iDigBio
Integrated Digitized Biocollections

Using museum data for species distribution modeling: The case of plants in Florida

Charlotte Germain-Aubrey, Julie Allen, Robert Guralnick, Shawn Laffan, Brent Mishler, Kurt Neubig, Douglas Soltis, Lucas Majure, Pamela Soltis

UF UNIVERSITY of FLORIDA · FLORIDA STATE UNIVERSITY 1851 · FLORIDA MUSEUM OF NATURAL HISTORY · NSF

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

Niche Modeling: A widespread tool

- PRACTICAL
  - Invasive species
  - Disease
  - Data deficient species
  - Conservation/Land Management
- THEORETICAL
  - Diversity through time
  - Evolutionary and ecological patterns

Modeling the Distribution of Species

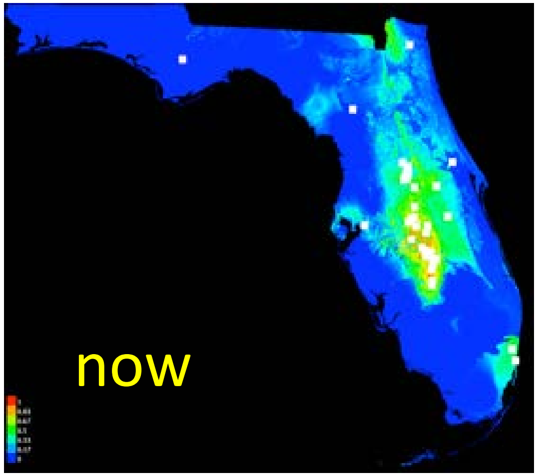
- Location information and environmental data
- Software to model the range of each species
- Project onto future climate conditions
- For Florida plants:
  - 1548 plant species (of 4100 species)
  - >511,000 georeferenced points (GPS)
  - Environmental features: temperature, precipitation, soil, etc.

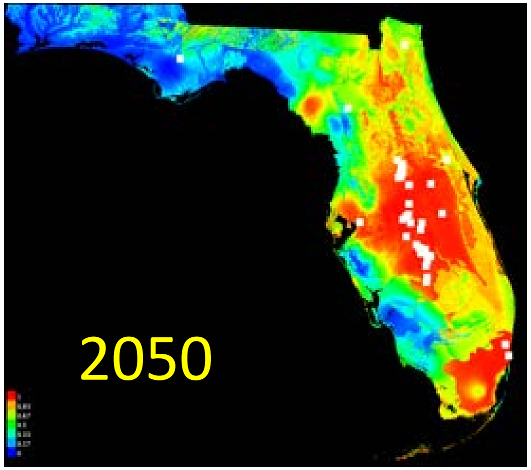
Niche Modeling: A widespread tool

- PRACTICAL
  - Invasive species
  - Disease
  - Data-deficient species
  - Conservation/Land Management
- THEORETICAL
  - Diversity through time
  - Evolutionary and ecological patterns

Response to Climate Change

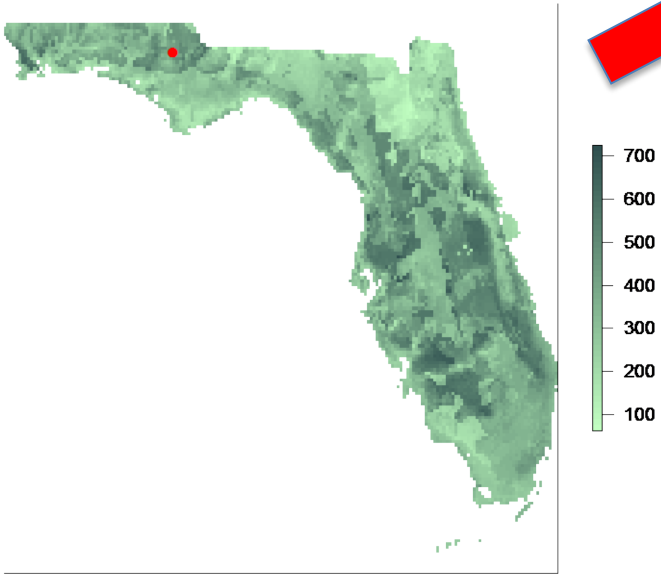
Prunus geniculata (scrub plum) – Lake Wales Ridge

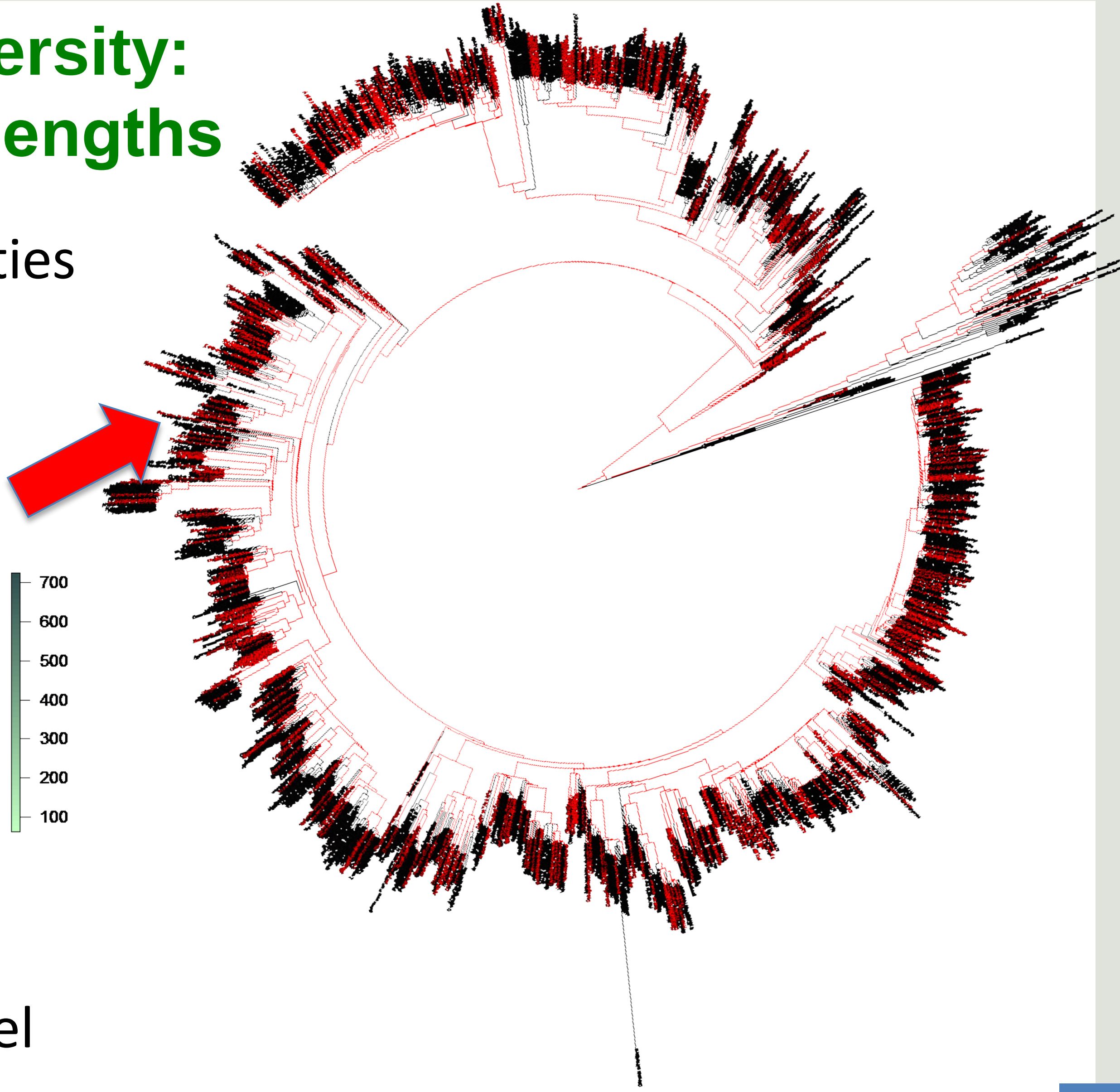




Phylogenetic Diversity: ≈ sum of branch lengths

8,045 pixels/communities
16 km^2 per pixel





Species list at each pixel generated from ENMs

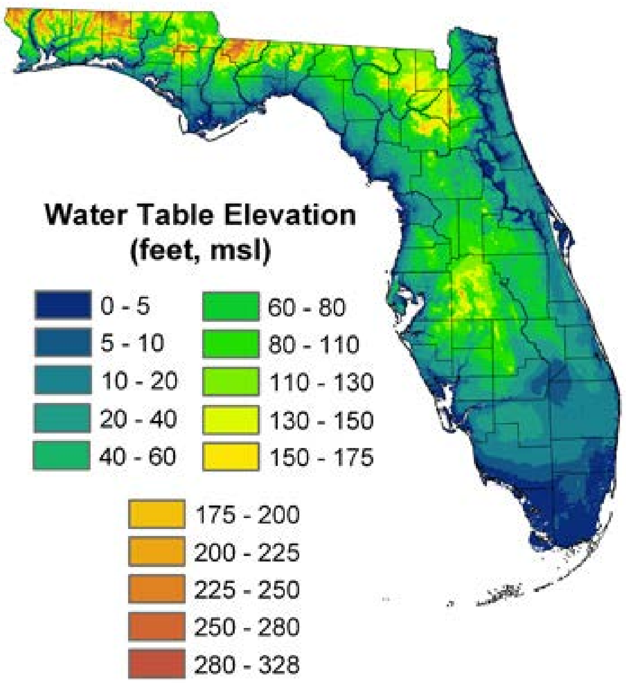
How does MaxEnt work?

- Uses presence only
- Using Maps:
  - Extract values from environmental layers
  - Creates background points as pseudo-absences
- Using Values:
  - You can feed MaxEnt the environmental values for each point
  - Allows one to extract values from different maps/inputs

Parameters to explore:

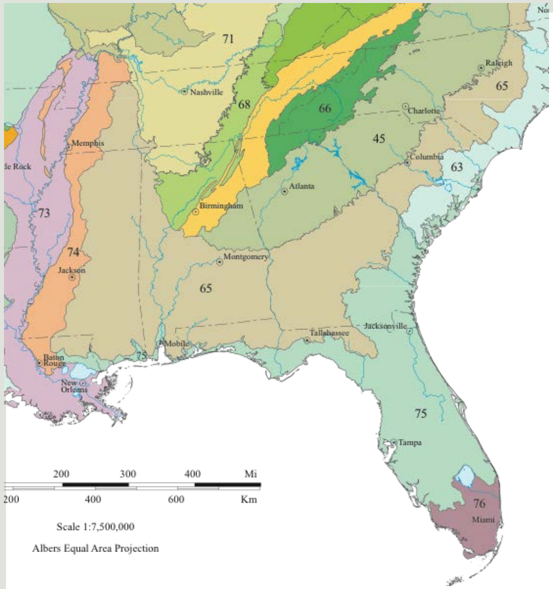
1. Using **yearly climate data** from time of collection to improve niche models using museum specimens
2. Using **reduced area** to train the models
3. **Smoothing response curve** to mimic more realistic response of organism
4. Using **fewer background points** to increase computing efficiency

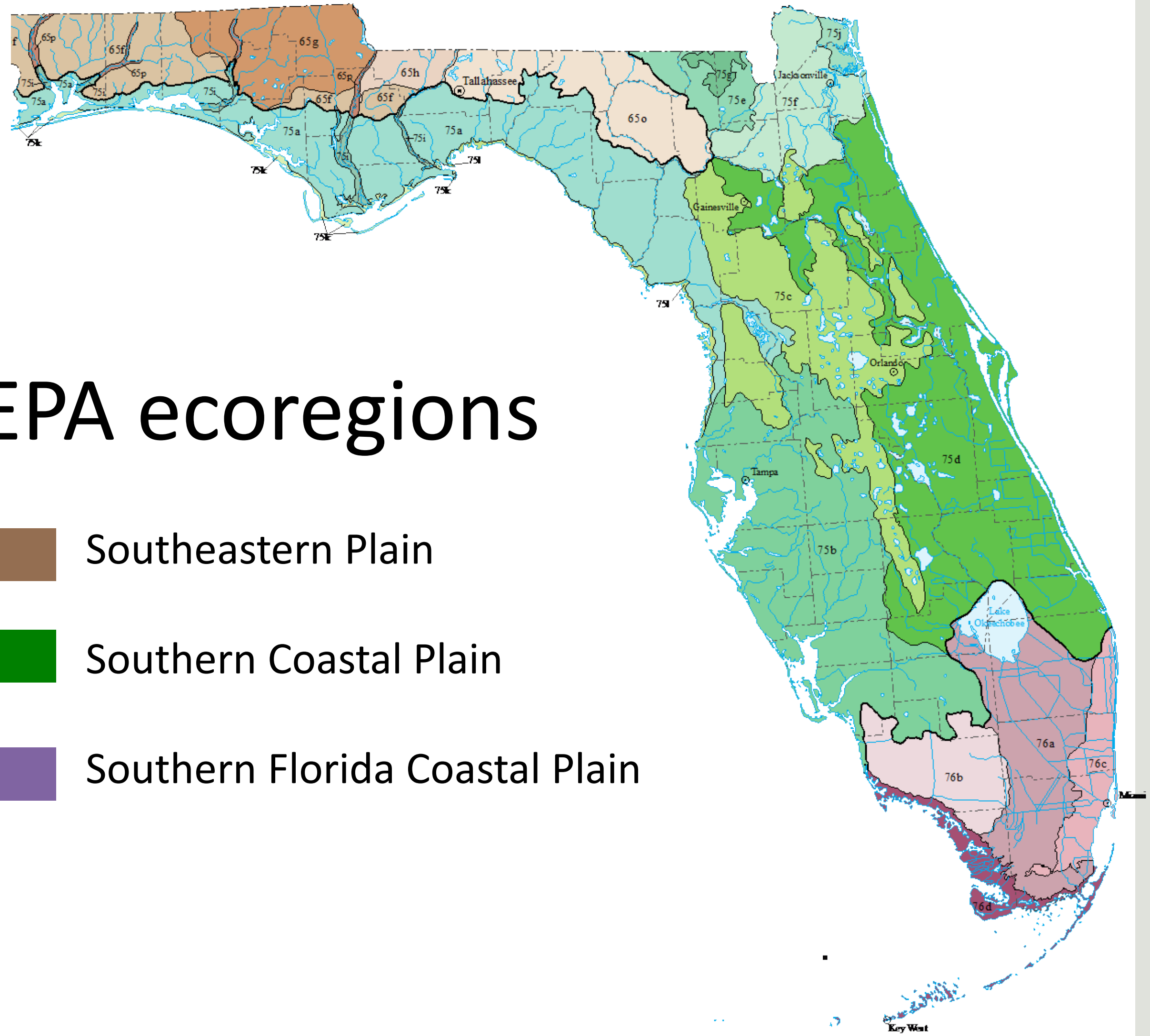
Florida


Water Table Elevation
(feet, msl)
0 - 5
5 - 10
10 - 20
20 - 40
40 - 60
60 - 80
80 - 110
110 - 130
130 - 150
150 - 175
175 - 200
200 - 225
225 - 250
250 - 280
280 - 328

iDiGBio Integrated Digitized Biocollections

Florida





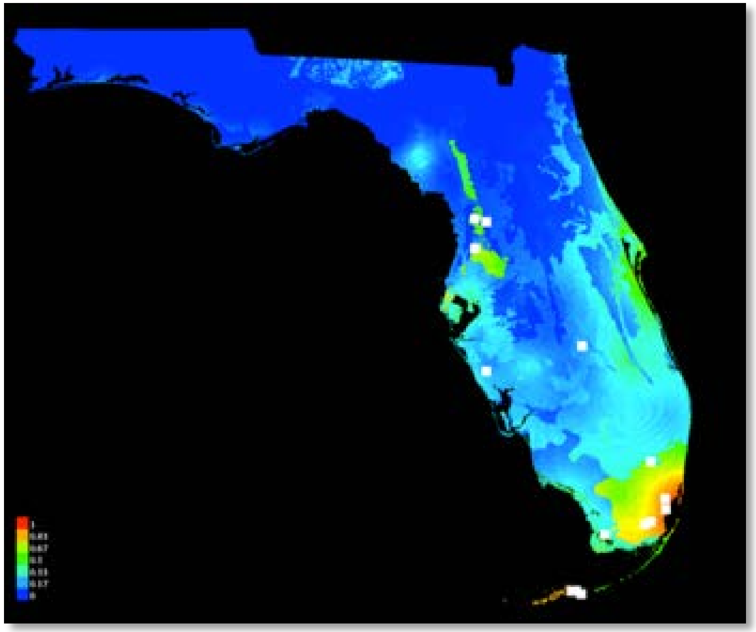
EPA ecoregions

- Southeastern Plain
- Southern Coastal Plain
- Southern Florida Coastal Plain

~ 4,100 plant species

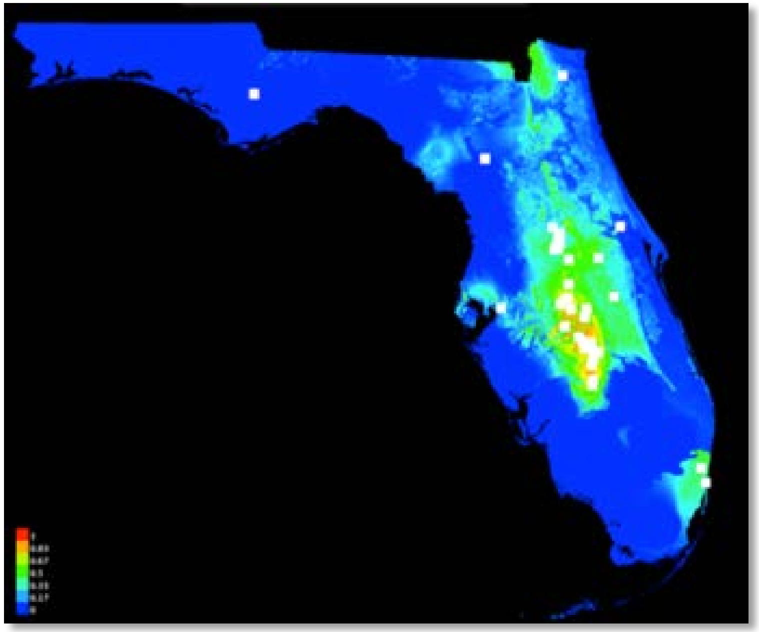
Flatspike Sedge

Abildgaardia obovata

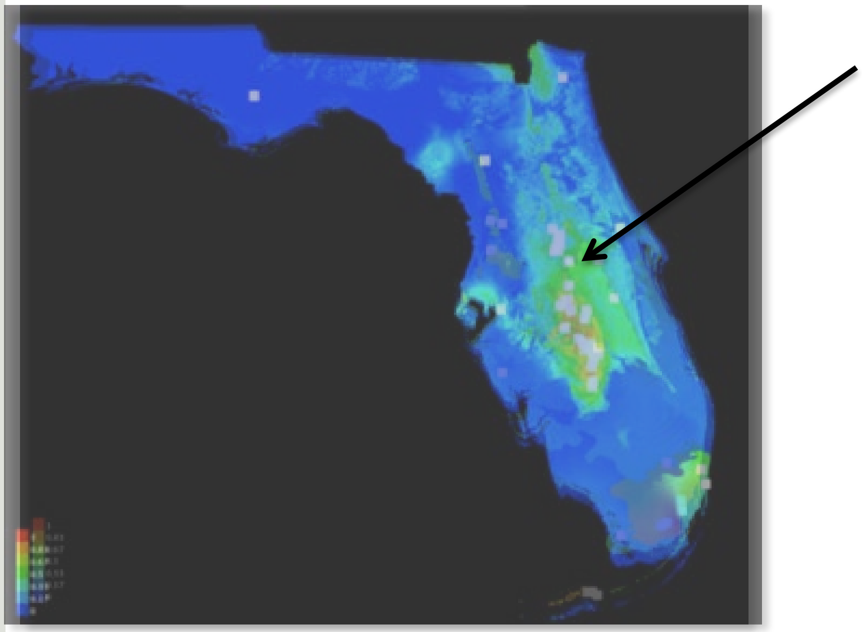


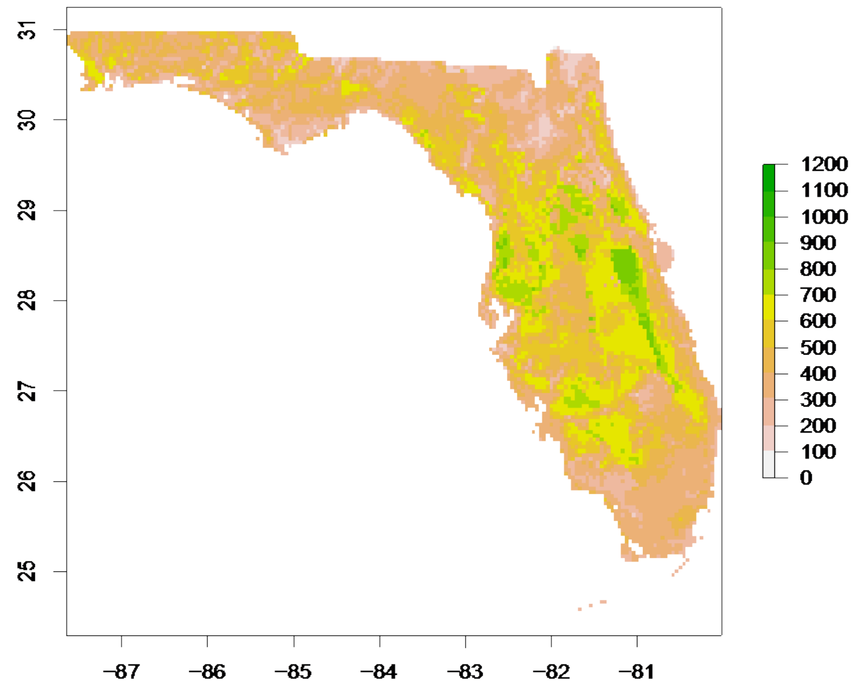
Scrub Plum

Prunus geniculata



What species are predicted to reside in this point?





Species Distribution Modeling

- Reconstructing species distributions
  - museum specimens
  - long-term monitoring data
  - *to generate ranges for use in biodiversity and PD analyses*

- Sampling in 3 EPA regions present in FL

- Taxonomic Name Resolution

- >30 points per species (10-20 for rare endemics)

  372,241 unique points

1,145 species models (30%)

Parameters to explore:

1. Using **yearly climate data** from time of collection to improve niche models using museum specimens
2. Using **reduced area** to train the models
3. **Smoothing response curve** to mimic more realistic response of organism
4. Using **fewer background points** to increase computing efficiency

Using Yearly Climate Data to reconstruct a species niche

- Dated (year)
- Reflect distribution of plants through recent history
- Can use the climate conditions at the time of collection
- Use PRISM monthly data
  - Reconstruct Bioclim layers for each year of collection
    - ➢ Run MaxEnt model

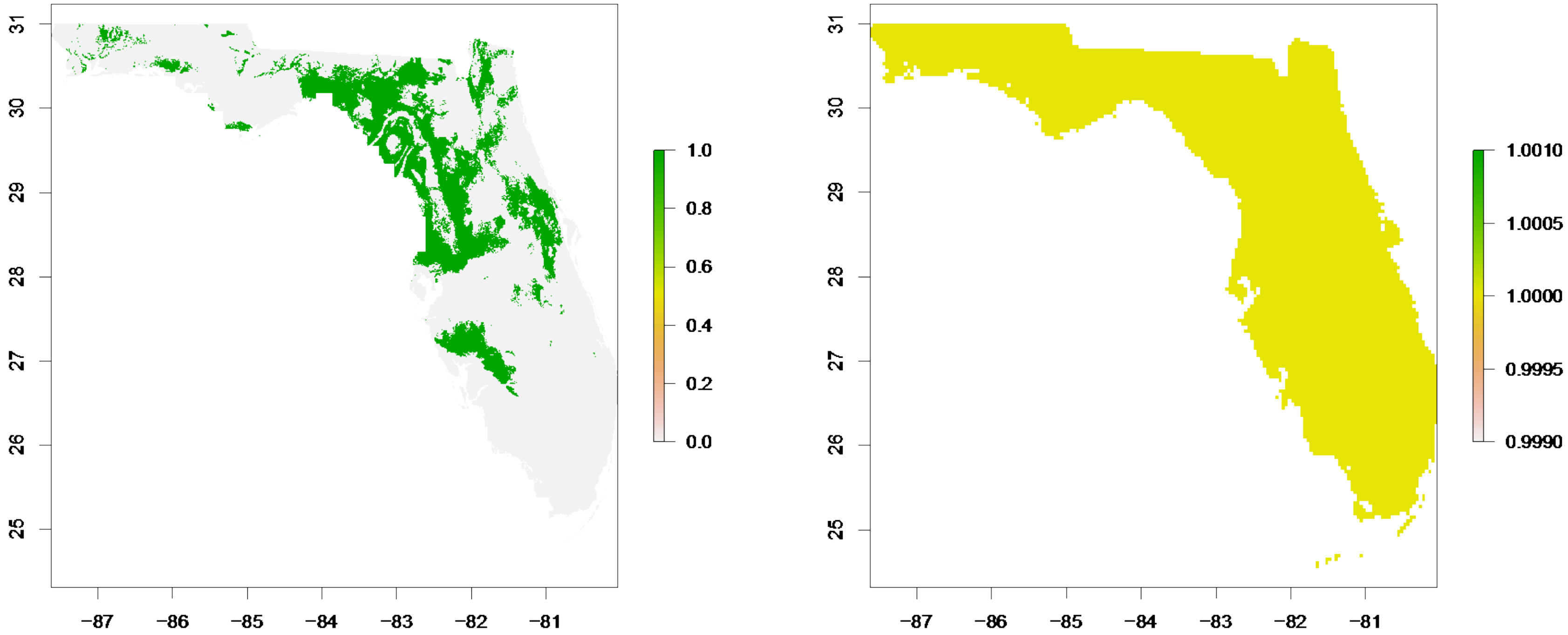
Collections: The Library of Life

>1600 natural history collections
in the US alone
1-2 billion specimens in the US
3-4 billion specimens worldwide

Long-lived species

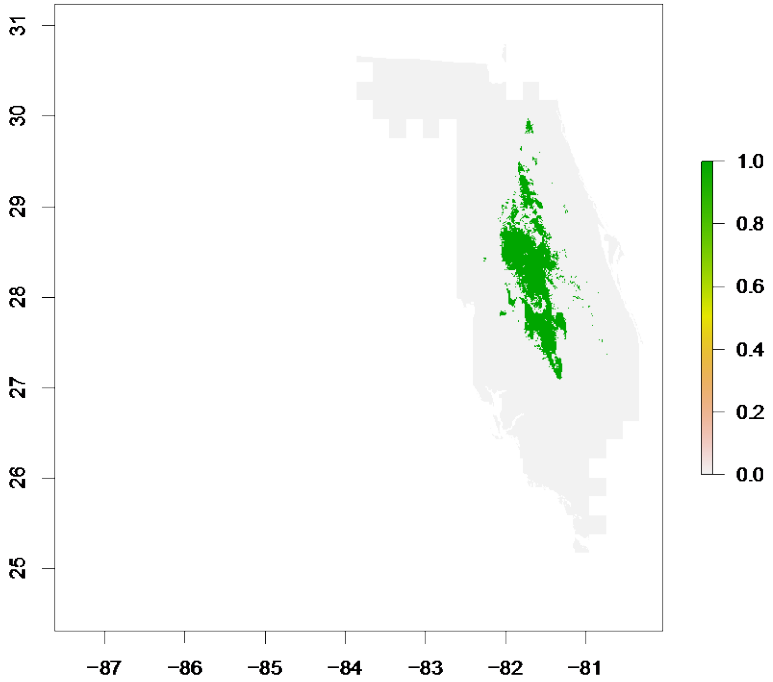
Average Climate

Yearly Climate

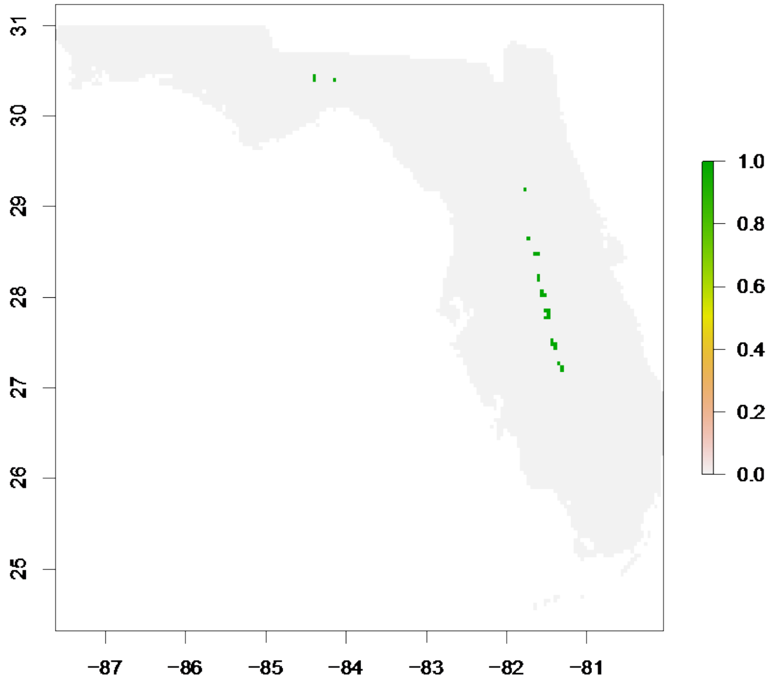


Short-lived narrow species

Average Climate



Yearly Climate



Parameters to explore:

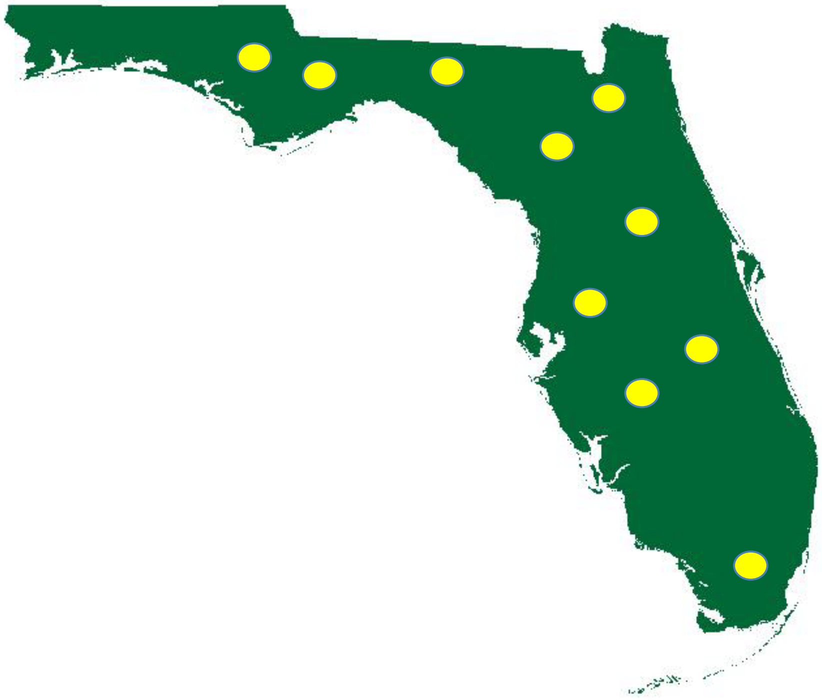
1. Using **yearly climate data** from time of collection to improve niche models using museum specimens
2. Using **reduced area** to train the models
3. **Smoothing response curve** to mimic more realistic response of organism
4. Using **fewer background points** to increase computing efficiency

Selecting an area to train the model

- Limiting the area in which the model trains
- If you want to recover suitable niche
  - – Include an area where biotic interactions might be excluding the species
    - Large mask
- If you want to recover species distribution
  - – Consider abiotic factors as part of the limiting factor
  - – Delimit mask from presence points
    - Tighter mask

Selecting a meaningful area to train the model

NO mask

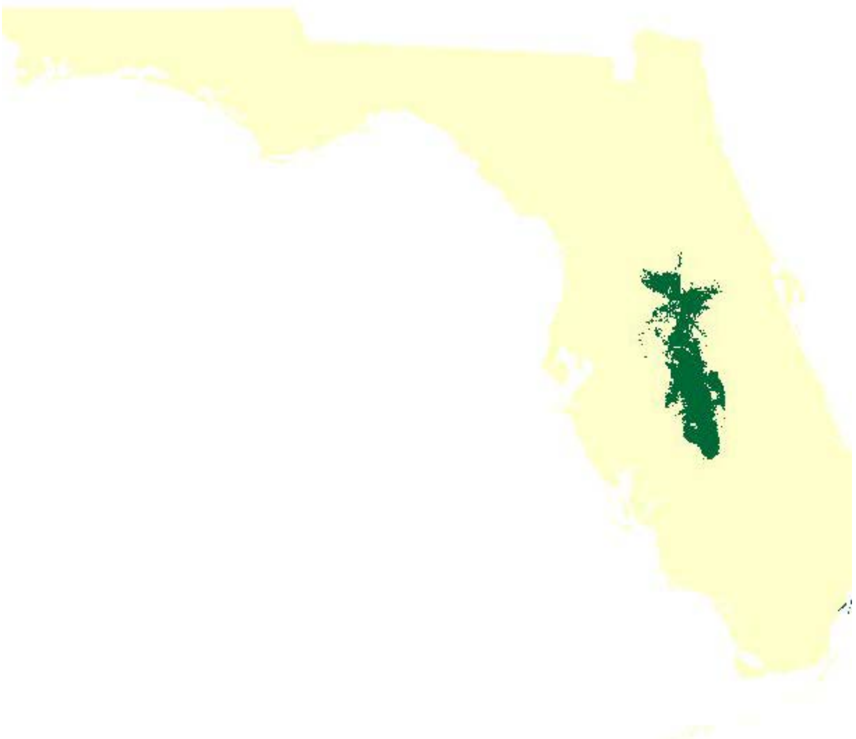


WITH mask

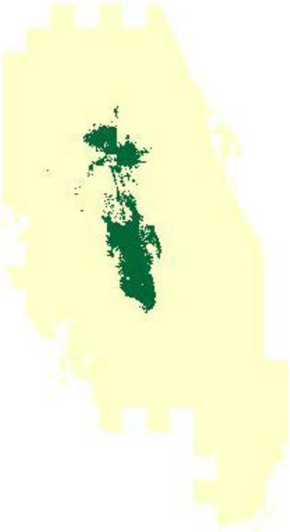
Prunus geniculata

Credit: fws

NO mask

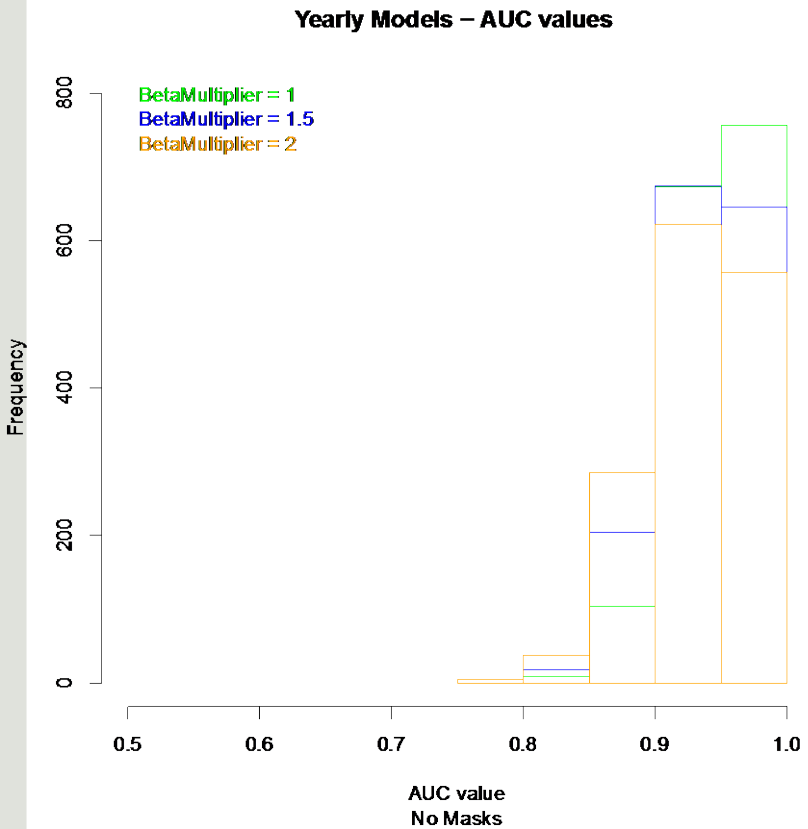


WITH mask

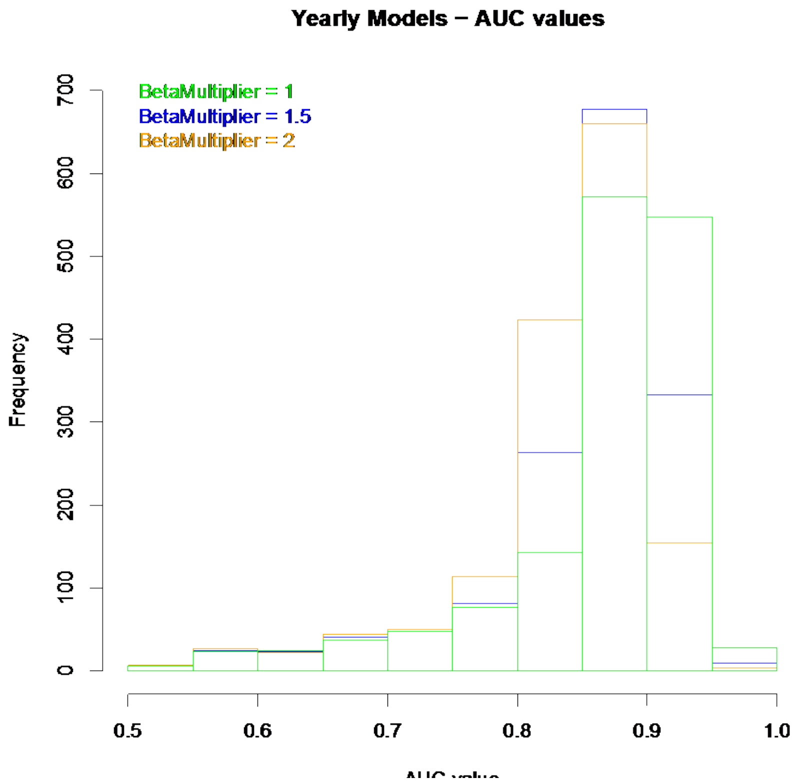


AUC scores

NO MASKS



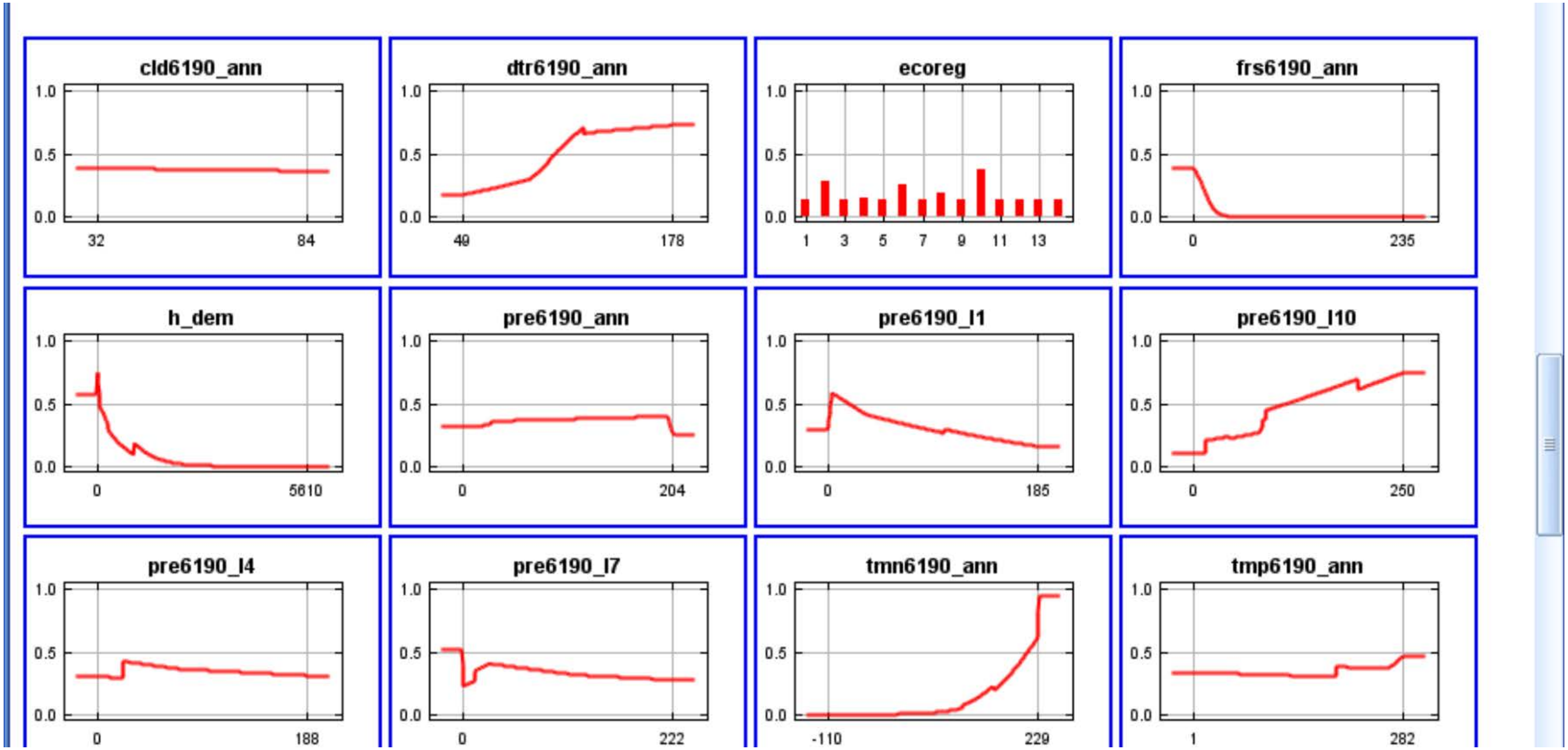
MASKS



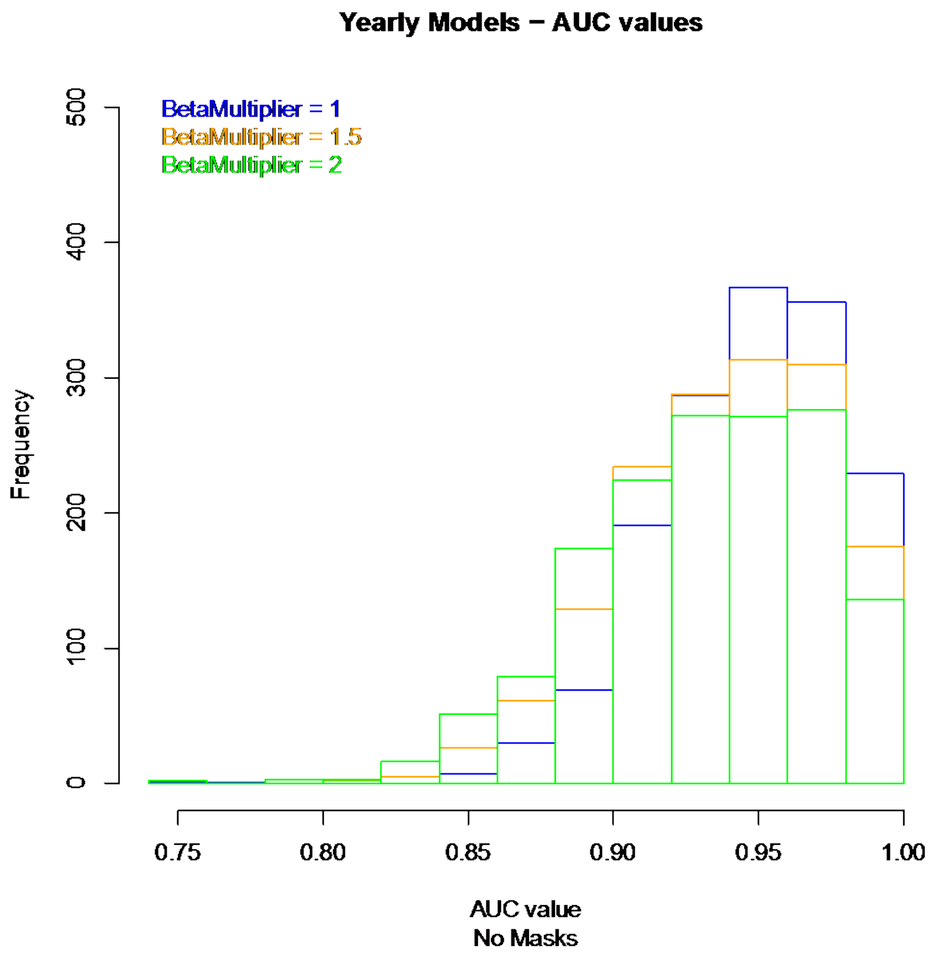
Parameters to explore:

1. Using **yearly climate data** from time of collection to improve niche models using museum specimens
2. Using **reduced area** to train the models
3. **Smoothing response curve** to mimic more realistic response of organism
4. Using **fewer background points** to increase computing efficiency

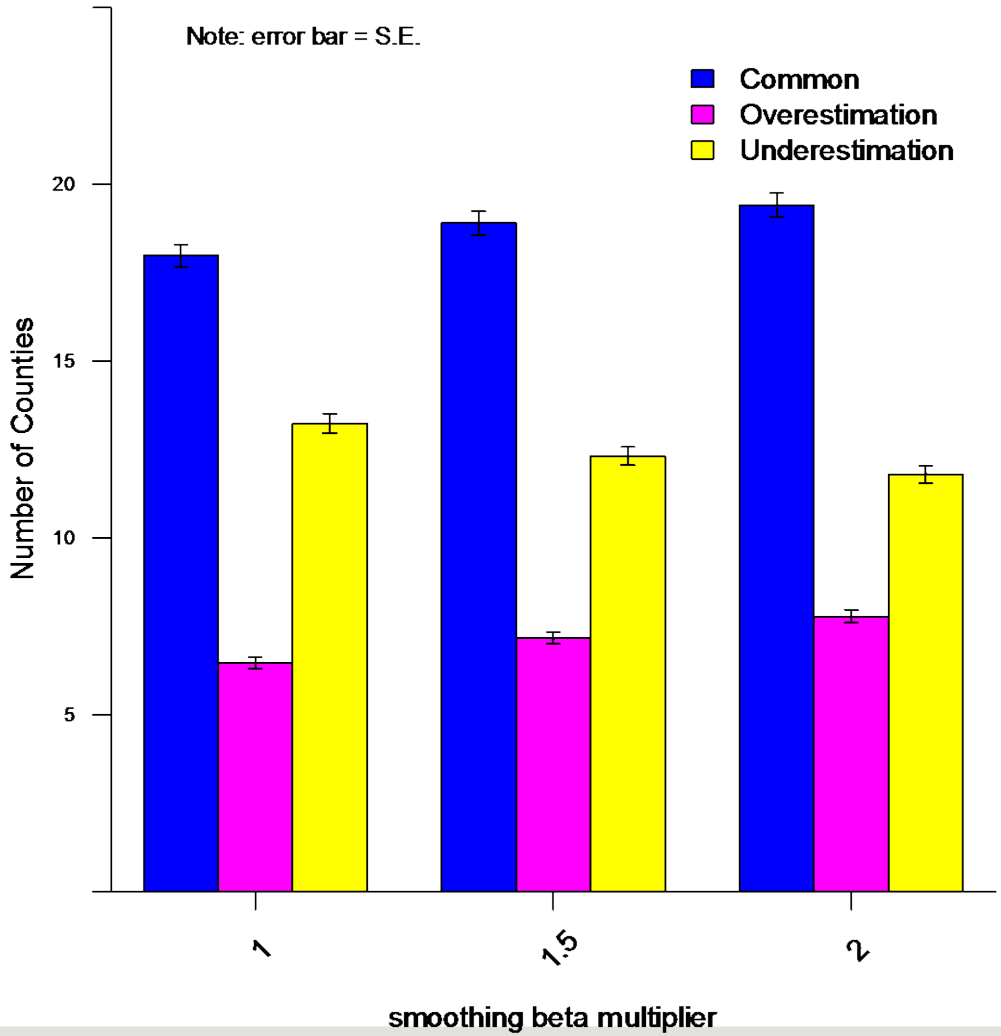
Beta parameter: smoothing environmental response



AUC values for different beta



Model performance



iDigBio Integrated Digitized Biocollections

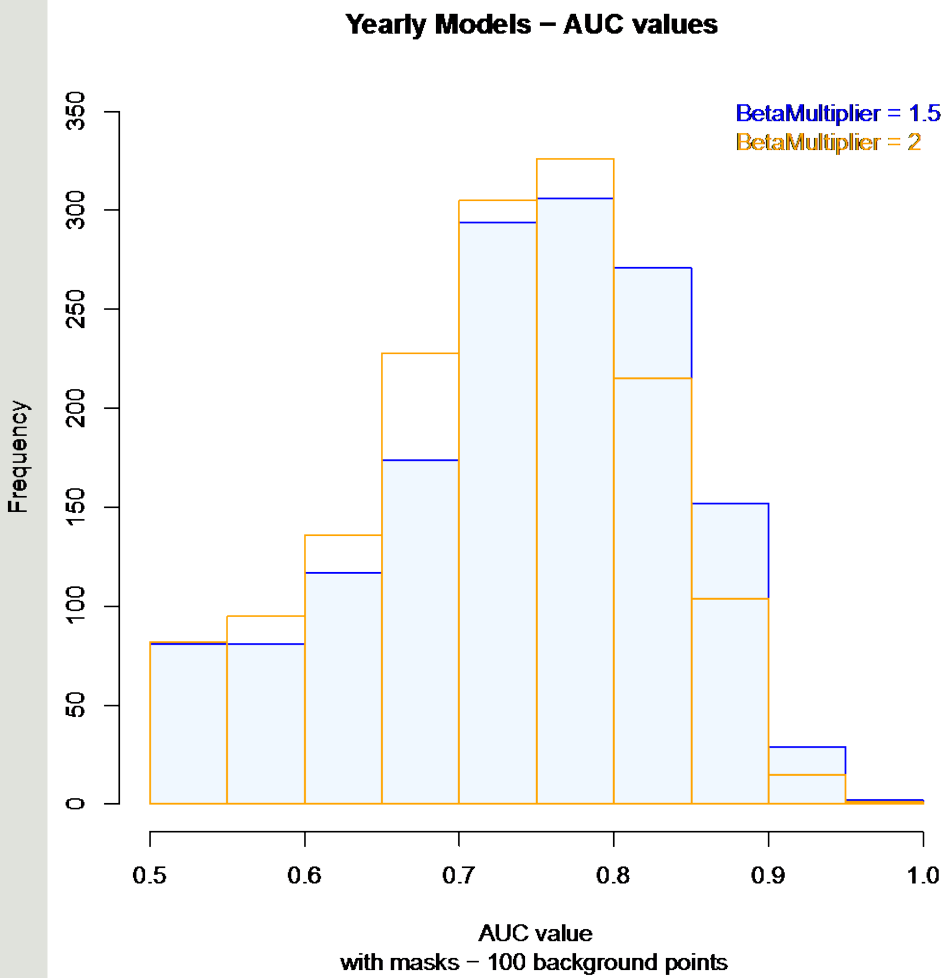
Parameters to explore:

1. Using **yearly climate data** from time of collection to improve niche models using museum specimens
2. Using **reduced area** to train the models
3. **Smoothing response curve** to mimic more realistic response of organism
4. Using **fewer background points** to increase computing efficiency

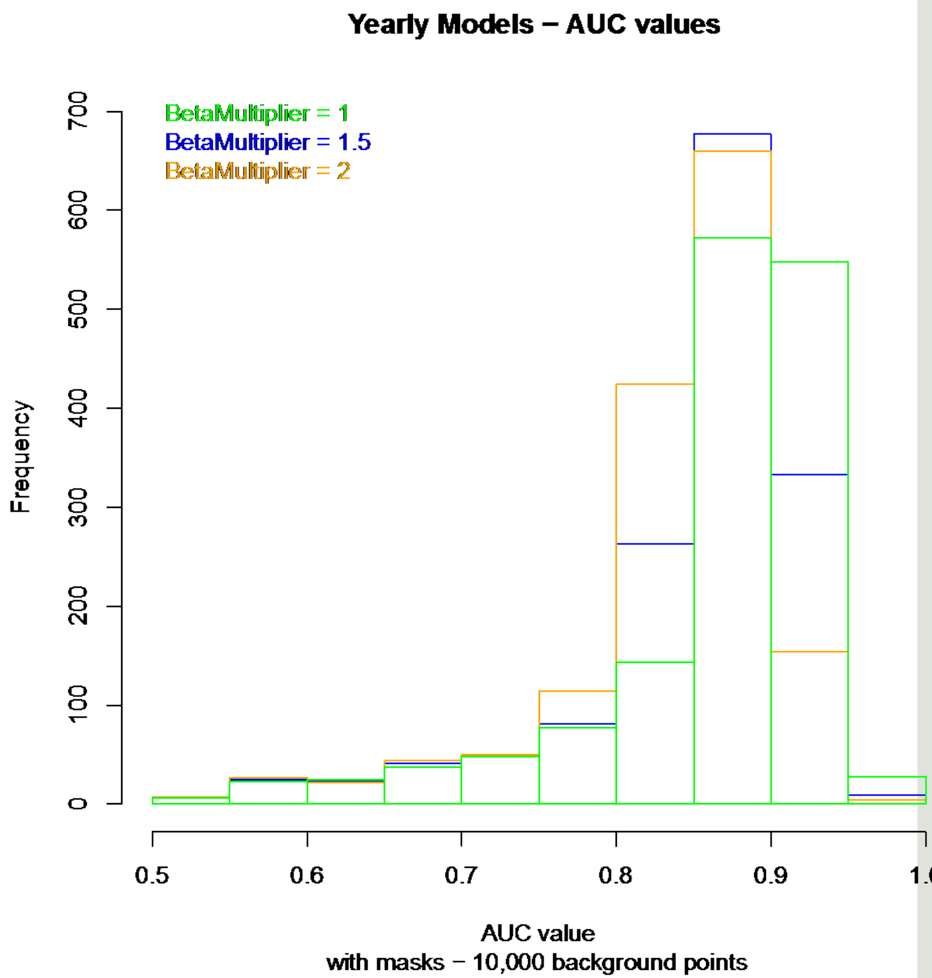
Importance of the number of background points

- Pseudo absences are ‘soft’ absences:
  - for the evaluation of the model
  - not the fitting
- Fewer background points
  - assign a higher probability to each cell
  - Hard to distinguish regions of high probability
- Large number of background points is computationally more demanding

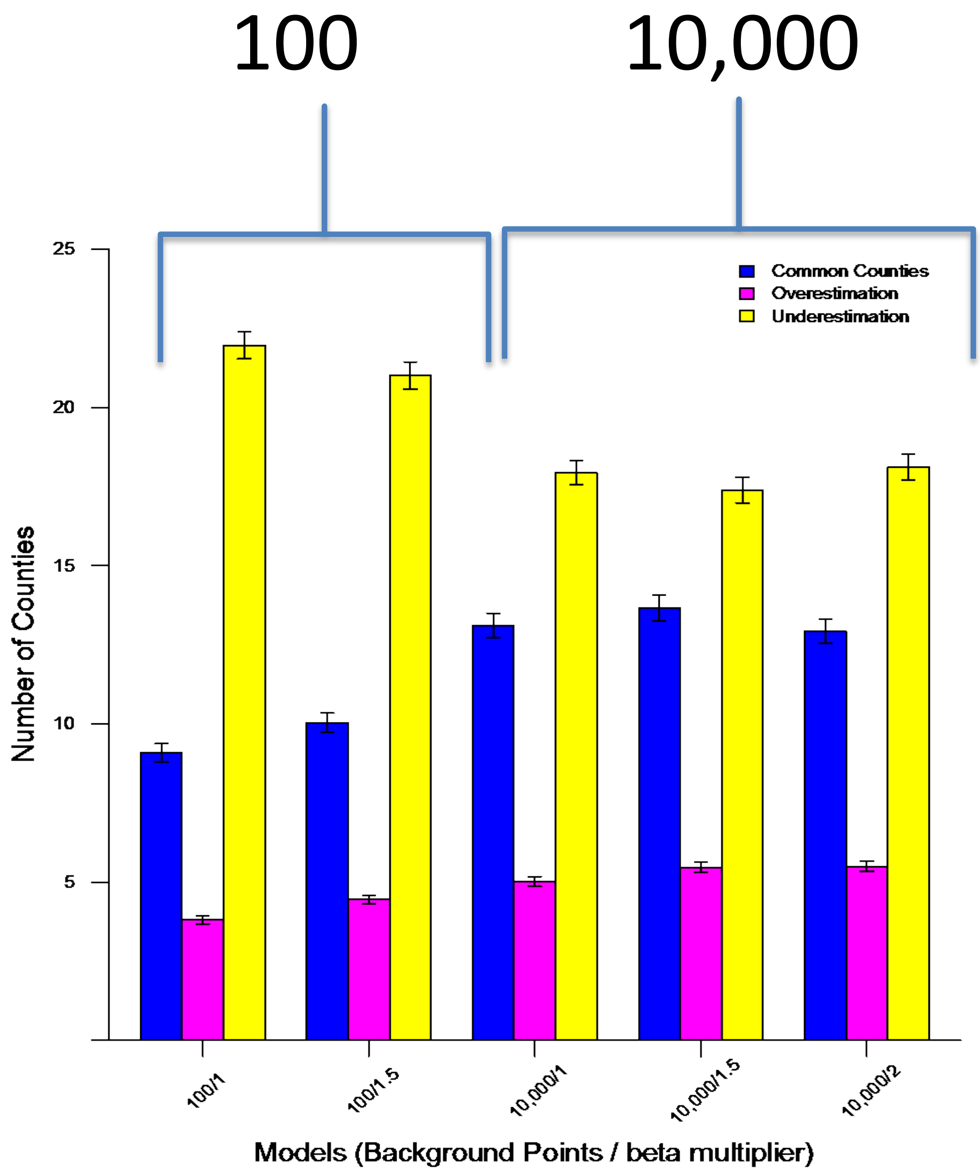
AUC



Yearly Models – AUC values
BetaMultiplier = 1.5
BetaMultiplier = 2
Frequency
0
50
100
150
200
250
300
350
0.5
0.6
0.7
0.8
0.9
1.0
AUC value
with masks – 100 background points



Yearly Models – AUC values
BetaMultiplier = 1
BetaMultiplier = 1.5
BetaMultiplier = 2
Frequency
0
100
200
300
400
500
600
700
0.5
0.6
0.7
0.8
0.9
1.0
AUC value
with masks – 10,000 background points



100
10,000
Common Counties
Overestimation
Underestimation
Number of Counties
25
20
15
10
5
100/1
100/1.5
10,000/1
10,000/1.5
10,000/2
Models (Background Points / beta multiplier)

Parameters to explore:

1. Using **yearly climate data** from time of collection to improve niche models using museum specimens
2. Using **reduced area** to train the models
3. **Smoothing response curve** to mimic more realistic response of organism
4. Using **fewer background points** to increase computing efficiency

CONCLUSIONS

- Yearly Climate Data:
  - Annual, small distribution performs better
  - Perennials and/or widespread doesn't
- Use of Masks:
  - Improves models, refines the local delimitation
    - Mask has to be carefully chosen
- Smoothing parameter
  - Test with a subset of species to determine
- Number of background points:
  - Worth the extra time !!!

✧ Think about why you are modeling the species niche (fundamental or realized?)

✧ Beware of statistical performance vs. biological accuracy

THANK YOU !!!!

Atlas of Florida Vascular Plants

Alabama Plant Atlas
Alabama Herbarium Consortium
& The University of West Alabama

GLOBAL BIODIVERSITY INFORMATION FACILITY

Université de Montréal

NSF

UNSW AUSTRALIA

ILLINOIS NATURAL HISTORY SURVEY
PRAIRIE RESEARCH INSTITUTE

FLORIDA STATE UNIVERSITY 1851

STATE

FLORIDA MUSEUM OF NATURAL HISTORY

FLORIDA Natural Areas INVENTORY

LSU

iDigBio
Integrated Digitized Biocollections

THANK YOU !!!

www.idigbio.org

facebook.com/iDigBio

twitter.com/iDigBio

vimeo.com/idigbio

idigbio.org/rss-feed.xml

webcal://www.idigbio.org/events-calendar/export.ics

UNIVERSITY of FLORIDA · FLORIDA STATE UNIVERSITY 1851 · FLORIDA MUSEUM OF NATURAL HISTORY · NSF

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.