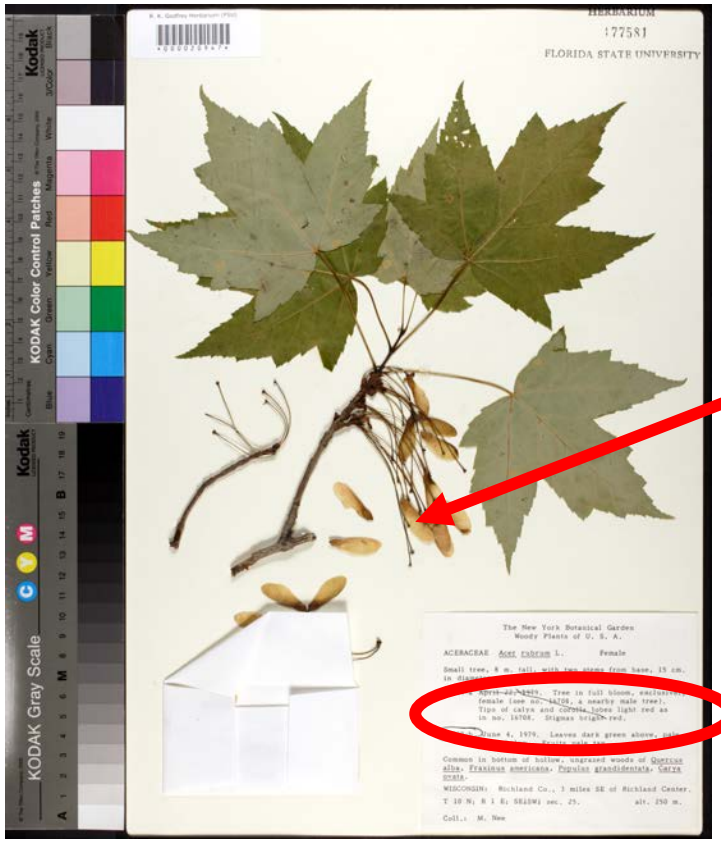
Integrating herbarium specimens with field notes, images, etc.



Willis et al. 2017

Traits from Labels and Images

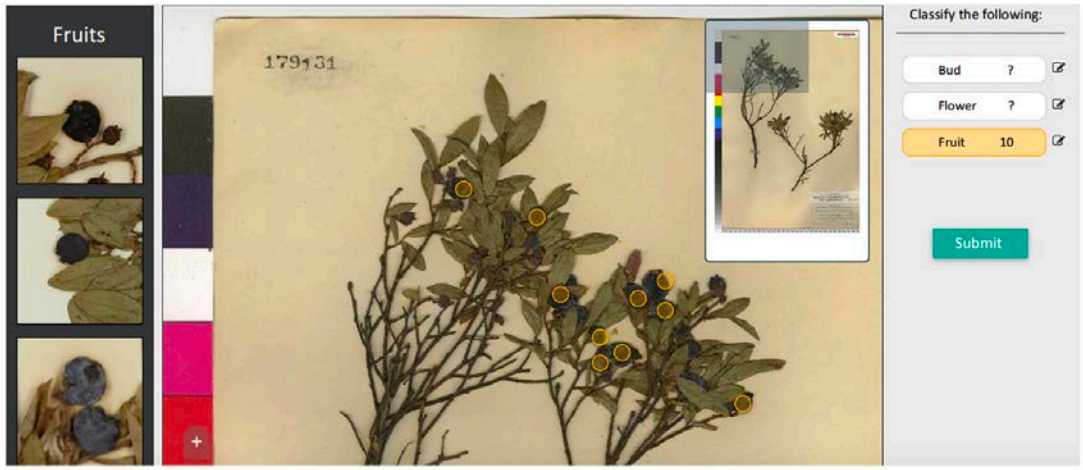
Phenological data – as described in label notes



"tree in full flower…"

or from image itself:

CrowdCurio, in Willis et al. 2017



Traits from Labels and Images

The need for standards and ontologies…
Plant Phenology Ontology

Willis et al. 2017

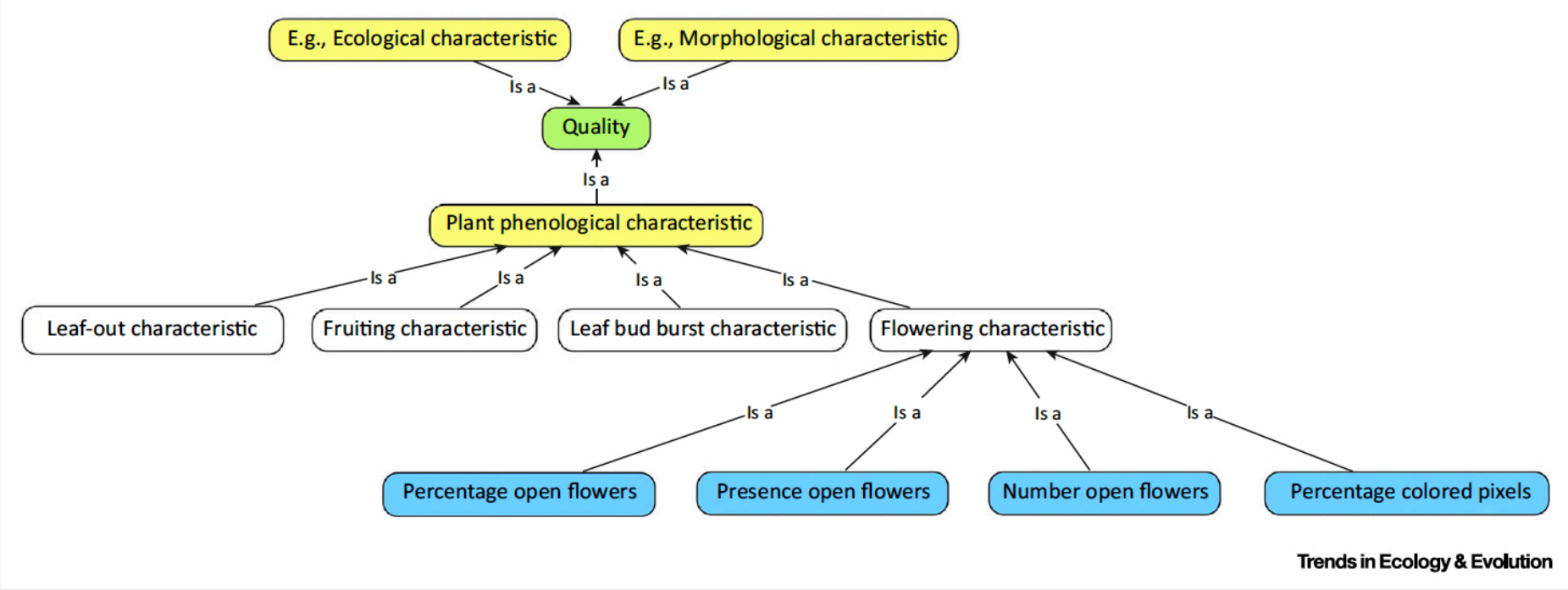


Figure I. Simplified Representation of Ontological Classes and Logical Structure. In a complete ontology, each term or 'class' has a specific definition and is linked to any and all related classes via 'relation terms' such as 'is_a' or 'part_of'. These structured linkages between classes allow integration among different methods of measuring a class (represented in blue), different subclasses within a class (white), and other types of data (yellow), which are subclasses of the general term 'quality' currently defined by the Phenotypic Quality Ontology.

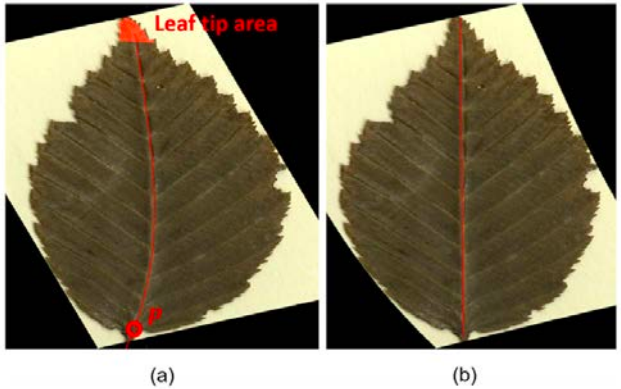
Traits from Labels and Images

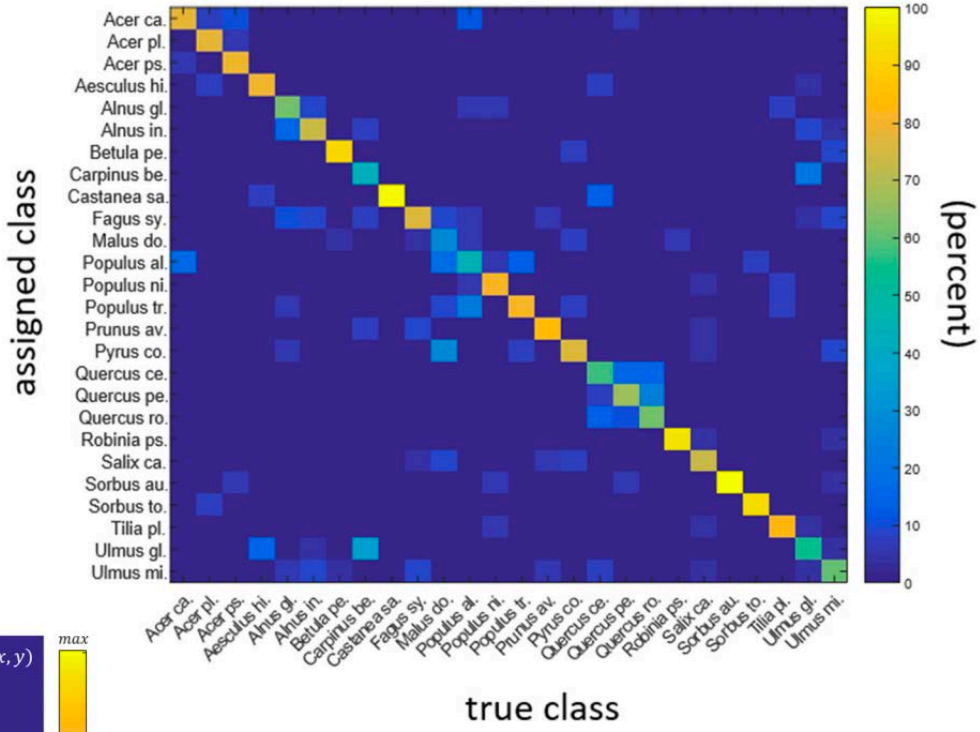
Machine Learning: Herbarium specimens

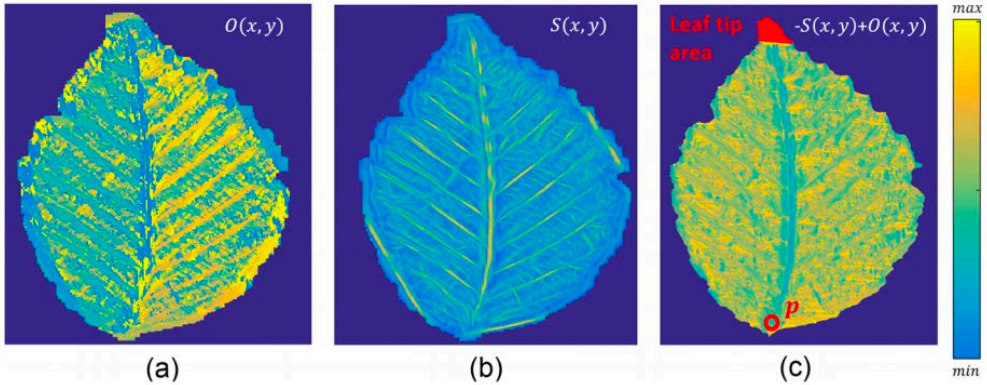
Classifying German trees to species

Leaf shape, venation

85% accuracy

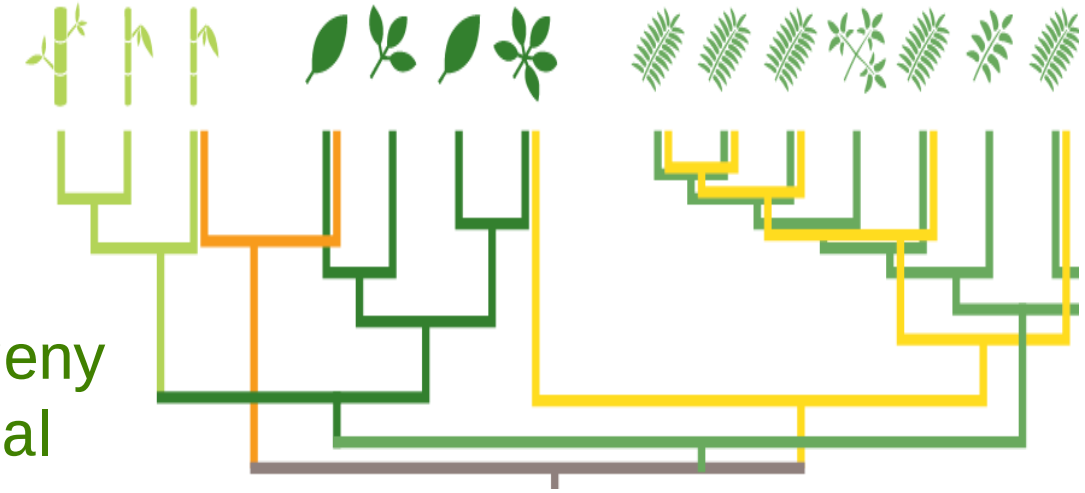






Unger et al. 2016

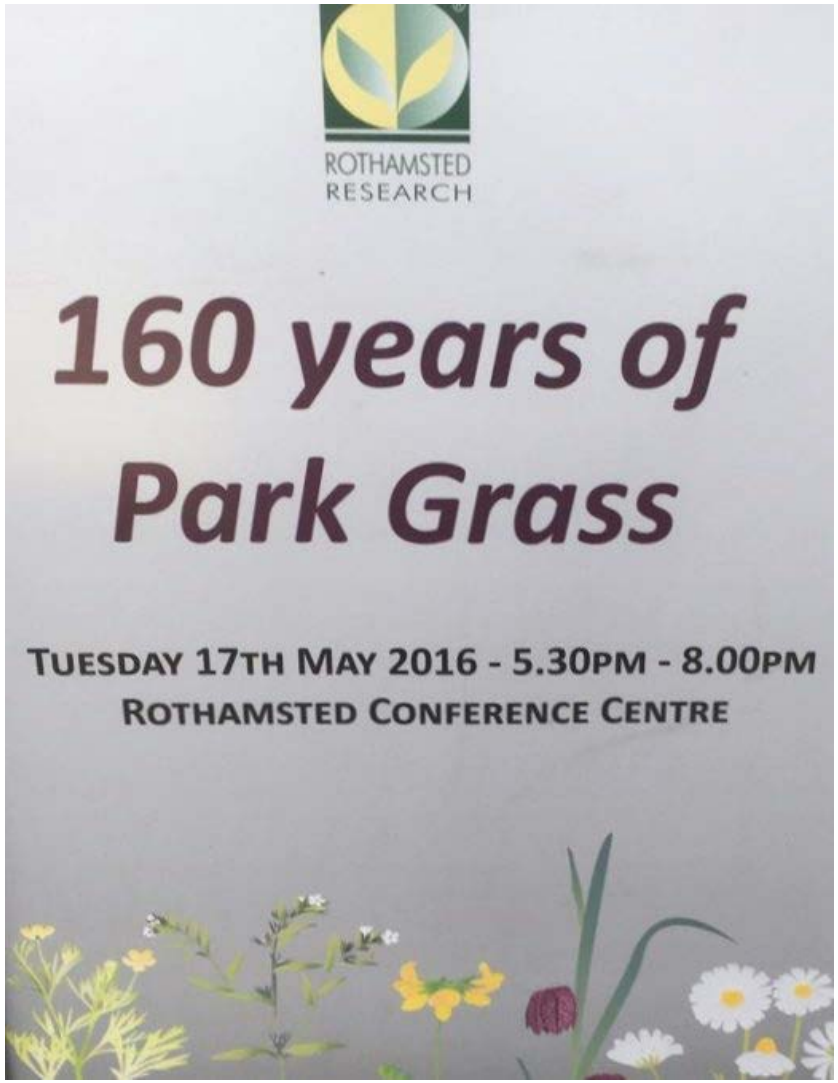
Traits from Labels and Images



Connect to ecology/phylogeny

Evolution of plant functional traits

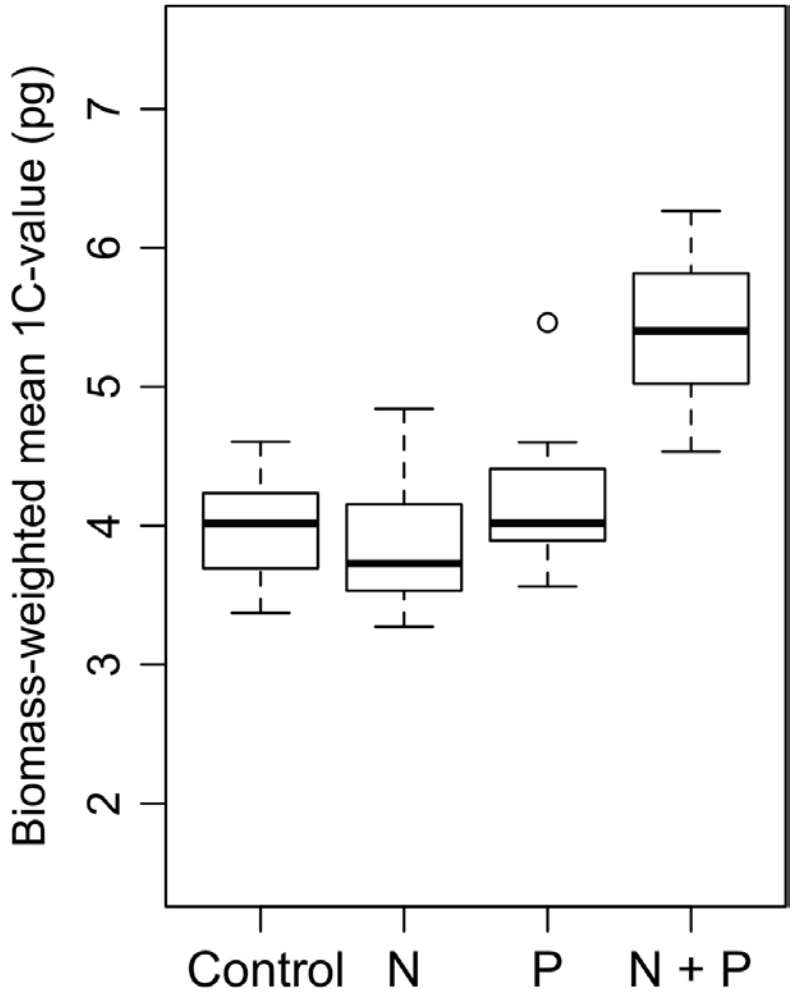
Spatial Distribution of Genome Sizes



From Tilman & Isbell 2015

Spatial Distribution of Genome Sizes

- Nucleic acids require large amounts of N and P; large genomes costly to build
- Plants with large genomes should be selected against on N- and P-poor soil, favored on high-N/P soil
- Park Grass Experiment used to test this hypothesis:
  - GS of plants in high N+P plots higher than in control, N, P plots
- Continental scale, GS related to soil geochemistry?



Guignard et al. 2016

Spatial Distribution of Genome Sizes

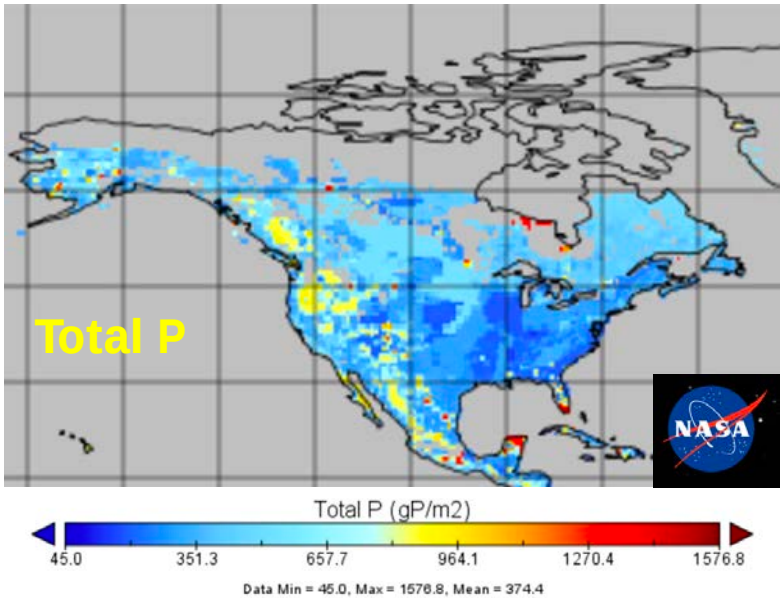
CCDB CHROMOSOME COUNT

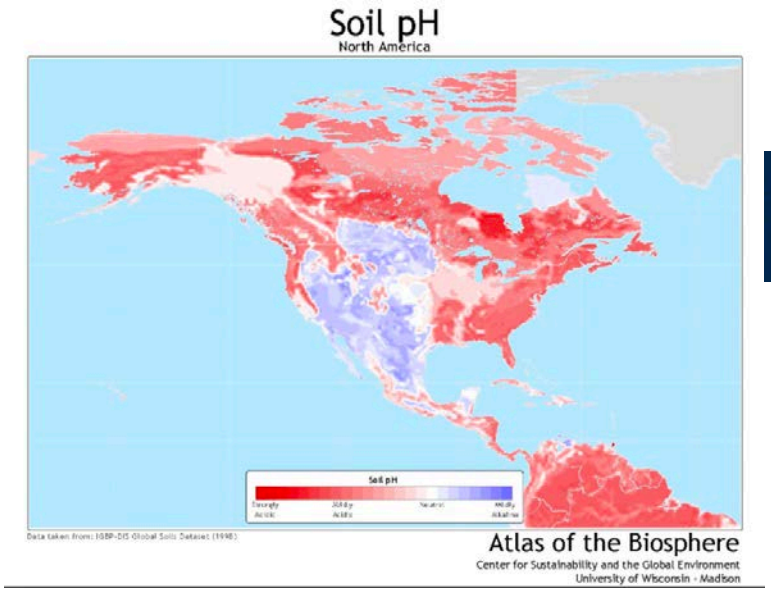
Home About Browse Services Contact

GeoEcoEvo:
USGS Powell Center Working Group
E. Bui, M. Goldhaber, PIs

iDigBio
Integrated Digitized Biocollections

Kew
ROYAL BOTANIC GARDENS
Plant DNA C-values Database





USGS
science for a changing world

USDA

Challenges in Linking Heterogeneous Data

- Assembling data
- Data management and sharing
- Taxonomic names
- Patchy data
- Issues of scale: resolution, analysis
- Data integration

Linking Heterogeneous Data: Connecting Specimens, Trees, Tools

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

D. Soltis, P. Soltis, J. Fortes,
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
revolutionary workflows

Linking Heterogeneous Data: Connecting Specimens, Trees, Tools

5 Possible Workflows

RESOURCES:

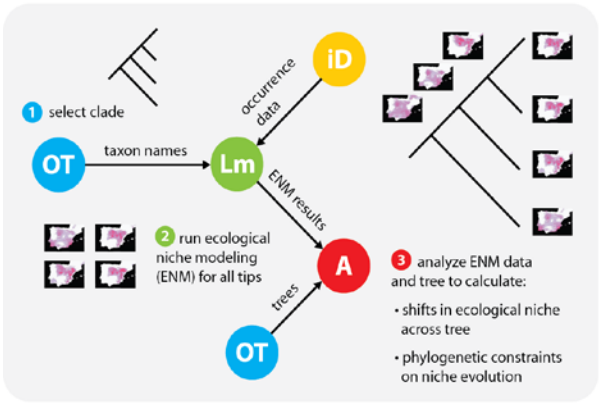
Lm Lifemapper
- ecological niche modeling
- biodiversity and range analysis
- visualization

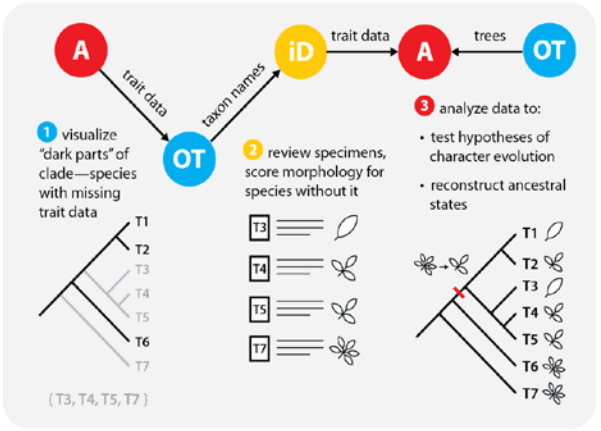
A Arbor
- evolutionary models
- comparative methods
- visualization

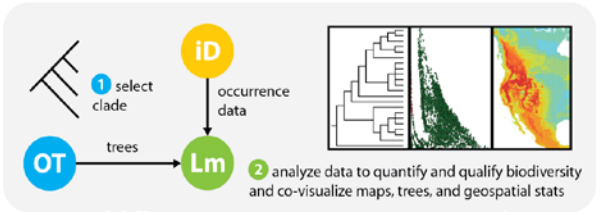
OT Open Tree of Life
- phylogenies
- taxonomy / names
- visualization

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

EXAMPLE WORKFLOWS:







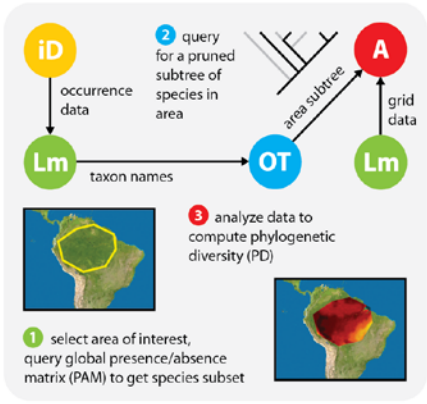
RESOURCES:

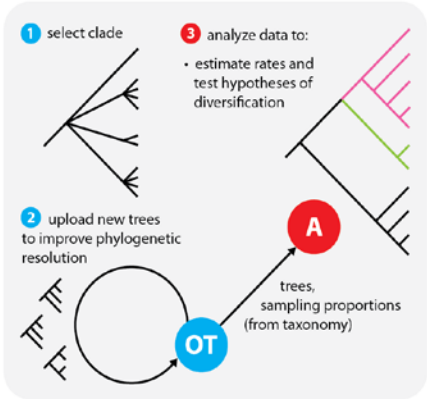
Lm Lifemapper
- ecological niche modeling
- biodiversity and range analysis
- visualization

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





Linking Heterogeneous Data: Connecting Specimens, Trees, Tools

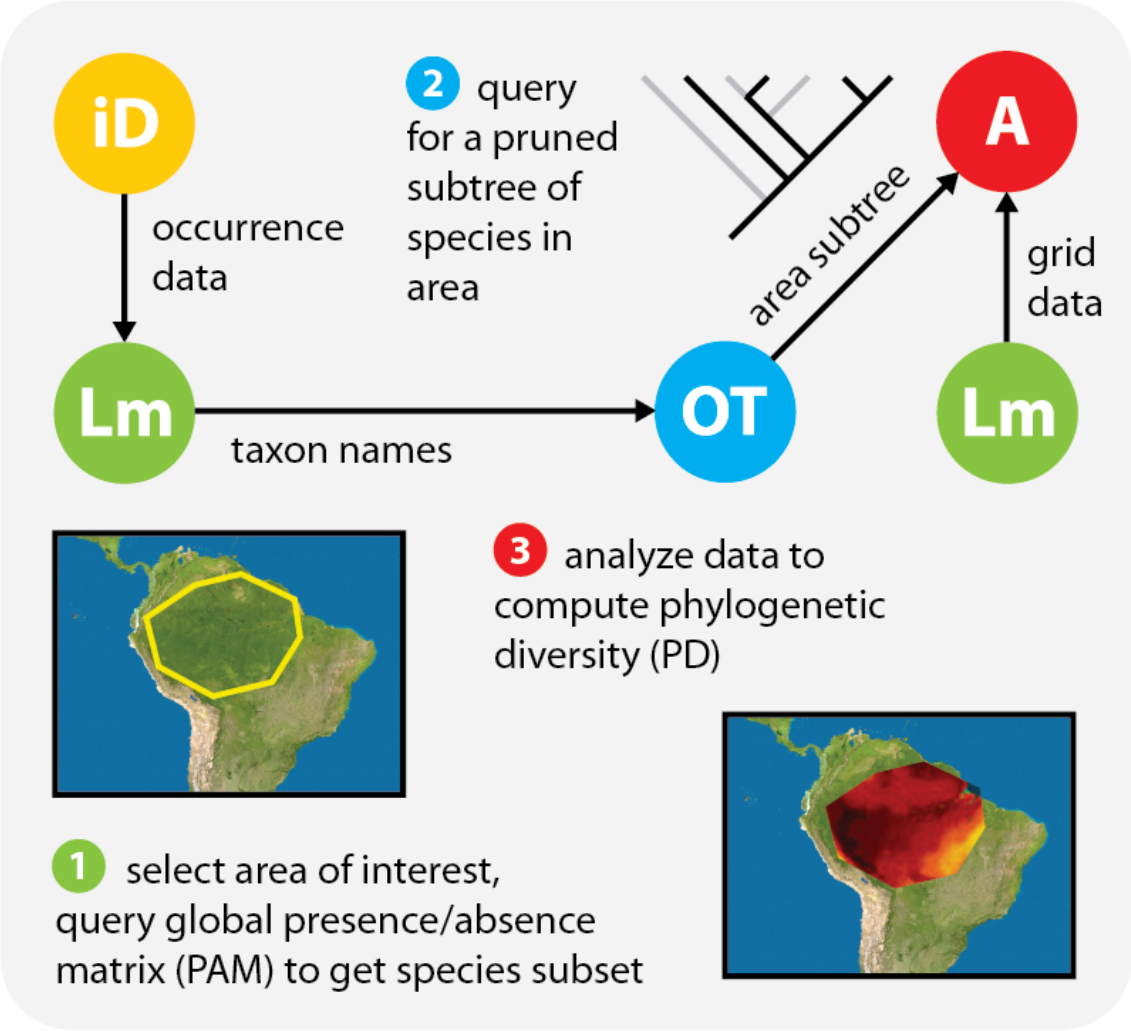
RESOURCES:

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

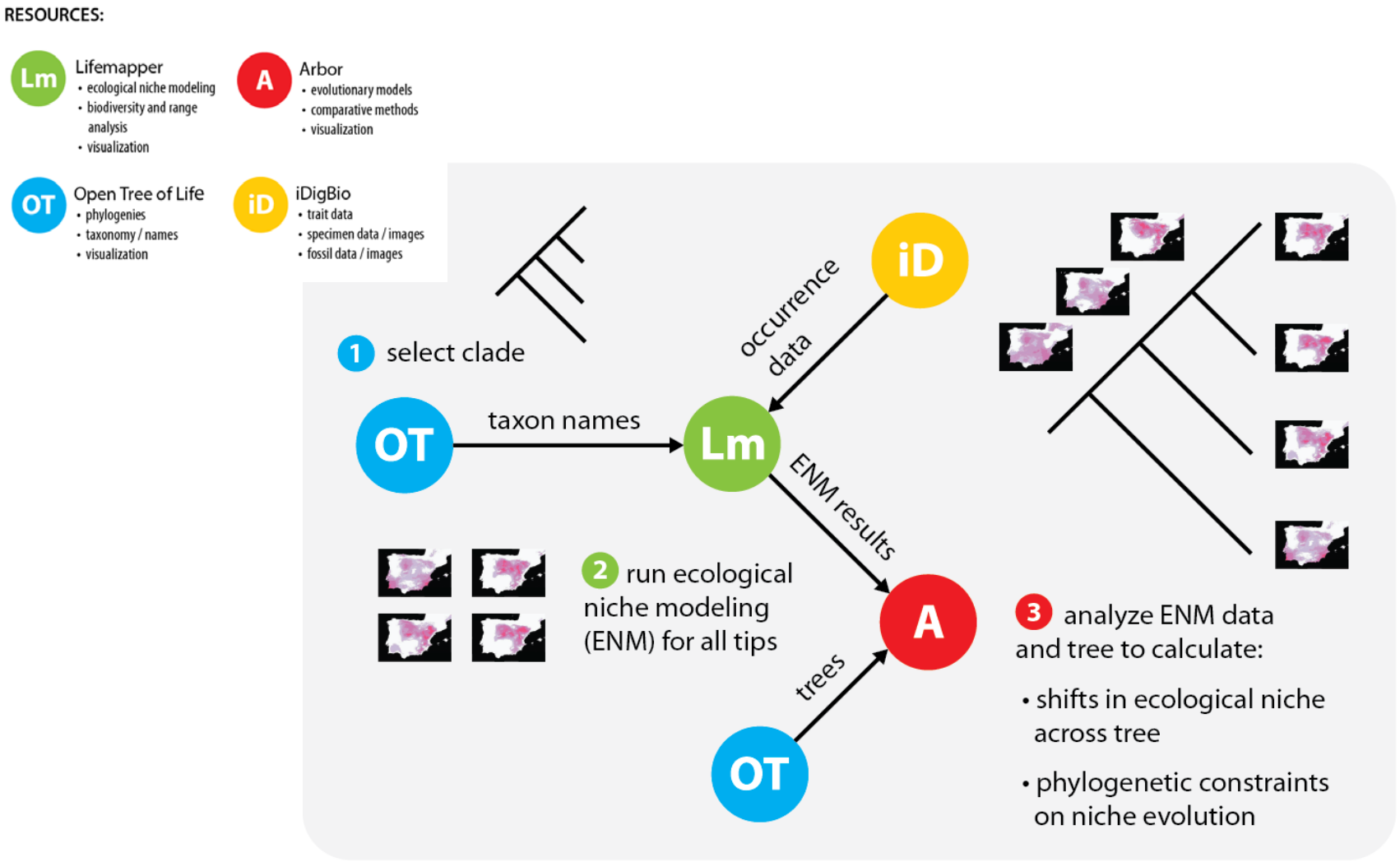
A Arbor
- evolutionary models
- comparative methods
- visualization

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

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



Linking Heterogeneous Data: Connecting Specimens, Trees, Tools



Summary

- Linking heterogeneous data challenging
  - Assembling data
  - Data management and sharing
  - Taxonomic names
  - Patchy data
  - Issues of scale: resolution, analysis
  - Data integration
- Value of spatial data from specimens
- Promise of images
- Need for tools, workflows
- Data-driven and hypothesis-driven research

Acknowledgments

iDigBio Team:

L. Page, J. Fortes, B. MacFadden, G. Riccardi, A. Mast, G. Nelson, D. Paul, S. James, C. Germain-Aubrey, B. Marchant

ABI Innovation: BiotaPhy Project:

Connecting resources to enable large-scale biodiversity analyses

D. Soltis, R. Folk, J. Fortes, A. Thompson, J. Beach, A. Stewart, J. Soberon, S. Smith

Florida Phylogenetic Diversity:

J. Allen, C. Germain-Aubrey, K. Neubig, L. Majure, R. Abbott, M. Whitten, N. Barve, H. Owens, J. M. Ponciano, B. Mishler, S. Laffan, R. Guralnick, J. Jantzen, D. Soltis

GeoEcoEvo Working Group, Powell Center, USGS:

E. Bui, M. Goldhaber, V. Funk, J. Miller, B. Edwards, C. Mason, B. Annaker, I. Pearse, J. Cartwright, J. Thompson, T. Nauman, M. Helmus

Genome Size & Geochemistry:

A. Leitch, I. Leitch, D. Soltis, A. Thompson, A. Stewart