Methods, Protocols, and Analytical Tools for Specimen-Based Research in the Biological Sciences
2019 Digital Data in Biodiversity Research Conference

No Collection Left Behind: Research Contributions of Small Collections

Yale, New Haven, CT
11 June 2019

Anna K. Monfils, Central Michigan University
John Bates, Field Museum
Blake Cahill, Central Michigan University
Erica Krimmel, Natural History Museum of Los Angeles County
Gil Nelson, iDigBio and University of Florida
Marci Revelez, Texas Tech University
Molly Phillips, iDigBio and University of Florida
Barbara Thiers, New York Botanical Garden
Jennifer Zaspel, Milwaukee Public Museum

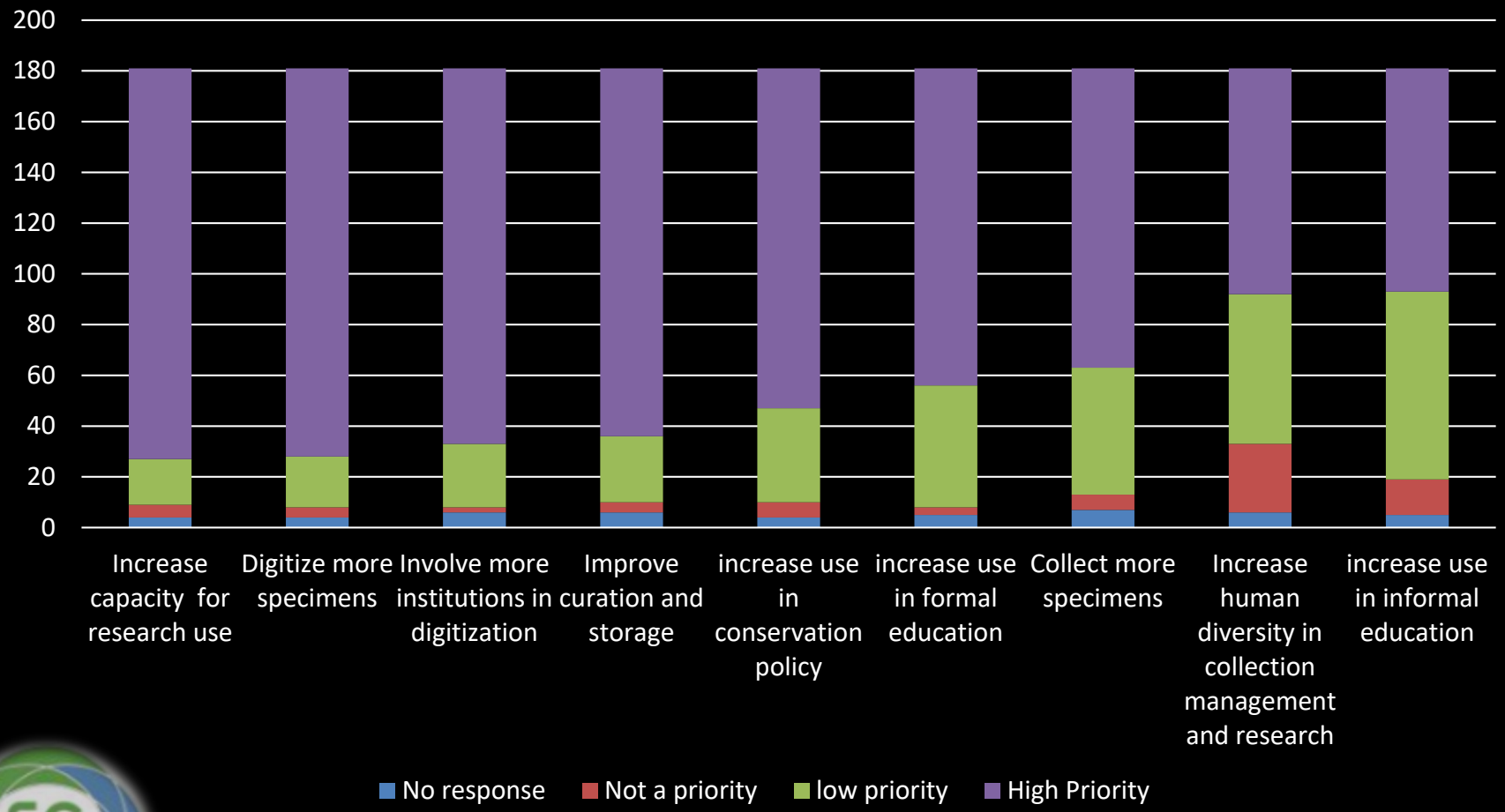
BIODIVERSITY COLLECTIONS NETWORK

SURVEY: Harnessing Biodiversity Collections Data for Addressing National Challenges

February 7th, 2018

This survey is intended to elicit information for a stakeholder vision of how to maximize the value of biodiversity collections data for collections management, research and education in the future

Priority Future Goals for Biodiversity Collections?



Involve More Institutions in Digitization

- 1600 US Natural History Collections
- 642 supplying data to iDigBio
- Almost 1000 collections currently not involved in the national digitization initiative.

https://www.idigbio.org/portal/collections

Extending U.S. Biodiversity Collections to Promote Research and Education Report

Report Release April 4th, 2019

Extended Specimen Recommendations

- Create an authoritative, comprehensive, and self-updateable index of U.S. collections institutions (similar to *Index Herbariorum* for global herbaria) with structured metadata to describe their holdings as a first step toward expediting the discovery of undigitized collections and revealing these to the research community.
- Continue digitization of existing material focused on underrepresented taxa (e.g., those in entomology and paleontology) and including incorporation of specimens held in small regional, personal, and individual researcher-based collections.

What is a “Small Collection?”

- ≤ 100,000 specimens (though varies by collection type)
- Regional in scope (typically)
- Ecological, taxonomic and geographic bias
- Often not included in inventory or monographic studies

Regional

What is a "~~Small~~ Collection?"

- ~~≤ 100,000 specimens (though varies by collection type)~~
- Regional in scope (typically)
- Ecological, taxonomic and geographic bias
- Often not included in inventory or monographic studies

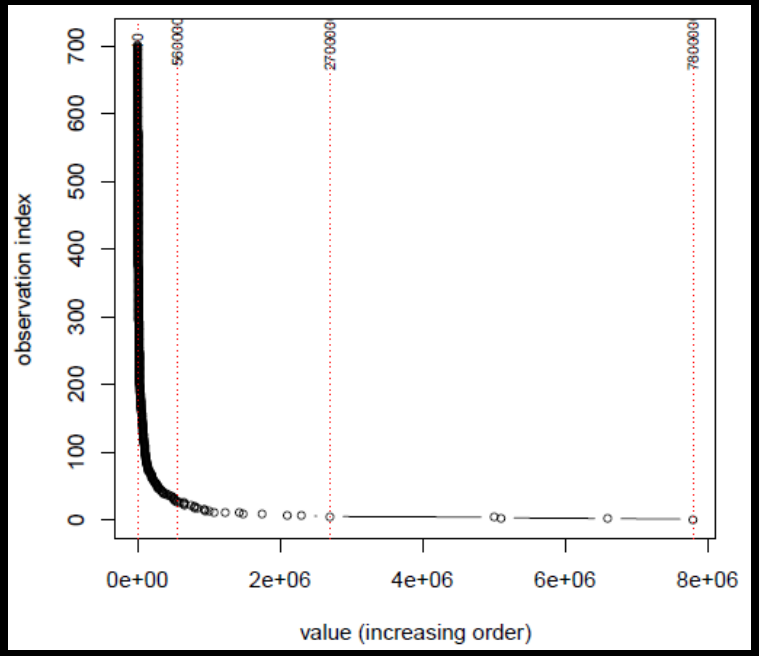
Regional collections serve a valuable role in documenting and monitoring global biodiversity

Central Michigan University Herbarium

Methods to Investigate Natural Breaks in Specimen Number

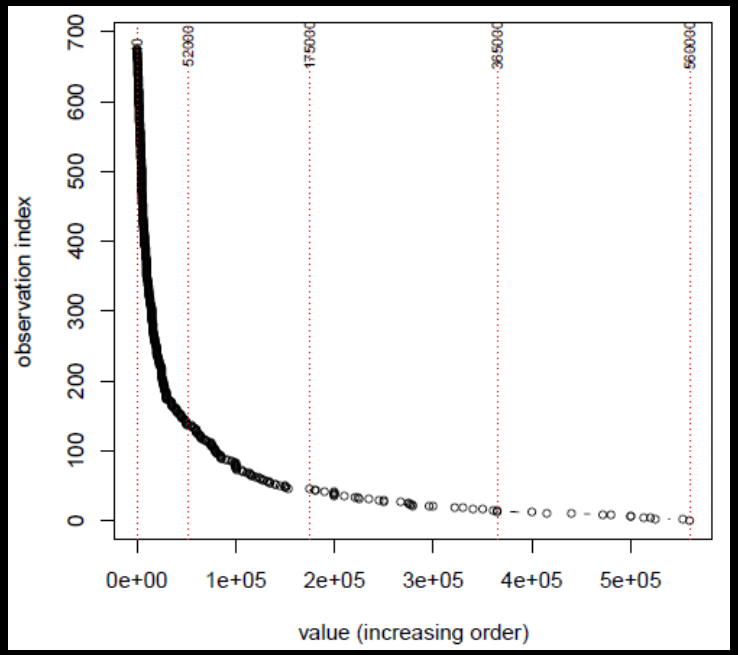
- US entries in Index Herbariorum (Jan. 19; 834)
  - Removed herbaria holding 0 specimens
  - Removed herbaria missing number of specimens
  - Left with 701 herbaria; 78,808,247 specimens
- Jenks Natural Breaks Classification
  - Set Goodness of Variance Fit (GVF) to 0.9

Natural Breaks In Specimen Number



- Sizes Classes (701 Total)
  - 10 – 560,000 (675)
  - 560,000 – 2,700,000 (22)
  - 2,700,000 – 7,800,000 (4)

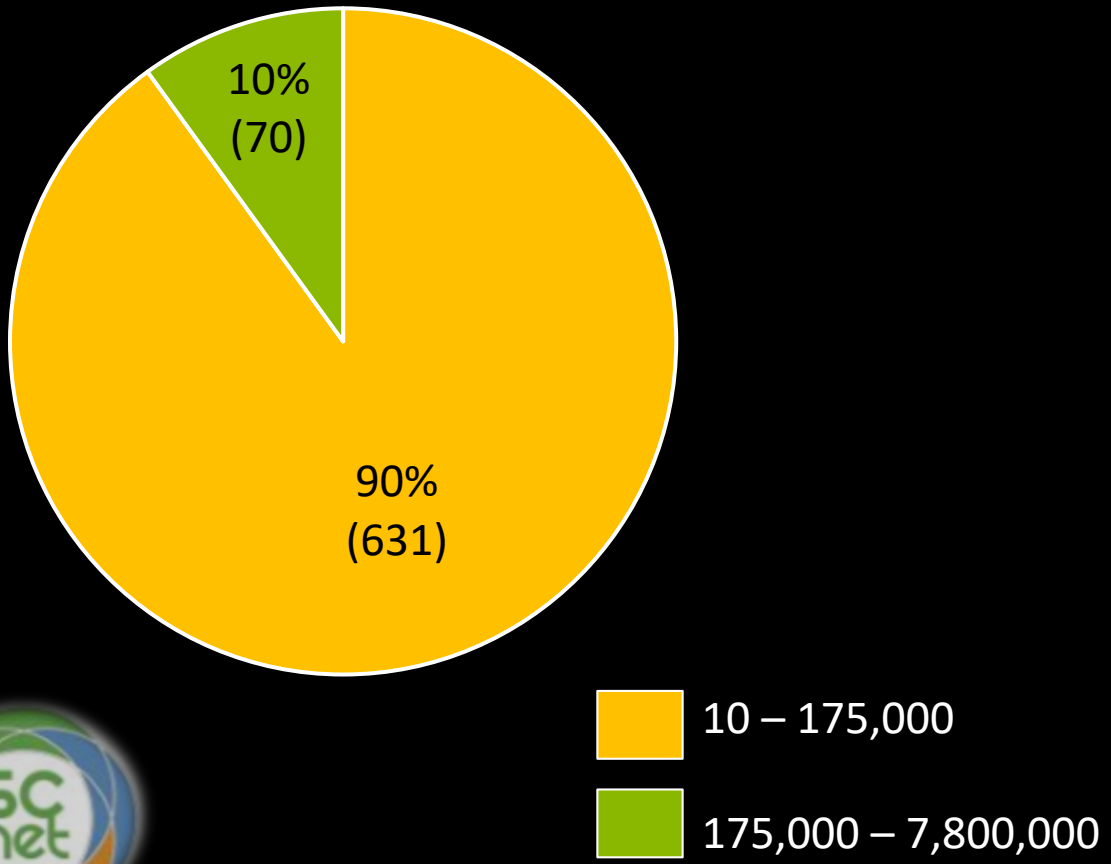
Natural Breaks In Specimen Number



- Sizes Classes (675 Total)
  - 10 – 52,000 (539)
  - 52,000 – 175,000 (92)
  - 175,000 – 365,000 (32)
  - 365,000 – 560,000 (12)
- Regional ≤175,000 (631)
- Large >175,000 (70)

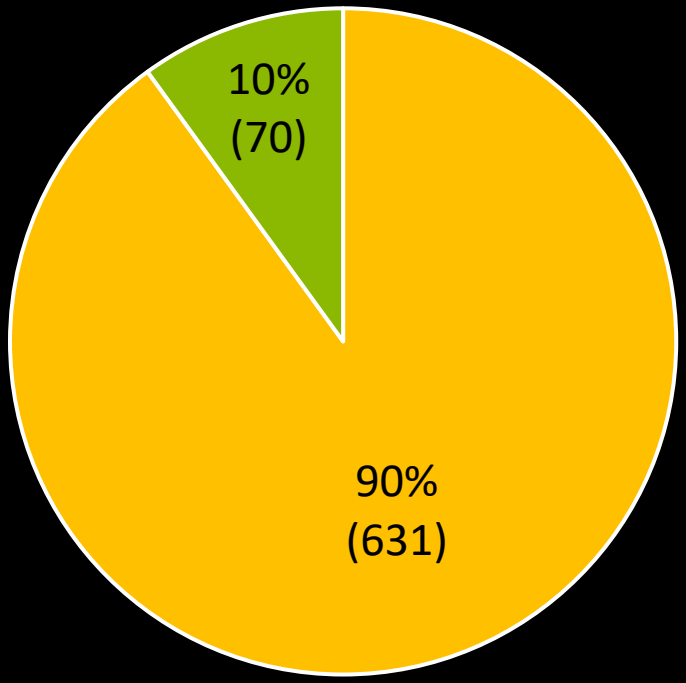
Regional Collections: Collections and Specimens

Percent of Collections

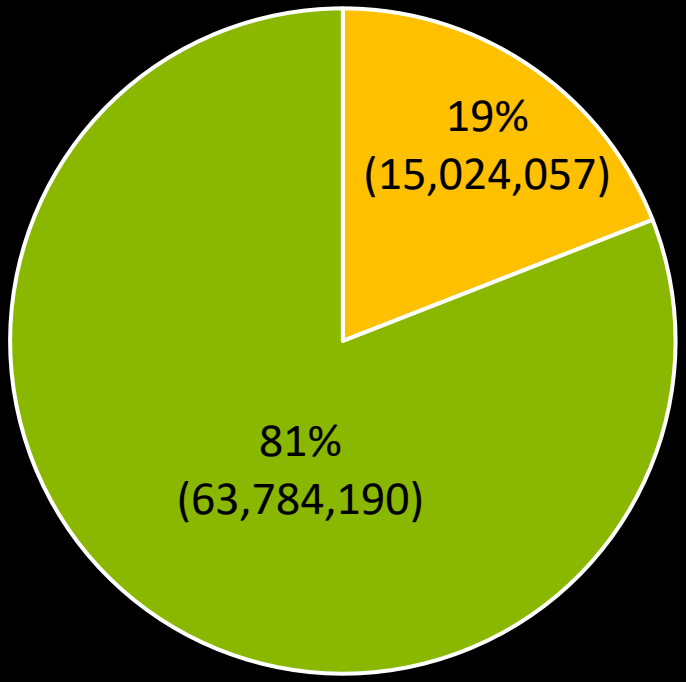


Regional Collections: Collections and Specimens

Percent of Collections



Percent of Specimens



10 – 175,000

175,000 – 7,800,000

Regional Herbaria: Unique contributions of specimen based records at the county-, locality-, and temporal-level

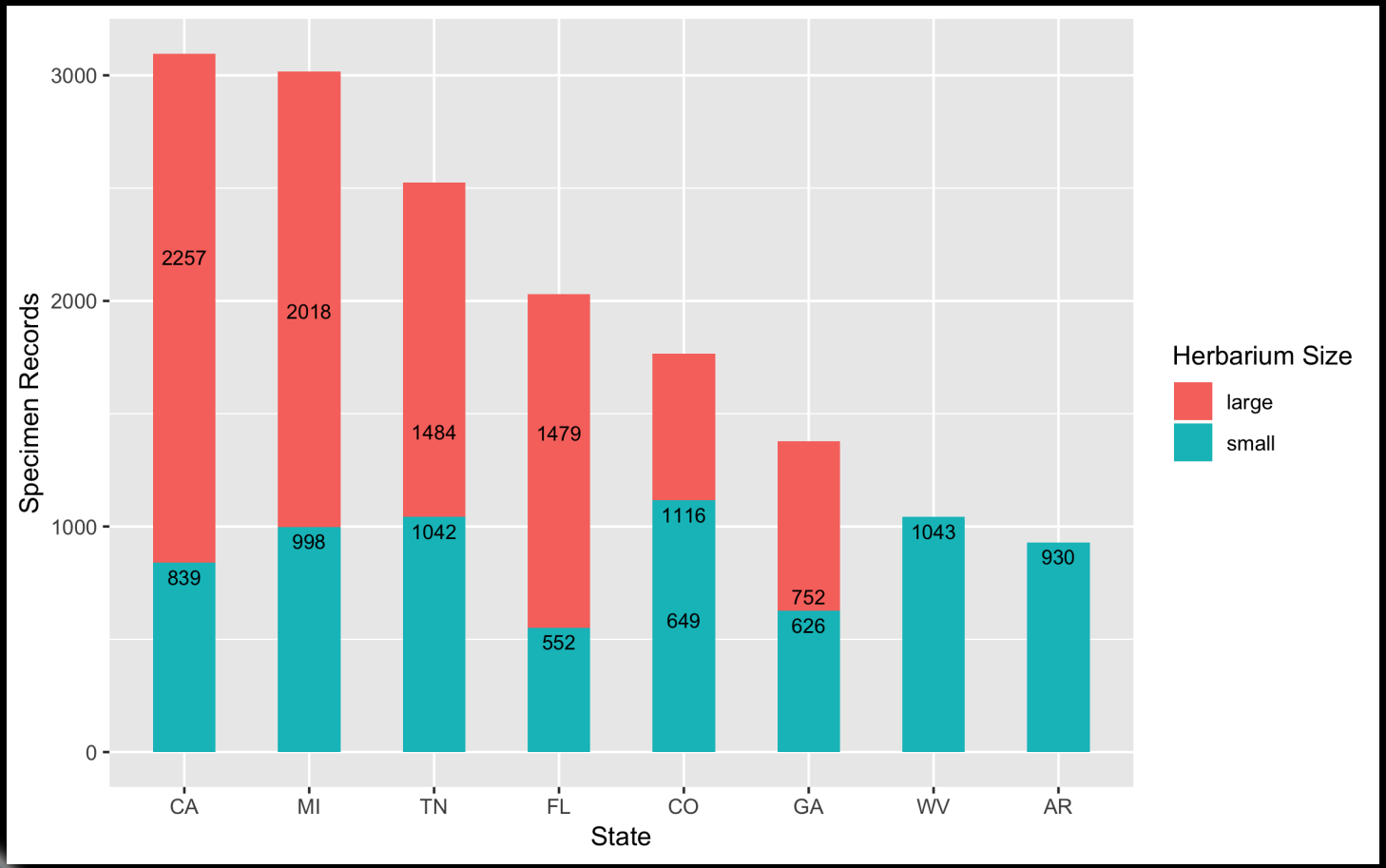
Travis D. Marsico[1], Erica Krimmel[2,3], Richard Carter[4], Emily Gillespie[5,6], Phillip D. Low[4], Ross McCauley[7], Ashley B. Morris[8], Gil Nelson[9], Michelle Smith[9], and Anna K. Monfils[10]

[1]Arkansas State University, [2]Sagehen Creek Field Station, [3]Natural History Museum of Los Angeles County, [4]Valdosta State University
[5]Marshall University, [6]Butler University, [7]Fort Lewis College,
[8]Middle Tennessee State University, [9]Florida State University,
[10]Central Michigan University

Specimen Data from Regional and Large Collections

- 8 States: Florida, California, Michigan, Georgia, Colorado, West Virginia, Tennessee, Arkansas
- 10 random species: S1, S2, Native, Introduced
- 21,546 specimens (10,381 regional, 11,165 large) in our initial dataset. (48%)
- 15,785 specimens (7,146 regional, 8,639 large) in our analysis dataset (excluding duplicates and low quality specimens) (45%)
- Regional herbaria: 19% of specimens held nationally; 23% of those held in state; 45% of specimens in study

Specimen Records by State
(excluding duplicates and low quality specimens)



Duplicates*

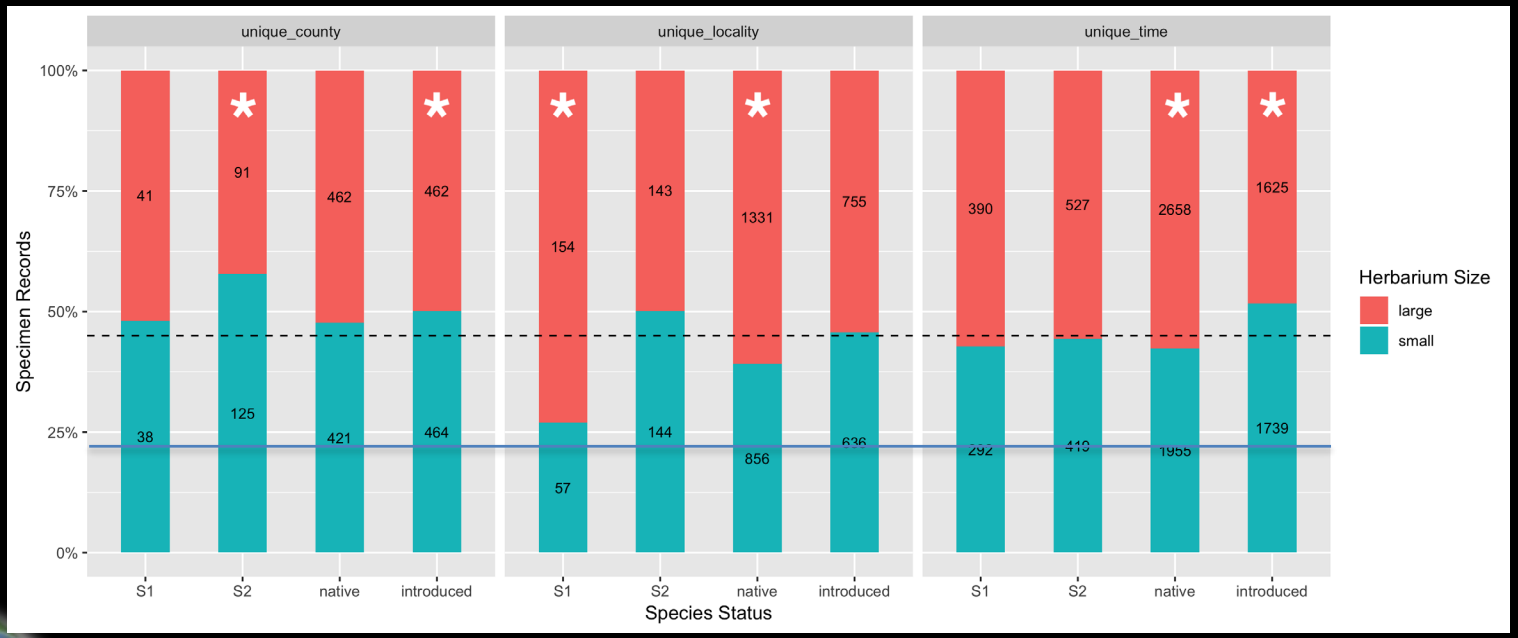
- 3% duplicates among regional and large
- 8% duplicates among regional
- 4% duplicates among large

Duplicate Status	Total	%
unduplicated specimens held by large herbaria	8002	49
unduplicated specimens held by regional herbaria	5869	36
duplicated specimens held only among large herbaria	693	4
duplicated specimens held only among regional herbaria	1358	8
duplicated specimens held by a large and regional herbarium	426	3

* 16,348 “unique” occurrences

Unique Contributions

- Solid line is 23%, represents expected contribution based on percent of specimens in state (χ^2 test; all with $p<0.05$ except Unique_Locality S1 that was not significantly different from expected)
- Dashed line is 45%; represents expected contribution based on percent of specimens for local flora (χ^2 test; asterisks denotes $p<0.05$)



Preliminary Results

- Regional collections and their relative contribution vary by state.
- Regional collections have a greater than expected number of regional specimens
- Specimens from regional collections provide comparable value to specimens from large collections
- Regional herbaria contain specimens that are not duplicated in large herbaria.

Ecological Informatics 42 (2017) 67–78

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

The contribution of small collections to species distribution modelling: A case study from Fuireneae (Cyperaceae)

Heather E. Glon[a,b,*], Benjamin W. Heumann[a,c], J. Richard Carter[d], Jessica M. Bartek[d], Anna K. Monfils[a,b]

[a] Central Michigan University, Institute for Great Lakes Research, Mount Pleasant, MI, USA
[b] Central Michigan University, Department of Biology, Mount Pleasant, MI, USA
[c] Central Michigan University, Department of Geography, Center for Geographic Information Science, Mount Pleasant, MI, USA
[d] Department of Biology, Valdosta State University, Valdosta, GA, USA

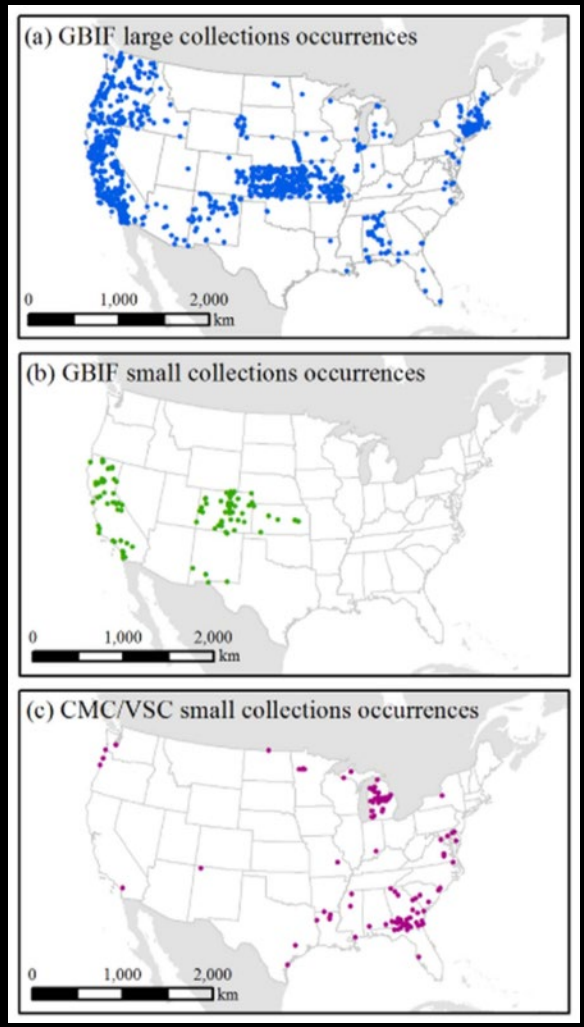
Contributions of Small Collections to Species Distribution Modeling

- Species distribution modeling to model potential suitable habitat for individual data sets
- Differences in geographic predictions by comparing habitat suitability

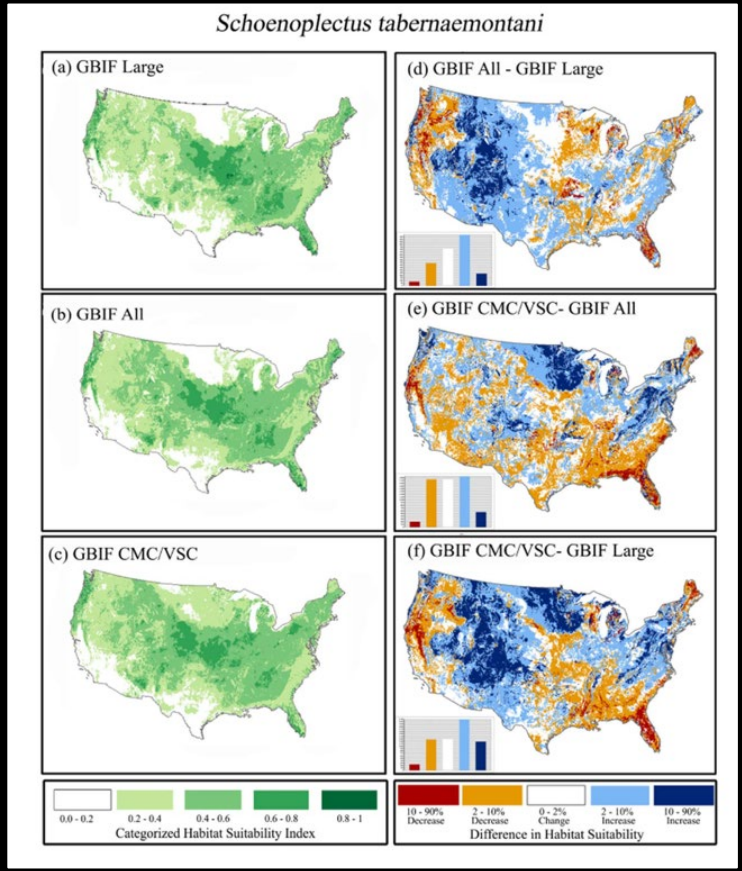
Specimen Based Occurrence Records

- Tribe Fuireneae (Cyperaceae; sedges)
  - GBIF – Large collections (1269)
  - GBIF – Small collections (122)
  - CMC and VSC Collections (127)

Species	GBIF large	GBIF small	CMC/VSC	Total
Fuirena squarrosa	25	n/a	44	69
Schoenoplectiella purshiana	45	n/a	15	60
Schoenoplectus acutus	434	52	13	499
Schoenoplectus pungens	413	32	26	471
Schoenoplectus tabernaemontani	352	38	29	419
Total	1269	122	127	1518



Geographic differences between the maps of the habitat suitability index



Habitat suitability predictions were significantly different among models based on datasets of large and small collections

Relative impact of regional collection data on species distribution models

- Models inclusive of small collections data result in more refined and robust predictions of ecological niche
- Small collections contribute unique occurrence data which enhance species distribution models:
  - bridges geographic collection gaps
  - shifts modelled predictions of suitable habitat.

Value of Regional Collections to Scientific/Collections Community

- Unduplicated specimens
- Intense regional sampling
- Unrepresented temporal sampling
- Focused sampling of community composition
- Critical source of data for biological hotspots (Biological Field/Research Stations)
- “Hidden source” of specimens representing curator’s expertise
- Contribute to sustaining an educated, diverse and inclusive, collaborative, and creative workforce.
- Distributed (“Decentralized”) effort provides resiliency and promotes sustainability for collection science and digitization

iDigBio
Integrated Digitized Biocollections

BIODIVERSITY COLLECTIONS NETWORK

http://scnet.acis.ufl.edu

To join SCNet listserve go to http://scnet.acis.ufl.edu/tags/listserv

Acknowledgements

- Shari Ellis, iDigBio and Assessment
- Ed Gilbert, SEINet and Symbiota
- North American Network of Small Herbaria (NANSH)
- Society for the Preservation of Natural History Collections (SPNHC)
- Small Collections Network (SCNet) through iDigBio
  - L. Page, L. Fortés, B. McFadden, G. Riccardi, & P. Soltis (NSF 1115210)
- CollectionsWeb
  - L.A. Prather, H. Bart, M. Blackwell, & J. Woolley (NSF 0639214)
- Biodiversity Collections Network (BCoN)
  - R. Gropp & A. Bentley, (NSF 144178)

Day 1

Day 2