Big Data and Biodiversity Studies:
Integrating Genetic, Evolutionary, and Ecological Data

- What is the history of life on Earth?
- How are species distributed in geographical and ecological space?
- What factors lead to speciation, dispersal, and extinction?
- What are the impacts of climate change likely to be?
- What are the effects of invasive species?
- Where have exotics been introduced, and how quickly are they spreading?
- What information is needed for effective conservation strategies?

Pamela S. Soltis
Florida Museum of Natural History
University of Florida

Biodiversity Data in Natural History Collections

- 1 billion specimens in 1600 natural history collections

- Important sources of information on past and present species distributions

- Location information and environmental data

- Associated metadata: GenBank accession numbers

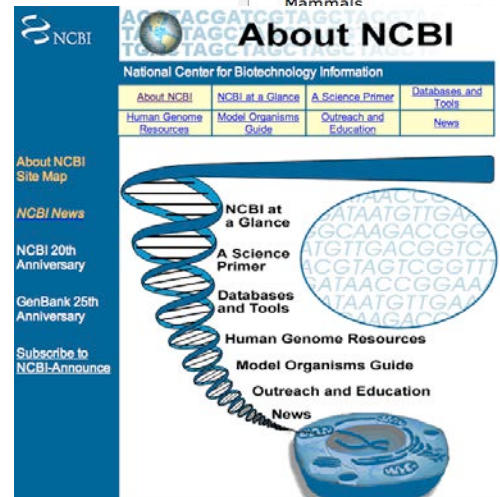- Associated collections: DNA, songs, calls
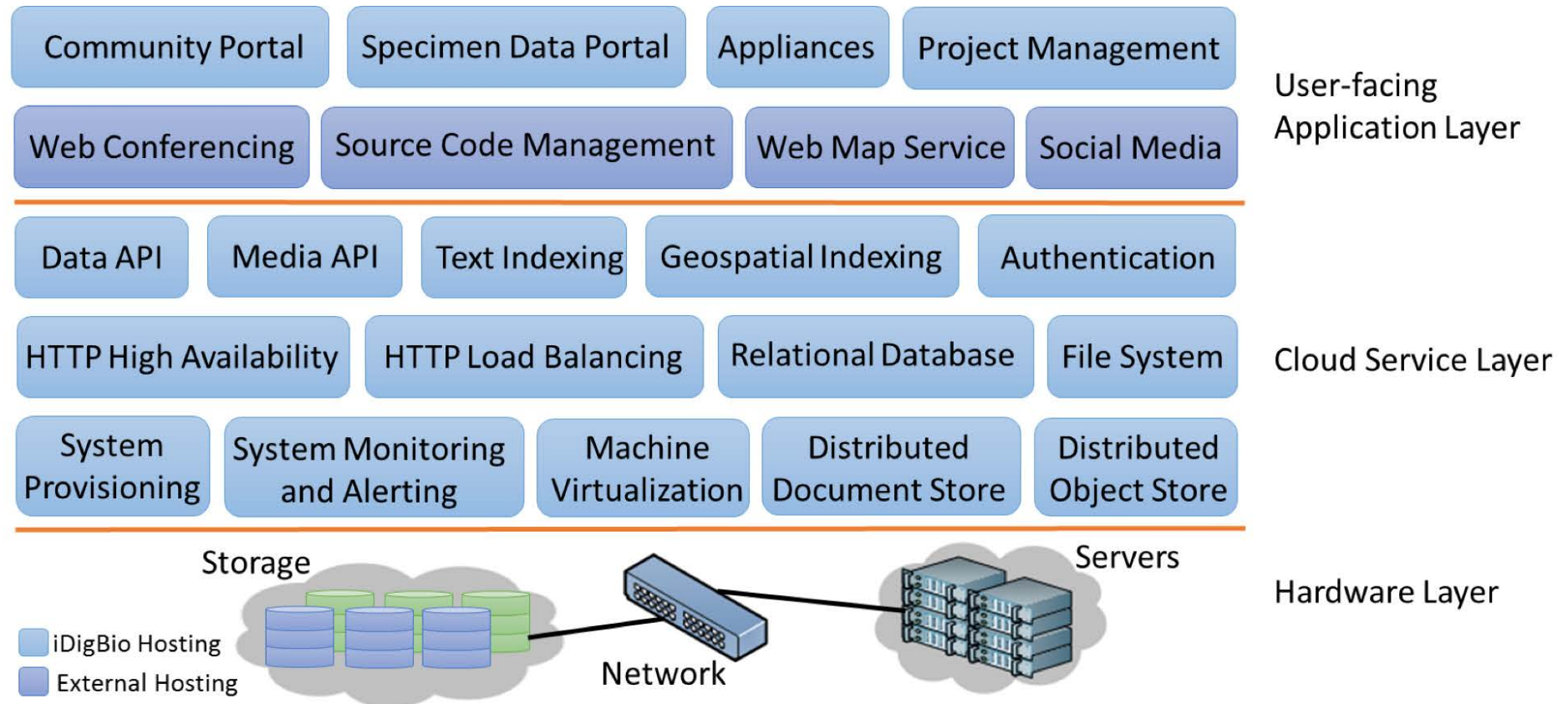
Linking Natural History Collections to…

- Ecology
- Paleontology
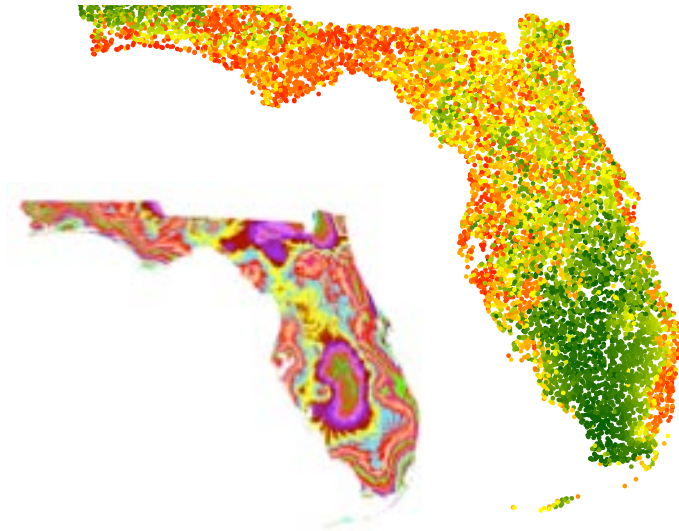- Genomics
- Phylogenies
- GIS data
- Climate Models
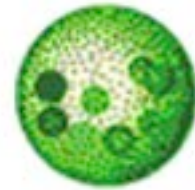
Building the iDigBio Cloud

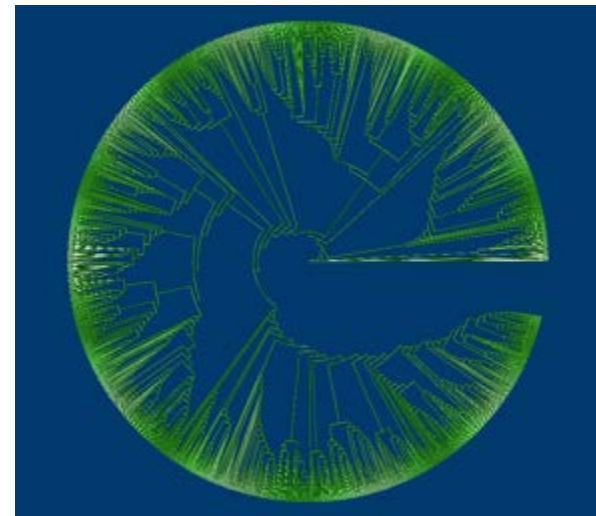L. Page, B. MacFadden, P. Soltis, FLMNH; J. Fortes, ECE; G. Riccardi, FSU

Florida Plant Diversity in a Changing Climate

Integrating specimen data, climate change models, and phylogeny

+

2609 species (of ~4200)
all included in phylogeny

Phylogenetic tree, 2609 species
GenBank, new (1000 spp)

C. Germain-Aubrey, J. Allen, K. Neubig, L. Majure, R. Abbott, G. Burleigh, D. Soltis
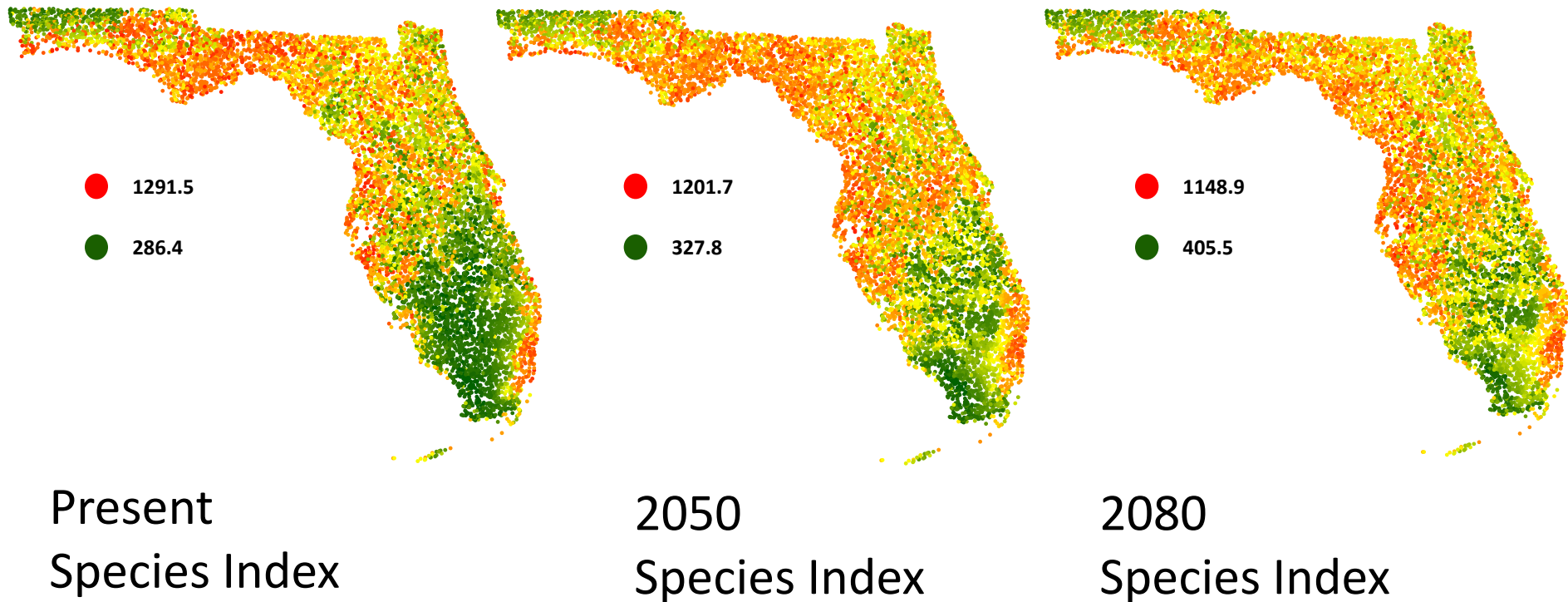
Using Museum Specimens and Computer Models in Biodiversity Studies

- Herbaria important sources of information on past and present species distributions

- Location information and environmental data
  - temperature, precipitation, soil

- Software to model the range of each species

- Project onto future climate conditions

- >2700 plant species (of 4200) >511,000 plant location records

Species Diversity: Future
species will move south
more precipitation than in north

🔴 1291.5	
🟢 286.4	

Present
Species Index

🔴 1201.7	
🟢 327.8	

2050
Species Index

🔴 1148.9	
🟢 405.5	

2080
Species Index

High species diversity   Low species diversity

Florida Plant Diversity in a Changing Climate

Computational & Data Requirements:

access to georeferenced specimen data

access to GenBank

access to Bioclim information

tools for georeferencing

software for phylogenetics

software for ecological niche modeling

ability to incorporate custom scripts

Formal development of workflow

application to FLMNH fish & Lepidoptera

C. Germain-Aubrey, J. Allen, K. Neubig, L. Majure, R. Abbott, G. Burleigh, D. Soltis