Using Digitized Specimen Data in Research: Applications for Ecology, Phylogenetics, and Biogeography

Pam Soltis
University of Florida

FLORIDA MUSEUM

UF UNIVERSITY of FLORIDA

iDigBio
Integrated Digitized Biocollections

NSF

BiotaPhy

Topics

- Intro to data
- Downloading data
- Cleaning data
- Georeferencing
  - GEOLocate
- Ecological Niche Modeling
  - Maxent
- ENM Analysis & Interpretation
- Scaling Up – BiotaPhy
- Additional uses of collections data
- Discussion/Q&A
- Manual and R-based options

Workshop Leaders

- Pam Soltis
- Makenzie Mabry
- Lauren Gillett
- Shelly Gaynor
- JT Miller
- Elizabeth White
- Maria Cortez
- Doug Soltis
- Malu Ore Rengifo
- Contributions from many previous lab members/colleagues:
  - Blaine Marchant, Charlotte Germain-Aubrey, Andre Naranjo, Anthony Melton, Tal Kinser, Mike Belitz, Rhett Rautsaw, João Vidal, Jr.

FLORIDA MUSEUM

UF UNIVERSITY of FLORIDA

iDigBio
Integrated Digitized Biocollections

BiotaPhy

Logistics

- Dropbox folder
  - https://www.dropbox.com/sh/gtvo3z20rw5j7wt/AADKEhCHgHxxM9gjT8MPq8c4a?dl=0
  - Word doc with schedule and instructions
  - Presentations folder
  - Demos folder
- Workshop evaluation survey
  - by email
- Pace/breaks


FLAS 22883


R. K. Godfrey Herbarium (FSU)
HERBARIUM
177581
FLORIDA STATE UNIVERSITY
The New York Botanical Garden
Woody Plants of U. S. A.
ACERACEAE Acer rubrum L. Female
Small tree, 8 m. tall, with two stems from base, 15 cm. in diameter.
16709-a April 22, 1979. Tree in full bloom, exclusively female (see no. 16708, a nearby male tree). Tips of calyx and corolla lobes light red as in no. 16708. Stigmas bright red.
16709-b June 4, 1979. Leaves dark green above, pale green below. Fruits pale tan.
Common in bottom of hollow, ungrazed woods of Quercus alba, Fraxinus americana, Populus grandidentata, Carya ovata.
WISCONSIN: Richland Co., 3 miles SE of Richland Center.
T 10 N; R 1 E; SE¼SW¼ sec. 25. alt. 250 m.
Coll.: M. Nee


HERBARIUM
UNIVERSITY OF FLORIDA
GAINESVILLE
139638

Herbaria Worldwide

NYBG STEERE HERBARIUM

Home Collections Discover Index Herbariorum Virtual Herbarium Loans Digitization

Index Herbariorum

3,400 herbaria

400,000,000 specimens!

Natural History Collections

1-2 billion specimens
in the US
3-4 billion specimens
worldwide

~1600 natural history collections in the US, thousands in the world

Smithsonian

Systematics & Taxonomy

The Linnean Collections

LINNEAN SOCIETY of London

Linnean Society Home

Linnaea borealis, 1732

Carl Linné, aka Carolus Linnaeus

Uses of Natural History Collections

…Species interactions
Phenology
Biogeography
More!

Genetics
Genomics
Chemistry…

Smithsonian

Uses of Natural History Collections

Most specimens locked
away in cabinets, unavailable
for general use.

Smithsonian

Uses of Natural History Collections

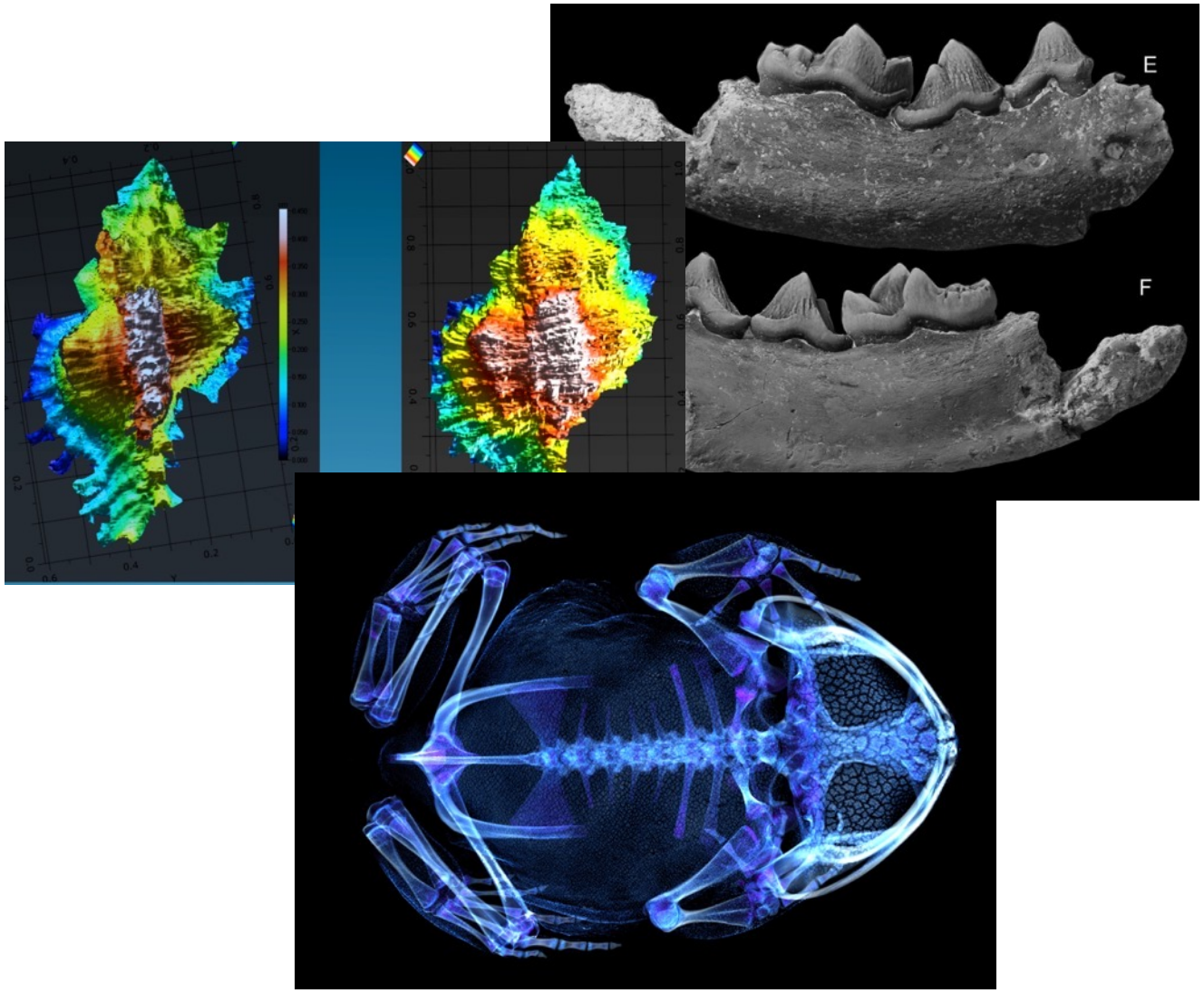
DIGITIZATION

Smithsonian

Digitization: Data and Images

- Scientific name –
  including authority
- Date
- Collector
- Location – state, county,
  specific site,
  GPS coordinates
- Associated species
- Notes

Digitization: Data and Images



iDigBio: www.idigbio.org

iDigBio
Integrated Digitized Biocollections

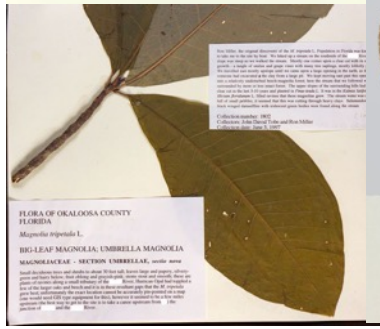
National Coordinating Center
For Digitization of Biodiversity Collections

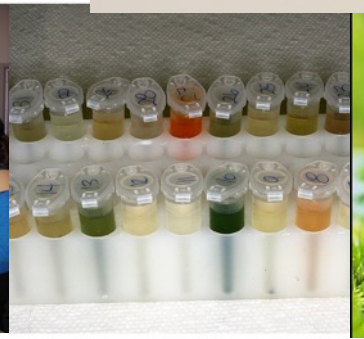
Ingest, serve, integrate data:

Localities

Dates

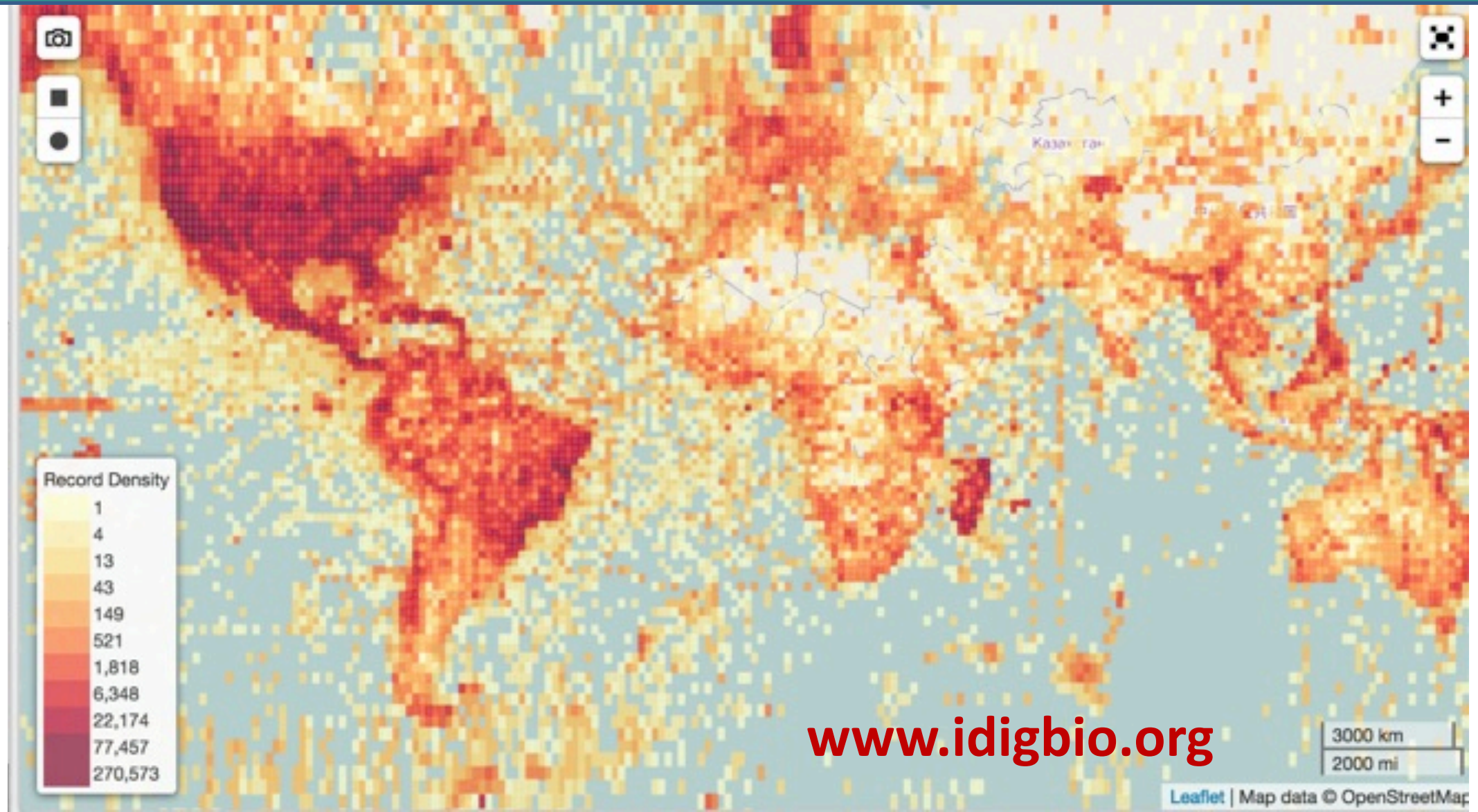
Images





iDigBio: A Source for Specimen Records

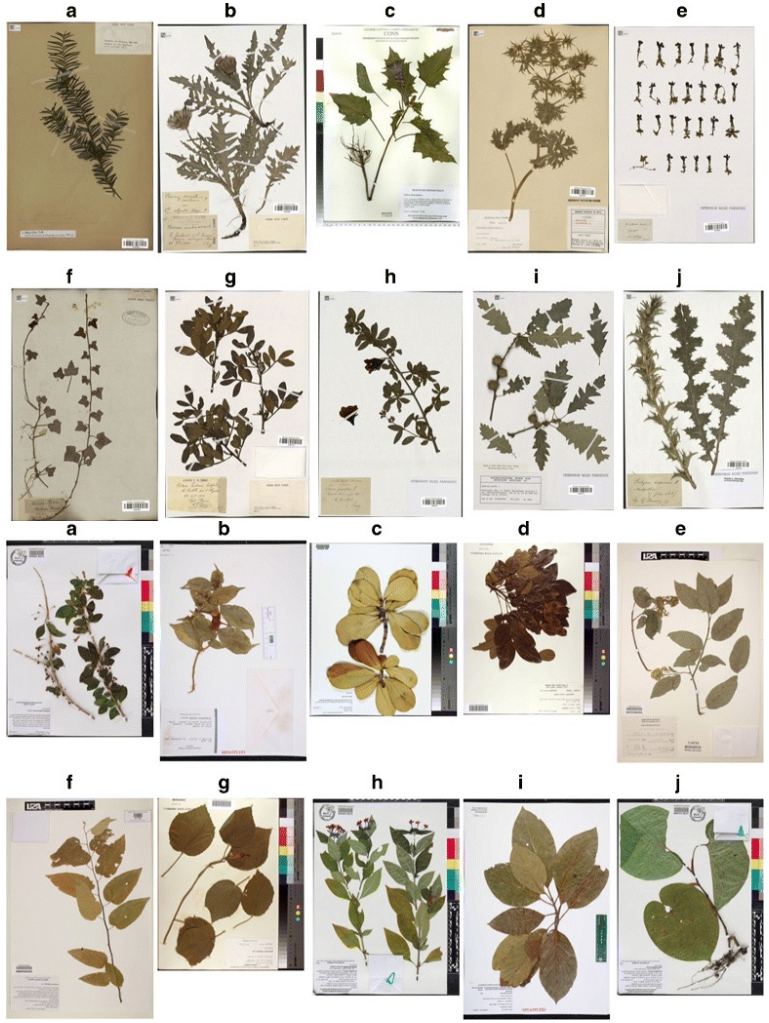
Specimen Occurrences in iDigBio



Label Data from Herbarium Specimens

- Scientific name –
  including authority
- Date
- Collector
- Location – state, county,
  specific site,
  GPS coordinates
- Associated species
- Notes

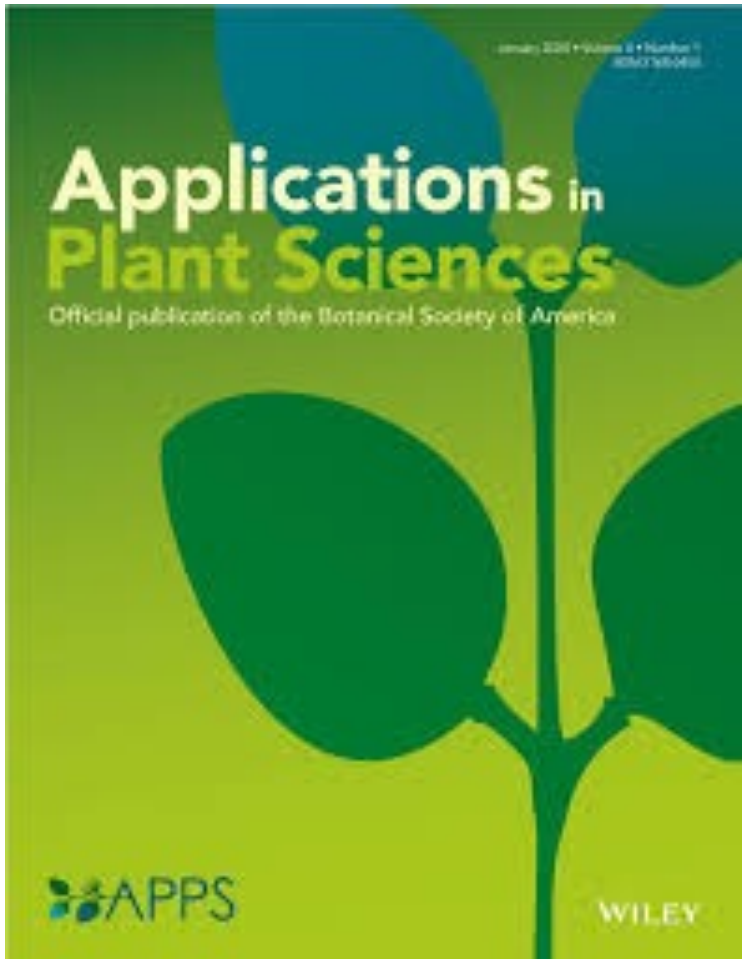
Machine Learning & Biodiversity Research



Carranza-Rojas et al. 2017

Pearson et al. 2020

Machine Learning & Plant Biology: Special Issue of *APPS* (2020)



June: Herbarium Specimen Images

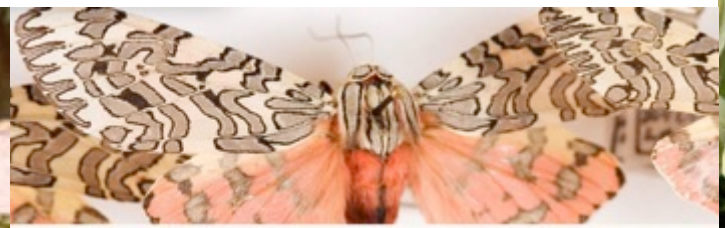
- Phenology (2)
- Morphology
- Herbivory
- Species Identification (2)
- Software: Leaf Detection
- Software: Leaf Recognition and Measurement

Edited by P. Soltis, G. Nelson, A. Zare, E. Meineke

Other Data Aggregators

‘Big Data’ Research in Biodiversity Science

- Monitoring shifts in biodiversity
- Tracking invasive species
- Ecological Niche Modeling, climate change
- Past movements and climate change
- Tracking phenological shifts
- Landscape genetics
- Integration of ENM with phylogeny
- Community phylogenetics/assembly
- Biogeography

Overview of Day's Activities

8:00	Welcome and Overview of the Workshop – Pam
8:15	Darwin Core and Data Fields – Pam
8:30	Data Downloading Overview – Makenzie
9:00	***Activity:*** Manual Data Downloads, iDigBio Portal – Lauren
9:30	***Activity:*** R-based Data Downloads – Shelly
9:45	Data Cleaning Overview – JT
10:00	**Break**
10:30	***Activity***: Data Cleaning (A) Manual (JT)
	(B) R-based (Shelly)
11:00	Georeferencing Overview – Lauren
11:15	***Activity*****:** Georeferencing (A) Manual (Lauren)
	(B) R-based (Makenzie)

Overview of Day's Activities

12:15	**Lunch**
1:00	Climatic Processing Overview – Shelly
1:20	***Activity***: Climatic Processing – (A) Manual (Elizabeth) (B) R-based (Shelly)
2:00	Ecological Niche Overview – Makenzie
2:15	ENM Settings Overview – JT
2:30	***Activity***: Ecological Niche Modeling (A) Manual (Elizabeth) (B) R-based (Shelly)
3:00	**Break**
3:30	Interpreting ENMs Overview – Shelly
3:40	***Activity***: Interpreting ENMs (A) Manual (Elizabeth/Lauren/Makenzie) (B) R-based (Shelly)
4:00	Post-ENM analysis – Shelly
4:15	BiotaPhy: An Overview – Doug/Maria
4:30	***Wrap-up and Q&A***
5:00	End