2018

Discovering Change Using Herbarium Specimens: Plant Phenology, Distributions, and Biological Outliers

Katelin Delight Pearson

FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

DISCOVERING CHANGE USING HERBARIUM SPECIMENS:

PLANT PHENOLOGY, DISTRIBUTIONS,

AND BIOLOGICAL OUTLIERS

By

KATELIN D. PEARSON

A Thesis submitted to the
Department of Biological Science
in partial fulfillment of the
requirements for the degree of
Master of Science

2018

Katelin D. Pearson defended this thesis on May 30, 2018.
The members of the supervisory committee were:

Austin Mast
Professor Directing Thesis

Scott Burgess
Committee Member

Roy “Gil” Nelson
Committee Member

Gregory Riccardi
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the thesis has been approved in accordance with university requirements.

To my first inspiration, Sheri Stanley, and my daily motivation,
Blair Pearson, with much love and gratitude

ACKNOWLEDGMENTS

I would like to thank my advisor, Austin Mast, for his support, encouragement, and inspiration through the trial and error that is research, as well his networking skills that have enabled me to connect with so many scientists. Thanks to my committee members, Scott Burgess, Gil Nelson, and Greg Riccardi for their willingness to answer research and professional questions and for their overall support. I am also grateful for comments, ideas, and support from Libby Ellwood and Brendan Scherer and to Loran Anderson for domain expertise.

For chapter 3, thanks to Kevin Love, Alex Thompson, and Matthew Collins for assistance in producing data for Figure 3.2. I thank Andrew Kratter, Gil Nelson, Max Nickerson, Larry Page, and David Reed for introducing Austin and me to collection and digitization practices in their respective domains in a series of interviews. I also thank Shari Ellis for helping develop the survey, and I am grateful to those who distributed the survey of collectors, including member societies of the Natural History Collections Alliance (NSCA). Thanks to all who participated in the survey.

Thanks to Keith Bornhorst at the University of South Florida for access to USF Atlas of Florida Plants data and to the United States Department of Agriculture Plant Data Team for assistance in accessing data from the USDA PLANTS database for chapter 4.

This research was supported through iDigBio, which is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (award number 1547229). Digitization of the Florida State University Robert K. Godfrey Herbarium was funded in part by NSF award number 1410288 supporting the Southeast Regional Network of Expertise and Collections (SERNEC). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Herbarium specimens and the professionals who collect them can be powerful resources for understanding biological change, and many opportunities to maximize this impact remain. In this thesis, I develop and apply novel approaches in specimen data analysis, collection and exploration to empower rich and timely research use of specimen data. First, I use a new method of assessing specimen phenology to investigate differing phenological sensitivities of asteraceous plant species in the U.S. Southeastern Coastal Plain—an under-studied region in the field of phenology. I found contrasting phenological responses of spring- and fall-flowering species to warming climate in this region that could have significant ecological and evolutionary effects on, e.g., pollinator and herbivore interactions. Second, I propose two avenues by which the collecting community can contribute in an even greater capacity to studying biotic change: (1) by documenting and reporting specimen outliers, which could be indicators of change, and (2) by more consistently noting taxa associated with the specimens they collect, which could enable augmentation of existing occurrence data by up to 18% according to an analysis of over 84,000 specimen records. The results of this research represent advances in the burgeoning field of biodiversity informatics and have the potential to improve our understanding of life on Earth and the changes it is undergoing.

CHAPTER 1

INTRODUCTION TO BIODIVERSITY SPECIMENS AS TOOLS FOR DISCOVERING CHANGE

Biodiversity specimens—collections of organisms across time and space for preservation and study—have long been indispensable sources of taxonomic and natural history data. More recently, the broad spatiotemporal scope of this rich resource has been leveraged to study biotic change in an increasing diversity of fields (James et al. 2018), from invasive species management (Martin et al. 2014) and conservation planning (Soberon et al. 2000) to modeling the effects of climate change (Gomez-Mendoza & Arriaga 2007; Jarvis et al. 2008) to evolutionary developmental biology (Hetherington et al. 2016). Herbarium specimens, in particular, have proven hugely impactful for studying and predicting change in plant life with time and space such as declines in abundance of ethnobotanically significant native species (Case et al. 2007), rapid expansion of allergen-producing invasive species (Lavoie et al. 2007), increases in heavy metal bioaccumulation with industrialization (Herpin et al. 1997), declines in pollinator activity in urban areas (Pauw & Hawkins 2011), changes in elevation and physiological traits of plants (Agnihotri et al. 2017), and significant shifts in phenology (i.e., timing of life history events such as flowering or leafing out) with climate change (Willis et al. 2017).

Mass digitization (i.e., imaging and data transcription) of herbarium specimens has empowered researchers to investigate many of these topics at much broader scales. Still, the quickly expanding field of biodiversity informatics—including analysis of large amounts of specimen-based and other biodiversity data—has much to explore and improve. In this thesis, I underline the importance of specimen data and propose three foci for advancement in biodiversity informatics and specimen collecting: (1) finer-scale phenological assessments of specimens for, e.g., elucidating trends in plant phenological sensitivities to climate, (2) renewed attention to specimen outliers and how they may be indicators of change, and (3) mining existing data for previously "hidden" species occurrence data and encouraging rich data collection during specimen collecting events. Each of these approaches can enable researchers to better track change in the past, during the present, and into the future.

In chapter 2, I focus on an area of past change: plant phenology. Despite a surge of herbarium- and observation-based studies since the early 2000s (Willis et al. 2017[1]), many questions concerning the timing of plants' life history events remain. Specifically, it is yet unknown how plant species in the highly biodiverse, warm-temperate to subtropical southeast United States respond phenologically to climate change and whether species' responses depend on their traits. Repeatable, interoperable methods of assessing phenology of herbarium specimens are just now being developed (Yost et al. 2018[*]), and this

[1]Authors include Katelin D. Pearson

chapter addresses not only pertinent biological questions, but also offers a novel method with which to answer them.

Changes in Earth's biota during the Anthropocene go far beyond phenological shifts, and many of these changes have the potential to be catastrophic for biodiversity. In my third chapter, I present a new approach to track biotic change in the present and into the future: discovering and reporting specimen outliers at the time of collection. My survey of biodiversity specimen collectors reveals that these professionals often have the experience and resources to quickly recognize outliers, yet collectors may not be empowered to document and report these anomalies in timely and accessible ways. In this chapter, I describe the current state of outlier detection, documentation, and reporting by collectors using survey results and an analysis of outlier term use in 75 million records, and I propose future steps to facilitate this potentially transformative mechanism of detecting change.

Similarly, in my fourth chapter I explore another potential way to improve change detection using existing specimen data. Specimen data present a wealth of opportunities to mine information on species habitats, traits, and distributions, yet few efforts have been made to do so using newly digitized specimen records. I show how specimen label data on associated species—species located near collected specimens at the time of collection—can be leveraged to increase our understanding of species distributions and how they change over space and time. This chapter models how the development of new tools may be able to shape future data collection, optimizing our methods for maximum data utility.

With these chapters, I demonstrate the value of specimen data for the past, present, and future of detecting and documenting change and emphasize the importance of innovative research methods (e.g., chapter 2), frameworks (e.g., chapter 3), and tools (e.g., chapter 4) for enabling this discovery. With substantial innovation, improved data accessibility, and attention to data quality, specimen-based research using digital data will continue to make dramatic advances in biology and environmental science.

CHAPTER 2

DIVERGING PHENOLOGICAL RESPONSES TO CLIMATE IN SOUTHEAST U.S. SUNFLOWERS (ASTERACEAE)

Introduction

Plant phenological shifts (e.g., earlier flowering dates) are known consequences of climate warming that may significantly alter ecosystem functioning (Parmesan 2006; Calinger et al. 2013), productivity (Richardson et al. 2010), and ecological interactions such as those between plants and pollinators (Kharouba & Velland 2015; Forrest 2015). Despite a myriad of studies since the turn of the century investigating the effects of climate on plant phenology (i.e., "phenological sensitivity" of plants) using observational data (e.g., Fitter & Fitter 2002; Ellwood et al. 2014; Tansey et al. 2017), herbarium specimens (e.g., Primack et al. 2004; Lavoie & Lachance 2006; Munson & Long 2016), experiments (e.g., Price & Waser 1998; Pan et al. 2017; Posledovich et al. 2017), and combinations of data sources (e.g., Miller-Rushing et al. 2006; Panchen et al. 2012), significant gaps in our understanding of these phenomena and their potential consequences remain. In this chapter, I use thousands of newly available digitized herbarium specimen records and a new method of assessing the phenology of specimens to address critical geographic and trait-related knowledge gaps.

Most studies have discovered negative relationships between temperature and phenological events such as flowering and leaf-out; that is, plants flower or leaf out earlier with increased temperatures in the 2-3 months preceding the phenological event (Sparks et al. 2006). However, our understanding of these relationships has largely relied upon studies in temperate, boreal, or subalpine climates such as the Northeastern U.S. and north-central Europe (reviewed in Willis et al. 2017), though some efforts have focused on Mediterranean climates in Spain (Gordo & Sanz 2010) and California (Cleland et al. 2006), subtropical China and India (Hart et al. 2014; Gaira et al. 2014; Chen et al. 2017), coastal Australia (Rumpff et al. 2010), and xeric regions in the western U.S. (Neil et al. 2010; Munson & Long 2016). In this study, I examine plant phenological sensitivity to climate in the U.S. Southeastern Coastal Plain (SECP). Few studies have examined the phenological effects of climate in such warm, humid yet temperate climates (Pau et al. 2011; Willis et al. 2017; but see Park & Schwartz 2015; Chen et al. 2017), though these regions may provide unique insights into mechanisms of phenological change. Some researchers hypothesize that phenological sensitivities of organisms in cooler regions are constrained by an increased risk of frost damage (Inouye 2008; Gezon et al. 2016) or may be at the limits of their phenological plasticity (Scranton & Amarasekare 2017). Lacking these constraints, plants in warmer regions like the SECP may exhibit stronger phenological responses to temperature than those in cooler regions (Menzel et al. 2006). Furthermore, because the warm, humid climates like the SECP experience

fewer frost days than in cool climates, plants in this region may be less likely to have strong chilling requirements for phenological events to occur, which could otherwise moderate the effects of temperature on phenology (Chmielewski et al. 2011). Plant phenological sensitivities to temperature have yet to be investigated in warm, humid climates, and the SECP provides an ideal opportunity to do so.

Similarly, the SECP offers an opportunity to examine the effects of precipitation on phenology in a climate that shares characteristics of both temperate (e.g., temperature seasonality) and subtropical (e.g., high humidity) climates. While studies in temperate (Abu-Asab et al. 2001; Sparks et al. 2006), alpine (Hart et al. 2014), and some Mediterranean climates (Gordo & Sanz 2005) have discovered no significant effect of precipitation on phenology, precipitation cycles are critical to phenology in tropical regions (Sahagun-Godinez 1996; Zalamea et al. 2011), and may even outrank temperature in phenological importance in subtropical regions (Peñuelas et al. 2004). The influence of precipitation in the SECP was not addressed in the one study of phenological shifts in this region (Park & Schwartz 2015), leaving a gap in our understanding of climatic drivers of phenological events.

The SECP ecoregion, stretching from east Texas to east Massachusetts and south through Florida (Figure 2.1), is a biodiversity hotspot and home to over 6000 species of native plants, over 25% of which are endemic to the region (Sorrie & Weakley 2006). Nonparallel phenological shifts of plants and interacting taxa such as pollinators with the globally warming climate can lead to phenological asynchrony, which may detrimentally alter plant vital rates (Kudo & Ida 2013), cause local extirpation of pollinator species (Burkle et al. 2013), or produce novel trophic interactions (Liu et al. 2011). In such a highly biodiverse region already faced with critical threats (Nordman et al. 2014), the need to understand the potential threat of phenological shifts is clear. Doing so in this ecosystem may provide insight into local phenological sensitivities to climate warming and thus help predict future challenges for endemic and threatened species.

Examining overall trends in plant phenological sensitivities to temperature and precipitation in the SECP fills a critical geographic knowledge gap in a region of high biodiversity. This study also aims to strengthen our predictive power by determining how differences in plant traits impact phenological change. Species-specific responses may differ in magnitude and even direction (Calinger et al. 2013; Ellwood et al. 2013; Hart et al. 2014), which suggests that the effects of climate on plant phenology is moderated by other factors. Phylogenetic relatedness (Molnár et al. 2012; CaraDonna & Inouye 2015), diet (in insects; Diamond et al. 2011), habit (Calinger et al. 2013), nativity (Miller-Rushing & Primack 2008; Calinger et al. 2013; Bertin 2015), time and duration of phenological stages (Fitter & Fitter 2002; Miller-Rushing & Primack 2008; Bertin 2015), pollination mode (Fitter & Fitter 2002; Molnár et al. 2012), and other traits have been investigated in some systems, though results do not always agree between studies.

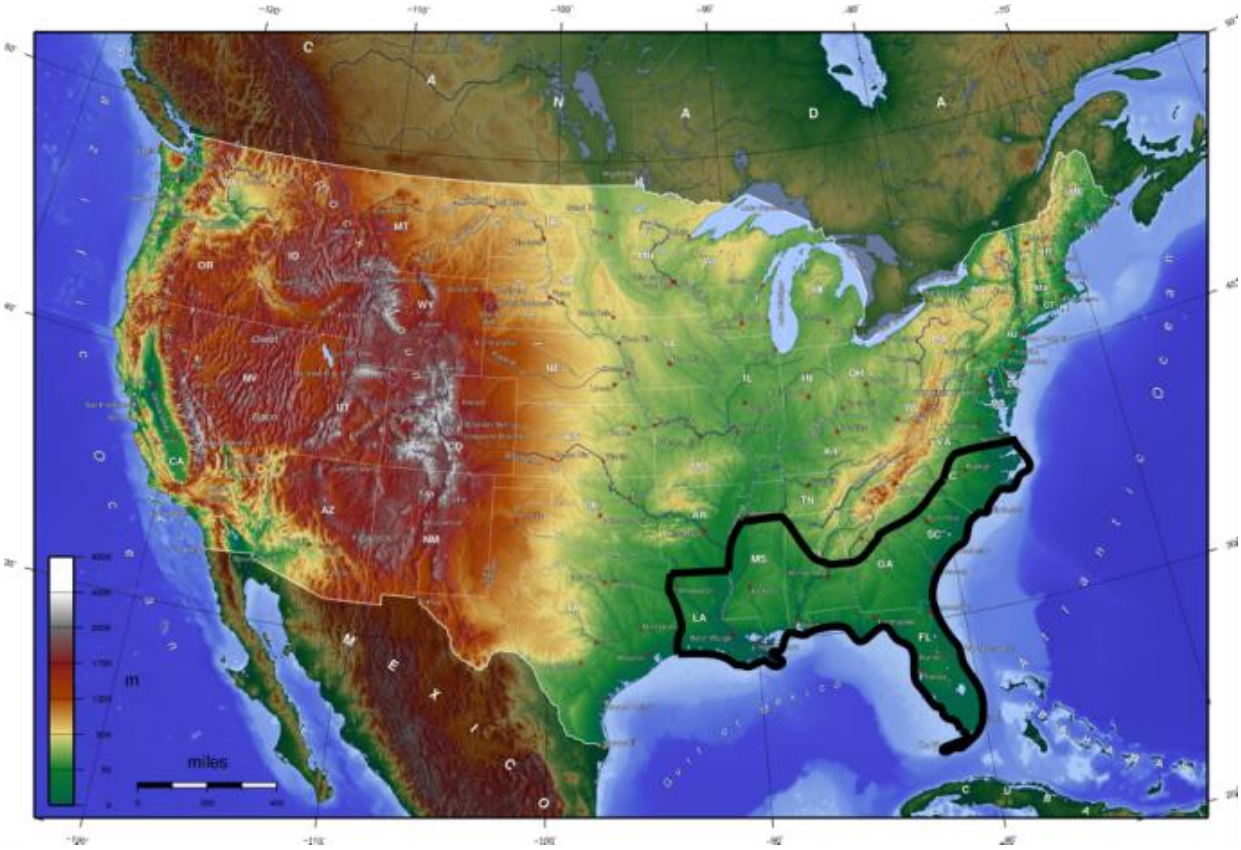

Figure 2.1 U.S. Southeastern Coastal Plain region selected for sampling of herbarium specimen records (outlined in black). Note the relatively flat topography. Although it is not generally considered within the SECP, south Florida is included to enable analysis of phenological shifts across a wide range of annual temperatures. Map created by User:Captain Blood using Generic Mapping Tools (http://gmt.soest.hawaii.edu/) and made available via https://commons.wikimedia.org/wiki/File:USA_topo_en.jpg.

One potentially important factor affecting phenological sensitivity in the SECP—and perhaps other regions with extended growing seasons—is flowering guild (i.e., spring-flowering or fall-flowering). Many species in the SECP flower either in the spring or late summer to fall, with fewer flowering in the hot summer months (Wunderlin & Hansen 2011). These distinct flowering guilds may respond to different climate cues or possess dissimilar physiological or evolutionary constraints. Previous studies have indicated that spring and autumn phenological events may have contrasting responses to climate, with fall phenology showing slight to moderate delays in contrast to strong advancement in spring-flowering species (Sparks et al. 2000; Gordo & Sanz 2005; Sherry et al. 2007; Jeong et al. 2011; Gill et al. 2015), though some have found opposite (Høye et al. 2013) or no trends for spring- versus fall-flowering species (Bock et al. 2014). Diverging phenological responses to climate warming across seasons could influence associated species such as pollinators by creating gaps in in which floral resources are scarce, and shifts in phenological timing could affect inter- and intraspecific competition

between both plant and pollinator species. I examine phenological responses to climate between spring- and fall-flowering species to determine whether this possibility poses a potential risk in this region.

Two additional traits also warrant investigation in this region: habitat preference and pollination mode. Habitat could play a critical role in moderating or exacerbating phenological sensitivity to climate, though little attention has been paid to this factor. Plants in moist habitats, for example, may respond weakly to climate warming because local temperatures are moderated by soil moisture (e.g., Mäkiranta et al. 2018). Alternatively, these plants may respond strongly to climate warming because warmer temperatures instigate drought conditions to which moist-habitat species are not well adapted. The SECP is a mosaic of habitats, including mesic and xeric hammocks, longleaf pinelands, hardwood forests, swamps, and other habitat types with unique plant communities and varying levels of moisture (NatureServe 2017), providing an excellent system in which to study the effect of habitat preference on plant phenology. Determining the role of habitat in phenological sensitivity to climate will improve our understanding of the differential effects of climate warming for different species and regions, allowing us to prioritize conservation for those that need it most.

Furthermore, species with different pollination modes may differ in phenological sensitivity to climate. For example, animal-pollinated species may have evolved greater phenological sensitivity to temperature to track temperature-sensitive emergence of pollinators. Alternatively, wind-pollinated species may use higher sensitivity to temperature to exploit temperature-related fluctuations in wind. Both of these hypotheses have received some support: Calinger et al. 2013 found that wind-pollinated species were more sensitive to temperature than animal-pollinated species in Ohio, yet Fitter & Fitter 2002 found the opposite trend in the UK. This investigation will further elucidate the importance of pollination mode on phenological response to climate variables, which may have implications for patterns of seed dispersal or seed availability under future climate change.

In this study, I investigate the phenological sensitivity of plant species to climate using the rich data source of digitized herbarium specimen records. Herbarium specimens are collections of plants from across the globe that have been dried, pressed, and preserved for hundreds of years in natural history collections (herbaria). Although not necessarily collected with the intent to document phenological events, herbarium records have proven to be reliable sources of phenological data that are vital to advancing our understanding of plant phenology on wide temporal and spatial scales (Davis et al. 2015; Willis et al. 2017). With the large amount of digitized specimen data—including specimen images—now available (e.g., online at idigbio.org), obtaining the statistical power necessary to distinguish phenological trends is more tractable than ever.

Understanding regional, taxon-specific, and trait-specific effects of climate change on plant phenology will allow a more nuanced ability to infer mechanisms, predict phenological trajectories, and

form hypotheses for future study. In this chapter, I investigate (1) how peak flowering times of asteraceous plants change with temperature and precipitation in the SECP, (2) how this relationship differs between spring-flowering and fall-flowering species, and (3) how shifts in peak flowering time with climate depend on two additional plant traits: habitat preference and pollination mode.

If shifts in flowering time with temperature are conserved among climate types, I expect spring-flowering species, if not fall-flowering species, in the SECP to flower earlier under greater than average temperatures at a rate near 2-3 days/°C (Calinger et al. 2013). If phenological sensitivity to temperature depends more strongly on climate type, I do not expect to see such a trend. Given the impact of precipitation on phenology in subtropical and tropical regions (Peñuelas et al. 2004; Zalamea et al. 2011), I expect that species in the warm, humid SECP will exhibit a relationship between peak flowering time and precipitation. However, it also is possible that, because most of this region experiences colder winter temperatures than the subtropics and tropics, plant phenology in the SECP may remain more tightly linked to temperature regimes. Regardless of how phenological sensitivities to climate compare between climate types, I expect to find differing phenological responses between spring-flowering and fall-flowering species (Sherry et al. 2007; Gill et al. 2015). The effect of habitat preference and pollination mode on phenological sensitivity is not well predicted by the literature and thus remains an open question.

Methods

Dataset selection and cleaning

I selected a subset of species from 11 genera in the sunflower family (Asteraceae; Figure 2.2) in which to investigate the questions outlined above. This family provides an ideal system for this study because of its abundance and diversity in the SECP (Figure 2.3); iDigBio, a national aggregator of specimen records, reports over 48,000 Asteraceae specimens collected in Florida alone since 1842 (idigbio.org), and the Atlas of Florida Plants reports over 430 Florida asteraceous species (florida.plantatlas.usf.edu). Species in the Asteraceae are found in a wide variety of habitats, can be either animal- or wind-pollinated (Judd et al. 2002), and many bloom either during the spring flowering peak (Feb-May) or the fall flowering peak (Aug-Oct) in the SECP (Wunderlin & Hansen 2011), allowing examination of the effects of climate on flowering among species of different habitats, pollination modes, and flowering guilds (Figures 2.2 and 2.3). The 11 genera were selected to maximize (1) number of specimens per genus, (2) representation of taxonomic (i.e., tribal) diversity within the Asteraceae, and (3) diversity of flowering guild, habitat preference, and pollination mode, particularly among species within genera. Within these genera, I selected 87 species to include, maximizing the number of specimens per species and representation of different trait values (Appendix A).

Figure 2.2 Examples of Asteraceae species in the dataset found in the U.S. Southeastern Coastal Plain in the field (top) and as herbarium specimens (bottom): (a) spring-flowering, animal-pollinated *Coreopsis lanceolata*, (b) spring-flowering, animal-pollinated *Gamochaeta coarctata*, (c) fall-flowering, animal-pollinated *Liatris tenuifolia*, and (d) fall-flowering, wind-pollinated *Eupatorium compositifolium.* Field images were taken by the author, and specimen images are courtesy of the Robert K. Godfrey Herbarium database (herbarium.bio.fsu.edu).

I downloaded all herbarium specimen records of the 11 selected genera collected in states comprising the U.S. Southeastern Coastal Plain (Figure 2.1: Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina) from the iDigBio online portal (idigbio.org). I removed records from counties not located within the SECP as informed by state ecoregion maps (Georgia Department of Natural Resources, North Carolina Department of Public Instruction, Riekerk n.d., University of Alabama Department of Geography). Restricting data selection to this ecoregion not only ensures results are relevant to understanding phenology in the SECP, but also reduces the potential influence of elevation on my analyses (Gugger et al. 2015), as altitudinal variation is low across this landscape (Figure 2.1).

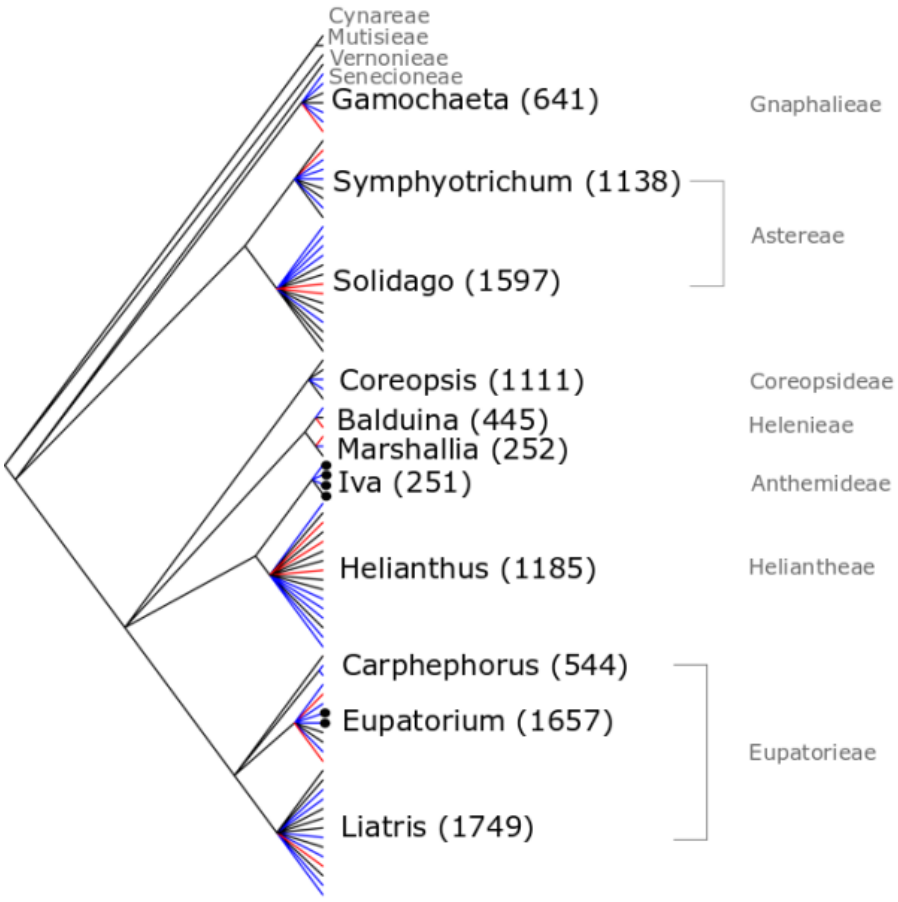

Figure 2.3 Genera selected for analysis shown on a phylogenetic tree. The number of branches per genus corresponds to the number of species included in the dataset per genus, and the number in parentheses after the genus name is the number of specimens of that genus in the final dataset. Tribal ranks are indicated in gray. The four asteraceous tribes also present in the southeast U.S. but not represented in this dataset are also shown. Blue branches indicate taxa with moist habitat preferences, and red branches indicate taxa with dry habitat preferences. Taxa indicated by black dots at branch tips are classified as wind-pollinated, and the remaining taxa are classified as animal-pollinated. The phylogeny was created using the Interactive Tree of Life (Letunic & Bork 2016).

I standardized the taxonomic names of the specimens using the iPlant Collaborative Taxonomic Name Resolution Service (Boyle et al. 2013) and removed specimens that were not of one of the 87 selected species. After determining the phenophase of each specimen (see below), I identified duplicate specimens, defined as specimens of the same species collected in the same county on the same date, assigned an average phenophase to one record in each set of duplicates, and removed the other duplicate record(s) from the dataset. I removed records of specimens that lacked flowers, including specimens in 100% budding and 100% fruiting phases. Finally, spatial (i.e., latitudinal or longitudinal) outliers were identified using Cleveland dot plots and removed to prevent disproportional effect of these points on subsequent models (Zuur et al. 2010).

The cleaned dataset consisted of 10,589 specimen records of 87 species in the 11 genera. Collections spanned a relatively wide temporal range (1891 – 2014) and spatial range (25.08611 – 36.4491° N; -94.02306 – -75.79799° W).

Determination of specimen phenophases

Many studies of phenology using herbarium specimens have assessed phenophase on a binary scale (e.g., flowers present or absent; Willis et al. 2017), yet this approach results in coarse estimates of plant flowering times, especially when the flowering duration of the species is long. Similarly, other metrics of phenology such as first flowering date have been shown to be unreliable (Moussus et al. 2010). I chose to compare peak flowering dates of specimens, as these are likely to be near the mean flowering time of the population, which has been implicated as the most reliable metric, even for small sample sizes (Miller-Rushing et al. 2008; Moussus et al. 2010). Since many specimens were not collected during peak flower, I assigned numerical phenophases (1 – 9) of each specimen based on the percentage of buds, flowers, and fruits present on the specimen using the nearest quartile values (0, 25, 50, 75, 100%) such that the total of all reproductive structures on a specimen equaled 100%. For example, specimens with 100% buds were assigned phenophase 1, specimens with 75% buds and 25% flowers were assigned phenophase 2, and so on according to the schema outlined in Figure 2.4. I then used flowering duration data (described below) to add or subtract days from the collection day of year (DOY; from 1 to 365) of the specimen to estimate its day of peak flowering (i.e., phenophase 5).

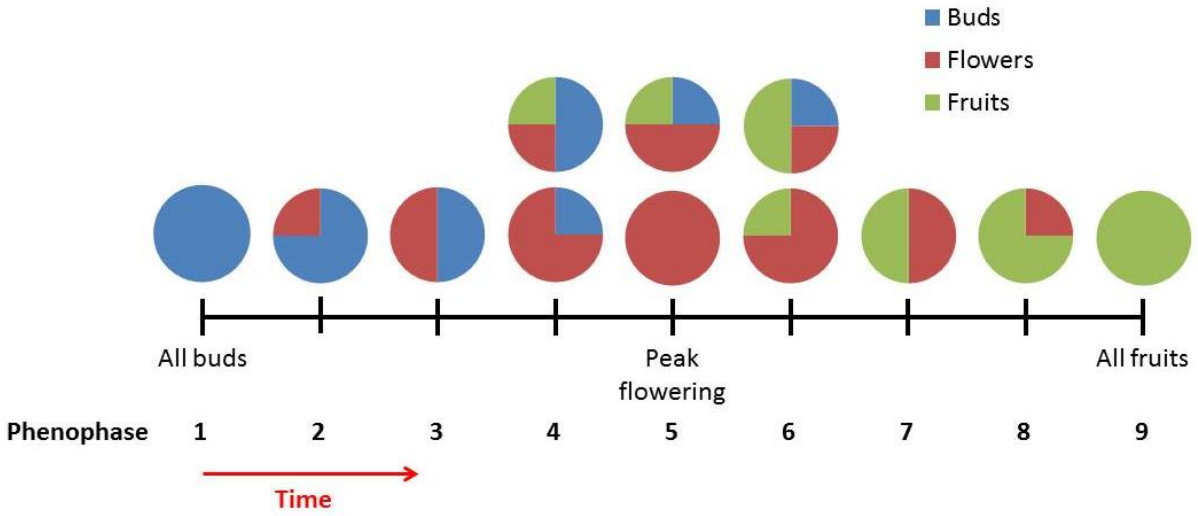

Figure 2.4 Graphic illustration of assignment of phenophases to specimens based on relative percentages of buds (blue), flowers (red), and fruits (green). Phenophases 4, 5, and 6 can be coded by either of two possibilities because these phases are mutually exclusive yet expected to take place at approximately the same time during the flowering season.

To determine how many days to add or subtract to estimate day of peak flowering, I located wild populations of one species from each genus in Leon county, Florida, USA during spring (1 species) and late summer to fall 2017 (10 species), and I marked at least 11 individuals of each species prior to or near the beginning of the flowering period. The quartile percentages of buds, flowers, and fruits on the plant were recorded every 3 – 4 days until the end of the flowering period (100% fruits), and these precentages were converted into phenophases following the same schema applied to herbarium specimens (Figure 2.4). For each species, I used a linear mixed effects model to determine the number of days elapsed per phenophase while taking into account different individual starting dates (DOY~phenophase + (1|individual)). The slope of this model was used to adjust the day of flowering for each specimen record to reflect estimated date of peak flowering. For example, the estimated length of each phenophase in the genus *Liatris* was two days, and thus the date of peak flowering for a *Liatris* specimen in phenophase 8 would be estimated by subtracting six days (2 days/phenophase $\times$ 3 phenophases) from the collection date.

This method operates under three main assumptions: (1) the relationship between time and phenophase is linear; (2) flowering duration does not vary significantly with location, climate, time, or population; and (3) flowering duration is similar among species within a genus. The data suggest that assumption 1 is reasonable in these species, as a significant linear relationship between time and phenophase was discovered for all monitored species (Appendix B). I also examined whether flowering duration of each species was better modeled by a curvilinear relationship, and this was only true for species of two genera, *Eupatorium* and *Marshallia*. When I used the curvilinear relationship to calculate peak flowering date for specimens of these genera and re-ran subsequent analyses, however, overall results did not substantially change (see Results). With regard to the second assumption, flowering duration is expected to be moderately shorter in warmer regions (Sherry et al. 2011; Bock et al. 2014). Consequently, I expect my measures of flowering duration estimated in Florida, the warmest state in the SECP, to be conservative and thus not introduce a large amount of variance. Although data on the effects of climate on flowering duration is lacking, some comparative (Kang & Jang 2004) and experimental (Gillespie et al. 2016) studies have found no correlation between warmer temperatures and flowering duration. There is some evidence that flowering duration has changed over time in some regions (Bock et al. 2014); however, Bock et al. (2014) only investigated flowering duration on the population level, which provides little evidence that individual rates of progression between phenophases have changed over time. Regarding assumption 3, even between genera, durations of the observed species ranged from 1.6 – 4.4 days per phenophase (Table B.1), indicating that flowering durations are similar within the family, and intrageneric variation in flowering duration is therefore expected to be mild.

Specimens flowering significantly out of season (before DOY 150 for fall-flowering species or after DOY 150 for spring-flowering species) were excluded from per-season analyses, as these individuals were likely responding to climate-independent cues such as fire (Conceicao & Orr 2012) or other disturbance.

Climate data

Bias-averaged monthly and annual average temperature and total precipitation data for all available years was obtained from the United States Climatology Network Version 2 in February 2017 (Menne et al. 2009). I used the R packages *sp* and *rgeos* to determine the nearest meteorological station to either the collection coordinates provided for each specimen or, if the specimen lacked coordinates, the centroid of the county in which the specimen was collected. Specimens that could not be assigned to a county from label data were excluded. For each specimen, I associated climate data in the collection year from the nearest meteorological station. Specimens with year + station combinations lacking climate data were excluded.

Previous studies have indicated that plants are most responsive to climate during the months immediately prior to a phenological event (Menzel et al. 2006; Munson and Sher 2015). Thus, I investigated spring-flowering species' sensitivities to climate in March and fall-flowering species' sensitivities to climate in July.

Because latitude may influence flowering time independently of temperature (Molnár et al. 2012; Bjorkman et al. 2017) and temperature was strongly correlated with latitude in my dataset, I used temperature deviation rather than absolute temperature as a fixed effect in my models. Temperature deviation was calculated as the difference between the actual value of temperature at the latitude of measurement (climate station) and an expected value calculated using a linear regression of monthly temperature versus latitude. All data were pooled in this linear regression, so the resulting expected values were those for all years and longitudes. In the temperature deviation metric, negative values reflect colder-than-average years and positive values reflect warmer-than-average years. Calculating precipitation deviation was not appropriate because precipitation did not vary as predictably with latitude, and I instead used annual and per-month (March or July) total precipitation values.

To determine whether climate has changed with time in this already warm region, I used simple linear regression models to determine the relationships of each of the climate variables (annual or per-month temperature deviation or total precipitation) with year. I created separate models for the entire range of dates (1894 – 2014) and for the range of dates beginning in 1970, which has been suggested as the onset of the most recent, rapid climate warming (Hodgkins et al. 2003). Climate and year data were those associated with specimens in the phenological dataset.

Trait data

Traits of flowering guild (spring, summer, fall, winter, or combinations of two or more of these) and habitat preference were assigned for each species using the Flora of North America (efloras.org) and Wunderlin and Hansen's Guide to the Vascular Plants of Florida (2011). Spring was defined as March-May, summer as June-August, fall as September-November, and winter as December-February in accordance with local temperature trends. Moist habitat preference was assigned to a species if its habitat description included words such as "moist," "mesic," "wet," or "swamp" and did not imply that the species also occurred in dry habitats. Similarly, dry habitat preference was assigned if its habitat description included words such as "dry" or "xeric" and did not indicate occurrence in moist habitats as well. Pollination mode was assigned to each species as either wind-pollinated or animal-pollinated according to floral morphology, field observations, and other resources.

Statistical analyses

I used linear mixed effects (LME) models (*lmer* function of *lme4* package in R; R Core Team 2016) to model the relationship between estimated peak flowering DOY and climate (continuous fixed effect), accounting for differences among species (random effect). As described above (see Climate data), climate variables were either annual/monthly average temperature deviation or total annual/monthly precipitation. LMEs allowing both slopes and intercepts to vary between species with temperature deviation (DOY~TemperatureDev + (TemperatureDev |Species)) had no better fit than LMEs allowing only intercepts to vary between species (DOY~ TemperatureDev + (1|Species); Appendix C). LMEs allowing slopes and intercepts to vary with precipitation failed to converge, so I was not able to properly assess their fit. Thus, for both climate variables (temperature deviation and precipitation), I used only variable intercept models. Fall-flowering and spring-flowering species were modeled separately because they are likely responding to different temporal cues. In these models, negative values of estimated model slope indicate an advance in peak flowering date, while positive values of estimated slope indicate a delay in flowering date. Confidence intervals (CI) reported are 95% confidence intervals calculated using the *confint* function in the *stats* package of R.

Modeling phenological sensitivity to climate in this way assumes that species respond similarly across the large spatial range covered by this dataset. This assumption seems reasonable in light of current evidence. Phillimore et al. 2012 discovered no differences in phenological sensitivities to climate among locations. Toftegaard et al. 2016 found that only 1 of 5 cruciferous species studied in Sweden showed a slight difference in phenological sensitivity among latitudes, though the potentially important effect of photoperiod (Tooke & Battey 2010) was not accounted for in this study. Plants at the same latitude in the UK and Poland demonstrated dissimilar phenological responses to climate (Tryjanowski et al. 2006), but

contrasting conditions at these two sites (i.e., island vs. mainland climates) may be driving this difference. Climatic conditions within the SECP are, in contrast, similar even across the latitudinal and longitudinal range of this study.

To examine whether traits affect phenological sensitivity to climate (i.e., peak flowering DOY), I tested for interactive and additive effects of flowering guild, pollination mode, dry habitat preference, and moist habitat preference in the LME models of phenology vs. climate described above. Only fall-flowering species were included in the latter three models to avoid the strong effect of flowering guild on climate sensitivity (see Results).

Climatic outliers were identified using Cleveland dot plots and removed from the dataset prior to model fitting, as they were likely to represent data quality problems rather than actual climatic conditions. All models were examined for homogeneity of variance and normal distribution of within- and between-group residuals. Statistically significant improvement of model fit was assessed by comparing small-sample-size corrected Akaike information criterion (AICc) values calculated with the *AICc* function in the *MuMIn* package in R.

Results

Climate change in the SECP

Annual, July, and March temperature deviation showed significant, positive relationships with year over the whole time period (1894 – 2014; Table 2.1). In the 1970 – 2014 time period, the rate of change tripled for annual and March temperature deviation and doubled for July temperature deviation such that the average temperature deviation increased 0.18 – 0.24° C per year (Table 2.1; Figure 2.5). July and March precipitation, but not annual precipitation, decreased over time over the whole time period. Conversely, total annual precipitation decreased dramatically over time in the 1970 – 2014 time period at a rate of nearly one inch per year, and March precipitation decreased 0.2 inches per year. July precipitation did not change significantly with time between 1970 and 2014.

Table 2.1 Relationships of climate variables with year over the entire period of specimen collection (1894 – 2014) and during the accelerated period of recent climate warming (1970 – 2014). Statistical significance is indicated as follows: * = $0.05 \geq p \geq 0.01$, ** = $0.01 \geq p \geq 0.001$, *** = $p < 0.001$.

Climate variable	**Climate change (1894 – 2014)** (°C / decade or in. / decade)	**Climate change (1970 – 2014)** (°C / decade or in. / decade)
Annual temperature deviation	0.068***	0.24***
July temperature deviation	0.10***	0.20***
March temperature deviation	0.058***	0.18***
Total annual precipitation	0.37	-9.8***
Total July precipitation	-0.57***	-0.61
Total March precipitation	-0.25*	-2.0***

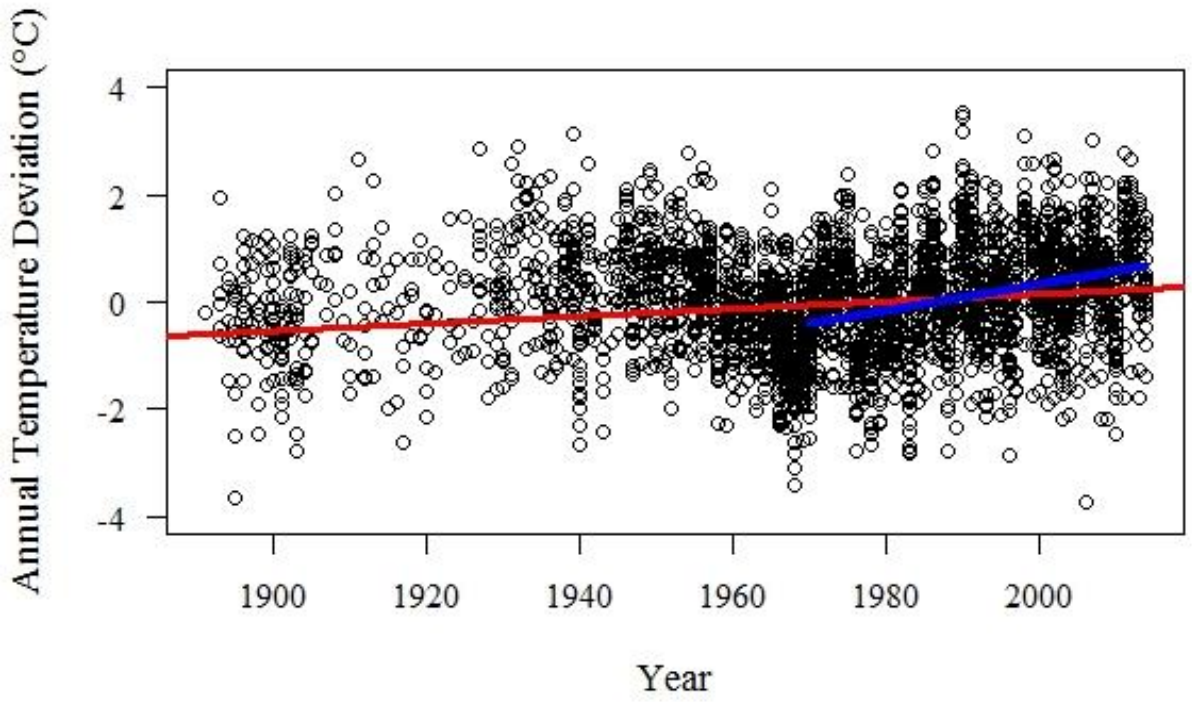

Figure 2.5 Change in annual temperature deviation with time in the specimen dataset. The red line shows linear regression of annual temperature deviation with year across the entire period (1894 – 2014), and the blue line indicates linear regression during the period of recent, rapid climate warming (1970 – 2014; Hodgkins et al. 2003).

Phenological sensitivities to climate

Spring- and fall-flowering species showed marked differences in flowering date change with climate. Flowering date of spring-flowering species (494 specimens, 6 species) advanced 2.2 days with every +1°C deviation in annual temperature (95% CI: 0.44, 4.2) and advanced 2.3 days per +1°C deviation in March temperature (95% CI: 1.4, 3.1; Figure 2.6).

In contrast, flowering dates of fall-flowering species (7106 specimens, 60 species) did not change with annual temperature (95% CI: -0.51, 0.67) and advanced only 0.49 days per +1°C deviation in March temperature (95% CI: 0.23, 0.76). Flowering dates of fall-flowering species were instead delayed 1.2 days per +1°C deviation in July temperature (95% CI: 0.68, 1.6; Figure 2.6).

Flowering date did not change with annual precipitation in spring-flowering species and was delayed by a mere 0.05 days per inch increase in annual precipitation in fall-flowering species (95% CI: 0.0053, 0.10). However, spring-flowering species' flowering dates were delayed 1.1 days for each inch increase in March precipitation (95% CI: 0.439, 1.8), and fall-flowering species' flowering dates were delayed by 0.49 days for each inch increase in July precipitation (95% CI: 0.32, 0.67).

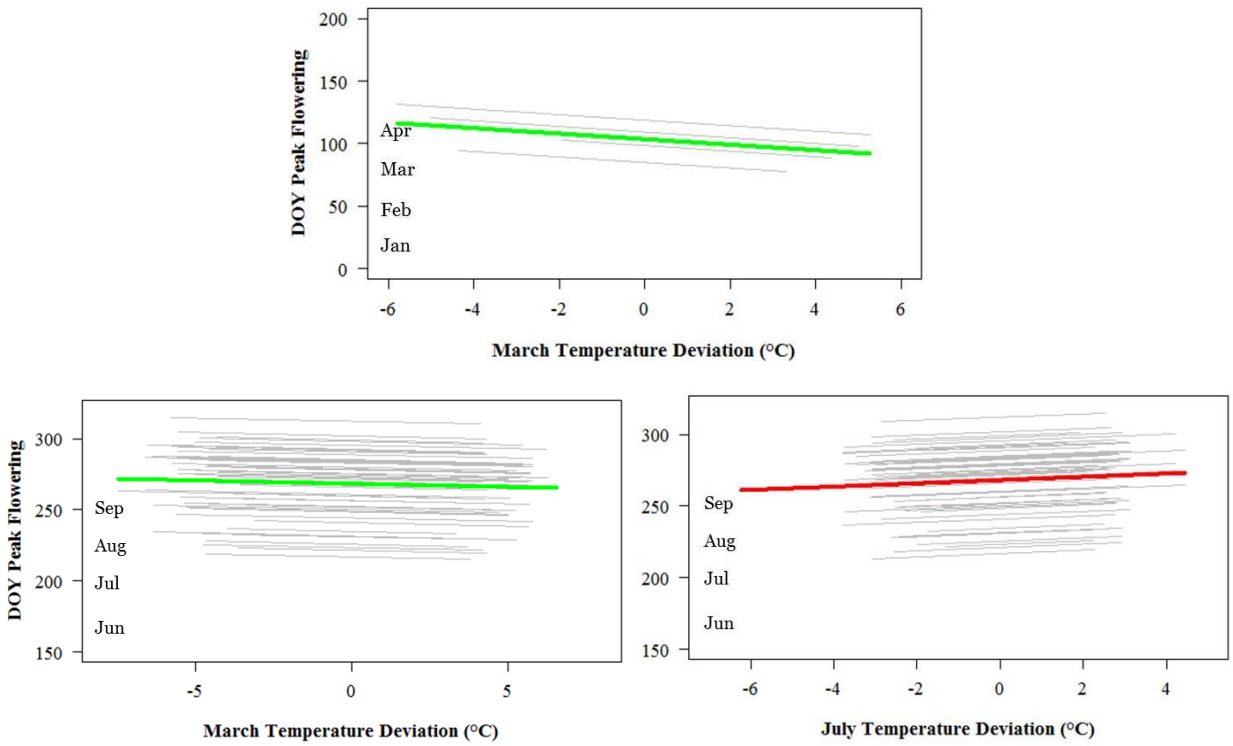

Figure 2.6 LME models of relationship between day of peak flowering (1 – 365) and temperature deviation for spring-flowering species (top) and fall-flowering species (bottom). Gray lines indicate regressions for each species across the range of temperatures in which it occurred in the dataset, and bold green or red lines indicate the average for all species across the total range of temperatures. Green lines indicate an advance in flowering time with increased temperature, and the red line indicates a delay in flowering time with increased temperature. Both spring-flowering and fall-flowering species showed significant advancement of flowering time with warmer-than-average March temperatures (top and bottom left), yet fall-flowering species demonstrated flowering time delays with warmer-than-average July temperatures (bottom right).

Effects of traits on phenological sensitivities to climate

Consistent with the differing sensitivities of spring- and fall-flowering species to climate as indicated above, a significant additive effect—but no interactive effect—of flowering guild was observed in a model of flowering date vs. annual average temperature deviation ($p < 0.001$).

No significant additive or interactive (climate*trait) effects of pollination mode (wind-pollinated vs. animal-pollinated) or dry habitat preference were indicated in a model of fall-flowering species' flowering date vs. July temperature deviation (Table 2.2). No interactive effect of moist habitat preference with July temperature deviation was observed; however, there was a slight trend toward a significant additive effect of this trait ($p = 0.09$). Similar results were obtained for all traits in models of fall-flowering species' flowering date vs. total July precipitation with the exception of pollination mode, which trended toward a marginally significant interactive effect ($p = 0.05$). The significance of this effect

must be interpreted with caution, as it was non-significant (p = 0.097) when peak flowering times of *Eupatorium* and *Marshallia* specimens were calculated according to curvilinear rather than linear phenophase/time relationships.

Table 2.2 Results of likelihood-ratio tests comparing LMEs with fixed effects of traits (FE1) and climate (FE2) to models with fixed effect of climate alone. Note that estimating p-values for mixed effects models is known to be imprecise. The significance of p-values close to 0.05 should be interpreted with caution. P-values less than 0.001 are in bold.

Response	**Fixed effect 1 (trait)**	**Fixed effect 2 (climate)**	**Effect type (FE1)**	**p**
Peak flowering time, all species	Flowering guild	Annual average temperature deviation	Additive	**<0.001**
			Interactive	0.13
		Annual total precipitation	Additive	**<0.001**
			Interactive	0.096
Peak flowering time, fall-flowering species	Moist habitat preference	July average temperature deviation	Additive	0.089
			Interactive	0.55
		July total precipitation	Additive	0.09
			Interactive	0.17
	Dry habitat preference	July average temperature deviation	Additive	0.14
			Interactive	0.79
		July total precipitation	Additive	0.13
			Interactive	0.99
	Pollination mode	July average temperature deviation	Additive	0.82
			Interactive	0.97
		July total precipitation	Additive	0.82
			Interactive	0.051

Discussion

Contrary to some previous predictions (e.g., Pau et al. 2011), plant species in the warm temperate climate of the SECP responded to temperature in ways similar to those in cold temperate climates. The 2.2 day phenological advancement of spring-flowering species per degree March warming shows striking agreement with estimates in, for example, north-central North America (2.4 days/°C; Calinger et al. 2013), northeast North America (3.07 days/°C; Miller-Rushing and Primack 2008), and the UK (1.4 – 3.4 days/°C; Sparks et al. 2000). Also somewhat unexpectedly, phenological responsiveness was identified in the Asteraceae, a plant family that has been suggested to track climate less strongly than other groups (Davis et al. 2010). These findings suggest that phenological trends—and perhaps many phenological

cues—may be reasonably generalizable among climate regions and some taxa, and even warm-adapted species like those in the southeast U.S. are not immune to the potential phenological effects of climate warming.

The flora and fauna of the SECP may instead be uniquely threatened given the phenological trends and evidence of climate warming discovered in this study. Flowering times of fall-flowering species were delayed by 1.2 days per 1°C July warming, suggesting that these species flower later with warmer-than-average summer temperatures. This response was consistent among fall-flowering species, as allowing slope (i.e., phenological sensitivity) to vary among species did not improve model fit. A delay in fall flowering is consistent with general trends in fall phenological shifts (Walther et al. 2002; Ibanez et al. 2010); however, this study is the first, to my knowledge, to detect this effect within a large number of fall-flowering species rather than in, e.g., trends of leaf senescence of deciduous trees (Gill et al. 2015).

This study demonstrated that climate warming is evident in the SECP and may have accelerated in the 1970's (Figure 2.5), indicating that warming-induced phenological delays and advances inferred here may be currently coming to bear. These phenological shifts could have numerous ecological and evolutionary consequences. Especially when coupled with advances in spring flowering events, delays in fall flowering could have negative consequences for associated species such as pollinators by creating a longer summer "dead zone" in which floral resources are scarce. Species that depend on the availability of flowers between peak blooms may experience increased competition for floral resources and potentially suffer from decreased fitness and population declines. Similarly, plants that flower at disparate ends of the flowering season may experience changes in abundance and diversity of floral visitors, which could affect fitness and alter selective pressures on phenological traits. Flowering later could also affect the phenological overlap of plants with herbivores (Liu et al. 2011), fruit dispersal patterns, or temporal overlap with climatic conditions. For example, assuming winter months are not delayed in the same way, plants that flower later may be more susceptible to flower or fruit damage due to cold conditions in coming winter months, just as flowering too early can predispose spring plants to frost damage (Inouye 2008). Fall phenological events may be just as critical to monitor as the story of spring.

An alternative explanation of the results of this study is that the flowering period of fall-flowering species is becoming longer, causing a delay in mean flowering time yet not affecting the onset of the fall flowering period. In this scenario, the effects on individuals and populations could be similar to those outlined above. However, the hypothesis of an extended growing season was not supported by these data, as variance in flowering date did not change significantly with temperature (Appendix D). Thus, at least in the southeast U.S., the relationship of fall-flowering species' flowering date with temperature is most likely due to delays in their flowering seasons.

In addition to delays with warmer-than-average summer temperatures, fall-flowering species experienced a small advance in flowering time in warmer-than-average springs. This contrasting response to different seasonal cues highlights the importance of understanding changes in climate if we are to predict climate change effects on phenology in this region. For instance, if this region experiences uniform warming within a year, warm springs may moderate the delaying effect of warm summers for fall-flowering species. Conversely, if plants are exposed to both spring cooling and summer warming, delays in flowering time may be compounded, potentially exacerbating effects on plants, pollinators, and higher trophic levels. Accurately predicting phenological responses of plants and monitoring potential effects will require careful attention to temperature cues across seasons.

Another important consideration resulting from this study is the impact of precipitation on both spring and fall flowering events. Unlike in many temperate, alpine, and Mediterranean climates (Abu-Asab et al. 2001; Hart et al. 2014; Gordo & Sanz 2005), precipitation was implicated as an important phenological cue in the SECP, with both spring- and fall-flowering species blooming later with increased spring or summer precipitation, respectively. Because temperature and precipitation are strongly correlated, it is difficult to determine whether the effect of precipitation on phenology is independent of temperature; however, precipitation decreased with increased temperature in this dataset (data not shown), and thus the effect of precipitation on phenology would be expected to be negative (i.e., advanced flowering date with increased precipitation) if it were strictly the result of correlation with temperature. While this was true for spring-flowering species, the opposite result was discovered for fall-flowering species, which may indicate an independent effect of precipitation on phenology in this region, potentially for spring events but more likely for fall events.

For both temperature and precipitation, the season of flowering proved critical to explaining phenological sensitivity, underlining the importance of considering seasonal phenological events separately rather than assuming a uniform response. Other species traits were not as informative in these analyses: including fixed effects of moist habitat preference, dry habitat preference, or pollination mode in LME models did not significantly improve fit. This might suggest that plants do not respond to climate differently depending on their habitat or pollination mode; however, it could also be the case that the coarse method of assigning traits to whole species prevents elucidation of key trends. Future investigations should focus on individual traits, particularly the habitat of each specimen, and perhaps more nuanced designations of pollination mode than a binary animal-pollinated versus wind-pollinated scale. In Asteraceae, some species—or even individuals within a species—may be self-pollinated or apomictic (Loran Anderson, pers. comm.), precluding their categorization using this method.

One potential exception to the lack of effect of traits on phenology sensitivity is the slight trend toward an interactive effect of pollination mode in response to precipitation. Wind-pollinated species

advanced their flowering times with increased precipitation in this model, while animal-pollinated species were delayed. This result may indicate that animal-pollinated and wind-pollinated species are responding to different phenological cues, which makes biological sense as the requirements of each to achieve pollination (i.e., pollinators versus favorable wind conditions) are quite different. Wind-pollinated species may have evolved sensitivity to precipitation to avoid having their pollen washed away in the violent convection storms that are frequent in late summer to early fall in the SECP. Alternatively, precipitation may precede or otherwise indicate wind conditions favorable for pollination and trigger a phenological response. Our understanding of the mechanisms by which wild, wind-pollinated species respond to precipitation in this region is limited. Given the biased representation of animal-pollinated versus wind-pollinated species (54 and 6 fall-flowering species, respectively) in this dataset, discovering an effect of pollination mode is remarkable, and the effect of this trait on phenological responsiveness warrants additional study.

Determining phenological sensitivities to each of these potential cues and among species with different traits is important to assessing the potential for phenological asynchrony in this region. Although some studies have suggested that the consequences of phenological asynchrony may not be as dire as once believed (Miller-Rushing et al. 2010; Forrest 2015), temporal mismatches with pollinators (Burkle et al. 2013; Kudo & Ida 2013) and increased overlap with herbivores (Liu et al. 2011) may decrease floral fitness (Thomson 2010; Miller-Rushing et al. 2010; Forrest 2015) and negatively impact pollinator populations (Burkle et al. 2013). The high biodiversity of the SECP may make it even more vulnerable to species loss due to such change, thus the threat of negative effects of asynchrony must be taken seriously. Critical to the species studied here, Rafferty et al. (2015) predicted that more generalized mutualisms with brief seasonal interactions—characteristics of many asteraceous species in the SECP—are more likely to become unsynchronized with other ecologically important species and may be in greater peril of possible detrimental effects. Timing with abiotic factors such as frost, storms (e.g., hurricanes, which are common in the SECP), and wind may also play a key role in determining population success. Rapid evolution of phenological traits under such potentially strong selective processes is possible (Franks et al. 2007), yet the capacity of these taxa and their interacting species to adapt quickly enough to avoid substantial fitness losses is uncertain.

Assessing phenological sensitivities of plant species to climate using herbarium specimen data has limitations due to, for example, the coarse spatial granularity of some specimen locality data and the lack of repeated measures of phenology at identical sites. Experimental and observational studies are needed to further examine the effect of traits and different climatic cues on plant phenological change. Nevertheless, this and similar studies provide critical data from which hypotheses can be formed and spatiotemporal trends can be extracted on a much larger scale than is feasible for most experimental and

observational studies (Willis et al. 2017). Efforts to obtain a similar scale of data via citizen science are underway (e.g., National Phenology Network, usanpn.org, Schwartz et al. 2012; Project Budburst, budburst.org; European Phenology Campaign, globe.gov/web/european-phenology-campaign), and combining observational and specimen-based records may prove a powerful way forward for understanding phenological change (Spellman & Mulder 2016). Still, these records lack the historical record of phenological events that herbarium specimens possess. With increasing availability of specimen data through digitization, development of protocols and standards for better integration of specimen-based phenological data (Yost et al. 2018), and development of statistical techniques to account for data limitations (Pearse et al. 2017), specimen data present ever-increasing opportunities to examine phenological trends and direct mitigation of adverse biotic change.

CHAPTER 3

ALL HANDS ON DECK: MOBILIZING THE BIODIVERSITY SPECIMEN COLLECTION COMMUNITY FOR EFFECTIVE OUTLIER DETECTION IN THE ANTHROPOCENE

Introduction

The current era of global biotic change requires rapid identification of the onset of potentially adverse processes (e.g., arriving invasive species) to facilitate effective countermeasures. Early detection and rapid response systems are quickly expanding beyond ecological monitoring (e.g., Anderson-Teixeira et al. 2014) and formal detection networks (e.g., FICMNEW 2003) toward ever more creative solutions, such as using new technologies (e.g., NEON's phenocam network, Brown et al. 2016) or enlisting the help of indigenous populations (Lauer & Aswani 2010), citizen scientists (Prince & Zuckerberg 2014; Scyphers et al. 2015), and even social media users (Daume 2016), to discover and track biotic change. Clearly, now is the time for all-hands-on-deck, but the mixed successes of these unconventional detectors of change (Lauer & Aswani 2010) underscore the need to mobilize new groups of experienced professionals to step into this role. One such group well-suited to the task of observing, documenting, and reporting change is collectors and preparators of biodiversity specimens (henceforth referred to here as "collectors"). However, I have found in a community survey, a search of collector training materials, and an exploration of identified "outlier terms" in over 75 million specimen records, that the collector community is hampered by several factors. Here, I present my findings and seek to mobilize the community as effective sentinels of change.

Collectors are a diverse group of researchers who collect organisms or parts of organisms for identification, study, documentation, and preservation in natural history collections. I began with the observation that collectors often have extensive personal experience, a network of scholarly resources, detective skills honed to recognize differences among organisms, and the motivation to resolve incongruities with existing resources when identifying their collections, especially since unique specimens may indicate new species. Nevertheless, collectors' observations of outliers appear to be regularly lost or, at most, noted on a label where the observation is left to the vagaries of future digitization and data mining. Indeed, most outlier detection occurs in an *ad hoc*, *post hoc* analysis of patterns among historically collected specimens, rather than using the collector's observations at time of collection (Peñuelas & Filella 2002; Meyers et al. 2009; Ożgo & Schilthuizen 2012; Bucher & Aramburu 2014). Meanwhile, significant biotic change can occur over stunningly brief periods of time (Franks et al. 2007; Stuart et al. 2014), revealing a need for rapidity that traditional, retrospective methods of outlier detection lack. Leveraging the experience and *in situ* activity of collectors could become a transformative step for global change biology if collectors were encouraged and empowered to immediately report their observations to stakeholders in an effective way.

Biotic change is a constant over long spatial and temporal scales (Tennant et al. 2017), but failure to identify the beginnings of adverse processes can have significant implications for biodiversity loss (Cardinale et al. 2012), costly exotic species invasions (USFWS 2012), and declines in ecosystem services such as pollination (Potts et al. 2010). Adverse processes may be initially marked by outliers (otherwise described as anomalies, oddities, etc.): observations that fall outside of a previously understood norm. I identified six outlier types that could be apparent to collectors and for which they might have resources or personal experience with which to construct a norm: distributional, phenological, ecological, morphological, behavioral, and genetic outliers. These outlier types are often interrelated due to cascading effects. For example, newly arrived species might lead to genetic outliers (e.g., hybrids) that are first revealed by morphological and behavioral changes, which may then produce altered ecological interactions. Determining the exact causes of biological outliers is often beyond the purview of the collector and can be quite complicated to determine. For example, morphological, anatomical, and phenological outliers might be due to previously unexpressed phenotypic plasticity (Torres-Dowdall et al. 2012), physiological stress due to environmental change (Harper & Wolf 2009), significant mutations such as polyploidy (Otto & Whitton 2000), hybridization (López-Caamal & Tovar-Sánchez 2014) or the fingerprint of selective processes (Grant & Grant 1995). If anything, this complexity argues for a smooth and timely hand-off of information between those who first notice the outlier and the stakeholders who use that information for future study, policy, and management. When properly empowered, biodiversity specimen collectors can play a critical role in quickly detecting and reporting biotic change, yet the collecting community remains largely untapped for this purpose. With this survey and subsequent analyses, I aim to illuminate impediments to effectively mobilizing this community and catalyze future action.

Materials and Methods

I designed an 18-question survey of biodiversity specimen collectors to determine (1) whether collectors detect outliers other than new taxa, (2) how and with what frequency collectors detect outliers, (3) how collectors document and report outliers, (4) what resources were employed during their training as collectors, and (5) what factors impede outlier documentation and reporting (S1). I distributed a weblink and brief description of the survey via natural history collection listservs (ECN-L, HERBARIA, iDigBio, NHCOLL-L, SERNEC, TAXACOM, TDWG), iDigBio social media (Facebook, Twitter), and with the help of the Natural Science Collections Alliance (NSCA). At least four professional societies within the NSCA also distributed the link through social media or email lists. The survey was open for 36 days, and I sent a reminder email via the same listservs two weeks prior to the close of the survey. All respondents gave their informed consent to participate online prior to completing the survey, and ethical

approval for the survey and the study methods was obtained from the Institutional Review Board of the Florida State University prior to distribution (Appendix E). Not all participants answered every survey question, and each question received an average response rate of 82%. Results presented in this chapter exclude responses from collectors of fossils (10 responses) because I am focusing on recent biotic change.

In the survey, I asked collectors to list words and phrases that they would use to describe outliers. I queried all available fields of the 75,569,035 specimen records (as of January 30, 2017) served by iDigBio, the United States National Science Foundation's National Resource for Advancing Digitization of Biodiversity Specimens and a major aggregator of biodiversity specimen records, for these "outlier terms" (some were phrases, such as "than normal"). I sought to determine their frequency and distribution of use in digitized specimen data with special attention given to the use of those words that alone could communicate outlier status and therefore support a stakeholder's search of the data using existing search functionality of the iDigBio web portal and API.

The survey also asked collectors to name literature reference they used when learning to collect, and I received 84 references. Many of these, such as monographs and field guides, proved to be sources other than instructional materials and were thus considered out-of-scope. Only 22 of the provided references were considered in-scope, and seven of these were inaccessible to us. Starting with these 15 resources, I assembled additional collector training literature (e.g., institutional manuals and written protocols) from works cited, collection websites, and other sources (see Appendix F for methods and bibliography). The search produced an additional 33 books, articles, protocols, manuals, and websites (Appendix F). I examined all materials for any reference to methods of detecting, documenting, or reporting biological outliers.

Results and Discussion

Community survey

I designed and distributed an 18-question survey of biodiversity specimen collectors to evaluate the current state of outlier detection and documentation by collectors and how collectors can be better empowered in this role. The survey received 222 responses with representation from collectors of 10 major groups of organisms and additional groups in the "Other" category. A wide range of collecting experience (0–31+ years) was represented among survey respondents. The survey also asked collectors to provide terms that they would use to describe outliers. Collectors provided 170 unique words and phrases (Figures 3.1 and 3.2), and 70% of these were unique to the respondent. Some of these words alone could communicate outlier status (e.g., "atypical," "strange," and "aberrant"), whereas most would only do so in a longer phrase (e.g., "fruiting," "small," and "new").

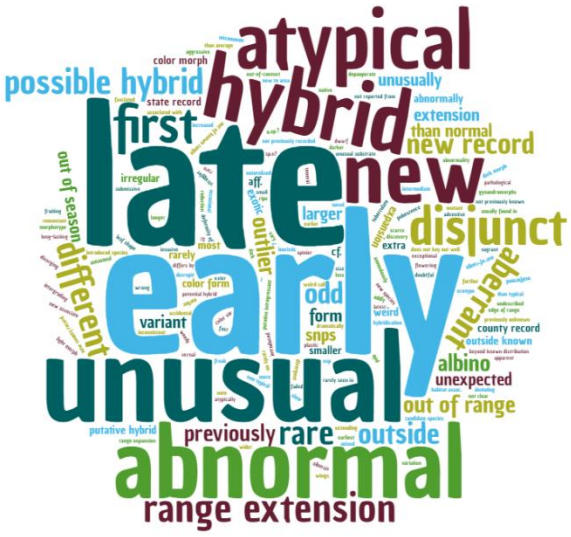

Figure 3.1 Word cloud of terms and phrases used by collectors to describe outliers, as reported in the survey. Relative word sizes correspond to relative number of collectors who listed the term.

Collectors discover some outliers frequently and easily

The survey results corroborate my main premise: collectors detect specimen outliers, and they do so using a rich suite of resources. Most frequently, collectors reference other specimens of the same taxon when identifying outliers (91% of respondents), but they also use personal experience (88%), monographs and other highly vetted resources (81%), taxonomic experts (71%), online specimen data aggregators (53%), and other resources (S1).

About 80% of respondents (excluding responses of "Does Not Apply") indicated that they discover distributional and morphological outliers at least "occasionally," and almost half of respondents considered those two types of outlier "Easy" or "Very Easy" to detect. On the other hand, more than half of respondents reported that they rarely or never discover behavioral, genetic, and anatomical outliers, and more than half deemed them "Difficult" or "Very Difficult" to detect. Over half of respondents indicated discovering phenological outliers at least "occasionally," and phenological outliers were considered equally "Difficult" and "Easy" to detect.

But I identified relatively few outlier reports among the specimen records

More than half of respondents (58%) reported noting outliers on specimen labels, and nearly half (49%) reported noting outliers in field notes/journals. Data from both of those sources might be expected to eventually appear in digital specimen records after some time lag. Fewer than 17% reported that they "generally do not make note of outliers."

However, when I queried all available fields of the 75,569,035 specimen records (as of January 30, 2017) served by iDigBio, a major aggregator of specimen records, for "outlier terms" provided by collectors, very few specimen records contained even the most frequently cited outlier terms. Survey respondents most frequently cited "abnormal" as used to document both morphological and anatomical outliers, "early" to document phenological outliers, "extension" to document distributional outliers, "unusual" to document ecological outliers, and "hybrid" to document genetic outliers. However, queries of specimen records indicate that these terms may not be used in these contexts consistently. "Hybrid" appeared most frequently in the iDigBio-aggregated data (in as many as 19,556 of 75 million records, about 0.03%), followed by "early" (16,218 after removing fields that were references to geological time periods; about 0.02%) and "extension" (13,177; about 0.02%). Other terms appeared much less frequently than 0.01% of the records. "Unusual" appeared in just 2,862 field values and "abnormal" in just 623. Other seemingly promising terms that could communicate outlier status without being contained in a longer phrase appeared infrequently as well: "odd" in 1,730 field values, "atypical" in 741, "strange" in 498, "vagrant" in 250, "aberrant" in 161, "weird" in 100, "straggler" in 60, and "unexpected" in 35. "Outlier," the term that I use here to describe the topic, appeared in 1,085 field values, but most of these were references to the locality (e.g., "SW outlier of the Sierra de Manantlán"), rather than a quality of the specimen.

Furthermore, using the terms to search particular fields for specific outlier types is not straightforward. This arises because some terms have multiple meanings (e.g., the use of "early" to describe phenology and geological time periods), some terms can flag different types of outliers (e.g., a specimen can have "abnormal" morphology, anatomy, behavior, etc.), all terms are found in many fields (Figure 3.2), and most term-by-field combinations did not produce results of single outlier types. Terms that appeared in at least one database field were present in an average of 15 fields (median=13); 32 terms did not appear in any records. Of the highly cited terms, "early" appeared in 38 fields (most frequently in occurrenceRemarks after earliestEpochOrLowestSeries), "hybrid" appeared in 32 fields (most frequently in scientificName), "unusual" in 25 (most frequently in occurrenceRemarks), "abnormal" in 17 fields (most frequently in occurrenceRemarks), and "extension" in 17 (most frequently in dynamicProperties). Other seemingly promising terms appear most frequently in occurrenceRemarks except "strange" and "outlier" (which occur most frequently in locality; e.g., "Strange Rd." and "SW outlier"), "vagrant" (most frequently in habitat), "straggler" (most frequently in taxonRemarks), and "unexpected" (most frequently in fieldNotes).

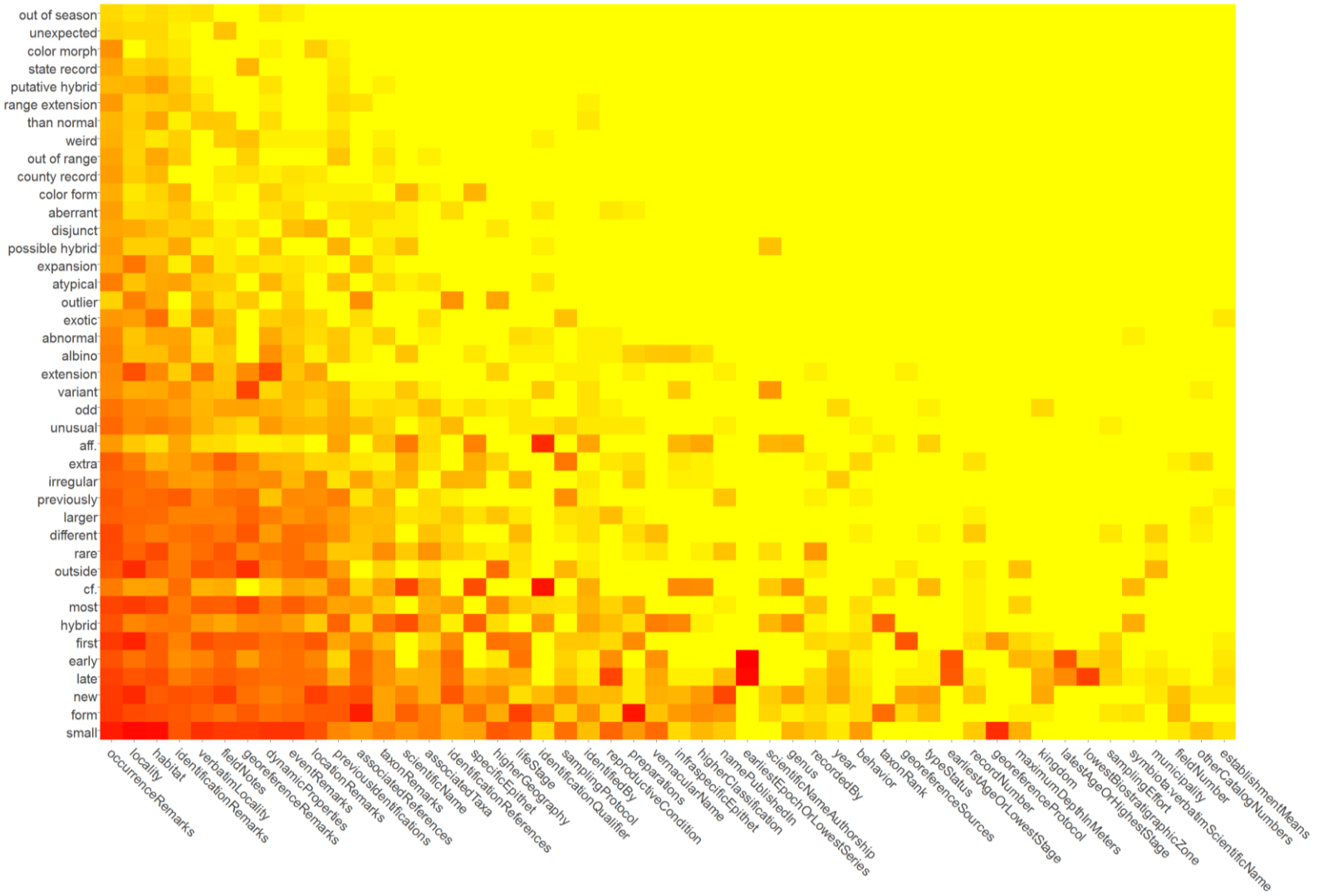

Figure 3.2 Heat map representing the frequency of terms provided by two or more survey respondents to describe outliers (y-axis) in data fields (x-axis) of 75,569,035 specimen records (as of January 30, 2017) accessed through iDigBio. Only the fifty data fields containing the highest frequency of outlier term occurrences are shown for simplicity.

Impediments to reporting are diverse but include lack of protocols and training

What might be impeding outlier reporting, as is suggested by the rarity of records with outlier terms among the 75 million examined? About a quarter (26%) reported not generally being impeded in noting or reporting outliers, suggesting that three-quarters of respondents to the question recognize impediments. The greatest number of respondents reported that lack of time served as an impediment (46%), followed by low confidence in the outlier status (43%), and lack of standard community protocols for reporting (30%).

Half of survey participants indicated that they were not taught or were self-taught to document outliers. About 17% indicated they were taught on the job or by the example of mentors or other professionals. Only 3% of participants indicated that they learned to document outliers by reading literature, and just 2% indicated that they learned through courses. Accordingly, only one of the 45 training references (RIC 1999) I searched included recommendations for documenting outliers or anomalies. None described the important role of collectors in documenting change or provided best practice recommendations for the activity.

The community seems receptive to mobilization as sentinels of change

This study suggests that collectors should be an easy group to mobilize for outlier detection and documentation. A large percentage (81%) of collectors view themselves as at least somewhat being "on the front lines of observing and documenting change in Earth's biota," and the majority of respondents (59%) have seen their attitudes on the topic change at least somewhat during their career. Given the latter, I was somewhat surprised that I did not find evidence that outlier terms were becoming more frequently used (data now shown). For example, the percentages of records that cited "unusual" and "odd" peaked in the 1990's and have since declined.

Many collectors report outliers other than new taxa to close colleagues (75%), experts on the respective taxon (65%), and in journal articles or reports (50%), though fewer do so to other potential stakeholders (e.g., land managers or park staff, 25%; government agencies, 22%; and enthusiast groups, 14%) or to online community resources that those stakeholders might use (e.g., iNaturalist or EDDMapS, 9%).

Moving forward

I seek to make outlier detection, documentation, and reporting core activities of the collection community, and I view community buy-in and collaborative progress as critical to moving beyond what might be described as abundant but unfocused goodwill towards the activities. I recognize three steps in the collector community's outlier-related data flow that merit further attention given my findings (Figure 3.3): (1) immediate reporting to a broader set of stakeholders, (2) reporting on the specimen label, and (3) digitization of specimen label data. This study identified a near-complete absence of training resources

related to these steps. I suggest that a first step will be to develop best practices for each in such settings as workshops and working groups, at which times requirements of missing resources such as data standards and cyberinfrastructure can additionally be addressed.

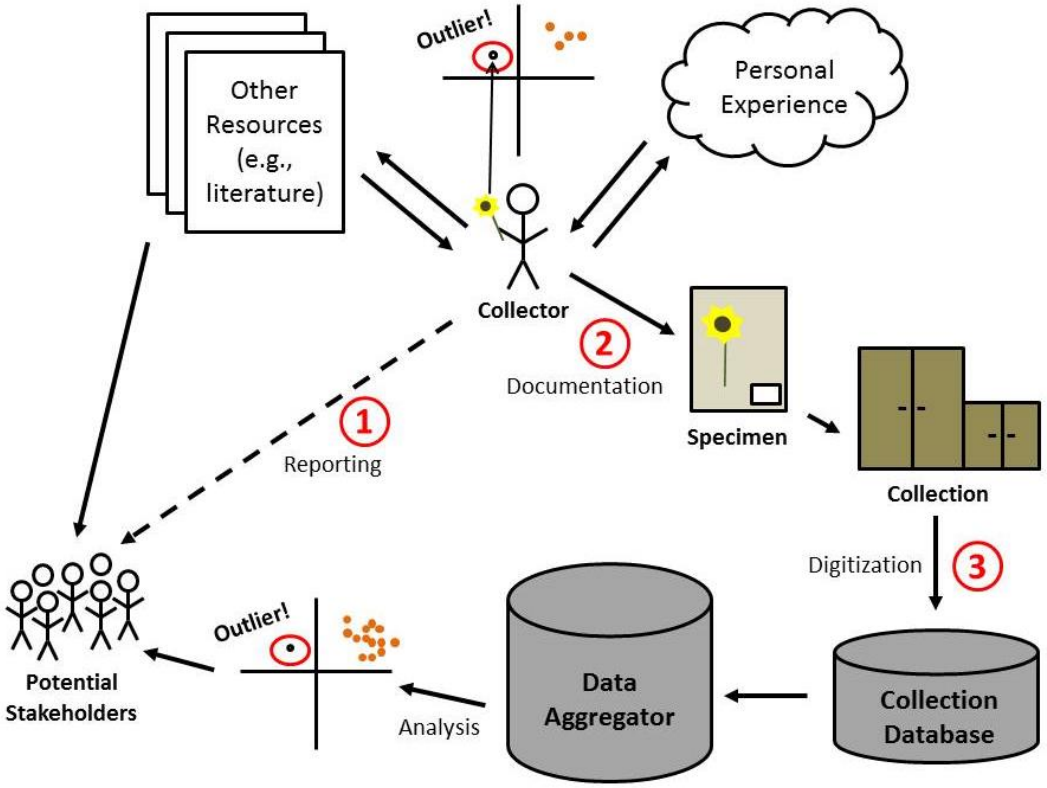

Figure 3.3 Simplified model of the current process of outlier documentation and reporting by collectors (black arrows). The steps I identified as needing further attention are highlighted by red numbers: (1) immediate reporting to a broader set of stakeholders, (2) documenting on the specimen label, and (3) digitization of specimen label data. The survey suggests that few collectors report outliers directly to consumers (e.g., land managers), indicated above by the dashed arrow. Some collectors note outlier information on the specimen label or in field notes; however, they often do so using unstandardized, largely descriptive terms that may impede data accessibility. This problem is compounded upon digitization of the specimen data and aggregation in e.g., iDigBio or GBIF, as unstandardized language is stored in a variety of database fields, making outlier data difficult to distinguish from other types of data.

Best practice recommendations are outside the scope of the current chapter, but I will make a few observations that could be useful when framing those community-driven activities. First, I recognize that if collection data were "born digital," as is likely in the future, Figure 3.3 would look different. That is, were collection data regularly entered directly into a specimen data management system by the collector (e.g., on a mobile device) with subsequent creation of the specimen label from that entry, then best practices for outlier description could be supported in the data entry forms and publication of the outlier information automated with the specimen data management system as the provider. The three numbered

steps in the figure could then occur simultaneously or in close succession.

However, it is useful to recognize the three steps since they reflect the current procedure of most collectors, and each highlights a different challenge for the community to address. Step (1) might involve the construction or tailoring of a subscription service for outlier observations, which has parallels with trailblazing projects like FilteredPush (for digital annotations of specimen data; Wang et al. 2009) and FreshData (for occurrence data more generally; Hammock & Poelen 2016). Ideally, this step would result in the data efficiently arriving in the "go-to" sites that stakeholders might visit to find relevant types of outlier data, such as the Early Detection and Distribution Mapping System (eddmaps.org; Wallace & Bargeron 2014) for invasive species and National Phenology Network (usanpn.org; Schwartz et al. 2012) for phenology. Step (2) involves greater semantic sophistication in the description of outlier observations, which should take into account data discovery mechanisms (e.g., at iDigBio and GBIF), data models at the go-to sites for stakeholders like those mentioned above, and work on relevant ontologies, such as the Plant Phenology Ontology (Stucky et al. 2017; Willis et al. 2017) and the Open Annotation Ontology (e.g., Morris et al. 2013). Step (3) involves careful consideration of the downstream discovery and reuse of the collector's observation, enabling data providers (e.g., iDigBio and GBIF) to optimize discovery and delivery of this type of data, and needs to engage the deep expertise on the topic represented by the Biodiversity Information Standards (TDWG) group. Future work on these steps will also benefit from the foundation that the community survey results and subsequent analyses provide.

In the meantime, I encourage collectors to report outliers in a timely and effective way to a broad audience of potential stakeholders in ways appropriate to the outlier type, taxon, and location and then tell the collector community how they did it. Twitter is an increasingly effective mechanism for community conversation (e.g., @iDigBio has over 2,600 followers as of December 2017), and I suggest tweeting the information with the hashtag #ODDbyCollectors (shorthand for Outlier Detection and Documentation by Collectors). iDigBio (@idigbio) will tweet an example specimen record in which the collector noted the outlier status of the specimen most Wednesdays in 2018 using that hashtag and others as appropriate (e.g., #iseechange). I invite you to join this conversation and encourage those with a deeper interest to join the iDigBio Outlier Detection and Documentation by Collectors Working Group that I co-chair with Austin Mast.

Collectors are ideally suited to provide early warnings of biotic change in the Anthropocene for the benefit of science and society. Let's work together to ensure that our community is successfully mobilized for this purpose.

CHAPTER 4

RAPID ENHANCEMENT OF BIODIVERSITY OCCURRENCE RECORDS USING UNCONVENTIONAL SPECIMEN DATA

Introduction

In this era of anthropogenic influence, the need to understand past and present species distributions to track biotic change has never been greater. Understanding geographical and temporal distributions of species is central to biogeography (Brown et al. 1996; Lomolino et al. 2016), biodiversity research (Gaston 2000; Ricklefs 2004), evolution (Sexton et al. 2009), and ecology (Weins & Graham 2005; Parmesan 2006), among other disciplines, and is vital for biodiversity conservation and planning (Ferrier 2002; Mota-Vargas & Rojas-Soto 2012), yet our knowledge of where and when species occur is incomplete. Biodiversity specimens, such as dried, pressed plants housed in herbaria, are a significant source of species distribution data (e.g., Otero-Ferrer et al. 2017), as each specimen represents a verifiable occurrence of a species at a certain place and time. Recent efforts to digitize biodiversity specimen data have made millions of specimen records and images publically available on online portals (e.g., idigbio.org). However, even en masse, specimen data can be incomplete and geographically, temporally, or taxonomically biased, especially in under-studied regions (Tobler et al. 2007; Stropp et al. 2016; Daru et al. 2017). Observational occurrence datasets such as those aggregated by Global Biodiversity Information Facility (GBIF; gbif.org) and iNaturalist (inaturalist.org) are also rapidly expanding our knowledge of species distributions, but because historical records are often rare, observational datasets often cannot answer essential questions such as how species distributions may shift in time and space with changes in climate and land use.

One potentially transformative resource for obtaining reliable historical occurrence data remains relatively untapped: records of associated taxa. "Associated taxa," taxa co-occurring with a biodiversity specimen at the time and place of collection, are often documented on specimen labels in addition to standard date, locality, and collector data (Anderson 1965; Radford et al. 1974), and these data can serve as occurrence records of the associated taxon (Figure 4.1). Like biodiversity specimen records, these observational records have the advantage of traversing time and space, and because collectors are usually experienced professionals, associated taxon records are likely to be reliable. Associated taxon records represent what the collector did not collect, perhaps because of time, resource, or technical restraints such as collecting permits, and therefore, once aggregated, they may help fill the gaps left by collecting biases. Moreover, many more associated taxon records can be created in the time that it takes to collect one biodiversity specimen, which suggests that associated taxon data, if consistently recorded, can rapidly expand current occurrence data.

Figure 4.1 Example herbarium specimen with associated taxa noted on the label

To explore the potential for associated taxon data to augment current occurrence data, I developed R code (R Core Team 2016) to isolate associated taxon records from 84,328 digitized specimen records available from the Florida State University Robert K. Godfrey Herbarium as of September 2017. In this chapter, I report on the quantity and quality of mined data, explore their usefulness in expanding known species distributions, and discuss challenges and considerations for producing and using these data.

Materials and Methods

Observational dataset generation

All 84,328 available digitized herbarium specimen records (as of September 13, 2017) of the Florida State University Robert K. Godfrey Herbarium (henceforth "FSU herbarium") were downloaded using the data portal provided by iDigBio, the U.S. National Science Foundation's National Resource for Advancing Digitization of Biodiversity Specimens and a major aggregator of biodiversity specimen records. The FSU herbarium is a large (220,000+ specimen) herbarium located within the North

American Coastal Plain biodiversity hotspot (including the U.S. Southeastern Coastal Plain; Noss et al. 2015) in Tallahassee, Florida, USA. Digitization efforts as of September 2017 have primarily focused on the flora of Florida, though the downloaded dataset contained specimens from around the world. This dataset was chosen because associated taxon records are consistently stored in the "habitat" database field in accordance with FSU databasing protocol; however, the method developed here can be applied to any database field or multiple fields. Duplicate specimen records, defined as records of the same species collected in the same county on the same date, were removed, reducing the dataset to 72,120 unique occurrence records.

The code developed for this study uses the Global Names Recognition and Discovery application programming interface (GNRD API; Myltsev and Mozzherin 2016) to distinguish scientific names in the "habitat" database field of the downloaded dataset. The GNRD tool is a web-based application that recognizes families, genera, species, and even abbreviated binomial names (e.g., *E. elatus*) in images, documents, or text strings, and the GNRD RESTful API parses submitted text strings or websites. For each recognized scientific name in the habitat field, the code created a new observational occurrence record with relevant data (e.g., locality, date, habitat) copied from the original specimen record.

The resulting associated taxon dataset was cleaned by removing duplicate records (as defined above), records that had been created from words that the GNRD API misinterpreted as taxonomic names (e.g., Apalachicola, Wakulla), and a handful (8) of records that included the word "no" in front of the associated taxon name. Another R script was developed to resolve the likely identity of observational records with abbreviated binomial names (1,510 records) by matching the abbreviated genus letter to the genus of the original specimen record or, if the genus letter did not match the genus of the original record, the first genus listed in the habitat field. This algorithm was able to correctly infer the binomial name of the associated taxon for 89% of the records. All records with inferred genera were hand-checked for accuracy.

Because some collectors collect species that they also list as associated taxa, I combined the original specimen records with the associated taxon records, standardized all scientific names using the Taxonomic Name Resolution Service v4.0 (Boyle et al. 2013), and again removed duplicates as defined above. The Taxonomic Name Resolution Service also identified misspellings and flagged unknown taxonomic names, which were manually resolved prior to duplicate removal. Resolving misspellings was particularly important for associated taxon data since these data are manually transcribed into a database field rather than chosen from a pick list and are thus prone to typographic errors. Duplicate removal reduced the combined dataset from 86,669 records to 85,493 records.

Identification of range extensions

Potential extensions of known species distributions were identified using an R script that compared the counties in which associated species were found to known county-level species distributions according to each of three databases: the Atlas of Florida Plants (for Florida specimens only; Wunderlin et al. 2017), the United States Department of Agriculture PLANTS database (for U.S. specimens only; USDA 2017), and iDigBio specimen records using the iDigBio API via the *ridigbio* package. Purported range extensions according to the Atlas of Florida Plants were manually verified to ensure each was not an artifact of incongruent taxonomy or other errors. Because the purpose of this chapter is to examine the potential for associated taxon data to expand known taxon distributions rather than produce a full report of new county records, only a subset (100) of non-Florida range extensions of both the USDA PLANTS-based new county records and iDigBio-based new county records were examined to estimate the number of "true" potential new county records that were not the result of errors.

Comparison of specimen data and associated taxon data

The habits and native statuses of taxa reported in original specimen record and associated taxon records were compared to determine whether certain plant types are more frequently documented as associated taxa rather than collected as specimens or vice versa. Plant habit (herb/forb, tree, shrub, or graminoid) and native status (native or introduced) were assigned to each taxon using the USDA PLANTS database (USDA 2017), the Flora of North America (efloras.org; eFloras 2008), and the Atlas of Florida Plants (Wunderlin et al. 2017). For these comparisons, "original specimen records" are only those from which the R script recovered associated taxon records in their habitat fields, and "associated taxon records" are the recovered observational records after data cleaning—including primary duplicate removal—but prior to combination with original specimen records and final duplicate removal.

The R script developed to produce associated taxon records and the dataset generated during this study are deposited on the Florida State University Digital Repository (code: http://diginole.lib.fsu.edu/islandora/object/fsu%3A539055; data: http://diginole.lib.fsu.edu/islandora/object/fsu%3A539064).

Results

After data cleaning and both duplicate removal steps, 13,372 associated taxon records were extracted from the initial dataset of 72,120 unique herbarium specimen records, representing an 18.5% increase in total occurrence records (Figure 4.2). Nearly two-thirds of these records (61.1%) were identified at least to species, and all but two of the remaining records were identified to genus. Of the associated taxon dataset, 1,262 records (8.6%) had abbreviated scientific names (e.g., *E. elatus*) that were inferred to species using specimen data and geographic context.

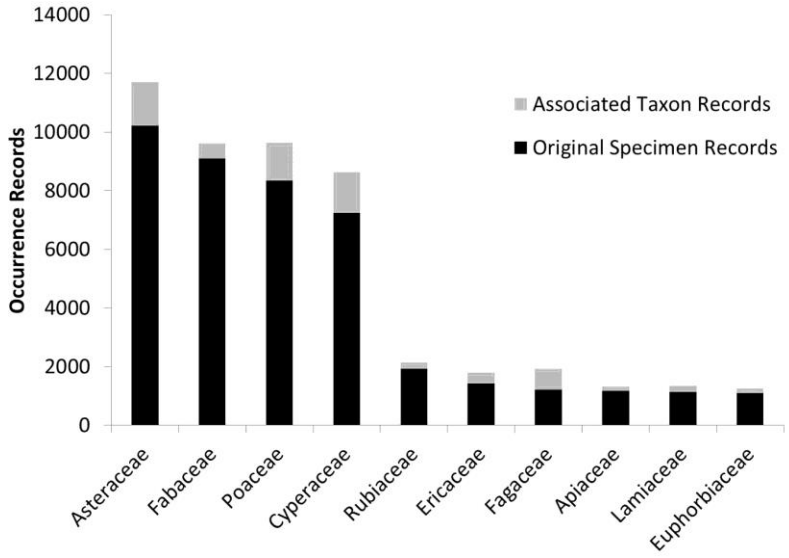

Figure 4.2 Increases in occurrence records due to extraction of associated taxon records. The 10 most specimen-rich families in the original dataset of digitized herbarium specimen records are shown. These 10 families account for nearly 60% of the total specimen records.

Associated taxon records consisted of 2,973 taxa, 207 plant families, and one family of lichen, while the original specimen dataset contained 9,685 taxa and 317 plant families. Occurrences of the sunflower family (Asteraceae) were most frequent in both the associated taxon dataset and the original specimen dataset; however, the top ten most occurrence-rich plant families differed between datasets (Figure 4.3). Notably, families containing dominant canopy and shrub taxa in this region—the oaks (Fagaceae), pines (Pinaceae), magnolias (Magnoliaceae), and palms (Arecaceae)—comprised 4.9%, 2.3%, 1.9%, and 1.7% of the associated taxon dataset, respectively, while only comprising 1.7%, 0.3%, 0.3%, and 0.2%, of original specimen records.

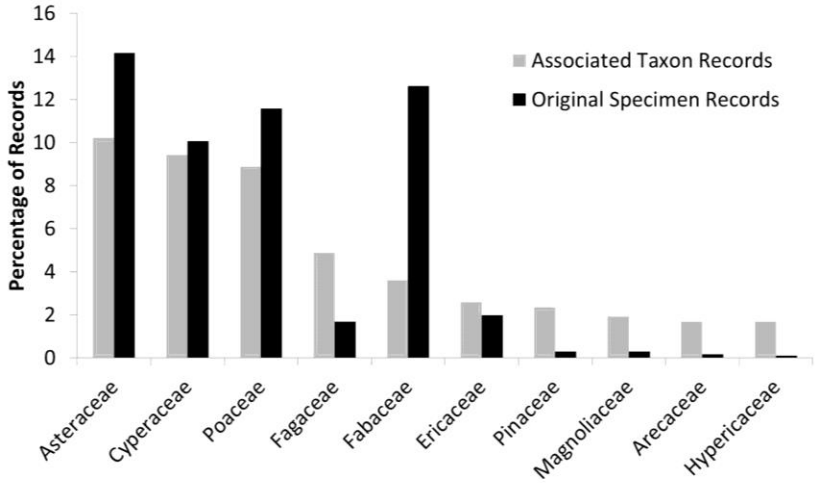

Figure 4.3 Comparison of relative family composition of the associated taxon dataset and the original specimen dataset. The 10 most occurrence-rich families in the associated taxon dataset are shown.

Associated taxon records consisted of a greater percentage of trees (22.2%) and shrubs (13.4%) when compared to original specimen records for which associated taxa had been found (9.9% trees, 11.7% shrubs). Conversely, specimen records consisted of more herbs/forbs and graminoids (51.6%, 26.8%) than associated taxon records (43.5%, 20.9%).

Temporal trends in associated taxon data did not closely follow specimen collecting trends (Figure 4.4). Associated taxon records spanned a narrower range of time (1937 – 2016) compared to specimen records (1880 – 2016), with the majority of associated taxa recorded during the mid-1980s. In one year (1988), the number of associated taxon records exceeded the number of collected specimens. This peak could reflect changes in cultural norms of collecting, perhaps facilitated by advances in technology (e.g., printed labels) or increased activity of a few collectors who regularly documented associated species.

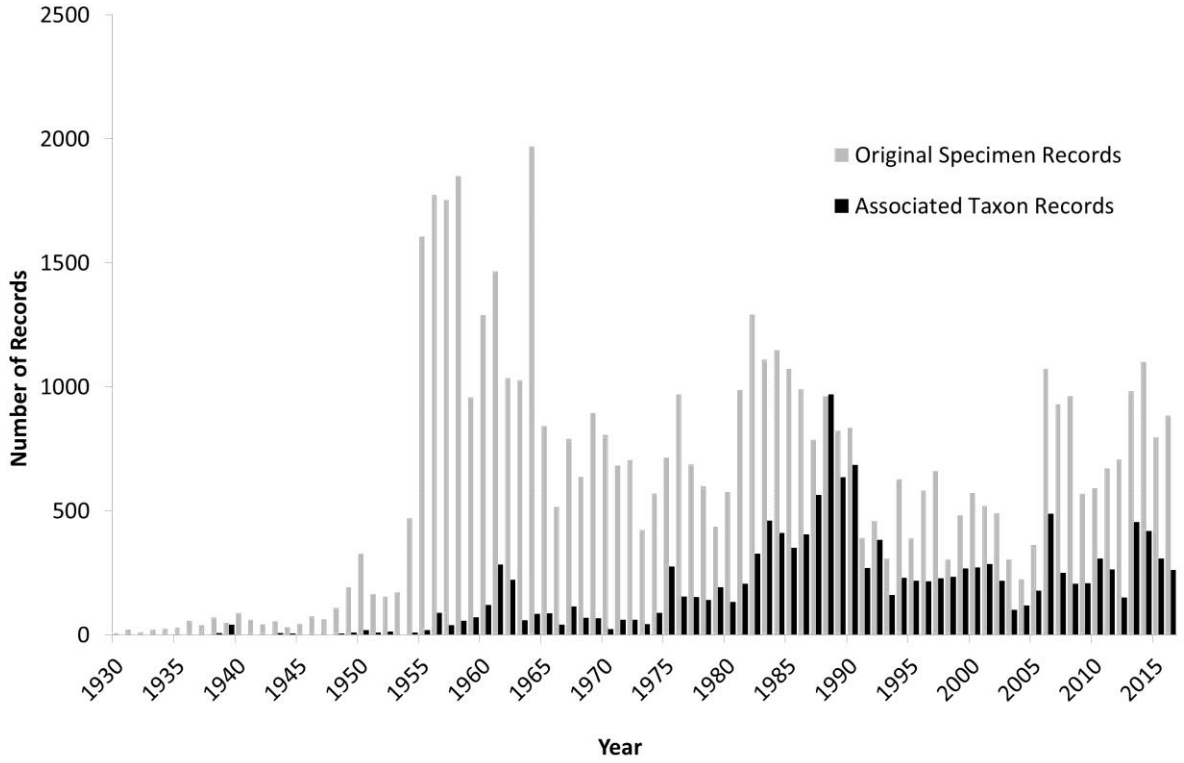

Figure 4.4 Histogram of original specimen records (gray) and associated taxon records (black) excluding duplicate records

Conversely, spatial density of associated taxon records did correspond with specimen collecting locations (Figure 4.5). The areas of highest record frequency for both associated taxon records and specimens were in Florida counties near the FSU herbarium: Leon, Franklin, Liberty, Wakulla, Gadsden, and Jackson. However, unlike the specimen dataset, the associated taxon dataset had an abundance of records from Escambia County that exceeded even those of Leon County, the location of FSU.

The spatial partitioning of associated taxon records can largely be explained by the data collection habits of the collectors in these regions. For example, although his specimens compose less than 1% of specimen records in the FSU dataset, James R. Burkhalter of Escambia County, Florida was responsible for over 4% of the resulting associated taxon records, recording an average of 1 associated taxon per specimen. In contrast, 20% of the specimens in the original dataset were collected by Robert K. Godfrey, a prolific historical collector in the central panhandle of Florida (e.g., Leon, Franklin, Liberty counties) and the namesake of the FSU herbarium, but fewer than 6% of the associated taxon records were from his specimens (0.05 associated taxa per specimen). Another influential collector, Loran C. Anderson, recorded an average of 0.4 associated taxa per specimen, with collections throughout Florida but primarily near the FSU herbarium.

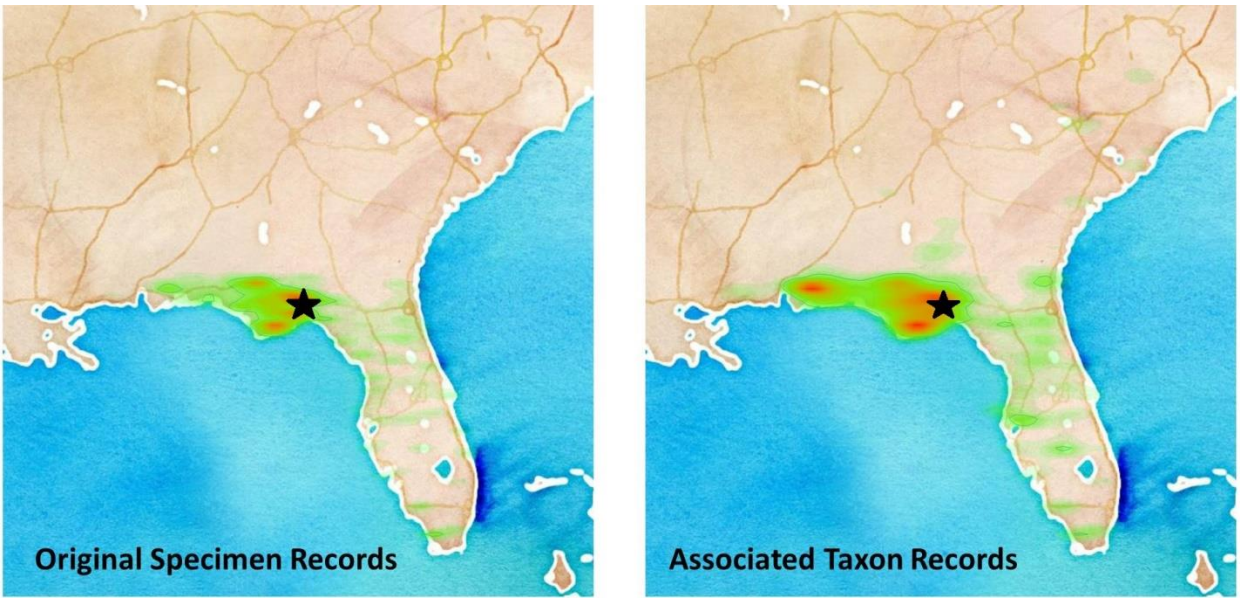

Figure 4.5 Heatmap comparison of record densities for the original dataset (left) and the associated taxon dataset (right). Colors indicate the density of records relative to each respective dataset independent of the other: red (darker) indicates higher record density and green (lighter) indicates lower record density. Black stars show the location of the FSU herbarium. Although the original dataset included all digitized specimen records from the FSU herbarium, which span the globe, only the state of Florida is shown in this figure since record density was highest in this region. The heatmap overlays were produced using identical settings for both datasets in the R packages ggplot2 and ggmap, and the background map is courtesy of the Google API accessed using the same R packages.

The associated taxon dataset contained 25 records of 7 federally threatened species, 223 records of 52 state threatened species, 41 records of 14 federally endangered species, and 326 records of 108 state endangered species.

Range extensions

The cleaned associated taxon dataset contained 247 new county records for 217 Florida plant species when compared to the Atlas of Florida Plants (Wunderlin et al. 2007). When compared both to the

USDA PLANTS database and specimen records in the iDigBio portal, the associated taxon dataset produced 2,371 and 1,193 new county records, respectively. An estimated 66% of USDA PLANTS new county records and 75% of iDigBio new county records could be confirmed as apparent range extensions rather than, for example, taxonomic inconsistencies. By these estimates, the newly generated observational dataset may provide as many as 894 to 1,564 "true" potential new county records for these databases from the original 72,120 specimen dataset.

Discussion

Increasing our understanding of species distributions is crucial to many scientific aims, including assessing the impact of anthropogenic effects such as climate and land use changes. This analysis of FSU herbarium data demonstrates that accessing the relatively untapped resource of associated taxa noted on biodiversity specimen labels can significantly augment current distribution data. Extracting associated taxon data from 72,120 records resulted in 247 new county records for the state of Florida when compared to the Atlas of Florida Plants, 2,371 (estimated 1564 true records) for the U.S. when compared to the USDA PLANTS database, and 1,193 (estimated 894 true records) new county records for the U.S. compared to digitized herbarium records hosted on iDigBio. Furthermore, these records spanned multiple decades (1937 – 2016), providing an irreplaceable historical record of species' past distributions, potentially in locations where the species can no longer be found. These data can be invaluable to, for example, conservation managers in determining pre-disturbance conditions or researchers seeking to understand spatiotemporal biodiversity change.

The results of this study further suggest that associated taxon records can augment data for a wide variety of taxa. Over 2,900 taxa from over 200 plant families were represented in the final dataset. Trees and shrubs were overrepresented by 124% and 14%, respectively, relative to specimens with associated taxon data, which may indicate a tendency of collectors to record dominant and canopy species. Indeed, the grass (Poaceae), sedge (Cyperaceae), oak (Fagaceae), pine (Pinaceae), magnolia (Magnoliaceae), and palm (Arecaceae) families were among the top 10 families in the associated taxon dataset, even though pines, magnolias, and palms were not even in the top 50 families in the specimen dataset. Data on these often dominant (in the southeast United States), habitat-shaping taxa can improve our knowledge of the distribution of ecosystems over space and time, especially in highly heterogeneous, disturbance-reliant regions such as the North American Coastal Plain. Still, common species may be systematically under-represented in herbarium collections in comparison with their natural abundances (Garcillan et al. 2008), and associated taxon records may help fill in the gaps left by this and other collecting biases.

Imperiled species may also be under-collected due to their protected status (Daru et al. 2017), and their distributions may be poorly understood because they are rare. The associated taxon dataset contained 449 records of 161 state or federally threatened or endangered species and may therefore provide much-

needed insight into the distributions of data-depauperate taxa of high conservation interest. Moreover, associated taxon records may provide a broader spatial and temporal range of data for these taxa, which is critical for species facing immediate anthropogenic threats.

On a more basic level, associated taxon records gleaned from biodiversity specimen records increase the quantity of data at hand, which is becoming increasingly important in an era of large-scale analytical methods. For instance, Environmental Niche Models have proven most effective with a high number of training points (i.e., large amount of starting data; Loiselle et al. 2008). Leveraging associated taxon records from digitized specimens from the FSU herbarium increased the size of the usable dataset by 18.5% over a significant temporal and spatial distribution, demonstrating that this method can substantially boost species occurrence data across time and space.

Limitations

Associated taxon records may offer a new frontier for gaining valuable biodiversity data; however, like all datasets, they are subject to certain coverage, quality, and usage limitations. First, the spatiotemporal range of retrievable data from associated taxon records is limited by the coverage of specimen records. While these data may fill gaps in individual species distributions, they will not be able to address systematic temporal and spatial collecting biases such as lower data collection during World Wars (Delisle et al. 2003) and may instead introduce new biases such as increased occurrences in regions or time periods wherein collectors have been trained to record associated taxa (see Figure 4.4). For this reason, associated taxon data are best combined with additional data sources to reduce spurious trends.

Second, associated taxa may be misidentified, and because associated taxon records are purely observational, they lack the verifiability of specimen records. For this reason, associated taxon records should be treated with the same level of caution that is applied to observations. Associated taxon identifications are expected to be reasonably accurate since collectors are often taxonomic experts and are likely to document associated taxa that they have confidently identified in the field; however, further investigation into the reliability of associated taxon records and methods to overcome this limitation is needed. Misidentifications are not a new problem for users of specimen data (see Goodwin et al. 2015) and can be handled through outlier identification and other data quality control methods, or, in some cases, on-site verification. In the case of potential new county records, for example, collectors should re-examine the original collection sites for the purported taxa.

Third, the methods developed in this study assume that the appropriate genus of abbreviated associated taxon names (e.g., *E. elatus*) could be found in the original specimen record or in the habitat field from which the associated taxon was gathered. This assumption appeared reasonable for 89% of records, and the remaining 11% could be corrected by hand using regional taxonomic knowledge. If

employed on a large scale or without careful curation of the output, this method may be inefficient or cause data quality issues similar to those of misidentifications.

Future directions

This study explores the potential for associated taxon records from specimen data to broaden our understanding of species distributions. The methods developed to tap this potential could be improved for efficiency, thoroughness, and universality. Because the web-based Global Names Recognition and Discover API (GNRD) was used to identify associated taxon records, each specimen record took slightly more than 4 seconds to parse, which could add up to a substantial amount of time for large datasets. Furthermore, the GNRD is not designed to identify common names from the given text, which limited the output of the code and may have caused underrepresentation of particularly common species (e.g., oak, wiregrass, longleaf pine). With improvement on these and other fronts, as well as development of further data cleaning processes, similar methods could unlock massive amounts of associated taxon data with even greater ease.

The focus of this study was herbarium specimen label data, but other types of collections may offer similarly rich—or even greater—opportunities. For example, it is common practice when collecting insects (Martin 1977) and fungi (Leonard 2010) to record the host plant or animal of the collected individual. Similarly, collectors of vertebrate specimens may record ecto- or endo-parasites or gut contents (RIC 1997; ISLES 2011). Thus, delving into the data of many types of biodiversity specimens may reveal additional, previously "hidden" occurrence data, even for taxonomically distant groups (e.g., insects and plants) and potentially for groups that are under-collected or difficult to preserve such as parasites.

Finally, examining trends in nearly a century of documenting associated taxa at time of collection can aid the development of better data creation practices. Results from this study suggest that collectors of plants most often record dominant and canopy taxa. These data are indeed useful for determining local habitat types and the distributions of characteristic species, yet our understanding of species distributions could be broadened that much more if collectors included non-dominant taxa as well. Collecting specimens is a time- and labor-intensive activity that may become rarer in periods of decreased funding for basic biodiversity research, making the collection of rich data at each event increasingly important. Recording even one or two associated taxa when making a collection could be a simple and efficient way to double or triple the return of every investment in field work and avoid over-crowding in collections spaces.

The recent push for digitization of biodiversity specimens is making a vast amount of specimen data publically accessible, and we have the increasing opportunity to leverage these resources to produce new types of data. Extracting associated taxon data from existing specimen records may improve our

knowledge of species and community distributions, as well as enable collectors and other biodiversity researchers to better identify data gaps, prioritize future collecting events, and optimize methods of data collection. Broadening our knowledge of species distributions and improving data- and specimen-collection practices may be as simple as examining the data we already have.

CHAPTER 5

CONCLUSIONS

Herbarium collections comprise a wealth of high-quality, spatiotemporally diverse data useful for botanical, ecological, evolutionary, and many other applications. In this thesis, I have presented the potential for the collecting process and its resulting data to inform our understanding of plant phenology, species distributions, and biotic change in the past, present, and future. In chapter 2, using thousands of "snapshots" of plant phenological states under differing climatic conditions and a new method of assessing phenology of herbarium specimens, I showed that spring-flowering asteraceous species in the U.S. Southeastern Coastal Plain flower earlier with warmer spring temperatures, while fall-flowering species delay flowering with warmer summer temperatures. These phenological shifts have potentially significant eco-evolutionary consequences, such as altering plant-pollinator and plant-herbivore interactions (Burkle et al. 2014; Liu et al. 2011), that may have implications for conservation in this biodiversity hotspot. Notably, the diverging responses of spring and fall phenology may widen a summer "dead zone" in which flowers are scarce, which could detrimentally affect species that rely on floral resources or the phenological cue their presence provides. The results of this chapter also indicated that the impact of plant traits, especially pollination mode, on phenological sensitivity to climate requires further investigation. Thus, herbarium specimens can provide insight into changes of the past that lead into the present and ultimately direct us toward studies of the future.

In chapter 3, I investigated the collecting practices of specimen collectors, focusing on how they detect and document outliers, which may be indicators of biotic change. I demonstrated how current practices may not be optimized for outlier discoverability, particularly after digitization. Collectors of biodiversity specimens indicated that they note outliers, but often do so using a myriad of words and phrases. My query of over 75 million specimen records suggested that outlier records are then stored in many database fields across the community, and the lack of standardization and designated outlier "flag" terms prevents future data discovery that would be critical for further analyses. I posited that community discussion of how and where collectors should document outliers are two important next steps in increasing the accessibility of these data. Pursuing this goal will enable us to discover change as it occurs, either prompting or obviating post hoc analyses like that in chapter 2, which may further empower us to stem adverse change before it is too late.

Finally, in chapter 4, I showed how new data and applications of specimen data need not be restricted to future collections, but rather, our current holdings provide untapped resources that can still change our understanding of life on Earth. Using only the records from a single collection, the FSU Robert K. Godfrey Herbarium, I was able to significantly augment current distribution data with

thousands of occurrence records, hundreds of which were potential extensions of previously documented ranges. With careful data curation, cleaning, and analyses that take into account the limitations of collections data, the research potential of specimen data is great.

The digitization of millions of specimen records worldwide throws open doors for rich analyses of large datasets as demonstrated in chapter 2, not only for challenging topics such as phenology, but also for processes less frequently studied using specimens such as patterns of herbivory, competition, and evolutionary change over time and space. Collecting, mining, and aggregating such data constitute major challenges of the future for specimen-based research, and they may be overcome by the implementation of standardized best practices among collectors and data aggregators. This is my hope for the research in chapters 3 and 4: that presenting the enormous potential of herbarium specimens to inform our understanding of Earth's biota will spur action toward thoughtful discussion and eventual application of unified, optimized strategies of data collection and aggregation.

APPENDIX A

SPECIES INCLUDED IN CHAPTER 2 DATASET

Table A.1 Species included in chapter 2 dataset, their tribal rank, number of specimens in the final dataset, and traits documented for each species. In the habitat preference column, positive preference is indicated by "1" and lack of preference is indicated by "0". Flowering guilds are listed as P = spring, S = summer, F = fall, W = winter, or some combination of these. Pollination modes are categorized as animal-pollinated (A) or wind-pollinated (W).

Species	Specimens in Dataset	Moist Habitat Preference	Dry Habitat Preference	Flowering Guild	Pollination Mode
Balduina angustifolia	243	0	1	PSF	A
Balduina atropurpurea	46	1	0	F	A
Balduina uniflora	156	0	0	F	A
Carphephorus corymbosus	154	0	1	F	A
Carphephorus odoratissimus	223	1	0	F	A
Carphephorus paniculatus	167	1	0	FW	A
Coreopsis gladiata	443	1	0	F	A
Coreopsis grandiflora	30	0	0	S	A
Coreopsis lanceolata	258	0	0	PSF	A
Coreopsis leavenworthii	275	1	0	S	A
Coreopsis major	105	0	0	S	A
Eupatorium album	195	0	1	S	A
Eupatorium capillifolium	214	1	0	F	W
Eupatorium compositifolium	159	1	0	F	W
Eupatorium fistulosum	95	1	0	SF	A
Eupatorium hyssopifolium	176	0	1	F	A
Eupatorium leptophyllum	140	0	0	F	A
Eupatorium mohrii	287	1	0	S	A
Eupatorium semiserratum	188	1	0	F	A

Table A.1 - continued

Species	Specimens in Dataset	Moist Habitat Preference	Dry Habitat Preference	Flowering Guild	Pollination Mode
Eupatorium serotinum	202	0	0	F	A
Gamochaeta antillana	107	1	0	PSF	A
Gamochaeta argyrinea	42	0	0	P	A
Gamochaeta chionesthes	20	0	0	P	A
Gamochaeta coarctata	65	1	0	P	A
Gamochaeta pensylvanica	200	1	0	P	A
Gamochaeta purpurea	175	0	1	P	A
Gamochaeta stagnalis	32	1	0	P	A
Helianthus agrestis	30	1	0	F	A
Helianthus angustifolius	319	1	0	F	A
Helianthus annuus	49	0	0	SF	A
Helianthus atrorubens	37	0	0	F	A
Helianthus carnosus	10	1	0	P	A
Helianthus debilis	197	0	0	Y	A
Helianthus divaricatus	40	0	1	S	A
Helianthus floridanus	70	0	0	F	A
Helianthus heterophyllus	102	1	0	F	A
Helianthus hirsutus	67	0	1	F	A
Helianthus microcephalus	45	0	0	F	A
Helianthus radula	96	1	0	F	A
Helianthus resinosus	31	0	1	F	A
Helianthus simulans	41	1	0	F	A
Helianthus strumosus	48	0	0	F	A
Helianthus tuberosus	33	0	0	F	A
Iva annua	44	1	0	F	W

Table A.1 - continued

Species	Specimens in Dataset	Moist Habitat Preference	Dry Habitat Preference	Flowering Guild	Pollination Mode
Iva frutescens	64	1	0	F	W
Iva imbricata	47	0	0	F	W
Iva microcephala	96	1	0	F	W
Liatris aspera	45	0	0	F	A
Liatris chapmanii	114	0	0	F	A
Liatris elegans	150	0	1	F	A
Liatris garberi	37	1	0	S	A
Liatris gracilis	291	1	0	F	A
Liatris laevigata	132	0	0	F	A
Liatris pauciflora	85	1	0	F	A
Liatris pilosa	69	1	0	F	A
Liatris provincialis	31	0	0	F	A
Liatris pycnostachya	94	1	0	F	A
Liatris spicata	233	1	0	S	A
Liatris squarrosa	196	0	0	S	A
Liatris squarrulosa	76	0	0	F	A
Liatris tenuifolia	196	0	0	F	A
Marshallia caespitosa	20	0	0	S	A
Marshallia graminifolia	193	1	0	F	A
Marshallia obovata	39	0	1	S	A
Solidago altissima	120	0	0	F	A
Solidago arguta	124	0	0	F	A
Solidago brachyphylla	33	0	0	F	A
Solidago caesia	120	0	0	F	A
Solidago fistulosa	186	1	0	F	A
Solidago leavenworthii	83	1	0	W	A
Solidago patula	44	1	0	F	A
Solidago petiolaris	49	0	0	F	A
Solidago puberula	46	0	0	F	A

Table A.1 - continued

Species	Specimens in Dataset	Moist Habitat Preference	Dry Habitat Preference	Flowering Guild	Pollination Mode
Solidago rugosa	154	1	0	F	A
Solidago sempervirens	186	0	0	F	A
Solidago stricta	327	1	0	F	A
Solidago tortifolia	82	0	1	F	A
Solidago ulmifolia	43	0	1	F	A
Symphyotrichum concolor	156	0	0	F	A
Symphyotrichum dumosum	338	1	0	F	A
Symphyotrichum elliottii	76	1	0	F	A
Symphyotrichum lateriflorum	148	0	0	F	A
Symphyotrichum pilosum	169	0	0	F	A
Symphyotrichum praealtum	50	1	0	F	A
Symphyotrichum simmondsii	87	0	0	W	A
Symphyotrichum tenuifolium	63	1	0	F	A
Symphyotrichum undulatum	51	0	1	F	A

APPENDIX B

MODELS TO DETERMINE SPECIES' RATE OF TRANSITION BETWEEN PHENOPHASES

Table B.1 Outputs of linear mixed effects models used to determine species' rates of transition between phenophases. The slopes of these models were used to calculate the estimated day of peak flowering of herbarium specimens in each corresponding genus. Confidence intervals (CI) reported are 95% confidence intervals calculated using the *confint* function in R. Reported p-values are results of likelihood ratio tests performed via the *anova* function in R.

Species	Number of Individuals	Intercept (DOY)	Slope (days/phenophase)	95% CI	p-value
Balduina uniflora	20	223	2.8	2.4, 3.1	<0.001
Carphephorus paniculatus	17	278	1.6	1.4, 1.8	<0.001
Coreopsis gladiata	12	277	1.7	1.5, 1.9	<0.001
Eupatorium mohrii	25	246	4.4	4.1, 4.7	<0.001
Gamochaeta coarctata	11	135	3.4	3.0, 3.8	<0.001
Helianthus angustifolius	16	271	1.9	1.7, 2.1	<0.001
Iva microcephala	14	262	2.5	2.1, 3.0	<0.001
Liatris spicata	17	244	2.1	1.8, 2.3	<0.001
Marshallia graminifolia	14	197	4.1	3.8, 4.4	<0.001
Solidago fistulosum	15	267	1.7	1.5, 1.8	<0.001
Symphyotrichum dumosum	18	284	3.0	2.6, 3.5	<0.001

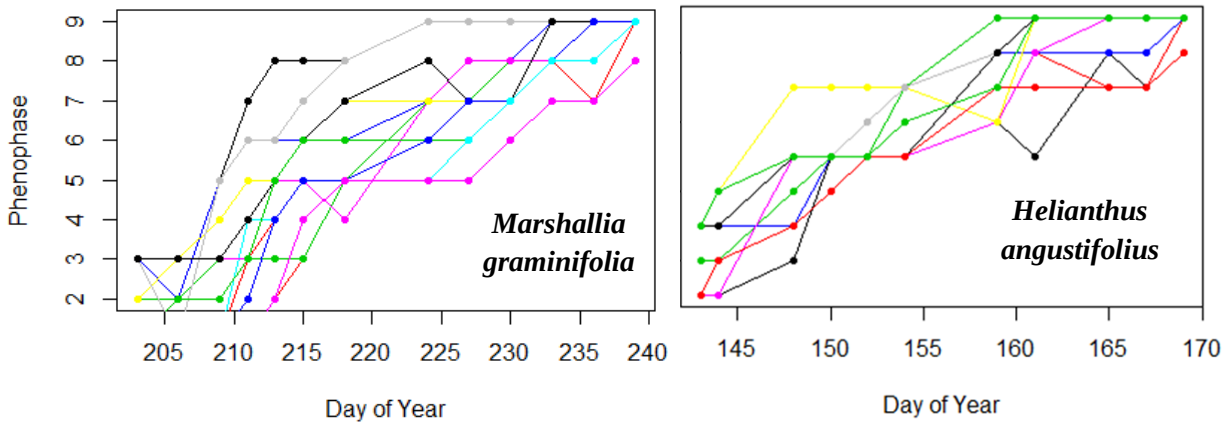

Figure B.1 Example graphs showing the transitions of individual plants of species between phenophases over time. Different line and point colors indicate different individuals on each graph. Note that although they are shown here, observations of individuals during phenophase 1 (100% buds) or phenophase 9 (100% fruits) were omitted when fitting LME models.

APPENDIX C

COMPARISON OF LMES WITH VARYING INTERCEPTS VS. VARYING SLOPES AND INTERCEPTS

Table C.1 Comparison of linear mixed effects (LME) models allowing intercepts to vary with species and those allowing both slopes and intercepts to vary with species. AICc values were calculated using the *AICc* function of the *MuMIn* package in R. These relatively small differences in model fit indicate that LMEs with variable slopes and intercepts do not fit the data better than LMEs with only variable slopes. LMEs with the fixed effect of precipitation failed to converge and are therefore not shown.

Response	Fixed effect	Random Effects	ΔAICc
Peak flowering time, fall-flowering species	July temperature deviation	Variable intercept only (1|Species)	0
		Variable intercept and slope (JulTempDev | Species)	0.91
Peak flowering time, fall-flowering species	March temperature deviation	Variable intercept only (1|Species)	0
		Variable intercept and slope (MarTempDev | Species)	2.91
Peak flowering time, spring-flowering species	March temperature deviation	Variable intercept only (1|Species)	0
		Variable intercept and slope (MarTempDev | Species)	2.66

APPENDIX D

ASSESSMENT OF VARIANCE IN PHENOLOGICAL MODELS

To determine whether variance changed predictably with climate, linear mixed effects (LME) models with differing variance functions were fit to the data using the *lme* function of the *nlme* package and compared by calculating AICc values calculated via the *AICc* function of the *MuMIn* package in R. As in all reported models, the response variable was estimated peak flowering date, the fixed effect was a climate variable, and the species name was added as a random effect. Comparing models with and without variance functions indicates no significant improvement of model fit with added variance components (Table D.1).

Table D.1 Results of phenological LME models with differing variance functions

Response	Fixed effect	Variance Function	ΔAICc
Peak flowering time, fall-flowering species	July temperature deviation	None	0
		Linear	7041.27
		Constant power	2.16
Peak flowering time, fall-flowering species	March temperature deviation	None	0
		Linear	7604.29
		Constant power	0.91
Peak flowering time, spring-flowering species	March temperature deviation	None	0
		Linear	398.06
		Constant power	1.98

APPENDIX E

HUMAN SUBJECTS RESEARCH CONSENT LETTER, APPLICATION, AND APPROVAL FORMS

The purpose of this survey is to determine current practices followed by those who collect and/or prepare biological specimens with regard to recording "outliers." For our usage, outliers are defined as individual specimens that differ from a previously documented or perceived general norm within a taxon in any biological characteristic such as morphology, anatomy, distribution, behavior, phenology, or ecology. Other terms that may be used to describe our use of "outliers" include "anomalies" or "oddities." The discovery and description of a new species is a type of outlier detection and documentation that is well developed across the community, and this is not the focus of our study. We ask that you instead focus your answers on outliers within a previously described taxon. Examples of this include, but are not limited to, a new population of an introduced species or earlier initiation of a reproductive stage than has been previously observed.

The information you share will be used to inform the work of the iDigBio Outlier Detection and Documentation by Collectors (ODD Collectors) Working Group. The survey will take about 10 - 15 minutes. Please respond by March 22, 2016.

By completing this survey, you are giving consent for the researchers to use your responses. Your participation is voluntary. There are no direct benefits or risks to you for participating and no compensation. You may quit at any time or skip any item. You may withdraw your consent to participate at any time without penalty. If you respond electronically, your IP address will be registered; however, your responses will remain anonymous. There is a minimal risk that security of any online data may be breached, but our survey host (Qualtrics) uses strong encryption and other data security methods to protect your information. Only members of the ODD Collectors Working Group will have access to your information on the Qualtrics server. Thank you for your help.

If you have any questions about this survey, contact Austin Mast (REDACTED) or Katelin Stanley [Pearson] (REDACTED). If you want more information about your rights as a research participant, contact the FSU Human Subjects office (850-644-7900, jth5898@fsu.edu). By clicking the box below you acknowledge that you have read the information and agree to participate in this survey. If you do not wish to participate, please close your browser at this time.

☐ I agree

Human Subjects Application For Full IRB and Expedited Exempt Review

Print

1. Project Title and Identification

1.1 Project Title

Outlier Detection and Documentation by Collectors of Biodiversity Specimens

Project is: Research

1.2 Principal Investigator (PI)

Name(Last name, First name MI): Mast, Austin	Highest Earned Degree: Doctorate
Mailing Address: 4295	Phone Number: [redacted]
	Fax:
University Department: BIOLOGICAL SCIENCE	Email: [redacted]
The training and education completed in the protection of human subjects or human subjects records: FSU Training Module	Occupational Position: Faculty

1.3 Co-Investigators/Research Staff

Name(Last name, First name MI): Stanley, Katelin ; Co-Investigator	Highest Earned Degree: Bachelor's Degree
Mailing Address: [redacted]	Phone Number: [redacted]
	Fax:
University Department: BIOLOGICAL SCIENCE	Email: [redacted]
The training and education completed in the protection of human subjects or human subjects records: NIH	Occupational Position: Student

1.4 Faculty Advisor/Department Chair/Dean Information

Name(Last name, First name MI): Levitan, Don ; Chair	Highest Earned Degree:
Mailing Address: 4295	Phone Number: [redacted]
	Fax:
University Department: BIOLOGICAL SCIENCE	Email: [redacted]
The training and education completed in the protection of human subjects or human subjects records:	Occupational Position:

2. Funding

2.1 Is this research funded by an internal (FSU) or external agency?

Yes

This project has been submitted to the following funding agency:

Name of Sponsor:	National Science Foundation
OMNI Number:	029490
SRS/Research Fuondation Contact Person:	Emily Nerona
The funding decision:	Awarded
Type of funding source:	Federal grant

3. Institutional Oversight

3.1 Is this research proposal being reviewed by any other institution or peer review committee?

No

4. Conflict of Interest

Federal guidelines encourage Institutions to assure there are no conflicts of interest in research projects that could adversely affect the rights and welfare of human subjects. If this proposed research study involves a potential conflict of interest, additional information will need to be provided to the IRB. Examples of potential conflicts of interest may include: any sort of compensation, in cash or other form, for services to an individual and his or her immediate family, the value of which exceeds $10,000 in a one-year period or an equity interest which exceeds $10,000 or which exceeds a five percent ownership interest.

4.1 Do any of the Investigators or personnel listed on this research have a potential conflict of interest associated with this study?

No

5. Payment or Other Compensation for Research Subjects

5.1 Will you give subjects gifts, payments, compensation, reimbursement, services without charge or extra credit/class credit?

No

6. Protocol Description and Other Detail

6.1 Describe the objective(s) of the proposed research including purpose, research question, hypothesis, method, data analysis, research design and relevant background information etc.

Background/Purpose/Questions
In this era of global change, it is increasingly important for collectors of biological specimens (e.g., plants, insects, mammals) to detect and report change at the collecting event, rather than delaying until specimen data is digitized or later used for research. Biological outliers (e.g., in phenology, distribution, morphology/anatomy, etc.) may indicate the beginnings of significant, transformative change that merits immediate attention. Specimen collectors are on the front lines of observing that change and need the encouragement, training, cyberinfrastructure, and semantics to report change more effectively to stakeholders who can use the information. The overall purpose of our study is to bring attention to this issue within a critical era of burgeoning biodiversity informatics. To do so, we will first investigate the extent of the problem by gauging opinions and determining current practices followed by those who collect and/or prepare biological specimens with regard to noting or recording outliers. We are also interested in the following questions: (1) How do collectors determine whether something they have collected is an outlier, (2) What are the impediments to detecting and documenting outliers?, and (3) What words would collectors use to note outliers when documenting them?
Hypotheses
From our experience in the collecting community, we believe that collectors may not view themselves as being on the front lines of observing and noting change. We also hypothesize that collectors have not traditionally noted outliers in practical or accessible ways, and are generally impeded from doing so by a lack of standards and protocols. We have assembled a list of potential vocabulary that may be used to describe outliers, which we wish to expand with knowledge from experts in the field.
Method
To gain a deeper understanding of the community of biological collectors, we will distribute a carefully crafted, anonymous survey (see attached documents) via listservs and emails of professional societies.
Data Analysis
Data analysis will include descriptive statistics (e.g., means, percentages) and testing for

differences between groups (e.g., early vs. later career professionals). Responses to essay and other open-ended questions will be coded based on emergent themes and used to guide our description of current opinion on outlier detection and documentation.
Literature or resources suggested by the participants will be examined for instructions or protocols regarding outlier or character documentation.
Words and phrases provided by participants in question 9 of the survey will be used to guide searches of biological databases for specimens that have been catalogued using such terminology. We will also assemble a "vocabulary of outliers" that will include these words and phrases and be used to describe how collectors have traditionally treated outliers.

6.2 Following categories will apply for the evaluation of the project:

- Questionnaires or Surveys to be administered
- Exclusion of Women or Children Subjects (must explain why they are being excluded)
- Subjects studied at FSU
- Subjects studied at non_FSU location(s)

6.3 Survey Techniques: the only involvement of human subjects will be in the following categories:

- Research involving survey or interview procedures

Research involving survey or interview procedures:

1. *Responses will be recorded in such a manner that human subjects cannot be identified, by persons other than the researcher, either directly or through identifiers linked to the subjects.* Yes

2. *Would subjects' responses, if they became known outside the research, reasonable place the subject at risk of criminal or civil liability or be damaging to the subjects' financial standing or employability.* No

3. *The research deals with sensitive aspects of the subject's own behavior, such as illegal conduct, drug use, sexual behavior, or use of alcohol.* No

6.4 This study will include following methods:

- Descriptive

6.5 Describe the tasks subjects will be asked to perform.
Upload surveys, instruments, interview questions, focus group questions etc. Describe the frequency and duration of procedures, psychological tests, educational tests, and experiments; including screening, intervention, follow-up etc. (If you intend to pilot a process before recruiting for the main study please explain.)

The survey will be distributed via email and can be filled out at any time or location. The survey consists of fifteen questions, not including a consent form and box for additional comments, and should take no longer than 10-15 minutes depending on the thoroughness of replies. Participants are asked to fill out the survey only once.

6.6 How many months do you anticipate this research study will last from the time final approval is granted?

24

7. Participant (Subject) Population

7.1 Expected number of participants

Number of male: 200 Number of female: 200
Expected number of participants: 400

7.2 Expected Age Range

- 18-65
- 65 and older

7.3 Inclusion/Exclusion of Children in this Research

Exclusion

If this study would ***exclude*** children, NIH guidelines advise that the exclusion be justified, so that potential for benefit is not unduly denied. Indicate whether there is potential for direct benefit to subjects in this study and if so, provide justification for excluding children. Note that if inclusion of children is justified, but children are not

seen in the PI's practice, the sponsor must address plans to include children in the future or at other institutions.

- No direct benefit established (exclusion of children permissible)

Provide justification for exclusion of children:

Children will be excluded from this study because the purpose of the study is to determine opinions and practices of biological collectors, who are exclusively adults. Research products will not directly affect children.

7.4 Other Protected Populations to be Included in this Research

- Elderly Subjects -- 65+

7.5 Inclusion and Exclusion of Subjects in this Research Study
Describe criteria for inclusion and exclusion of subjects in this study

Inclusion Criteria:

Survey participants are expected to be collectors or preparators of biological specimens of any taxonomic group.

Exclusion Criteria:

Survey participants are expected to be collectors or preparators of biological specimens of any taxonomic group.

7.6 Location of subjects during research activity or location of records to be accessed for research

- Florida State University
- Other: Subject's institution
- University Campus (non-clinical): Subject's office
- Subject's Home: If desired
- Other Special Institutions: Subjects can take the survey at any location

7.7 Describe the rationale for using each location checked above
Upload copies of IRB approvals or letters of cooperation from other agencies or sites, if it has been granted or the application submitted if approval has not been granted.

For the convenience of the participants, the survey may be taken at any location with access to a computer with Internet.

8. Recruitment of Participants (Subjects)

8.1 Describe the recruitment process to be used for each group of subjects
Upload a copy of any and all recruitment materials to be used e.g. advertisements, bulletin board notices, e-mails, letters, phone scripts, or URLs.

Participants will be recruited via advertising through listservs and membership lists from professional societies. The survey link will not be specific to the recipient and will thus be anonymous. A transcript of the recruiting email is shown below:

Calling all specimen collectors, curators, and preparators!
Calling all biodiversity specimen collectors and preparators!

iDigBio's Outlier Detection and Documentation by Collectors (ODD Collectors) Working Group would be grateful for 10–15 minutes of your time to complete an anonymous survey related to your experience with "outliers" (or "anomalies" or "oddities") in biodiversity collections. For our usage, outliers are defined as individual specimens that differ from a previously documented or perceived general norm within a taxon in any biological characteristic such as morphology, anatomy, distribution, behavior, phenology, or ecology. To participate, follow the link below. We hope to see your response by March 15, 2016.

LINK PROVIDED HERE

Please direct questions to Katelin Stanley ([redacted] or Austin Mast [redacted]. Thanks for considering this request!

With best regards,

Katelin Stanley, Austin Mast, and the rest of iDigBio's ODD Collectors Working Group

8.2 Explain who will approach potential subjects to take part in the research study and what will be done to protect individuals' privacy if required in this process

Participants will be recruited via advertising through listservs and membership lists from professional societies. The survey link will not be specific to the recipient and will thus be anonymous.

8.3 Are subjects chosen from records?

No

8.4 FSU policy prohibits researchers from accepting gifts for research activities. Is the study sponsor offering any incentive connected with subject enrollment or completion of the research study (i.e. finders fees, recruitment bonus, etc.) that would be paid directly to the research staff?

No

8.5 Is the study going to be posted on the Research Studies at Florida State University recruiting website?

No

9. Risks and Benefits

9.1 The research may involve following possible risks or harms to subjects:

9.2 Does the Research Involve Greater Than Minimal Risk to Human Subjects?

"Minimal Risk" means that the risks of harm anticipated in the proposed research are not greater, considering probability and magnitude, than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.

No

9.3 Explain what steps will be taken to minimize risks or harms and to protect subjects' welfare. If the research will include protected populations (see question 7.4) please identify each group and answer this question for each group.

Senior citizens: Participants may withdraw from the survey at any time and may skip any or all survey questions. Participation is voluntary.

9.4 Describe the anticipate benefits of this research for individual subjects in each subject group. If none, state "None".

Biological specimen collectors may benefit from this research if the results raise awareness of the importance of collectors to scientific research. Outcomes of further research efforts by the Outlier Detection and Documentation Working Group will also focus on improving the resources available to collectors for outlier detection and documentation.

9.5 Describe the anticipated benefits of this research for society, and explain how the benefits outweigh the risks.

With an increased awareness of the importance of detecting and documenting outliers, as well as eventual products to encourage this process, we hope to better track changes that may impact environmental health and thus human society. Discovering trends such as invasive species invasion or major shifts in plant reproduction, for example, may allow us to stem harmful effects to nature and to agriculture.

10. Confidentiality of Data

10.1 Will you record any direct identifiers, names, social security numbers, addresses, telephone numbers, email addresses, cookies etc.?

No

10.2 Will you retain a link between study code numbers and direct identifiers after the data collection is

complete?

No

10.3 Will you provide the link or identifier to anyone outside the research team?

No

10.4 Where, how long, and in what format (such as paper, digital or electronic media, video, audio, or photographic) will data be kept? In addition, describe what security provisions will be taken to protect this data (password protection, encryption, etc.)

Data will be kept on the investigators computers in the form of text and spreadsheet files for as long as possible. This data does not require protection because it is not associated with any personal identification information.

10.5 Will you place a copy of the consent form or other research study information in the subjects' record such as medical, personal or educational record?

No

10.6 *If the data collected contains information about illegal behavior, please refer to the NIH Certificates of Confidentiality Kiosk for information about obtaining a Federal Certificate of Confidentiality.*

10.7 Will you be given or have access to personal information regarding employee, customer, student, parent and/or patient accounts with Florida State University?

No

11. Use of Protected Health Information (PHI): HIPAA Requirements

In the course of conducting research, researchers may desire to obtain, create, use, and/or disclose individually identifiable health information. Under the HIPAA Privacy Rule, covered entities (healthcare providers, health plans, employer or healthcare clearinghouses) are permitted to use and disclose protected health information for research with individual authorization, or without individual authorization under limited circumstances set forth in the Privacy Rule.

11.1 As part of this study, will you be accessing PHI from a covered entity for research purposes?

No

12. Informed Consent Process

12.1 Recognizing that consent itself is a process of communication, please expand on your responses to questions 8.1 and 8.2 and describe what will be said to the subjects to introduce the research.

The email that was provided in question 8.1 will be sent via listservs and to membership lists of professional societies. We will address questions as they arise in an ad hoc manner.

12.2 In relation to the actual data gathering, when will consent be discussed and documentation obtained? (e.g., mailing out materials, delivery of consent form, meetings)

Participants must give consent before beginning the survey questions.

12.3 Please name the specific individuals who will obtain informed consent and include their job title/credentials and a brief description of your plans to train these individuals to obtain informed consent and answer subject's questions:

Participants must give consent before beginning the survey questions.

12.4 What questions will you ask to assess the subjects' understanding of the risks and benefits of participation?

Participants will be asked whether they understand and agree to the terms of the survey (see the survey for full explanation/consent form).

12.5 Informed Consent Waivers

☐ Request waiver of documentation of consent.

☐ The only record linking the subject and the research would be the consent form and the principle risk of the research would be the potential harm from a breach of confidentiality (If Checked, explain below):

☐ The research involves minimal risk and includes no procedures for which written consent is normally required outside the research context.

☐ Request waiver of some or all elements of consent.

☐ The research involves no more than minimal risk to the subjects.

☐ A waiver will not adversely affect the rights and welfare of the subjects.

☐ The research could not practicably be carried out without waiver or alteration.

☐ Where appropriate, the subjects will be provided with additional pertinent information after participation (If checked, explain below):

Appendix: Inclusion of Vulnerable Populations

The targeting or inclusion of potentially vulnerable populations (other than children, pregnant women/fetuses and prisoners) in research requires special considerations. Provide information on the following populations, if applicable, in this research. *Note: 1-4 not all required but at least one must be filled out.

1. Mentally/Emotionally/Developmentally Disabled

Provide justification:

Explain how competency to provide consent will be determined and plan for obtaining surrogate consent:

2. Minority Group(s)/Non-English Speakers

Provide justification:

Provide plan for obtaining consent:

3. Elderly (65+)

Included

Provide justification:

Elderly collectors and preparators are included in the survey if they wish to participate, as they will likely have experience in the realms that we wish to address.

If competency to provide consent may be an issue, describe how competency will be determined and plan for obtaining consent:

These participants are assumed to be able to competently use a computer and read the consent documentation, as they will be receiving the email, read it, and click on the link to participate.

4. Gender Imbalance

If all or more of one gender are targeted, provide justification for this:

The Florida State University
Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673, FAX (850) 644-4392

APPROVAL MEMORANDUM
Date: 2/16/2016
To: Austin Mast
Address: 4295
Dept.: BIOLOGICAL SCIENCE
From: Thomas L. Jacobson, Chair

Re: Use of Human Subjects in Research
Outlier Detection and Documentation by Collectors of Biodiversity Specimens

The application that you submitted to this office in regard to the use of human subjects in the proposal referenced above have been reviewed by the Secretary, the Chair, and one member of the Human Subjects Committee. Your project is determined to be Expedited per 45 CFR Â§ 46.110(7) and has been approved by an expedited review process.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals, which may be required.

If you submitted a proposed consent form with your application, the approved stamped consent form is attached to this approval notice. Only the stamped version of the consent form may be used in recruiting research subjects.

If the project has not been completed by 2/14/2017 you must request a renewal of approval for continuation of the project. As a courtesy, a renewal notice will be sent to you prior to your expiration date; however, it is your responsibility as the Principal Investigator to timely request renewal of your approval from the Committee.

You are advised that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report, in writing any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department, and should review protocols as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Human Research Protection. The Assurance Number is FWA00000168/IRB number IRB00000446.

Cc: Don Levitan, Chair
HSC No. 2016.17558

The Florida State University
Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673, FAX (850) 644-4392

RE-APPROVAL MEMORANDUM

Date: 2/27/2017

To: Austin Mast

Address: 4295
Dept.: BIOLOGICAL SCIENCE

From: Thomas L. Jacobson, Chair

Re: Re-approval of Use of Human subjects in Research
Outlier Detection and Documentation by Collectors of Biodiversity Specimens

Your request to continue the research project listed above involving human subjects has been approved by the Human Subjects Committee. If your project has not been completed by 2/26/2018, you must request renewed approval by the Committee.

If you submitted a proposed consent form with your renewal request, the approved stamped consent form is attached to this re-approval notice. Only the stamped version of the consent form may be used in recruiting of research subjects. You are reminded that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report in writing, any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor are reminded of their responsibility for being informed concerning research projects involving human subjects in their department. They are advised to review the protocols as often as necessary to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

Cc: Don Levitan, Chair
HSC No. 2017.20408

The Florida State University
Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673, FAX (850) 644-4392

RE-APPROVAL MEMORANDUM

Date: 2/5/2018

To: Austin Mast

Address: 4295
Dept.: BIOLOGICAL SCIENCE

From: Thomas L. Jacobson, Chair

Re: Re-approval of Use of Human subjects in Research
Outlier Detection and Documentation by Collectors of Biodiversity Specimens

Your request to continue the research project listed above involving human subjects has been approved by the Human Subjects Committee. If your project has not been completed by 2/4/2019, you must request renewed approval by the Committee.

If you submitted a proposed consent form with your renewal request, the approved stamped consent form is attached to this re-approval notice. Only the stamped version of the consent form may be used in recruiting of research subjects. You are reminded that any change in protocol for this project must be reviewed and approved by the Committee prior to implementation of the proposed change in the protocol. A protocol change/amendment form is required to be submitted for approval by the Committee. In addition, federal regulations require that the Principal Investigator promptly report in writing, any unanticipated problems or adverse events involving risks to research subjects or others.

By copy of this memorandum, the Chair of your department and/or your major professor are reminded of their responsibility for being informed concerning research projects involving human subjects in their department. They are advised to review the protocols as often as necessary to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

Cc: []
HSC No. 2018.23097

Bottom of Form

APPENDIX F

METHODS AND BIBLIOGRAPHY OF SURVEY OF TRAINING MATERIALS USED FOR LEARNING COLLECTING METHODS AND PROTOCOLS

To assess collector training materials, I first surveyed all in-scope literature cited by collectors in the survey. I considered literature in-scope if it appeared to be either reasonably general (e.g., *Introductory mycology*), and thus might include some reference to specimen collection, or if the resource appeared to describe the collection and preservation of biodiversity specimens (e.g., *Methods of collecting and preserving vertebrate animals*). I then scanned the bibliographies of several of these works for any other literature that might contain methods, protocols, or suggestions for collecting methods (i.e., were not regional or taxon-specific floras or faunas). I also conducted a Google search for additional references using this selection criterion.

*Starred references indicate literature cited by collectors in the survey

*Alexopoulos CJ, Mims CW, Blackwell M. 1965. Introductory mycology. 4th ed. John Wiley & Sons, Inc.

*Anderson RM. 1965. Methods of collecting and preserving vertebrate animals. 4th ed. National Museum of Canada; 69.

*Bridson D, Forman L, eds. 1992. The herbarium handbook. Revised ed. Great Britain: Royal Botanic Gardens, Kew.

*Borrer DJ, DeLong DM, Triplehorn CA. 1981. An introduction to the study of insects. Philadelphia: Saunders College Pub.

British Columbia Ministry of Forests. 1996. Techniques and procedures for collecting, preserving, processing, and storing botanical specimens. Res. Br., B.C. Min. For., Victoria, B.C. Work. Pap. 18.

Bulbert M, Gollan J, Donnelly A, Wilkie L. 2007. The invertebrate collection manual. Australian Museum.

*Carter D, Walker AK. 1999. Care and conservation of natural history collections. Oxford: Butterworth Heinemann.

Carter R, Bryson CT, Darbyshire SJ. 2007. Preparation and use of voucher specimens for documenting research in weed science. Weed Technology. 21:1101-1108.

*Cockrum EL. 1955. Laboratory manual of mammalogy. Minneapolis: Burgess Publishing Company.

Cook JA, Dunnum JL, Bogan M, Gannon WL, Ramontnik CA, Yates TL. 2016. Division of Mammals Collection Management Procedures Manual. Museum of Southwestern Biology, University of New Mexico.

Cota AP. Técnicas de coletas, herborização e inventário florístico de arbóreas. Manejo Florestal – DEF/UFV. Available from: ftp://ftp.ufv.br/def/disciplinas/ENF344/MANEJOFLORESTASNATIVAS/ManualFloristica/Apostila-ManFlo.pdf.

Croft J. 1999. A guide to collecting herbarium specimens of ferns and their allies. Available from: http://www.anbg.gov.au/fern/collecting.html.

Davis AP. 2011. Collecting herbarium vouchers. In: Collecting Plant Genetic Diversity; Technical Guidelines, L Guarino, V Ramanatha Rao, E Goldberg (eds.) published by Biodiversity International. Available from: http://cropgenebank.sgrp.cgiar.org/index.php?option=com_content&view=article&id=663.

De Groot SJ. 2011. Collecting and processing cacti into herbarium specimens, using ethanol and other methods. Systematic Botany. 36:981-989.

Dransfield J. 1986. A guide to collecting palms. Annals of the Missouri Botanical Garden. 73:166-176.

*Fidalgo O, Bononi VLR. 1984. Técnicas de coleta, preservação e herborização de material botânico. São Paulo: Instituto de Botânica.

Frank MS, Perkins KD. 2015. Preparation of plant specimens for deposit as herbarium vouchers. Available from: https://www.floridamuseum.ufl.edu/herbarium/voucher.htm

Funk VA, Gostel M, Devine A, Kelloff C, Van Neste K, Gabler S, Wurdack K, Tuccinardi C, Radosavljevic A, Peters M, Coddington J. 2016. Botany Best Practices. (US National Herbarium): A product of GGI-Gardens. National Museum of History. Available from: https://ggi.si.edu/sites/default/files/resources/GGI-Botany%20Expedition%20Protocols%20%5BDRAFT%5D%2016June2016.pdf

Hahn J. 2016. Collecting and preserving insects. University of Minnesota Extension. Available from: <http://www.extension.umn.edu/youth/mn4-H/projects/environment/entomology/collecting-and-preserving-insects/>

Hall ER. 1962. Collecting and preparing study specimens of vertebrates. Lawrence, KS: University of Kansas, Museum of Natural History.

Island Surveys to Learn about Endemic Species (ISLES). 2011. Instructions for the Field Collection and Preservation of Mammals. Museum of Southwestern Biology. Available from: http://msb.unm.edu/isles/Instructions%20for%20the%20field%20collection%20%20and%20preservation%20of%20mammals.pdf.

Kathyrn Kalmbach Herbarium. Plant collection protocol. Denver Botanic Gardens. Available from: https://www.botanicgardens.org/sites/default/files/khdprotocols2012.pdf.

Kaul TN. 1997. Introduction to mushroom science. Enfield, NH: Science publishers, Inc.

Knowlton FH. 1891. Directions for collecting recent and fossil plants. Smithsonian Institution, United States National Museum. Washington: Government Printing Office.

Lacey J, Short S, Mosley J. 2001. How to collect, press and mount plants. Bozeman, MT, USA: Montana State University Extension Service. Available from: http://store.msuextension.org/publications/agandnaturalresources/mt198359ag.pdf

Langeron M. 1965. Outline of mycology. Second edition revised by R. Vanbreuseghem. Springfield, IL: Charles C. Thomas Publisher.

MacFarlane RB. 1985. Collecting and preserving plants for science and pleasure. New York: Dover Publications, Inc.

Manton WP. 1882. Field botany: handbook for the collector, containing instructions for gathering and preserving plants and the formation of the herbarium. Boston: Lee and Shepard Publishers.

*Martin JEH. 1977. The insects and arachnids of Canada, Part 1: Collecting, preparing, and preserving insects, mites, and spiders. Ottawa, Ontario: Biosystematics Research Institute. Available from: http://esc-sec.ca/aafcmonographs/insects_and_arachnids_part_1_eng.pdf

*Martin RE, Pine RH, DeBlase AF. 2001. A manual of mammalogy: With keys to families of the world. 3rd ed. McGraw-Hill Science/Engineering/Math.

*Metsger DA, Byers SC, eds. 1999. Managing the modern herbarium: An inter-disciplinary approach. Vancouver, BC, Canada: Society for the Preservation of Natural History Collections.

*Mueller G, Foster M, Bills G. 2004. Biodiversity of fungi, 1st ed. Academic Press.

National Standards for Phytosanitary Measures. 2014. Standard technical protocols for collection and handling of insect samples. Available from: http://www.npqpnepal.gov.np/

Oldroyd H. 1958. Collecting, preserving, and studying insects. New York: The Macmillan Company.

Oman PW, Cushman AD. 1946. Collection and preservation of insects. U.S. Department of Agriculture. 601.

Queensland Herbarium. 2013. Collecting and preserving plant specimens, a manual. Department of Science, Information Technology and Innovation.

*Radford AE, Dickison WC, Massey JR, Bell CR. 1974. Vascular plant systematics. Harper & Row, Publishers, Inc., New York.

Resources Inventory Committee. 1997. Fish collection methods and standards, Version 4.0. Ministry of Environment, Lands and Parks Resources Inventory Branch, Terrestrial Ecosystems Task Force, Resources Inventory Committee, The Province of British Columbia.

Resources Inventory Committee. 1999. Voucher specimen collection, preparation, identification and storage protocol: Plants & fungi (Standards for components of British Columbia's biodiversity no. 4b). Version 2.0. Ministry of Environment, Lands and Parks Resources Inventory Branch, Terrestrial Ecosystems Task Force, Resources Inventory Committee, The Province of British Columbia.

Robertson KR. 1980. Observing, photographing and collecting plants. Illinois Natural History Survey Circular. 55. Available from: http:/hdl.handle.net/2142/73357.

Ross TS. 1996. Herbarium specimens as documents; purposes and general collecting techniques. Crossosoma. 22:3-39.

Rotta E, Caminha de Carvalho e Beltrami L, Zonta M. 2008. Manual de prática de coleta e herborização de material botânico. Colombo, Brazil: Empresa Brasileira de Pesquisa Agropecuária, Embrapa Florestas, Ministério da Agricultura, Pecuária e Abastecimento. Available from: https://www.infoteca.cnptia.embrapa.br/bitstream/doc/315636/1/Doc173.pdf.

Schauff ME, ed. Collecting and preserving insects and mites: techniques and tools. Washington DC: National Museum of Natural History. NHB-168.

*Simmons JE. 2002. Herpetological collecting and collections management. Herpetological Circular. 31:1-153.

Soderstrom TR, Young SM. 1983. A guide to collecting bamboos. Annals of the Missouri Botanical Garden. 70:128-136.

*Vandel A. 1965. Biospeleology: the biology of cavernicolous animals. Pergamon Press Ltd.

*Webster J. 1970. Introduction to fungi. Cambridge University Press, New York.

REFERENCES

Abu-Asab MS, Peterson PM, Shetler SG, Orli SS. 2001. Earlier plant flowering in spring as a response to global warming in the Washington, DC, area. Biodiversity and Conservation. 10:597-612.

Agnihotri P, Husain T, Shirke PA, Sidhu OP, Singh H, Dixit V, Amla DV, Nautiyal CS. 2017. Climate change-driven shifts in elevation and ecophysiological traits of Himalayan plants during the past century. Current Science. 112(3):595-601.

Anderson RM. 1965. Methods of collecting and preserving vertebrate animals. National Museum of Canada. 69(18).

Anderson-Teixeira KJ, Davies SJ, Bennett AC, Gonzalez-Akre EB, Muller-Landau HC, Joseph Wright S, Abu Salim K, Almeyda Zambrano AM, Alonso A, Baltzer JL, Basset Y, Bourg NA, Broadbent EN, Brockelman WY, Bunyavejchewin S, Burslem DF, Butt N, Cao M, Cardenas D, Chuyong GB, Clay K, Cordell S, Dattaraja HS, Deng X, Detto M, Du X, Duque A, Erikson DL, Ewango CE, Fischer GA, Fletcher C, Foster, RB, Giardina CP, Gilbert GS, Gunatilleke N, Gunatilleke S, Hao Z, Hargrove WW, Hart TB, Hau BC, He F, Hoffman FM, Howe RW, Hubbell SP, Inman-Narahari FM, Jansen PA, Jiang M, Johnson DJ, Kanzaki M, Kassim AR, Kenfack D, Kibet S, Kinnaird MF, Korte L, Kral K, Kumar J, Larson AJ, Li Y, Li X, Liu S, Lum SK, Lutz JA, Ma K, Maddalena DM, Makana J, Malhi Y, Marthews T, Mat Serudin R, McMahon SM, McShea WJ, Memiaghe HR, Mi X, Mizuno T, Morecroft M, Myers JA, Novotny V, Oliveira AA, Ong PS, Orwig DA, Ostertag R, Ouden J, Parker GG, Phillips RP, Sack L, Sainge MN, Sang W, Sri-ngernyuang K, Sukumar R, Sun I, Sungpalee W, Suresh HS, Tan S, Thomas SC, Thomas DW, Thompson J, Turner BL, Uriarte M, Valencia R, Vallejo MI, Vicentini A, Vrška T, Wang X, Wang X, Weiblen G, Wolf A, Xu H, Yap S, Zimmerman J. 2015. TFS-ForestGEO: a worldwide network monitoring forests in an era of global change. Global Change Biology. 21: 528-549.

Bertin RI. 2015. Climate change and flowering phenology in Worcester county, Massachusetts. International Journal of Plant Sciences. 176(2):107-119.

Bjorkman AD, Vellend M, Frei ER, Henry GHR. 2017. Climate adaptation is not enough: warming does not facilitate success of southern tundra plant populations in the high Arctic. Global Change Biology. 23:1540-1551.

Bock A, Sparks TH, Estrella N, Jee N, Casebow A, Schunk C, Leuchner M, Menzel A. 2014. Changes in first flowering dates and flowering duration of 232 plant species on the island of Guernsey. Global Change Biology. 20:3508-3519.

Boyle B, Hopkins N, Lu Z, Raygoza Garay JA, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, Mckay SJ, Lowry S, Freeland C, Peet RK, Enquist BJ. 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics. 14:16.

*Bradley NL, Leopold AC, Ross J, Huffaker W. 1999. Phenological changes reflect climate change in Wisconsin. Proceedings of the National Academy of Sciences USA. 96:9701-9704.

Brown JH, Stevens GC, Kaufman DM. 1996. The geographic range: Size, shape, boundaries, and internal structure. Annual Review of Ecology and Systematics. 27:597-623

Brown TB, Hultine KR, Steltzer H, Denny EG, Denslow MW, Granados J, Henderson S, Moore D, Nagai S, SanClements M, Sánchez-Azofeifa A, Sonnentag O, Tazik D, Richardson AD. 2016. Using phenocams to monitor our changing Earth: toward a global phenocam network. Frontiers in Ecology and the Environment. 14:84-93.

Bucher EH, Aramburu RM. 2014. Land-use changes and monk parakeet expansion in the Pampas grasslands of Argentina. Journal of Biogeography. 41:1160-1170.

Burkle LA, Marlin JC, Knight TM. 2013. Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. Science. 339:1611-1615.

Calinger KM, Queenborough S, Curtis PS. 2013. Herbarium specimens reveal the footprint of climate change on flowering trends across north-central North America. Ecology Letters. 16:1037-1044.

CaraDonna PJ, Inouye DW. 2015. Phenological responses to climate change do not exhibit phylogenetic signal in a subalpine plant community. Ecology. 96(2):355-361.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderi A, Srivastava S, Naeem S. 2012. Biodiversity loss and its impact on humanity. Nature. 486:59-67.

Case MA, Flinn KM, Jancaitis J, Alley A, Paxton A. 2007. Declining abundance of American ginseng (*Panax quinquefolius* L.) documented by herbarium specimens. Biological Conservation. 34:22-30.

Chen X, Wang L, Inouye D. 2017. Delayed response of spring phenology to global warming in subtropics and tropics. Agricultural and Forest Meteorology. 234-235:222-235.

Chmielewski F, Blümel K, Henniges Y, Blanke M, Weber RWS, Zoth M. 2011. Phenological models for the beginning of apple blossom in Germany. Meterologische Zeitschrift. 20(5):487-496.

Cleland EE, Chiariello NR, Loarie SR, Mooney HA, Field CB. 2006. Diverse responses of phenology to global changes in a grassland ecosystem. Proceedings of the National Academy of Sciences. 103(37):13740-13744.

Conceicao AA, Orr BJ. 2012. Post-fire flowering and fruiting in *Vellozia sincorana*, a caulescent rosette plant endemic to Northeast Brazil. Acta Botanica Brasilica. 26(1):94-100.

Cook BI, Wolkovich EM, Parmesan C. 2012. Divergent responses to spring and winter warming drive community level flowering trends. Proceedings of the National Academy of Sciences. 109(23):9000-9005.

Daru BH, Park DS, Primack RB, Willis CG, Barrington DS, Whitfeld TJS, Seidler TG, Sweeney PW, Foster DR, Ellison AM, Davis CC. 2017. Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytologist. https://doi.org/10.1111/nph.14855

Daume S. 2016. Mining Twitter to monitor invasive alien species – An analytical framework and sample information topologies. Ecological Informatics. 31:70-82.

Davis CC, Willis CG, Primack RB, Miller-Rushing AJ. 2010. The importance of phylogeny to the study of phenological response to global climate change. Philosophical Transactions of the Royal Society B. 365:3201-3213.

Davis CC, Willis CG, Connolly B, Kelly C, Ellison AM. 2015. Herbarium records are reliable sources of phenological change driven by climate and provide novel insights into species' phenological cueing mechanisms. American Journal of Botany. 102(10):1599-1609.

Delisle F, Lavoie C, Jean M, Lachance D. 2003. Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. Journal of Biogeography 30:1033-1042.

Diamond SE, Frame AM, Martin RA, Buckley LB. 2011. Species' traits predict phenological responses to climate change in butterflies. 92(5):1005-1012.

*Diskin E, Proctor H, Jebb M, Sparks T, Donnelly A. 2012. The phenology of *Rubus fruticosus* in Ireland: herbarium specimens provide evidence for the response of phenophases to temperature, with implications for climate warming. International Journal of Biometeorology.

eFloras. 2008. Missouri Botanical Garden, St. Louis, MO & Harvard University Herbaria, Cambridge, MA. http://www.efloras.org. Accessed 28 September 2017.

Ellwood ER, Temple SA, Primack RB, Bradley NL, Davis CC. 2013. Record-breaking early flowering in the eastern United States. PLoS ONE. 8(1):1-9.

Ellwood ER, Playfair SR, Polgar CA, Primack RB. 2014. Cranberry flowering times and climate change in southern Massachusetts. International Journal of Biometerology. 58:1693-1697.

*Estrella N, Sparks TH, Menzel A. 2007. Trends and temperature response in the phenology of crops in Germany. Global Change Biology. 13:1737-1747.

Federal Interagency Committee for the Management of Noxious and Exotic Weeds (FICMNEW). 2003. A National Early Detection and Rapid Response System for Invasive Plants in the United States: Conceptual Design. Washington, DC.

Ferrier S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? Systematic Biology. 51:331-363.

Fitter AH, Fitter RSR. 2002. Rapid changes in flowering time in British plants. Science. 296:1689-1691.

Flora of North America Editorial Committee, eds. 1993+. Flora of North America North of Mexico. 20+ vols. New York and Oxford.

Forrest JRK. 2015. Plant-pollinator interactions and phenological change: what can we learn about climate impacts from experiments and observations? Oikos. 124:4-13.

Franks SJ, Sim S, Weis AE. 2007. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proceedings of the National Academy of Sciences. 104:1278-1282.

Gaira KS, Rawal RS, Rawat B, Bhatt ID. 2014. Impact of climate change on the flowering of *Rhododendron arboretum* in central Himalaya, India. Current Science. 106(12):1735-1738.

*Gallagher RV, Hughes L, Leishman MR. 2009. Phenological trends among Australian alpine species: using herbarium records to identify climate-change indicators. Australian Journal of Botany. 57:1-9.

Garcillan PP, Ezcurra E, Vega E. 2008. Guadalupe Island: Lost paradise recovered? Overgrazing impact on extinction in a remote oceanic island as estimated through accumulation functions. Biodiversity and Conservation. 17:1613-1625.

Gaston KJ. 2000. Global patterns in biodiversity. Nature. 405:220-227.

Georgia Department of Natural Resources. *Geographic Regions of Georgia, with Counties*. Accessed at <http://georgiainfo.galileo.usg.edu/images/uploads/gallery/GeographicRegions.jpg>

Gezon ZJ, Inouye DW, Irwin RE. 2016. Phenological change in a spring ephemeral: implications for pollination and plant reproduction. Global Change Biology. 22:1779-1793.

Gill AJ, Gallinat AS, Sanders-DeMott R, Rigden AJ, Short Gianotti DJ, Mantooth JA, Templer PH. 2015. Changes in autumn senescence in northern hemisphere deciduous trees: a meta-analysis of autumn phenology studies. Annals of Botany. 116:875-888.

Gillespie MAK, Baggesen N, Cooper EJ. 2016. High Arctic flowering phenology and plant-pollinator interactions in response to delayed snow melt and simulated warming. Environmental Research Letters. 11:115006.

Gomez-Mendoza L, Arriaga L. 2007. Modeling the effect of climate change on the distribution of oak and pine species of Mexico. Conservation Biology. 21(6):1545-1555.

Goodwin ZA, Harris DJ, Filer D, Wood JRI, Scotland RW. 2015. Widespread mistaken identity in tropical plant collections. Current Biology. 25(22):R1066-R1067.

Gordo O, Sanz JJ. 2005. Phenology and climate change: A long-term study in a Mediterranean locality. Oecologia. 146(3):484-495.

Gordo O, Sanz JJ. 2010. Impact of climate change on plant phenology in Mediterranean ecosystems. Global Change Biology. 16:1082-1106.

Grant PR, Grant BR. 1995. Predicting microevolutionary responses to directional selection on heritable variation. Evolution. 49:241-251.

Gugger S, Kesselring H, Stöcklin J, Hamann E. 2015. Lower plasticity exhibited by high- versus mid-elevation species in their phenological responses to manipulated temperature and drought. Annals of Botany. 1-10.

Hammock J, Poelen J. 2016. Fresh Data: what's new and what's interesting? Symposium: Big Data Analysis Methods and Techniques as Applied to Biocollections. Taxonomic Database Working Group (TDWG) 2016 Annual Conference. 7 December 2016. Santa Clara de San Carlos, Costa Rica.

Harper C, Wolf JC. 2009. Morphologic effects of the stress response in fish. Institute for Laboratory Animal Research Journal. 50:387-396

Hart R, Salick J, Ranjitkar S, Xu J. 2014. Herbarium specimens show contrasting phenological responses to Himalayan climate. Proceedings of the National Academy of Sciences. 111(29):10615-10619.

Herpin U, Markert B, Weckert V, Berlekamp J, Friese K, Siewers U, Lieth H. 1997. Retrospective analysis of heavy metal concentrations at selected locations in the Federal Republic of Germany using moss material from a herbarium. The Science of the Total Environment. 205:1-12.

Hetherington AJ, Dubrovsky JG, Dolan L. 2016. Unique cellular organization in the oldest root meristem. Current Biology. 26(12):1629-1633.

Hill A, Guralnick R, Smith A, Sallans A, Gillespie R, Denslow M, Gros J, Murrell Z, Conyers T, Oboyski P, Ball J, Thomer A, Prys-Jones R, de la Torre J, Kociolek P, Fortson L. 2012. The notes from nature tool for unlocking biodiversity records from museum records through citizen science. ZooKeys. 2009:217-233.

Hodgkins GA, Dudley RW, Huntington TG. 2003. Changes in the timing of high river flows in New England over the 20th Century. Journal of Hydrology. 278:244-252.

Høye TT, Post E, Schmidt NM, Trøjelsgaard K, Forchhammer MC. 2013. Shorter flowering seasons and declining abundance of flower visitors in a warmer Arctic. Nature Climate Change. 3:759-763.

Ibañez I, Primack RB, Miller-Rushing AJ, Ellwood E, Higuchi H, Don Lee S, Kobori H, Silander JA. 2010. Forecasting phenology under global warming. Philosophical Transactions of the Royal Society B. 365:3247-3260.

Inouye DW. 2008. Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. Ecology. 89(2):353-362.

Island Surveys to Learn about Endemic Species (ISLES). 2001. Instructions for the Field Collection and Preservation of Mammals. Museum of Southwestern Biology http://msb.unm.edu/isles/Instructions%20for%20the%20field%20collection%20%20and%20preservation%20of%20mammals.pdf. Accessed 22 November 2017

James SA, Soltis PS, Chapman AD, Nelson G, Paul DL, Collins M. 2018. Herbarium data: Global biodiversity and societal botanical needs for novel research. Applications in Plant Sciences. 6(2): e1024.

Jarvis A, Lane A, Hijmans RJ. 2008. The effect of climate change on crop wild relatives. Agriculture, Ecosystems, and Environment. 126:13-23.

Jeong S, Ho C, Gim H, Brown ME. 2011. Phenology shifts at start vs. end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982-2008. Global Change Biology. 17:2385-2399.

Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ. 2002. Plant Systematics: A Phylogenetic Approach, Second Edition. Sinauer Associates, Inc.: Massachusetts, USA.

Kang H, Jang J. 2004. Flowering patterns among Angiosperm species in Korea: diversity and constraints. 47(4):348-355.

Kharouba HM, Velland M. 2015. Flowering time of butterfly nectar food plants is more sensitive to temperature than the timing of butterfly adult flight. Journal of Animal Ecology. 84:1311-1321.

Kudo G, Ida TY. 2013. Early onset of spring increases the phenological mismatch between plants and pollinators. Ecology. 94(10):2311-2320.

Lauer M, Aswani S. 2010. Indigenous knowledge and long-term ecological change: Detection, interpretation, and responses to changing ecological conditions in Pacific Island communities. Environmental Management. 45 :985-997.

Lavoie C, Lachance D. 2006. A new herbarium-based method for reconstructing the phenology of plant species across large areas. American Journal of Botany. 943(4):512-516.

Lavoie C, Jodoin Y, Goursaud de Merlis A. 2007. How did common ragweed (*Ambrosia artemisiifolia* L.) spread in Quebec? A historical analysis using herbarium records. Journal of Biogeography. 34: 1751-1761.

Leonard P., ed. 2010. A guide to collecting and preserving fungal specimens for the Queensland Herbarium. Queensland Herbarium, Department of Environment and Resource Management, Brisbane.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for display and annotation of phylogenetic and other trees. Nucleic Acids Research. 44(W1):W242-245

*Li Z, Wu N, Gao X, Wu Y, Oli KP. 2013. Species-level phenological responses to ‘global warming’ as evidenced by herbarium collections in the Tibetan Autonomous Region. Biodiversity Conservation. 22:141-152.

Liu Y, Reich PB, Li G, Sun S. 2011. Shifting phenology and abundance under experimental warming alters trophic relationships and plant reproductive capacity. Ecology. 92(6):1201-1207.

Loiselle BA, Jorgensen PM, Consiglio T, Jimenez I, Blake JG, Lohmann LG, Montiel OM. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? Journal of Biogeography. 35(1):105-116.

Lomolino MV, Riddle BR, Whittaker RJ. 2016. Biogeography: Biological diversity across space and time, 5th ed. Sinauer Associates, Sunderland, MA.

López-Caamal A, Tovar-Sánchez E. 2014. Genetic, morphological, and chemical patterns of plant hybridization. Revista Chilena de Historia Natural. 87:1-14.

Mäkiranta P, Laiho R, Mehtätalo L, Straková P, Sormunen J, Minkkinen K, Penttilä T, Fritze H, Tuittila E. 2018. Responses of phenology and biomass production of boreal fens to climate warming under different water-table level regimes. Global Change Biology. 24:944-956.

Martin JEH. 1977. The insects and arachnids of Canada, Part 1: Collecting, preparing, and preserving insects, mites, and spiders. Ottawa, Ontario: Biosystematics Research Institute. http://esc-sec.ca/aafcmonographs/insects_and_arachnids_part_1_eng.pdf. Accessed 22 November 2017

Martin MD, Zimmer EA, Olsen MT, Foote AD, Gilbert MTP, Brush GS. 2014. Herbarium specimens reveal a historical shift in phylogeographic structure of common ragweed during native range disturbance. Molecular Ecology. 23:1701-1716.

Matthews WJ. 2015. Basic biology, good field notes, and synthesizing across your career. Copeia. 3:495-501.

Menne MJ, Williams CN, Vose RS. 2009. The U.S. historical climatology network monthly temperature data, version 2. Bulletin of the American Meterological Society. 90(7):993-1007.

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-Kübler K, Bissolli P, Braslavska O, Briede A, Chmielewski FM, Crepinsek Z, Curnel Y, Dahl Å, Defila C, Donnelly A, Filella Y, Jatczak K, Mage F, Mestre A, Nordli Ø, Peñuelas J, Pirinen P, Remišova V, Scheifinger H, Striz M, Susnik A, Van Vliet AJH, Wielgolaski F, Zach S, Zust A. 2006. European phenological response to climate change matches the warming pattern. Global Change Biology. 12:1969-1976.

Meyers P, Lundrigan BL, Hoffman SMG, Poor Haraminac A, Seto SH. 2009. Climate-induced changes in the small mammal communities of the Northern Great Lakes Region. Global Change Biology. 15:1434-1454.

Miller-Rushing AJ, Primack RB, Primack D, Mukunda S. 2006. Photographs and herbarium specimens as tools to document phenological changes in response to global warming. American Journal of Botany. 93(11):1667-1674.

Miller-Rushing AJ, Primack RB. 2008. Global warming and flowering times in Thoreau's Concord: A community perspective. Ecology. 89(2):332-341.

Miller-Rushing AJ, Inouye DW, Primack RB. 2008. How well do first flowering dates measure plant responses to climate change? The effects of population size and sampling frequency. Journal of Ecology. 96:1289-1296.

Miller-Rushing AJ, Høye TT, Inouye DW, Post E. 2010. Effects of phenological mismatches on demography. Philosophical Transactions of the Rotyal Society B. 365:3177-3186.

Molnár V A, Tökölyi J, Végvári Z, Sramkó J, Barta Z. 2012. Pollination mode predicts phenological response to climate change in terrestrial orchids: a case study from central Europe. Journal of Ecology. 100:1141-1152.

Morris RA, Dou L, Hanken J, Kelly M, Lowery DB, Ludäscher B, Macklin JA, Morris PJ. 2013. Semantic annotation of mutable data. PLoS ONE. 8.

Mota-Vargas C, Rojas-Soto OR. 2012. The importance of defining the geographic distribution of species for conservation: The case of the bearded wood-partridge. Journal for Nature Conservation. 20(1):10-17.

Moussus J, Julliard R, Jiguet F. 2010. Featuring 10 phenological estimators using simulated data. Methods in Ecology and Evolution. 1:140-150.

Munson SM, Long AL. 2016. Climate drives shifts in grass reproductive phenology across the western USA. New Phytologist. 1-11.

*Munson SM, Sher AA. 2015. Long-term shifts in the phenology of rare and endemic Rocky Mountain plants. American Journal of Botany. 102(8):1268-1276.

Myltsev A, Mozzherin D. 2016. Global Names Parser. https://github.com/GlobalNamesArchitecture/gnparser. Accessed 14 September 2017

*Nagai S, Nasahara KN. 2017. Seasonal leaf phenology data for 12 tree species in a cool-temperate deciduous broadleaved forest in Japan from 2005 to 2014. Ecological Research. 32:107-107.

NatureServe. 2017. East Gulf Coastal Plain. *LandScope.org*. Accessed 11 March 2017. Available at: <http://www.landscope.org/explore/natural_geographies/ecoregions/East%20Gulf%20Coastal%20Plain/>

Neil KL, Landrum L, Wu J. 2010. Effects of urbanization on flowering phenology in the metropolitan phoenix region of USA: Findings from herbarium records. Journal of Arid Environments. 74:440-444.

Nordman CW, Pyne M, Smyth RL, White R. 2014. Threats to ecological systems in the South Atlantic Landscape Conservation Cooperative area. Nature Serve. Available at: <http://data.southatlanticlcc.org/Threats_to_Ecological_Systems_in_SALCC.pdf>

North Carolina Department of Public Instruction. *North Carolina Regions*. Available at: <http://www.ncpublicschools.org/curriculum/socialstudies/elementary/studentsampler/20geography#maps>

Noss RF, Platt WJ, Sorrie BA, Weakley AS, Means DB, Costanza J, Peet RK. 2015. How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. Diversity and Distributions. 21:236-244.

Otero-Ferrer F, González JA, Freitas M, Araújo R, Azevedo JMN, Holt WV, Tuya F, Haroun R. 2017. When natural history collections reveal secrets on data deficient threatened species: Atlantic seahorses as a case study. Biodiversity Conservation. 26:2791-2802.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. Annual Review of Genetics. 34:401-437.

Ożgo M, Schilthuizen M. 2012. Evolutionary change in *Cepaea nemoralis* shell colour over 43 years. Global Change Biology. 18:74-81.

Pan C, Feng Q, Zhao H, Liu L, Li Y, Li Y, Zhang T, Yu X. 2017. Earlier flowering did not alter pollen limitation in an early flowering shrub under short-term experimental warming. Scientific Reports. 7(2795):1-5.

Panchen ZA, Primack RB, Anisko T, Lyons RE. 2012. Herbarium specimens, photographs, and field observations show Philadelphia area plants are responding to climate change. American Journal of Botany. 99(4):751-756.

Park IW, Schwartz MD. 2015. Long-term herbarium records reveal temperature-dependent changes in flowering phenology in the southeastern USA. International Journal of Biometeorology. 59:347-355.

Parmesan C. 2006. Ecological and evolutionary responses to recent climate change. Annual Review of Ecology, Evolution, and Systematics. 37:637-669.

Pau S, Wolkovich EM, Cook BI, Davies TJ, Kraft NJB, Bolmgren K, Betancourt JL, Cleland EE. 2011. Predicting phenology by integrating ecology, evolution, and climate science. Global Change Biology. 17:3633-3643.

Pauw A, Hawkins JA. 2011. Reconstruction of historical pollination rates reveals linked declines of pollinators and plants. Oikos. 120:344-349.

Pearse WD, Davis CC, Inouye DW, Primack RB, Davies TJ. 2017. A statistical estimator for determining the limits of contemporary and historic phenology. Nature Ecology & Evolution. https://doi.org/10.1038/s41559-017-0350-0.

Peñuelas J, Filella I. 2002. Metal pollution in Spanish terrestrial ecosystems during the twentieth century. Chemosphere. 46:501-505.

Peñuelas J, Filella I, Zhang X, Llorens L, Ogaya R, Lloret F, Comas P, Estiarte M, Terradas J. 2004. Complex spatiotemporal phenological shifts as a response to rainfall changes. New Phytologist. 161:837-846.

Phillimore AB, Stålhandske S, Smithers RJ, Bernard R. 2012. Dissecting the contributions of plasticity and local adaptation to the phenology of a butterfly and its host plants. The American Naturalist. 180(5):655-670.

*Polgar C, Gallinat A, Primack RB. 2014. Drivers of leaf-out phenology and their implications for species invasions: insights from Thoreau's Concord. New Phytologist. 202:106-115.

Posledovich D, Toftegaard T, Wiklund C, Ehrlén J, Gotthard K. 2017. Phenological synchrony between a butterfly and its host plants: Experimental test of effects of spring temperature. Journal of Animal Ecology. 87:150-161.

Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE. 2010. Global pollinator declines: trends, impacts and drivers. Trends in Ecology and Evolution. 25:345-353.

Poulickova A, Hajkova P, Kintrova K, Bat'kova R, Czudkova M, Hajek M. 2013. Tracing decadal environmental change in ombrotrophic bogs using diatoms from herbarium collections and transfer functions. Environmental Pollution. 179:201-209.

Powell K. 2016. The waiting game. Nature. 530:148-151.

Price M, Waser NM. 1998. Effects of experimental warming on plant reproductive phenology in a subalpine meadow. Ecology. 79(4):1261-1271.

Primack D, Imbres C, Primack RB, Miller-Rushing AJ, Del Tredici P. 2004. Herbarium specimens demonstrate earlier flowering times in response to warming in Boston. American Journal of Botany. 91(8):1260-1264.

Prince K, Zuckerberg B. 2014. Climate change in our backyards: the reshuffling of North America's winter bird communities. Global Change Biology. 21:572-585.

Radford AE, Dickison WC, Massey JR, Bell CR. 1974. Vascular plant systematics. Harper & Row, New York.

Rafferty NE, CaraDonna PJ, Bronstein JL. 2015. Phenological shifts and the fate of mutualisms. Oikos. 124:14-21.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ Version 3.3.2

Resources Inventory Committee (RIC). 1999 Voucher specimen collection, preparation, identification and storage protocol: Plants & fungi (Standards for components of British Columbia's biodiversity no. 4b). Version 2.0. Ministry of Environment, Lands and Parks Resources Inventory Branch, Terrestrial Ecosystems Task Force, Resources Inventory Committee, The Province of British Columbia.

Resources Inventory Committee (RIC). 1997. Fish collection methods and standards, Version 4.0 Ministry of Environment, Lands and Parks Resources Inventory Branch, Terrestrial Ecosystems Task Force, Resources Inventory Committee, The Province of British Columbia

Rhymer JM, Simberloff D. 1996. Extinction by hybridization and introgression. Annual Reviews in Ecology and Systematics. 27:83-109.

Richardson AD, Black TA, Ciais P, Delbart N, Friedl MA, Gabron N, Hollinger DY, Kutsch WL, Longodz B, Luyssaert S, Migliavacca M, Montagnani L, Munger JW, Moors E, Piao S, Rebmann C, Reichstein M, Saigusa N, Tomelleri E, Vargas R, Varlagin A. 2010. Philosophical Transactions of the Royal Society B. 365:3227-3246.

Ricklefs RE. 2004. A comprehensive framework for global patterns in biodiversity. Ecology Letters. 7:1-15.

Riekerk G. *South Carolina Coastal Plain*. SCDNR Marine Resources Research Institute. Accessed via <https://webapp1.dlib.indiana.edu/virtual_disk_library/index.cgi/4928836/FID1596/html/envicond/geomorph/gmmorph.htm#summary>

Rumpff L, Coates F, Morgan JW. 2010. Biological indicators of climate change: evidence from long-term flowering records of plants along the Victorian coast, Australia. Australian Journal of Botany. 58:428-439.

Sahagun-Godinez E. 1996. Trends in the phenology of flowering in the Orchidaceae of western Mexico. Biotropica. 28(1):130-136.

Schwartz MD, Betancourt JL, Weltzin JF. 2012. From Caprio's lilacs to the USA National Phenology Network. Frontiers in Ecology and the Environment. 10:324-327

Scranton K, Amarasekare P. 2017. Predicting phenological shifts in a changing climate. Proceedings of the National Academy of Sciences. 114(50):13212-13217.

Scyphers SB, Powers SP, Akins JL, Drymon JM, Martin CW, Schobernd ZH, Schofield PJ, Shipp RL,

Switzer TS. 2015. The role of citizens in detecting and responding to a rapid marine invasion. Conservation Letters. 8:245-250.

Sexton JP, McIntyre PJ, Angert AL, Rice KJ. 2009. Evolution and ecology of species range limits. Annual Review of Ecology, Evolution, and Systematics. 40:415-436.

Sherry RA, Zhou X, Gu S, Arnone III JA, Schimel DS, Verburg PS, Wallace LL, Luo Y. 2007. Divergence of reproductive phenology under climate warming. Proceedings of the National Academy of Sciences. 104(1):198-202.

Sherry RA, Zhou X, Gu S, Arnone III JA, Johnson DW, Schimel DS, Verburg PSJ, Wallace LL, Luo Y. 2011. Changes in duration of reproductive phases and lagged phenological response to experimental climate warming. Plant Ecology and Diversity. 4(1):23-35.

Soberon JM, Llorente JB, Oñate L. 2000. The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. Biodiversity and Conservation. 9:1441-1466.

Sorrie BA, Weakley AS. 2006. Conservation of the endangered Pinus palustris ecosystem based on Coastal Plain centres of plant endemism. Applied Vegetation Science. 9:59-66.

Sparks TH, Jeffree EP, Jeffree CE. 2000. An examination of the relationship between flowering times and temperature at the national scale using long-term phenological records from the UK. International Journal of Biometereology. 44:82-87.

Sparks TH, Huber K, Croxton PJ. 2006. Plant development scores from fixed-date photographs: the influence of weather variables and recorder experience. International Journal of Biometeorology. 50:275-279.

Spellman KV, Mulder CPH. 2016. Validating herbarium-based phenology models using citizen-science data. BioScience. 66:807-906.

Stropp J, Ladle RJ, Malhado ACM, Hortal J, Gaffuri J, Temperley WH, Skøien JO, Mayaux P. 2016. Mapping ignorance: 300 years of collecting flowering plants in Africa. Global Ecology and Biogeography. 25:1085-1096.

Stuart YE, Campbell TS, Hohenlohe PA, Reynolds RG, Revell LJ, Losos JB. 2014. Rapid evolution of a native species following invasion by a congener. Science. 346:463-466.

Stucky BJ, Walls R, Deck J, Denny EG, Bolmgren K, Guralnick R. 2017. An update on the plant phenology ontology and plant phenology data integration. Proceedings of the Taxonomic Database Working Group (TDWG). 1:e20487

Tansey CJ, Hadfield JD, Phillimore AB. 2017. Estimating the ability of plants to plastically track temperature-mediated shifts in the spring phenological optimum. Global Change Biology. 23:3321-3334.

Tennant JP, Mannion PD, Upchurch P, Sutton MD, Price GD. 2017. Biotic and environmental dynamics through the Late Jurassic-Early Cretaceous transition: evidence for protracted faunal and ecological turnover. Biological Reviews. 92:776-814.

Thomson JD. 2010. Flowering phenology, fruiting success and progressive deterioration of pollination in an early-flowering geophyte. Philosophical Transactions of the Royal Society B. 365:3187-3199.

Tobler M, Honorio E, Janovec J, Reynel C. 2007. Implications of collection patterns of botanical specimens on their usefulness for conservation planning: an example of two neotropical plant families (Moraceae and Myristicaceae) in Peru. Biodiversity and Conservation. 16:659-677.

Toftegaard T, Posledovich D, Navarro-Cano JA, Wiklund C, Gotthard K, Ehrlén J. 2016. Variation in plant thermal reaction norms along a latitudinal gradient – more than adaptation to season length. Oikos 125:622-628.

Tooke F, Battey NH. 2010. Temperate flowering phenology. Journal of Experimental Botany. 61(11):2853-2862.

Torres-Dowdall JT, Handeelsman CA, Reznick DN, Ghalambor CK. 2012. Local adaptation and the evolution of phenotypic plasticity in Trinidadian guppies (*Poecilia reticulata*). Evolution. 66:3432-3443.

Tryjanowski P, Panek M, Sparks T. 2006. Phenological response of plants to temperature varies at the same latitude: case study of dog violet and horse chestnut in England and Poland. Climate Research. 32:89-93.

United States Fish and Wildlife Service (USFWS). 2012. The cost of invasive species. Accessed at <https://www.fws.gov/verobeach/PythonPDF/CostofInvasivesFactSheet.pdf>

University of Alabama Department of Geography. *General Physiography of Alabama*. Accessed at <http://alabamamaps.ua.edu/contemporarymaps/alabama/physical/basemap6.pdf>

USDA, NRCS. The PLANTS Database. National Plant Data Team, Greensboro, NC, USA. http://plants.usda.gov. Accessed 25 January 2018

Wallace RD, Bargeron CT. 2014. Identifying invasive species in real time: Early detection and distribution mapping system (EDDMapS) and other mapping tools. Invasive Species and Global Climate Change. 4:219-231.

Walther G, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, Fromentin J, Hoegh-Guldberg O, Bairlein F. 2002. Ecological responses to recent climate change. Nature. 416:389-395.

Wang Z, Dong H, Kelly M, Macklin JA, Morris PJ, Morris RA. 2009. Filtered-Push: A map-reduce platform for collaborative taxonomic data management. Pp. 731–735 in: Burgin M, Chowdhury MH, Ham CH, Ludwig S, Su W, Yenduri S (eds.) (2009) WRI World Congress Computer Science and Information Engineering, vol. 3. Los Alamitos, California: IEEE Computer Society. http://dx.doi.org/10.1109/CSIE.2009.948

Weins JJ, Graham CH. 2005. Niche conservatism: Integrating evolution, ecology, and conservation biology. Annual Review of Ecology, Evolution, and Systematics. 36:519-539.

Willis CG, Ellwood ER, Primack RB, Davis CC, Pearson KD, Gallinat AS, Yost JM, Nelson G, Mazer SJ, Rossington NL, Sparks TH, Soltis PS. 2017. Old plants, new tricks: phenological research using herbarium specimens. Trends in Ecology and Evolution. 32:531-546.

Wunderlin RP, Hansen BF. 2011. Guide to the Vascular Plants of Florida, Third Edition. University Press of Florida: Gainesville.

Wunderlin RP, Hansen BF, Franck AR, Essig FB. 2017. Atlas of Florida Plants. http://florida.plantatlas.usf.edu/. Accessed 14 September 2017

Yost JM, Sweeney PW, Gilbert E, Gilson G, Guralnick R, Gallinat AS, Ellwood ER, Rossington N, Willis CG, Blum SD, Walls RL, Haston EM, Denslow MW, Zohner CM, Morris AB, Stucky BJ, Carter JR, Baxter DG, Bolmgren K, Denny EG, Dean E, Pearson KD, Davis CC, Mishler BD, Soltis PS, Mazer SJ. 2018. Digitization protocol for scoring reproductive phenology from herbarium specimens of seed plants. Applications in Plant Sciences. 6(2):e1022.

Zalamea P, Munoz F, Stevenson PR, Paine CET, Sarmiento C, Sabatier D, Heuret P. 2011. Continental-scale patterns in *Cecropia* reproductive phenology: evidence from herbarium specimens. Proceedings: Biological Sciences. 278(1717):2437-2445.

Zuur AF, Ieno EN, Elphick CS. 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution. 1:3-14.

BIOGRAPHICAL SKETCH

EDUCATION

Florida State University (FSU), College of Arts and Sciences, Tallahassee, FL. 2015 – July 2018 (projected).

- M.S. Biology, Ecology and Evolution, Department of Biological Sciences; GPA: 3.94
- Preparing Future Faculty Certificate

University of Portland College of Arts and Sciences, Portland, OR. 2011 – 2015.

- B.S. Biology, minor in chemistry; GPA: 4.0; Valedictorian
- B.A. German studies

PROFESSIONAL EXPERIENCE

Graduate herbarium curator. Florida State University R. K. Godfrey Herbarium, Tallahassee, FL. August 2016 – August 2017, January 2018 – present.

iDigBio graduate research assistant. Florida State University, Tallahassee, FL. August 2015 – May 2016.

Restoration ecology researcher. University of Portland, OR. April 2014 – October 2014.

Student researcher. REU program at Portland State University, OR. May 2012 – July 2012.

Program coordinator. Office of Studies Abroad, University of Portland, OR. August 2011 – May 2015.

RESEARCH

Grants and Awards

- Janie Leonard Bryan Native Plant Conference Endowment Award. July 2017.
- Botanical Society of America Graduate Student Research Award. May 2016.
- American Society of Plant Taxonomists Graduate Student Research Grant. April 2016.
- National Science Foundation Graduate Research Fellowship, Honorable Mention. April 2016.
- Robert K. Godfrey Endowment for Botany. November 2015.
- Florida State University Legacy Fellowship. February 2015.
- Marlene Moore Award for Academic Achievement in Biology. May 2015.
- Barry M. Goldwater Scholarship in Excellence and Education, Honorable Mention. March 2015.

Publications

Williams J [Undergraduate], **Pearson KD.** (Submitted). Examining collection biases across different taxonomic groups: Understanding how biases can compare across herbarium datasets. *American Journal of Undergraduate Research.*

Pearson KD. (Accepted). Rapid enhancement of biodiversity occurrence records using unconventional specimen data. *Biodiversity and Conservation.*

Pearson KD, Mast AR. (Submitted). On the front lines of discovering change: Biodiversity specimen collectors as the Anthropocene's unrecognized outlier detectors. *PLOS ONE.*

Yost JM, Sweeney PW, Gilbert E, Nelson G, Guralnick R, Gallinat S, Ellwood ER, Rossington N, Willis CG, Blum SD, Walls RL, Haston EM, Denslow MW, Zohner CM, Morris AB, Stucky BJ, Carter JR, Baxter DG, Bolmgren K, Denny EG, Dean E, **Pearson KD**, Davis CC, Mischler BD, Soltis BS, Mazer SJ. 2018. Digitization protocol for scoring reproductive phenology from herbarium specimens of seed plants. *Applications in Plant Sciences.* 6(2):e1022.

Ellwood ER, Kimberly P, Guralnick R, Flemons P, Love K, Ellis S, Allen JM, Best JH, Carter R, Chagnoux S, Costello R, Denslow MW, Dunckel BA, Ferriter MM, Gilbert EE, Goforth C, Groom Q, Krimmel ER, LaFrance R, Martinec JL, Miller AN, Minnaert-Grote J, Nash T, Oboyski P, Paul DL, **Pearson KD**, Petcheff ND, Roberts MA, Seltzer CE, Soltis PS, Stephens R, Sweeney PW, von Konrat M, Wall A, Wetzer R, Zimmerman C, Mast AR. 2018. Worldwide Engagement for Digitizing Biocollections (WeDigBio)—The Biocollections Community's Citizen Science Space on the Calendar. *BioScience.* bix143.

Willis CW, Ellwood ER, Primack RB, Davis CC, **Pearson KD**, Gallinat AS, Yost JM, Nelson G, Mazer SJ, Rossington NL, Sparks TH, Solits PS. 2017. Old plants, new tricks: phenological research using herbarium specimens. *Trends in Ecology and Evolution*. 32(7):531-546.

Stanley KD, Taylor DW. 2015. Effect of Manual Ivy Removal on Seedling Recruitment in Forest Park, Portland, OR. *American Journal of Undergraduate Research.* 12(4):31-41.

Presentations

*** = invited talk**

***Pearson KD**. Estimating peak flowering from specimen data calibrated by phenological observations. *Symposium: Tools, Standards, Techniques, and Methods for Using Herbarium Specimens in Phenological Research, Botany 2018*, Rochester, MN. 2018. (Accepted).

Pearson KD. Rapid enhancement of biodiversity occurrence records using unconventional specimen data. *Digital Data in Biodiversity Research*, Berkeley, CA. 2018. (Accepted).

*Nelson G, **Pearson KD**, Riccardi G. The impact of digital data on raising the profiles of natural history museums and ensuring that museum-based researchers remain at the forefront of science. *International Committee for Museums and Collections of Natural History Conference*, Pittsburgh, PA. October 26, 2017.

Pearson KD. WeDigFlowering: Engaging citizen scientists to explore plant phenology in a biodiversity hotspot. [poster] *Cullowhee Native Plant Conference*, Cullowhee, NC. July 20, 2017.

Pearson KD. Rapid enhancement of biodiversity occurrence records using unconventional herbarium specimen data. *Botany 2017*, Fort Worth, TX. June 27, 2017.

Pearson KD, Mast AR. On the front lines of discovering change: Biodiversity specimen collectors as the Anthropocene's unrecognized outlier detectors. [poster] *Society for the Preservation of Natural History Collections*, Denver, CO. June 22, 2017.

Pearson KD. Hole-y Plant Databases! Understanding and Preventing Biases in Botanical Big Data. *Digital Data in Biodiversity Research Conference*, Ann Arbor, MI. June 6, 2017.

Funaro M, Williams J, **Pearson KD**. Identifying collecting biases in biodiversity specimen databases. [poster] *Florida State University Undergraduate Research Symposium*, Tallahassee, FL. March 28, 2017.

Pearson KD. Ellis S, Ellwood ER, Nelson G, Paul D, Riccardi G, Mast AR. On the front lines of discovering change: Biodiversity specimen collectors as the Anthropocene's outlier detectors. *Botany 2016*, Savannah, GA. August 2, 2016.

***Pearson KD**, Ellis S, Ellwood ER, Nelson G, Paul D, Riccardi G, Mast AR. On the front lines of discovering change: Biodiversity specimen collectors as the Anthropocene's outlier detectors. *Ecological Society of America Annual Meeting*. Ft. Lauderdale, FL. August 10, 2016.

Stanley [Pearson] KD. Big Data for Big Problems: Using digitized biological collections to examine the effects of climate change on plant phenology. *Florida State University Fellows Society Research Sharing Luncheon*, Tallahassee, FL. November 12, 2015.

Stanley [Pearson] KD. Effect of Manual Ivy Removal on Seedling Recruitment in Forest Park, Portland, OR. [poster] *National Conferences for Undergraduate Research*, Cheney, WA. April 16-18, 2015.

Bauer H, **Stanley [Pearson] KD**. Monitoring the Ecological Recovery of the University of Portland River Campus Riparian Zone: Baseline Surveys Conducted Summer 2014 and Direction for the Future. [poster] *University of Portland Founder's* Day, Portland, OR. April 2015.

Bauer H, **Stanley [Pearson] KD**. Monitoring the Ecological Recovery of the University of Portland River Campus Riparian Zone, Portland, OR. [poster] University of Portland Summer Research Symposium. Portland, OR. November 2014.

Stanley [Pearson] KD. Relationship of Abdominal Length to Fecundity, Wolbachia Infection, and Location of a Threatened Coastal Butterfly. *Portland State University Research Experience for Undergraduates Symposium*, Portland, OR. August 2012.

Invited Workshops

- Worldwide Engagement for the Digitization of Biocollections (WeDigBio) Planning Workshop, Gainesville, FL. April 20 – 22, 2016.
- Developing Standards for Coding Phenological Data from Herbarium Specimens, Berkeley, CA. March 12 – 13, 2016.
- Plant Phenology Ontology Workshop, Ft. Collins, CO. January 12 – 16, 2016.
- Using Biodiversity Specimen-Based Data to Study Global Change, St. Louis, MO. December 2 – 3, 2015.

SERVICE

- **President.** Society of Herbarium Curators Early Career Section. May 2018 – present.
- **Co-editor.** Phenology special edition of *Applications in Plant Sciences*. February 2018 – present.
- **Outreach Committee member.** Ecology and Evolution Reading Discussion Group at FSU. March 2018 – present.
- **Committee member.** Society of Herbarium Curators Membership Committee. 2017 – present.
- **Organizer.** Field-to-Collections Bioblitz: Creating Lasting Record of Biodiversity. Ecological Society of America Annual Meeting, Portland, OR. August 4, 2017.
- **Co-founder.** Early Career Section, Society of Herbarium Curators. 2017.
- **Treasurer.** Ecology and Evolution Reading Discussion Group at FSU. August 2017 – 2018.
- **Founder and president.** Plant Club at FSU. December 2015 – 2018.
- **Event coordinator.** "Seek and Destroy" Invasive Plant Removal Events. 2017 – 2018.
- **Co-organizer.** Worldwide Engagement for Digitizing Biocollections Event. October 2015, 2016, 2017.
- **Science Fair Judge.** Trinity Catholic School. December 7, 2016
- **Outreach table presenter** for FSU's Robert K. Godfrey Herbarium. St. Marks National Wildlife Refuge Monarch Butterfly Festival. October 22, 2016.
- **Presenter.** Cornerstone Learning Community BioBlitz. September 23, 2016.
- **Coordinator and presenter**: Citizen Science in the Classroom: Discovering Biodiversity by Transcribing Specimen Labels. Tallahassee Community College. October 18, 2016
- **Presenter.** Florida State University School STEAM (Science, Technology, Engineering, Art, and Mathematics) Day. February 19, 2016.
- Presenter: **Augmented Reality for Biological Collections**. November 14, 2015
- **Visitor Center Volunteer. Hoyt Arboretum and Herbarium. May 2014 – May 2015.**

TEACHING AND MENTORING

- **Curriculum developer: Plants and Society Lab**. Florida State University. Spring 2018.
- **Directed Independent Study Mentor.** Florida State University. 2017 – 2018.
- **Teaching assistant: Field Botany (BOT3143).** Florida State University. August 2017 – December 2017.
- **Undergraduate Research Opportunity Program Mentor.** Florida State University. 2016 – 2017.
- **Organizer: Citizen Science in the Classroom.** Tallahassee Community College. October 18, 2016.
- **Curriculum developer** in association with CPALMS (State of Florida educational resource). 2016.
- **Lesson lead, curriculum volunteer, and assistant:** Student-led Afterschool Science Program. Shanks Middle School. September 2015 – September 2016.