PROJECT SUMMARY -- ***Digitization TCN: Mobilizing collection data to examine how environmental changes affect insect life history parameters and pollination*****.**

Overview

We propose the *Pollinator*+ TCN to digitize 6 million specimens collected over the past century for assessing and predicting the response of insects to climate change, land use, pollution, and other environmental factors. Our team is a consortium of researchers at 38 collections from across the United States. We will photograph specimens and their labels, and generate a high-quality, verifiable occurrence record for each specimen using a rapid new workflow. Our target taxa are an exemplar array of over 200 bee, butterfly, moth, fly, and beetle species with different biologies and interactions of varying specificity. We have selected them as an efficient means to better understand pollination, insect life history parameters, and changes in geographic distribution. *Pollinator*+ will expand upon our existing projects, active data capture pipelines, and public natural history data repositories, including Discover Life's citizen science *Mothing* project, which has fine-grained contemporary data on 3,000 species. We will make all images and associated data publicly available. Our team possess the necessary taxonomic expertise, access to collections, leadership, state-of-the-art technology, and experience to ensure project success.

We seek three years of funding to **1)** digitally image specimens and their labels for our target species, **2)** make images and associated data available through Darwin Core standards to web portals, including Discover Life, iDigBio, and GBIF, **3)** present georeferenced data with maps, phenological graphs, and tables so that the broader scientific and educational communities have a geographic and historic resource with which to analyze and visualize data for understanding the impact of environmental factors on biodiversity, **4)** generate sufficiently large sample sizes for comparative statistical analyses and predictive ecological and evolutionary modeling of insects, plants, and their interactions, **5)** train students in biodiversity science and informatics, and **6)** involve the general public through Discover Life's outreach.

Intellectual Merit

Pollinators, nearly 80% of which are insects, serve a central role in terrestrial ecosystems. Pollinator declines have been reported on nearly every continent, but many of these reports have not properly documented such changes, in part due to improper and incomplete data collection or because they have focused solely on particular species or a local region (e. g., city). Consequently, public policy and expenditures concerning pollinators are currently influenced by unreliable and/or incomplete studies. Broad synthetic studies of pollinators, with wide taxonomic and geographic coverage, are much needed.

Biological collections are a century-long record of species distribution, phenology, and ecology. The value of museum specimens within biotic, temporal, and geographic contexts motivates their digitization. Data harvests from historically relevant collections, individual taxon-focused projects, and databases integrated across multiple collections have demonstrated their utility in studies of climate change, habitat destruction, phenological change, and conservation prioritization. Here we will mobilize specimen data needed to determine how climate change, urbanization, and other factors affect exemplar insects.

Broader Impacts

This proposal digitizes insect specimens and shares data for studying climate change, pollination, phenology, conservation, systematics, and biogeography. We will teach undergraduate courses at three universities that highlight the importance of natural history collections, preservation methods, digitization workflows, and biodiversity informatics. At participating collections we will give hands-on workshops that train students in specimen handling, digitization, data extraction, and quality control. Students will learn how to capture data and understand issues surrounding their management. We will recruit students from local community colleges to provide opportunities for underrepresented groups in collections-based research. By providing images, data, and tools online, we will enable the public to contribute to our endeavor and explore how environmental factors affect species. Discover Life (discoverlife.org), an established web portal for public, educational, and professional access to biodiversity information, will also use the images and data to build and serve on-line identification systems.

PROJECT DESCRIPTION

Response to Previous Reviewers

Here we detail the significant improvements that we made to address the four main criticisms of our previous submission: **1)** Participating collections -- We have improved our selection of collections. *Pollinator*+ is now focused entirely on collecting much needed insect data. It leaves plant digitization to other TCNs. We have also eliminated previous plans to digitize some federal and private collections. Nevertheless, our regional support hubs, networked collections, and improved workflow strengthen our geographic coverage and yield much greater sample sizes for our exemplar taxa than before. To put our data in a larger perspective for hypothesis testing, we will mine iDigBio, GBIF, and other public repositories and integrate as much insect and plant data with ours as possible. **2)** Organizational structure -- We have streamlined our organization, reducing our previous 13 collaborative proposals to this single proposal from the Polistes Foundation, which serves as our legal and fiduciary umbrella with an indirect cost of 15% for distributing resources to our 38 participating collections. We can now more efficiently manage the project and ensure that more resources go to students and data capture. Our network is anchored on 5 regional/taxonomic support centers run by our PIs. These will provide training, computer services, and technical support to our highly distributed collections. **3)** Exemplar taxa -- Reviewers questioned how we selected our target taxa and how we will properly identify the specimens that we digitize. Our exemplar list now includes only insects, making concerns about our prior selection of plant species moot. Ecologists and systematic experts chose our current exemplar taxa to answer scientific questions regarding the impact of environment factors on insects, particularly pollinators. Our taxa are among the lowest-hanging fruit in insect collections. They have a broad array of biological characteristics and include species from the four major insect orders. They are all relatively easily to identify and are well represented in our participants' collections. And **4)** Crowdsourcing -- We used this term loosely before, raising in reviewers' minds the issue of poor quality-control that is too often associated with public participation in data collection. Here we more fully defined our use of the term 'crowdsourcing' as entering and correcting information with Discover Life's *Timemachine*. This software enables large numbers of participants to view images of specimens and their labels and enhance associated data records. Our 'crowd' is an established network of specialists, students, and technical support staff that we train and often fund. It is not untrained web users. Our 'crowd' will grow as our regional centers add capacity and our outreach and educational activities engage more people in high-quality, specimen-based museum science.

Introduction

Pollinators serve a central role in ecosystem stability, and nearly 80% of pollinators are insects. Pollinator declines have been reported on nearly every continent, but many of these reports have not properly documented such changes, in part due to improper and incomplete data collection (Potts et al., 2010) or because they have focused solely on particular species or a local region (e.g., cities, Biesmeijer et al., 2006; Burkle et al., 2013; LeBuhn et al., 2013). Consequently, impactful changes in public policy and expenditures are influenced by unreliable and/or incomplete studies (Bartomeus, 2013). Broad synthetic studies of pollinator-plant interactions, with wide taxonomic and geographic coverage, are much needed.

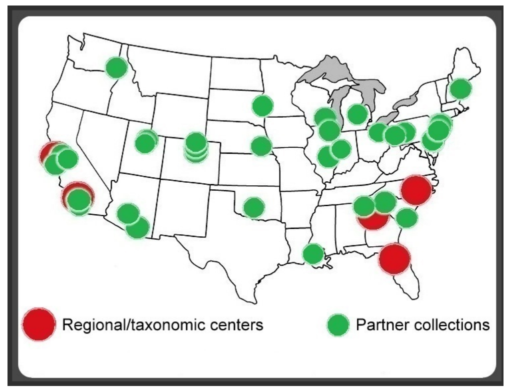
While honeybees are frequent subjects of pollination research, native species of insect pollinators warrant special interest. Native pollinators play a key role in the pollination process (Aebi et al., 2012) and are known to enhance pollination regardless of honeybee abundance (Garibaldi et al., 2013). Members of our collaborative team have conducted numerous assessments of environmental changes on insects and their associated plants within a single locality (e.g., Morris et al., 2010; Basset et al., 2012; Colla et al., 2012; Bartomeus, 2013; Bartomeus et al., 2013) and developed necessary pipelines for data capture, integration, and analysis. A major goal is to supplement the current understanding of invasive species' dynamics with an understanding of how environmental factors influence native pollinators on a continental scale.

Biological collections amassed over the past several centuries are widely considered the best and only available source of information on species, distribution, and abundance (Berendsohn and Seltmann, 2010;

Chavan et al., 2010; Scoble and Bourgoin, 2010; Vollmar et al., 2010; Colla et al., 2012). The value of specimen data within biotic, temporal, and geographic contexts motivates the rapid expansion of collection digitization (Baird, 2010; Beach et al., 2010; Beaman and Cellinese, 2012). Data harvests from historically relevant collections, individual taxon-focused projects, and databases that emerge from coordinated networks have demonstrated their utility in studies of climate change (Johnson et al., 2011), habitat destruction (Peterson et al., 2002), phenological change (Robbirt et al., 2011), and conservation prioritization (Kress et al., 1998). Although many large-scale digitization initiatives are focused on a single taxonomic group or a particular subregional locality, our approach is to target an ecological interaction – pollination – of high priority to agriculture and the health of many ecosystems (Berenbaum et al., 2007), for multiple taxa across the United States. We will mobilize collection data needed to determine how insect pollinators are affected by environmental factors such as climate change.

Objectives

We propose the *Pollinator*+ Thematic Collections Network (TCN) -- a national consortium of researchers at 38 participating collections (see map, plus one in Alaska). Collectively we possess the necessary taxonomic expertise, specimens, leadership, and technology to carry out this project. *Pollinator*+ will expand upon our consortium's existing projects, active data capture pipelines, and public natural history data repositories to make images and extracted georeferenced data available for 6 million insect specimens. We will process **1)** over 200 exemplar insect species selected to better understand how large-scale environmental factors affect pollinators and their life history parameters, **2)** other species collected at intensively-studied sites across multiple years, such as series of bees at NCSU collected by T. B. Mitchell, **3)** synoptic images of all well-determined moth species to support building web-based identification guides to moths world-wide, continuing what Moth Photographers' Group and Discover Life have already done for 12,000 North American moth species, and **4)** the added value of other relevant taxa prioritized for digitization by participating collections.



With prior NSF support, we have already built a successful collaborative network to digitize, mobilize, and analyze museum specimens. We already use this network to process millions of specimens, including 70,626 specimens of our exemplar taxa (~12% of targets), though most specimens of exemplar taxa are not yet imaged or the locality data are not georeferenced. For these partially complete taxa, we will add digital images and georeferenced data to complete specimen records. We will make full use of the data already accumulated by us, by our partners, and the public data on GBIF by synthesizing it with new data. Our effort will capture all label data and associated metadata, such as taxonomic information, phenological data, and locality, from public, non-Federal, collections spanning the continental US.

Table 1. *Pollinator*+ PI's, co-PI's, and senior personnel (Individual -- Institution -- primary responsibility)
Wiegmann, Brian -- North Carolina State University (NCSU) -- PI, Diptera, regional coordination
Daniels, Jaret -- Florida Museum of Natural History -- co-PI, Lepidoptera
Gillespie, Rosie -- University of California, Berkeley -- senior personnel, Education
Kawahara, Akito -- Florida Museum of Natural History -- senior personnel, Lepidoptera
Pickering, John -- Discover Life -- co-Pi, project & data manager
Roderick, George -- UC Berkeley Natural History Museums -- co-PI, regional coordination
Yanega, Doug -- University of California, Riverside (UCR) -- co-PI, Hymenoptera

Members of our team have demonstrated that long-term, geographically extensive data from collections can reveal important trends that typical short-term, geographically limited ecological studies cannot detect (e. g., Bartomeus et al., 2013; Bartomeus and Winfree, 2013). Although there have been numerous large-scale taxon or region-focused digitization efforts, our proposal is aimed at insects at a continental scale. Large sample sizes are necessary to analyze statistically geographic, taxonomic, and temporal trends of insect dynamics, and only museum data can provide us with historical information that dates back over a century. Our team of specialists (Table 1) has chosen our exemplar taxa from four insect orders that represent a broad spectrum of pollinators and life history parameters: Hymenoptera (bees, wasps), Lepidoptera (butterflies, moths), Coleoptera (beetles), and Diptera (flies). These are the most diverse orders in the animal kingdom, and they are well represented in U.S. collections.

TCN Theme

Environmental Change

It is widely accepted by the scientific community that the Earth's climate will change over the next century, with temperatures rising quickly (IPCC, 2013). This change will impact the morphology (e. g., body size), phenology, distribution, behavior, and survival of insects and their associated plants (e. g., Cleland et al., 2007; Lawler et al., 2009; Loarie et al., 2009). These impacts will have ecological ramifications, altering natural landscapes and agricultural, horticultural, and landscape management activities (Grimm et al., 2008). Phenological observations play an important role in understanding climate-based changes on native plants and animals (Amano et al., 2010; Wolkovich and Cleland, 2010). There is evidence that plants are now flowering earlier in the spring in response to warmer temperatures (Panchen et al., 2012; Ellwood et al., 2013) and that insects may be following this trend (Diamond et al., 2011; Bartomeus et al., 2013).

Insects are excellent candidates to study the effects of climate change and other driving environmental factors because historical data on them have been recorded for over a century. Plants often have short flowering times and the associated specialist pollinating insects have activity times tied closely to their plants (Dennis, 2012). While herbarium specimens can be used to document the time of plant flowering, historical data from insect specimens can be used to assess the time when particular pollinators are active as adults. Museum specimens also provide data for assessment of variation in voltinism, size, color, shape, and other valuable life history parameters that are potentially impacted by climate change. We will measure specimens so that their sizes can be easily used in future research projects.

Pollination

Pollinating insects are essential in shaping terrestrial ecosystems, communities, and biodiversity (Marks, 2005; Tylianakis, 2013). They are also intrinsically connected to many natural products consumed by humans; in 2009, the economic value of pollinators reached $11.68 billion in the United States alone (Calderone, 2012). Plant-pollinator associations have been a major component of terrestrial life on Earth for over 100 million years (Grimaldi and Engel, 2005); they enable reproduction of the world's nearly 300,000 or more flowering plant species (Kevan and Viana, 2003; National Research Council, 2007). Mutualism is common between specialist pollinators and their plants, so a pollinator decline can result in the decline of the pollinators' host plants or vice versa (Potts et al., 2010).

Out of all the pollinating insects, bees are often the most essential in both agricultural and natural ecosystems worldwide (Losey and Vaughan, 2006). They are a model group for documenting plant-insect interactions and are impressively diverse in their own right, with 20,000 valid described species (Ascher et al., 2008; Ascher and Pickering, 2013) and thousands more awaiting discovery and description. Although bees are the predominant insect pollinators in many habitats, other flower-visiting taxa such as flies, moths, butterflies, and beetles contribute to pollination, as either specialists or generalists (Winfree et al., 2011; Weiner et al., 2013), and play an important role in ecosystem stability, especially in fragmented habitats (Olesen et al., 2002; Potts et al., 2010).

Recent reports of global pollinator declines have generated considerable public and scientific interest (Potts et al., 2010; Burkle et al., 2013; Ghazoul, 2013), but some of the reports are unreliable, and the public interest is consequently misdirected. *Pollinator*+ will mobilize data for pollinators with an array of different interactions of varying specificity. These data will provide a foundation for future studies that test our current hypotheses on the impacts of climate change and other large-scale factors on pollination, life history strategies, species interactions, geographic ranges, and population dynamics. Many entomologists, botanists, and ecologists find themselves unable to implement certain study designs due to the absence of the appropriate data network (Winfree et al., 2011; Jauker et al., 2012; Willmer, 2012). For these researchers, *Pollinator*+ will be the final piece of the puzzle, enabling them to produce valid interpretations of pollinator data, so that the public's interest can be converted to true understanding.

Intersection of climate change and pollination

Plants and their associated insects are increasingly threatened by the impact of humans on the environment, especially those associated with unique habitats. The vulnerability of native pollinators is expected to increase in response to temperature changes and geographic shifts of their hosts (Memmott et al., 2007; Potts et al., 2010). Ecosystems cannot be managed effectively and biodiversity conserved, if the individual and collective responses of flora and associated insect fauna to environmental changes are poorly understood. Digital projects that prioritize relevant biological phenomena to humankind, such as pollination, are urgently needed to understand the potential impact of climate change, urbanization, and habitat loss on geographic range, phenology, abundance, and community diversity of plants and insects (Johnson et al., 2011; Robbirt et al., 2011; Bartomeus, 2013). While plants are now flowering earlier in the spring in response to earlier temperatures (Panchen et al., 2012; Ellwood et al., 2013), and insects may be following the trend (e.g., Diamond et al., 2011), studies of insects are scarce. Further, few of these test the effect of climate change on interactions between plants and insects (see Biesmeijer et al., 2006 for exceptions).

Studies on plant-pollinator interactions with respect to climate change typically focus either on a single site or small locality (e.g., Luedeling et al., 2011; Walter et al., 2012) or on a single low-level taxon (e.g., Netherer and Schopf, 2010; Barbraud et al., 2011; Primack and Miller-Rushing, 2012). Few have examined the impacts of climate change across a guild of closely interacting species. This project will collect data for specimens of adult pollinating insects across the continental US, including phenological data, and link them to existing data on the flowering times of their known plant hosts. We will also measure additional variables, such as insect body size. Although insect body proportions can shift postmortem, preliminary digitization efforts demonstrate that reliable body size data are obtainable from museum specimens (Johnson et al., 2013). We have already made preliminary use of such metadata by elucidating phenological and size trends that closely correlate with the activity time data obtained from associated label data, as well as correlations between phenology and wing length, which can vary across years and generations (Pickering, unpubl.).

Data from collections have demonstrated great promise to understand climatic effects on distributions of organisms (Newbold, 2010; Robbirt et al., 2011). They can play a major role in untangling the link between climate change and pollination and testing our current predictions regarding synergistic effects (Cameron et al., 2011; Bartomeus, 2013; Bartomeus et al., 2013). These data are not limited to information found on a specimen label; evidence of plant-insect interactions is often captured directly in the museum specimens. For instance, pollen is frequently found on the head (many bees, flies, moths) and legs or abdomen (bees) of specimens. Moreover, highly specialized pollination interactions, such as those between yucca moths and yucca plant, are reliably inferred from herbarium samples: Yucca fruits on a herbarium sheet are direct evidence for the prior presence of the yucca moth. Images of herbarium sheets have been used to locate and identify larvae trapped within the plant specimen (Lees et al., 2011). By capturing images of insect specimens, we can examine these indicators of plant-pollinator interactions in a historical context and correlate the observed patterns, when appropriate, to historical climate data.

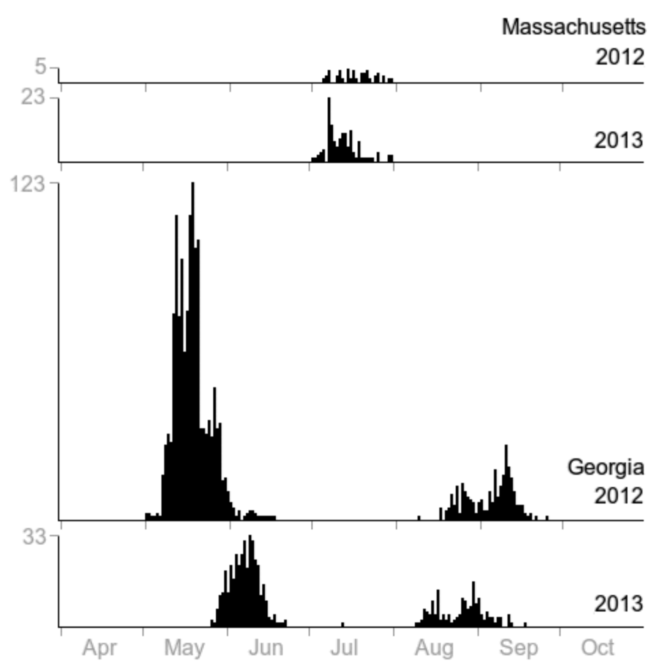
Hypotheses

This project will mobilize pollinator collection data that will allow for hypothesis testing on the historical distributions of pollinators and their hosts and the generation of model-based predictions on their future ranges. We will not test these hypotheses as part of the ADBC program. We outline them here to demonstrate the research potential of our data far beyond *Pollinator+*.

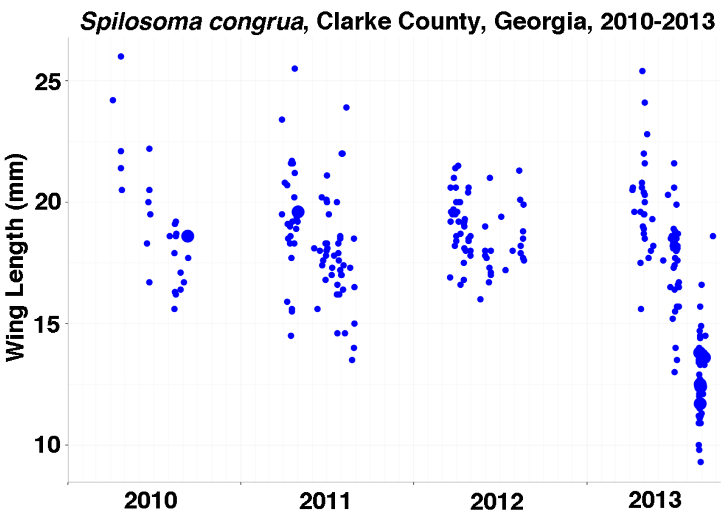
- Bartomeus et al. (2013) examined long-term bee distributions using museum collection data from the Eastern US. They discovered previously unknown declines in species diversity and large shifts in community composition. They hypothesized that certain traits are associated with a declining relative abundance including small dietary and phenological breadth and large body size. Our data, captured on a continental scale, can be used broadly to test whether similar processes are taking place at the continental scale and in non-hymenopteran taxa.
- Burkle et al. (2013) examined historical records of plant-bee interactions in western Illinois. They discovered that many interactions had ceased over the last two centuries and were able to correlate specific lost interactions with either local extinctions or phenological shifts of particular plant and bee species. They hypothesized that similar studies, using more geographically and taxonomically expansive sources of plant-pollinator interaction data, could derive similar conclusions to assist with efficacious conservation efforts on a continental scale across multiple orders of insects.
- Biesmeijer et al. (2006) examined bee and syrphid fly assemblages at two sites in Europe and found declines in specialist bees but increases in syrphid abundance over several decades. They predicted that complex networks of interactions between different pollinator taxa exist and that the decline in one pollinator species might influence the ecology of another. Our mobilized data will help examine whether potential interactions have changed over time.

Public involvement in science -- Moths

A key goal of our TCN is to reach out to schools and the general public, particularly youths in underserved communities, and vest them in our science at all levels. Moths are an ideal taxon for this purpose. They are biologically diverse, ecologically important, and can be studied efficiently and safely by virtually anyone with a digital camera and web access. Youths through retirees can photograph moths attracted to lights, use online resources to document and help identify them, and most importantly, learn to formulate their own questions and test ideas by analyzing and visualizing their own and our collective data.



Discover Life's *Mothing* project (discoverlife.org/moth) will engage the public in our TCN and provide high-quality, fine-grained contemporary data to supplement data from collections. Since 2010, *Mothing* participants have uploaded 430,000 photographs from 20 study sites in the United States and Costa Rica, documenting nightly differences in the seasonal activity and abundance of 3,000 species across years and sites. For example, the adjacent figure compares the 2012 and 2013 phenology of one of our exemplar species, *Hypoprepia fucosa (*Painted Lichen Moth*)*, at sites in Massachusetts and Georgia. It shows the difference in voltinism between the sites and an earlier 1st flight in Georgia after 2012's exceptionally warm winter. We anticipate that museum data will reveal where this species switches between one and two flights per year and how its populations respond across years to different weather patterns.



Size is an important life history parameter that affects insect fecundity. *Mothing*'s results document how it can change between generations. This figure for *Spilosoma congrua* (Agreeable Tiger Moth) shows that the wing length in the 3rd flight in 2013 averaged 12.5mm, a decrease of 6mm from the average length in the previous 11 flights. By measuring the size of museum specimens collected from sites and years with different weather patterns, *Pollinator+* will enable investigators to determine whether climate change will affect insect size.

Mothing's educational objective is to involve the public in hypothesis generation, data collection, identification, analysis, and presentation of results. Discover Life is developing Moth Math to teach students how to analyze real-time data. Using 38,000 diagnostic photographs from the Moth Photographers' Group, Discover Life now provides online identification guides to 12,000 North American moth species. These guides are customized by U.S. state or by Canadian province or territory. After *Pollinator+* photographs synoptic specimens of other species, Discover Life will use the images to build similar guides for the moths of the New World and eventually elsewhere.

Institutional collections, regional coverage, integrating existing databases, and training

The institutions participating in *Pollinator+* possess valuable data for many global regions and contain extensive, precise, and reliable Nearctic distribution data for our exemplar taxa. We have recruited collections widely to alleviate regional biases and to supply enough data for statistical analysis of broad patterns on a continental scale. We will support these principal institutions to reach out to other partners in their region to increase the sample sizes, as shown in Fig 2. Collaborative letters from the principal insitutions are included and available at discoverlife.org/adbc/los .

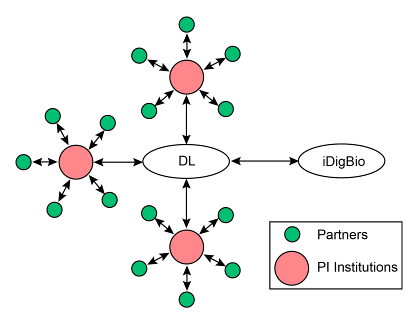


Fig. 2. Interaction between partners, PI institutions, DL, and iDigBio.

Some collections already have well-populated databases, such as the AMNH and UC Riverside bee databases, whereas others (e.g., FLMNH Lepidoptera collection) have few or no database records. Discover Life will provide *Pollinator+* with database services. Its staff will work with each participating collection to make sure that our data capture is compatible with local practices and will help integrate data extracted from images into their databases. Most notably, we will put unique identifiers on specimens that are the same as (or that are compatible with and will enhance) their existing labeling practices. We will also generate files with occurrence records in customized formats that collections can import easily into their local data management systems. Each night Discover Life's system will automatically update these files and the ones that we will use to export all data to iDigBio and GBIF.

In the past two years, Pickering and Discover Life's staff worked at 13 insect collections to improve our workflow. We have learned about infrastructural and logistical constraints and taken into consideration the concerns of the curators, collections managers, and database personnel about adopting new methods for managing and databasing their specimens. To help implement our methods, Pickering and/or Discover Life's staff will visit each collection, train local participants, and facilitate data collection and exchange. They will give extensive training to personnel at four regional/taxonomic centers to support digitization in the collections and oversee quality control. These four centers will be housed and run by our PIs at North Carolina State University, Raleigh; University of Florida, Gainesville; University of California, Berkeley, and University of California, Riverside. Discover Life will provide these centers and the collections that they support with technical support from the University of Georgia, Athens.

Taxonomic Coverage

We estimate that there are on the order of 200 to 500 million pinned insects in U.S. collections that have yet to be digitized. There are tens of millions in the collections participating in *Pollinator*+. As our starting point, our taxonomic experts have chosen 151 genera and 232 species of insects on the basis of previously documented pollination interactions (e.g., Motten et al., 1981; LaBerge, 1986; Pellmyr, 2003), distribution, abundance, ease of identification, and known life history parameters, such as changes in voltinism across latitude. These exemplar taxa include Coleoptera (beetles: 4 families, 10, genera, 10 species), Diptera (flies: 10 families, 36 genera, 49 species), Hymenoptera (bees and wasps: 8 families, 39 genera, 93 species), and Lepidoptera (moth and butterflies: 18 families, 66 genera, 80 species, including the focal species of Discover Life's *Mothing* project, for which we have extensive contemporary data, see www.discoverlife.org/moth/focal_species.html). In anticipation of finishing these taxa, our Steering Committee will choose additional ones based on broad input from participants at the proposed iDigBio workshop in 2016 and Entomological Collections Network meeting in November, 2017.

INFORMATICS

Insects and plants constitute an enormous diversity of species with different formats for documenting the provenance of specimens (Blagoderov et al., 2012). Fortunately, new technology and sampling strategies have greatly accelerated specimen digitization (Beaman and Cellinese, 2012; Nelson et al., 2012). Existing TCN's and other projects are making good progress at digitizing specimens in herbaria. We will rely on them for the plant data that we need to understand pollination. *Pollinator*+ will capture needed insect data.

Methodology and workflow

We will digitize pinned insects with Discover Life's tested new workflow and software that includes photographing specimens and their labels. Figure 3 illustrates this workflow, including our steps for **1)** staging, **2)** digitizing, **3)** image processing, **4)** data capture, **5)** management, and **6)** dissemination.

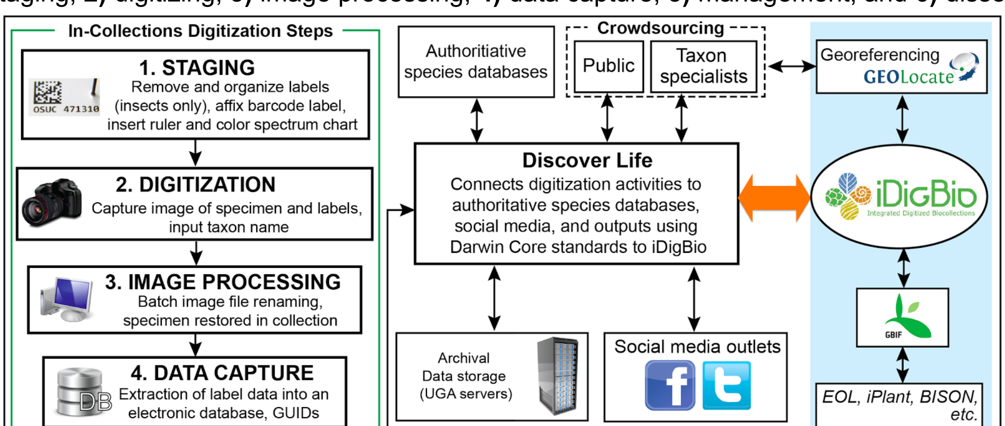


Fig. 3. Project workflow. The main digitization, data storage, and data transfer steps are illustrated.

Each night we will put images and associated data onto the web in ways that make them easily available to scientists, educators, policy makers, and the general public. For data aggregators, such as iDigBio and GBIF, we will include text files in Darwin Core archive format, a TDWG supported standard (Darwin Core Task

Group, 2009; Wieczorek et al., 2012). We will employ data cleaning and quality control steps at every stage to ensure high-quality results (Morris, 2005; Chapman, 2005a; Harpham, 2006). We will train staff and student at participating collections. In turn these individuals will transfer knowledge to additional students and volunteers, an approach that has been shown to be widely successful when conducting museum-based work (e.g., Munstermann and Gall, 2010). To provide regional support to participating collections, Pickering and Discover Life's staff will train a graduate student or staff member at each of the other 4 PI's institutions as regional/taxon coordinators. These coordinators will provide support, giving training, and oversee the quality of data from their assigned collections/taxa. Pickering and Discover Life's staff will support these coordinators and, in the case of tasks too complicated for them, work directly with the collections. Graduate students will select focal taxa for research projects that combine taxonomic revisionary study, biodiversity informatics, and ecological, evolutionary or biogeographic data collection.

We will customize our workflow at each museum collection to meet particular local needs (Morris, 2005). For example, some collections already have image and/or associated data that we will need to integrate into the process. At others, we will need to extract large amounts of information from field notebooks that are linked to specimens via lot numbers on their labels.

We will present our workflow at an iDigBio workshop at UF (see Year-1 timeline, below) to speed knowledge transfer between our collections and other TCNs and seek their input for improvements.

In-collections Digitization Steps

Step 1. Staging with unique labels

Using fine forceps we will carefully remove the labels from each pin and place them and the pinned specimen(s) on a stage before we photograph them. We will position each specimen so that its wings are at the same height as a millimeter ruler so that we can accurately measure wing length from the specimen image (see Morris et al., 2010; Nelson et al., 2012). If the existing labels of a specimen do not already include a label with an identifier that is unique within institution's collection, we add a globally unique label to the bottom of the series of labels so that the text image of the labels includes one.

In accordance with the wishes at each institution, these unique labels may or may not include a machine readable symbol (e. g., datamatrix symbol or Code-49 barcode). However, to encourage the adoption of unique, machine-readable labels, we will provide rolls of such pre-cut labels to all collections that agree to use them. Since 1992 Discover Life has used archival plastic polytag labels that it first purchased from Intermec Corporation and later printed in-house. Based on this experience, we highly recommend that collections use pre-cut labels that are printed by Intermec and do not succumb to the false economy and time-consuming process of printing labels in-house and cutting them up with scissors.

Within the constraints of existing practices, our unique labels will use either a Darwin Core triple with an institution code, collection code, and integer (e.g., FLMNH MGCL 1234567) or the older, shorter 'Abbreviations for Insect and Spider Collections of the World' maintained by the Bishop Museum (hbs.bishopmuseum.org/codens/codens-inst.html) (e. g., UGCA1234567).

After taking a specimen image and a text image we will replace the labels in their original order, facing the unique identifiers either up or down, in accordance with the wishes at the institution, and then place the specimen back in its original position in the drawer. While we will train participants to be as careful as possible, we are fully aware that we will damage some specimens and labels while handling them. We will follow procedures to mitigate such damage, for example, by putting abdomens that fall off into gelatin capsules that we will pin with the labels.

We will photograph color spectrum charts to standardize colors across images (Taylor, 2005).

Step 2. Imaging

For each of specimen that we will digitize, we will take a specimen image with a millimeter ruler and a text image of its labels, including the unique identifier. Depending on the taxon and how the specimen is mounted, we will take either a dorsal or lateral view of each specimen. For some taxa, such as the moth genus *Spilosoma* that has important identification characters on the underside, we will take more than one specimen image before photographing its labels. If labels have information on both sides, we will flip such labels and take a second text image, again including the specimen's unique identifier.

In addition to photographing specimens and labels, our workflow will require participants to photograph drawers, unit trays, determination labels, cell phones, color spectrum charts, and other pertinent factors. These control images help us to add taxonomic names to specimens, rapidly re-locate specimens, and track and improve the speed and quality of each participants' work.

Currently, we use Nikon D50 cameras with a AF Micro Nikkor 105mm lense to photograph pinned insects. This camera has a built-in flash and can be hand-held with the settings that we use. Because its flash burns out after about 40,000 actuations, we have budgeted to replace these cameras every 20,000 specimens. We will adopt newer camera technology when price and performance warrant it.

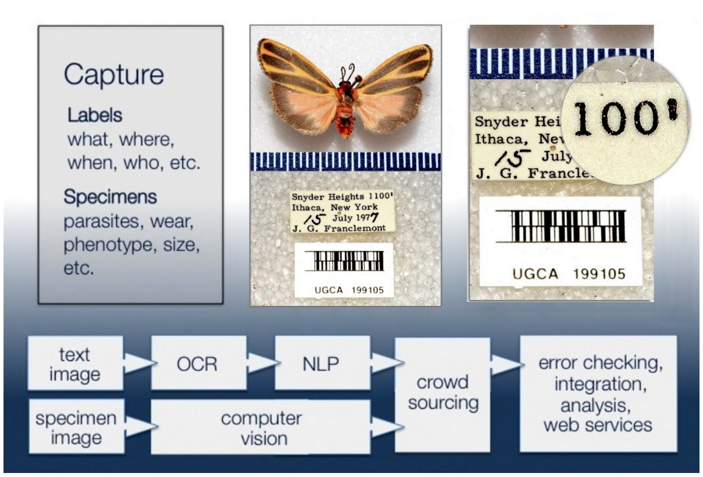
Our photographers will take high-resolution archival images in JPG format (Häuser et al., 2005; Nelson et al., 2012). Flemons and Berents (2012) have shown that such files are of sufficient quality for our purposes.

Step 3. Image processing, storage, and distribution

After each session photographers will transfer images from their camera to local computers and then start a batch upload using ssh/scp to Discover Life's servers for processing and storage. To handle the load rapidly, we will distribute processing across servers at the University of Georgia, North Carolina State University, University of Florida, University of California, Berkeley, and University of California, Riverside.

Once uploaded, our automated program called *web2db* assigns each JPG a unique identifier and stores the original (at maximum resolution) and 4 smaller resolutions (640, 320, 240, and 80 pixels across). At this stage the 5 associated files are named with just the unique identifier (e., g., I_UGCA1234) that is linked in its associated data record with the *_private_slide* filename assigned by the camera *(e. g., DSC_0123).* Once thus processed, photographers can view their images on the web in the sequence taken and rapidly associate data such as taxon name, photographers name, and institution name. After a taxon name is specified, *web2db* includes it in the image file names to aid Google and other search engines. Thus, www.discoverlife.org/IM/I_UCMC/0000/640/Spilosoma_virginica,I_UCMC49.jpg is image I_UCMC49 at resoluton 640; www.discoverlife.org/mp/20p?see=I_UCMC49&res=640 is a web service that displays this image with its associated data, and www.discoverlife.org/mp/20p?img=I_UCMC49&res=640 displays it without data.

Every night, Discover Life's automated program *mirror* assesses what files are on the network's servers and synchronises copying backup files across them. Currently, *mirror* manages 6 copies of 15 million files on Discover Life's servers in Georgia and Australia. We will use *mirror* to manage the 30 million jpg files and associated text and database files that *Pollinator+* will generate, making copies available through our servers at the five institutions listed above. This will ensure their protection from earthquakes, fires, and the like. We trust that our security procedures will protect them from malicious hackers.



Step 4. Data capture

We will structure label information and specimen size into an associated database record for each specimen. As diagrammed below, we will process each text image with **1)** Tesseract, a public domain OCR program, which Discover Life has automated successfully to capture much of the typed alphanumeric characters on labels, **2)** Discover Life's natural language processing (NLP) software that applies controlled vocabularies (gazetteers, taxonomic authority files, and collectors' names) to correct OCR errors and structure data into fields (what, where, when, who, etc.), **3)** Discover Life's *Timemachine*, a crowdsourcing program, which enables humans to enhance and correct records, capturing hand-written components of labels that Tesseract fails to read, and **4)** error checking by both automated programs and human review. We will use computer vision to detect the ruler in the specimen image and speed the measurement of specimen size.

Of particular concern during data capture is quality control. Through empirical observation at more than 10 insect collections, labels associated with specimens are prone to significant omissions of important data elements (approximately 20% of all legacy labels), and/or outright errors (approximately 5% of all legacy labels) (co-PI Yanega, unpublished). Spelling errors are the most common class of error, especially with scientific and locality names, but virtually all data elements are subject to other types of error: place names, dates, cardinal directions, distances, elevations, GPS readings, etc. As such, we will discourage participants from simply entering raw (i.e., potentially erroneous) label data. Our method enforces parsing data elements into standardized formatted fields, coupled with rigorous error-checking protocols. To reduce errors, we use drop-down menus and type-ahead features that have valid entries listed for a given field (e.g. all counties in a U.S. state) so that manual typing is not required.

Post-Digitization Processes

Authority files for taxon names

Typing errors often occur when individuals enter unfamiliar scientific names. We mitigate such errors by using authority files and controlled vocabularies during data entry. Discover Life uses taxon-specific authority files to ensure that specimens adhere to an updated nomenclature. These include Discover Life's nearly complete world catalog of 20,000 bees including synonyms and country occurrences. The valid names in this catalog were fully reviewed during compilation of the World Bee Checklist (Ascher et al., 2008). For other Hymenoptera we will use Hymenoptera Online (www.hol.osu.edu); for Diptera, Inouye (2013) and Pape & Thompson (2013); for Coleoptera, Arnett & Thomas (2000) and Arnett et al. (2010), and for Lepidoptera, the BAMONA checklist (Opler et al., 2013).

Data integration

Since inception in 1998, Discover Life has gained considerable experience importing, integrating, and exporting data between different databases, spreadsheets, and formats. It has gathered and integrated over 450 million occurrence records and 1.5 million images from over 1,000 sources, including GBIF. Pickering and Discover Life's staff will work with each collection, train participants to collect and manage data, and help them import data back into their local databases. Thus, we will process local data from many collections and then export them to iDigBio and GBIF.

Digital record identifiers

In a global data environment, it is essential to be able to identify unequivocally a piece of information, especially digital collection objects. While standard machine-readable labels are needed for each specimen, digital record identifiers are now being accepted by iDigBio as fundamental to tracking record data, as specimens may be destroyed, manipulated, or transferred to another institution (Goddard et al., 2011). Before sending records to iDigBio and GBIF, Discover Life will assign appropriate permanent, unique identifiers to each image and specimen record, so that these data can persist, be retrieved by the community, and never change. Thus, we will digitally link specimen images and text images with the physical specimens and labels in the collections.

Computer infrastructure and data storage

Discover Life has 12 servers in Georgia and 2 at the University of New South Wales. We will expand this capacity to meet our needs by adding equipment at North Carolina State University (Wiegmann), University of California (Berkeley: Roderick; Riverside: Yanega), University of Florida (Daniels), and University of Georgia (Pickering). At each site we will install a powerful server (Mac Pro), RAID disk array (24TB storage), and a workstation (iMac). These will run Mac OS, public domain software, and Discover Life's software. Dr. Long, Discover Life's system support guru, will manage them remotely.

Georeferencing

Georeferencing is an essential part of modern digitization protocol, especially if these data are to be used for understanding shifts in distribution ranges (Morris, 1996; Canhos et al., 2004; Chapman and Wieczorek, 2006). It requires the transformation of geographical data into X and Y coordinates with an accompanying estimate of precision (Chapman and Wieczorek, 2006). We propose to use Discover Life's gazetteers to georeference all legible labels to at least county centroids. We will use iDigBio's GeoLocate (www.museum.tulane.edu/geolocate) interface to help georeference localities more precisely. Through crowdsourcing under the oversight of specialists with requisite local or regional knowledge, we will correct problematic localities not captured properly by our automated processing of images.

Crowdsourcing and expert 'blessing'

Crowdsourcing is a powerful tool if implemented properly, especially if it is used by trained, highly-motivated staff, as we intend to do. We will use it in three ways. We will check data captured from images using Discover Life's crowdsourcing interface, *Timemachine*, which uses templates linked to gazetteers, dropdown boxes, and type-ahead controlled vocabularies. Because labels are written in numerous ways (e.g., county name sometimes absent, dates written in either American or British style, specific localities written in different orders, etc.) automated processing of images needs human oversight, especially for hand-written labels. *Timemachine* enables our staff, with help from the general public, to edit and annotate data captured by OCR/NLP and, in turn, for Discover Life's growing team of over 50 taxonomic experts, such as Charlie Covell at the McGuire Center, to review identifications and associated data and 'bless' them if they seem correct. Thus, we flag records that they have undergone a particular expert's review.

We will also use crowdsourcing to aid our georeferencing and in the measurement of specimen size.

Plans for interacting and integrating with the iDigBio HUB

Digitization efforts will yield better returns if data are linked across collections and projects, and an infrastructure is in place that facilitates collaboration, data automation, sustainable software development, and high performance computing (Chapman, 2005b; Donoghue et al., 2009; Hendry et al., 2010). Discover Life supports DwC archiving standards (Darwin Core Task Group, 2009) and has many tools to process non-standard formats, thereby making data easily transferrable to and from iDigBio and other initiatives in accordance with ADBC specifications to meet broad user needs. We will attend iDigBio meetings and share our workflows with them and other TCNs, integrating new methods into *Pollinator+* as iDigBio makes them available. While we will support participating institutions using their own established data formats, we will reformat their records to make them compliant with DwC standards. We propose to run a joint workshop with iDigBio at the University of Florida in Year-2 focused on digitizing pinned insects.

Training students and staff

For the duration of *Pollinator+*, Pickering proposes to take a semester leave of absence from UGA each year so that he can help Discover Life's staff to coordinate the project, set up the 4 regional centers, and train students and staff at participating collections. In addition to this informal training, he will teach 'biodiversity informatics' to undergraduates and graduate students as components of formal entomology and other natural history courses, including ones at the North Carolina State University, University of California, University of Florida, and University of Georgia..

Nancy Lowe, Discover Life's Outreach Coordinator, in consultation with iDigBio, will make training videos. These will document proper specimen handling, photography, and database techniques and the best practices that we require for each taxon. She will make these available via Discover Life and iDigBio (www.idigbio.org/tags/best-practice). She will also produce detailed written protocols tailored for each step of our workflow (e. g., see www.discoverlife.org/nh/id/20q/20q_help.html#timemachine). She will amended this material regularly to reflect emerging technologies and improved efficiencies (Kalms, 2012). She will seek electronic feedback and suggestions from participants and through discussion at our annual TCN meeting.

Student independent research projects

Throughout the project, we are dedicated to training both graduate and undergraduate students as researchers who use collections and specimen information as primary evidence in cutting-edge biodiversity science. The PIs will supervise graduate trainees on projects incorporating specimen data and digitization tools through their relevance to key questions in the systematics and evolutionary ecology of the specimens we study. Examples may include regional or faunistic studies of significant holdings in specific collections, taxon-focused surveys of pollinators or other key interactors and associated changes in life history linked to habitat change (eg., Youngsteadt et al 2014), or taxonomic research and development of identification systems using the latest tools in integrated, web accessible, taxonomic workflows (Schuh 2012). We will encourage faculty and staff at each collection to involve their undergraduates in independent research projects using the data that we collectively digitize. We intend that some of the funds that we will provide the collections will support their students presenting results at professional meetings.

Course development

In Year-3, Kawahara will teach a formal 2-credit, semester-long undergraduate course on the methods and objectives of collections digitization and the importance of collections preservation. There is currently no such formal undergraduate course at UF. The course will be under the umbrella of iDigBio and offered to local students, including ones at the University of Florida, Santa Fe College, Florida State University, Florida A&M University, Florida Atlantic University, University of Central Florida, Florida International University. It will teach modern databasing methods, explain software for data capture and analysis, allow students to experience firsthand the image capture techniques and databasing methods, and discuss how the HUB uses data. It is modeled after the Short Course on Biological Specimen Informatics, taught by K. Seltman and others through the TTD-TCN (R. Schuh PI), and on a multi-institutional graduate seminar offered by iDigBio on Biodiversity and Digitized Data (P. Soltis, instructor). We will advertise the course to the collections community via the ECN list serve, NHCOLL-L, Taxacom, and other social media. In Year-3, Wiegmann and Gillespie will teach similar courses at NCSU and UCB, respectively.

Student teams

We will employ 24 teams of students to photograph specimens and crowdsource data capture. Each team will have two cameras and get the training, supplies, and support that they need to digitize information rapidly and accurately. To avoid burn-out on what can become a boring, repetitive task, we will expect them to become vested in the scientific questions associated with the data that they are collecting, work in shifts not exceeding 4 hours, and work together in tandem with a fellow student. Senior personnel at each collection will recruit their students locally; the Polistes Foundation will employ them, and Discover Life's staff will train them. Once smaller collections finish digitizing, we will concentrate resources on larger collections but will 1) leave a camera at each smaller collection so that they can continue to digitize specimens beyond the ones that we are targeting and 2) provide technical and database support for them to do so.

Task analysis

Our budget justifies our costs to photograph specimens and capture data based on our experience taking nearly 30,000 images of moths and their labels at 13 collections since 2012. Our marginal cost per additional specimen is now approximately $0.35/specimen, including $0.27 for specimen handling, labeling, and photography, and $0.08/specimen for human involvement in data capture, error checking, and correction. This latter value should decrease over time as our NLP software learns places and collectors' names, becomes more efficient, and replaces the need for paid human labor..

Project sustainability

Our PIs are committed to the long-term maintenance of data collected by Pollinator+. In addition to transferring data to iDigBio and GBIF, Discover Life will store copies of all the images and associated data on servers at 5 major R-1 U.S. universities. Discover Life was started in 1998 and is run by the Polistes Foundation. We expect that it will sustain and serve the information produced by *Pollinator*+ long after ADBC funding ends, because it is supported via multiple grants, contracts, donations, and, above all, its growing community of contributors and users.

Governance and day-to-day management

A Steering Committee comprising our PIs and senior personnel will run *Pollinator*+ by consensus. PI Wiegmann will have ultimate authority to resolve issues that may arise. This committee will meet each fall at the Entomology Collections Network (ECN) meeting. ECN meets immediately before the Entomological Society of America and is typically attended by many curators and staff of our participating collections. Our annual meetings will cover project goals and targets, assess progress, adjust methods and strategies, and include training sessions. We will invite participants, including students, to report on progress and challenges, and seek their feedback. Wiegmann and Pickering together will provide daily project oversight and make immediate decisions as necessary. Each PI/co-PI will provide taxonomic/regional oversight and run one of our five centers. Pickering will oversee the informatics pipeline. Lowe, Discover Life's Outreach Coordinator, will work with everyone to help coordinate meetings and training sessions. Daniels and Kawahara are at UF and will coordinate our work with the iDigBio HUB. We will seek general guidance from external advisors, such as from Randall Schuh (AMNH, Tri-Trophic ADBC PI) and other leaders of previously funded ADBC proposals. We will hold quarterly conference calls, open to all interested parties, including students and collection. We will generate nightly metrics, prepare progress reports semi-annually, and submit annual reports to the NSF. Our benchmarks for evaluating progress will include **1)** the number of images taken, **2)** the number of labels transcribed, **3)** the number of records corrected, 'blessed' by experts, and exported to iDigBio and GBIF, **4)** Discover Life's metrics of web usage, **5)** external evaluation of our public outreach and student training, and **6)** dissemination of results through publications and research presentations. Weick and staff at the Polistes Foundation will pay our students and other employees and provide all other fiduciary services at an indirect cost of 15%.

Timeline [Activity -- Where -- Who -- When]

<u>Start-up</u>

Install Discover Life machines -- UGA -- Pickering & Long -- August, 2015
Install machines at regional centers -- NCSU, UF, UCB, UCR -- Pickering & Long -- September, 2015
Training video -- Discover Life -- Lowe -- September, 2015
Train graduate students/staff -- NCSU, UF, UCB, UCR -- Pickering & trainer -- Fall, 2015
Train undergraduate teams -- Collections -- Pickering, trainer, regional staff -- 2016 -> 2017

<u>Digitization workflow</u>

Imaging & data capture -- Regional centers -- Student teams & staff -- Fall, 2015 -> end
Imaging & data capture -- other collections -- Student teams, staff, public -- January, 2016 -> end
Fine-grain georeferencing -- Regional centers -- Staff, students, public -- final year
Data Export to iDigBio & GBIF -- Discover Life -- Database manager -- February, 2016 -> end

<u>Other activities</u>

Public involvement -- Discover Life -- Lowe -- 2016 -> end
Undergraduate course -- UF, NCSU, UCB -- Kawahara, Wiegmann, Gillespie -- 2017
Biodiversity informatics lectures in courses -- Regional centers, collections -- Pickering -- 2016, 2017
Steering Committee meetings -- ECN/ESA -- PIs + others -- 2015, 2016 (with workshop), 2017 (November)
Workshop with iDigBio -- UF -- PIs + others -- 2016

BROADER IMPACTS

This proposal creates a network to capture and integrate data from U. S. insect collections. It focuses on specimens relevant for studies of climate change, plant-pollination interactions, phenology, conservation, systematics, and biogeography. We divide our broader impacts into 'Research & Technology' and 'Outreach & Education'. We will recruit and employ an external evaluator to assess our success at the latter, using the Center for Advancement of Science Education's evaluator search tool.

Research & technology: Decades of historical information on pollinators remain untapped in entomological collections. We will mobilize 6 million specimens to enable: **1)** *Studies of Pollinators on Pollinated Food Crops:* Agricultural practices, which includes pest management strategies, has changed immensely in the last hundred years in order to meet the demands of our growing population (Costello et al. 2009). There is already ample evidence that these practices have impacted managed pollinators, especially honeybees. Our data will provide the opportunity to examine the ramifications of current cultivation practices on natural pollinators by examining land use changes and distribution records across both managed and natural systems. **2)** *Large-scale natural experiments:* Historical collections are snapshots through time. When associated with spatially explicit environmental data such as rainfall, temperature, pollution, and land use, they can be used to make predictions about future insect distributions, abundance, and life history parameters. These predictions are critical in mitigating impacts of global change on pollination and other plant/insect interactions. **3)** *Technology transfer:* Many entomological collections lack the capacity and resources to capture data digitally and database their specimens. Discover Life and our four regional/taxonomic centers will provide technology, training, and support to collections so that they can. Through training sessions and visits to partnering collections, we will implement our workflow for the rapid processing of pinned insects at multiple collections and advance the NSF's effort to facilitate large-scale digitization of our nation's collections.

Outreach & Education: 1) *Student Training:* Throughout this project, we will train students in museum databasing and digitization techniques at participating collections. Students trained will be at multiple levels of the educational pipeline, from graduate, undergraduate, to high school students. Discover Life's staff will training individuals at the collections; Pickering will give lectures by in courses at participating institutions, and Kawahara will teach a semester-long taught by Kawahara course at UF in conjunction with iDigBio. This 2-credit course will train undergraduate students in methods of collection databasing, digitization, and the importance of collections preservation. **2)** *STEM Recruitment in Community Colleges*: We will recruit students from local community colleges as a way to provide opportunities for underrepresented minority students. Some of our institutions (e.g., NC State, UF) have already implemented ways to incorporate minorities into collections-based projects from community colleges (e.g., Santa Fe College, Gainesville). Discover Life, UCB, UF, and other participants have outreach projects targeted at GK-12 students. We will use them to actively recruit GK-12 students and involve them in local tours of our collections. **3)** *Citizen Scientists:* Our nation's collections have vast holdings that are beneficial and interesting to everyone, but without digitization, most of these holdings are often only seen by professional researchers, not by the public. There is already evidence that the public is interested in observing and studying nature. Because we will disseminate images of specimens online, the public will be able to participate in, contribute to, and learn from a large-scale project to understand environmental factors affecting pollinators. Lowe will coordinate involving schools and the general public in our activities, including through her work with training videos.

PRIOR FUNDING

1. Digitization TCN: Collaborative Research: Plants, Herbivores and Parasitoids: A Model System for the Study of Tri-Trophic Associations. Schuh, Johnson, others. (EF 1115080) 07/01/2011- 06/30/2015; $1,194,534. Database ~1 million and image ~7K plant bugs and their parasitoids; database and image ~6K host plants; and integrate and share information using Discover Life and iDigBio aggregator. Two senior investigators, one project manager, eight digitizers, one REU student, four taxonomic specialists, one publication, one short course in Biodiversity Informatics for 24 students to date.
2. BRC: Collaborative Databasing of North American Bee Collections Within a Global Informatics Network. Ascher, Yanega, others. (DBI-0956388). $1,031,425. 06/01/10-05/31/13. This project digitized and consolidated specimen records from 10 bee collections across the United States. The project captured full label data, georeferenced and error-checked localities, and upload this information to publicly accessible databases. Using digital photography and rigorous research protocols, the 'Bee Hunt' outreach project empowered people at biological field stations, nature centers, parks, schools, and other sites.

3. ARTS: Phylogeny and systematic revision of the diverse and cryptic Euptychiina (Lepidoptera, Nymphalidae, Satyrinae). Willmott, Kawahara, Miller. (DEB 1256742). $458,104. 04/01/13-03/31/16. The focus of this grant is to complete taxonomic monographs for nine genera, construct phylogeny, and provide a broad training program for Lepidoptera systematics. Website being developed, two graduate students, several undergraduates, and a postdoc being trained.

4. Beyond Drosophila: Comparative transcriptomics of schizophoran flies to resolve a rapid radiation. Wiegmann; Trautwein, Meier, Yeates. DEB:1257960. 6/15/2013-5/31/2016. $599,975. This project addresses the currently intractable rapid radiation of schizophoran flies (~50,000 sp.) with transcriptome data. Fly biodiversity and evolution-based outreach is being developed and presented through collaboration with Nature Research Center in Raleigh, NC and the California Academy of Sciences, San Francisco.

5. Collaborative Research: BiSciCol Tracker: Towards a tagging and tracking infrastructure for biodiversity science collections. Cellinese, Roderick, Williams, others. NSF DEB-0956426, 09/01/10-08/31/14. To UCB: $663,859. The goal of this project is to build an infrastructure to tag and track scientific collections and all of their derivatives, particularly as enabled by globally unique identifiers (GUIDs). The project is improving data quality and quantity for non-scientists and scientists, and actively engages communities through training workshops, summer student internships, and community BioBlitz enhancements.

6. Collaborative Research: CalBug, and interactive database using arthropods to examine impact of climate change and habitat modification. Gillespie, Yanega, Kimsey, Roderick, others. NSF DBI-0956389, 08/01/10-07/31/15. To UCB, $975,253. This is a collaborative project among 9 major insect collections in California to digitize and geographically reference 1M specimens targeting groups of taxa and localities relevant to understanding changes in climate and habitat use.

DATA MANAGEMENT PLAN

Overview -- Discover Life is perfecting a new workflow to digitize pinned insects more rapidly, accurately, and inexpensively than previously accomplished. It has developed this workflow over the past two years at 13 insect collections. It involves humans photographing specimens and their labels separately, uploading and processing images in bulk, automatically capturing label information into standardized occurrence records with OCR and natural language processing software, and then enhancing data quality with sophisticated error checking programs that speed human review and enhancement. Currently, a novice digitizer can process 40 specimens per hour with this workflow; our most experienced digitizer has attained a speed of 109 per hour.

We will image all specimens and their labels, put them on the web, and enable researchers and volunteers to enhance information from the images into the associated digital records with crowdsourcing and secure editing software. In addition to its efficiency, our approach has advantages over directly databasing specimen information without imaging: **1)** It permits end users to examine specimens and capture additional measurements on size, morphology, color, specimen condition, number of associated pathogens, mites, etc. **2)** It greatly improves quality control by enabling experts to check and correct outliers flagged by error checking software and human feedback. Users can rapidly verify label transcription, for example, without returning to the physical collection. This overcomes a serious problem with mis-transcriptions inherent to digitization efforts that record labels directly into databases. **3)** It enables experts to review and bless the determination of specimens, correcting the 1-5% misidentifications that we have found among the exemplar moth species that we have photographed in the 13 collections. **4)** It transfers some of the cost of transcribing text, fine-scale georeferencing, and annotating records from *Pollinator+* to end users that we empower to contribute to the digitization process. And, possibly most importantly, **5)** the images make collections visible to the world so everyone can explore their diverse wonders and participate in science.

Best practices for imaging specimens --- Based on our considerable collective experience imaging specimens, we will develop and maintain state-of-the-art best practices for imaging specimens and their labels. We will use appropriate equipment for capturing high resolution images, switching to newer equipment when justified. In our first year, in consultation with iDigBio, we will make training videos of how to image insects based on our experience doing so.

Specimen identifiers --- We will use unique identifiers to associate specimens with their images. We will place a permanent, globally unique, machine readable, resolvable label on every specimen we image. If a collection already uses such labels, we will integrate their labelling system and databasing into our workflow. At other collections, we will work with collection manager to set up such a labelling system and associated database. If collection managers wish, they may choose to use Discover Life's specimen labelling system that databases and prints unique datamatrix labels via the web.

Image identifiers --- Discover Life assigns a unique identifier to each image. It associates each specimen and its image(s) through their identifiers. Currently Discover Life maintains over 1.5 million original images in 970 albums, each with a unique 'agent_id' starting with 'I_', for example, I_MGCL or I_MWS. We will create an album for each participating collection with an associated unique agent_id. Image identifiers are simply an agent_id followed by an integer unique to the album. The 29,000 images of specimens and labels taken at the 13 collections are in Discover Life's 'Entomology' albums (www.discoverlife.org/pa/ph/#e). They resolve through Discover Life's search box and web services.

Data transcription and annotation --- Discover Life is enhancing and simplifying its image management system so that we can rapidly and accurately transcribe labels and annotate each image's digital record. Our first priority is to capture taxon name(s), country, state, county, year, month, day, and collector name(s). Our second priority is to digitize and georeference the exact location of each occurrence. After images are uploaded and assigned unique identifiers, we **1)** use OCR to make a first pass at capturing text from labels, **2)** process the text with Discover Life's natural language processing software that uses controlled vocabularies, including large gazetteers to increase the speed and accuracy of georeferencing, and **3)**, use error checking software and humans in an iterative process enhance the quality and accuracy of the occurrence records.

Discover Life's *Timemachine* enables users to edit a limited number of fields, one record at a time. Privileged taxonomic experts and our staff, using passwords, can process additional fields and may work across multiple records and albums simultaneously. For security purposes, the system records all transactions, which can be reversed by staff members with the correct privileges.

Data integration --- Discover Life will work with each participating collection to integrate their data into our TCN and to help them import the data that we capture back into their data management systems. Discover Life has considerable experience importing and exporting data between different databases and spreadsheets. See 'Data Integration' in the Narrative for details. Public access to images, associated records, and contributors' datasets. With the exception of sensitive data that contributors wish to restrict, such as the location of rare and endangered species, we will make all information generated by *Pollinator*+ immediately available through web pages and services, cross-linking images and their unique specimen identifiers. Discover Life will format and share our contributors' datasets with the iDigBio HUB, GBIF, BISON, and others in accordance with ADBC specifications to meet broad user needs. We will support participating institutions using their own established data standards. However, we will encourage and help them to make their data compliant with Darwin Core (DwC) standards.

Data analysis and visualization --- Discover Life provides tools for researchers, students, and others to ask questions of data. These include our Global Mapper, Moth Math, phenology graphs, and plots to look at species interactions generated by automated R software programs. Currently, for example, the site serves 640,000 dynamic maps that it updates nightly. These are incorporated into the Encyclopedia of Life and reach a wide audience. It has online identification guides that we customize with local checklists.

Tracking progress --- Discover Life's automated software keeps copies of all changes in its contributors' datasets. From these we will build and display graphs showing our progress imaging specimens and annotating records. This will enable us to identify what is working well and what needs attention. Big Brother will be able to recognize institutions and individuals that are performing well and award them gold stars!

Servers, storage, backup, and processing capacity --- Discover Life has 12 servers in Georgia and 2 in Australia. The machines in Pickering's laboratory are connected with fiber optic cable to the building and campus network. The lab is wired with CAT6 cables that support transfer rates of 1 GB/s. The larger Sun servers have 8 disks and dual ports that are connected to two separate high-speed switches, one for frontend TCP/IP web traffic and the other for backend communication between machines. The 'original' server, pick18.pick.uga.edu, has 24TB of RAID6 disk storage. It runs Open Indiana with tcp wrappers, ipfilters, and is patched frequently, using ssh/scp for secure remote access and file transfer. Five production servers each have 14TB of disk storage. The network's considerable processing power enables nightly indexing of over 450 million records and rebuilding tables for a master index. File additions and changes are automatically mirrored within the lab and backed up nightly to the production machines in Australia. Images are uploaded to albums and served at five resolutions. The system assigns each image a unique identifier and builds an associated data record that includes the unique specimen identifier (or identifiers in instances of specimen-specimen interactions). We request funds to add servers and disks to process and store over 30 million files that *Pollinator*+ will produce.

Usage -- Since inception in 1998, Discover Life's computers in Georgia and Australia have served 2.9 billion pages and images to 27 million IP address. In September 2014, they served 57 million such hits to 716,000 IP addresses. This volume of use, in conjunction with that of partner websites, such at Moth Photographers Group (mothphotographersgroup.msstate.edu), gives us considerable potential to reach out and engage thousands of teachers, their students, and other individuals.

Persistence and Longevity --- Discover Life is run by the Polistes Foundation, a 501c3 nonprofit organization established in 2002. Should this foundation not be able to support Discover Life in the future, then it is legally required to transfer the website to a Federal agency or another nonprofit organization so that the site continues to fulfill its mission to "*assemble and share knowledge in order to improve education, health, agriculture, economic development, and conservation throughout the world.*"

NSF Opportunity 13-569
References

Aebi, A., B. E. Vaissiere, D. vanEngelsdorp, K. S. Delaplane, D. W. Roubik, and P. Neumann. 2012. Back to the future: Apis versus non-Apis pollination. *Trends Ecol Evol* 27:147–148.

Amano, T., R.J. Smithers, T.H. Sparks, and W.J. Sutherland. 2010. A 250-year index of first flowering dates and its response to temperature changes. *Proc Biol Sci* 277: 2451-2457.

Arnett, R. H., & Thomas, M. C. (Eds.). 2000. *American Beetles, Volume I: Archostemata, Myxophaga, Adephaga, Polyphaga: Staphyliniformia*. Boca Raton, FL: CRC Press.

Arnett, R. H., Thomas, M. C., Skelley, P. E., and Frank, J. H. (Eds.). 2010. *American Beetles, Volume II: Polyphaga: Scarabaeoidea through Curculionoidea*. CRC Press.

Ascher, J.S., C. Eardley, T. Griswold, G.A.R. Melo, A. Polaszek, M. Ruggiero, P. Williams, K. Walker and N. Warrit. 2008. World Bee Checklist. ITIS: http://www.itis.gov/beechecklist.html (accessed 17 October 2013).

Ascher, J.S. and J. Pickering. 2013. Discover Life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila). Discover Life. Draft 36 [20 July]: www.discoverlife.org/mp/20o?guide=Apoidea_species

Baird, R. 2010. Leveraging the fullest potential of scientific collections through digitization. *Biodiversity Informatics* 7: 130-136.

Barbraud, C., P. Rivalan, P. Inchausti, M. Nevoux, V. Rolland, and H. Weimerskirch. 2011. Contrasted demographic responses facing future climate change in Southern Ocean seabirds. *J Anim Ecol* 80: 89-100.

Bartomeus, I., J.S. Ascher, J. Gibbs, B.N. Danforth, D.L. Wagner, S.M. Hedtke, and R. Winfree. 2012. Historical changes in northeastern US bee pollinators related to shared ecological traits. *Proc Natl Acad Sci U S A* 110: 4656-4660.

Basset, Y., Eastwood, R., Sam, L., Lohman, D. J., Novotny, V., Treuer, T., Miller, S. E., Weiblen, G. D., Pierce, N. E., Bunyavejchewin, S., Sakchoowong, W., Kongnoo, P. Osorio-Arenas, M. A. 2013. Cross-continental comparisons of butterfly assemblages in tropical rainforests: implications for biological monitoring. *Ins Conserv Div,* 6: 223–233.

Beach, J., S. Blum, M. Donoghue, L. Ford, R. Guralnick, M. Mares, B. Thiers, M. Westneat, Q. Wheeler, and B. Wiegmann. 2010. A Strategic Plan for Establishing a Network Integrated Biocollections Alliance. http://digbiocol.files.wordpress.com/2010/08/niba_brochure.pdf (accessed 17 October 2013).

Berenbaum M, Bernhardt P, Buchmann S, Calderone N, Goldstein P, Inouye DW, Kevan P, Kremen C, Medellin RA, Ricketts T, Robinson GE, Snow AA, Swinton SM, Thein LB, Thomspson FC (2007) Status of pollinators in North America. National Academies Press, Washington

Bertone, M.A., R.L. Blinn, T.M. Stanfield, K.J. Dew, K.C. Seltmann, and A.R. Deans. 2012. Results and insights from the NCSU Insect Museum GigaPan project. *ZooKeys* 209: 115-132.

Biesmeijer, J.C., S.P.M. Roberts, M. Reemer, R. Ohlemüller, M. Edwards, T. Peeters, A.P. Schaffers, S.G. Potts, R. Kleukers, C.D. Thomas, J. Settele, and W.E. Kunin. 2006. Parallel Declines in Pollinators and Insect-Pollinated Plants in Britain and the Netherlands. *Science* 313: 351-354.

Blagoderov, V., I. Kitching, L. Livermore, T. Simonsen, and V. Smith. 2012. No specimen left behind: industrial scale digitization of natural history collections. *ZooKeys* 209: 133-146.

Burkle, L.A., J.C. Marlin, and T.M. Knight. 2013. Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science* 339: 1611-1615.

Calderone, N.W. 2012. Insect pollinated crops, insect pollinators and US agriculture: trend analysis of aggregate data for the period 1992–2009. *PloS One* 7: e37235.

Cleland, E.E., I. Chuine, A. Menzel, H.A. Mooney, and M.D. Schwartz. 2007. Shifting plant phenology in response to global change. *Trends Ecol Evol* 22: 357-365.

Colla, S.R., Gadallah, F., Richardson, L., Wagner, D., Gall, L. 2012. Assessing declines of North American bumble bees (Bombus spp.) using museum specimens. Biodiv Conserv 21 3585-3595.

Crawford, P.H.C. and B.W. Hoagland. 2009. Can herbarium records be used to map alien species invasion and native species expansion over the past 100 years? *J Biogeogr* 36: 651-661.

Dennis, R.L.H. 2012. *A Resource-Based Habitat View for Conservation: Butterflies in the British Landscape.* Wiley-Blackwell, Chichester, 420 pp.

Diamond, S.E., A.M. Frame, R.A. Martin, and L.B. Buckley. 2011. Species' traits predict phenological responses to climate change in butterflies. *Ecology* 92:1005–1012.

Dietrich, C.H., J. Hart, D. Raila, U. Ravaioli, N. Sobh, O. Sobh, and C. Taylor. 2012. InvertNet: a new paradigm for digital access to invertebrate collections. *ZooKeys* 209: 165-181.

Dolan, R.W., M.E. Moore, and J.D. Stephens. 2011. Documenting effects of urbanization on flora using herbarium records. *J Ecol* 99: 1055-1062.

Dyar, L.A., D.L. Wagner, H.F. Greeney, A.M.Smilanich, T.M. Massad, M. Robinson, M. Fox, R. Hazen, A. Glassmire, N. Pardikes, K. Fredrickson, C. Pearson, G.L. Gentry, and J.O. Stireman III. 2012. Novel insights into tritrophic interaction diversity and chemical ecology using 16 years of volunteer supported research. *Amer Entomol* 58: 15-19.

Ellwood, E.R., S.A. Temple, R.B. Primack, N.L. Bradley, and C.C. Davis. 2013. Record-breaking early flowering in the eastern United States. *PloS One* 8: e53788.

Feeley, K.J. and M.R. Silman. 2010. Land-use and climate change effects on population size and extinction risk of Andean plants. *Glob Chang Biol* 16: 3215-3222.

Franz, N.M. and D. Thau. 2010. Biological taxonomy and ontology development: scope and limitations. *Biodiversity Informatics* 7: 45-66.
Gardner, J.L., A. Peters, M.R. Kearney, L. Joseph, and R. Heinsohn. 2011. Declining body size: a third universal response to warming? *Trends Ecol Evol* 26: 285-291.

Garibaldi, L.A. et al. 2013. Wild Pollinators Enhance Fruit Set of Crops Regardless of Honey Bee Abundance. *Science* 339: 1608-1611.

Ghazoul, J. 2013. Pollination decline in context. *Science* 340: 923.

Grimaldi, D. and M. Engel. 2005. *Evolution of the Insects*. Cambridge University Press, New York. 755 pp.

Grimm, N.B., S.H. Faeth, N.E. Golubiewski, C.L. Redman, J. Wu, X. Bai, and J.M. Briggs. 2008. Global change and the ecology of cities. *Science* 319: 756-760.

Hill, A., R. Guralnick, A. Smith, A. Sallans, R. Gillespie, M. Denslow, J. Gross, Z. Murrell, T. Conyers, P.

Oboyski, J. Ball, A. Thomer, R. Prys-Jones, J. de la Torre, P. Kociolek, and L. Fortson. 2012. The notes from nature tool for unlocking biodiversity records from museum records through citizen science. *ZooKeys* 209: 219-233.

Inouye, D. 2013. Database of fly visitation to flowers. http://hdl.handle.net/1903/14424
Intergovernmental Panel on Climate Change (IPCC). 2013. Available at: http://www.ipcc.ch/. Accessed October 3, 2013.

Johnson, K.G., S.J. Brooks, P.B. Fenberg, A.G. Glover, K.E. James, A.M. Lister, E. Michel, M. Spencer, J.A.

Todd, E. Valsami-Jones, J.R. Young, and J.R. Stewart. 2011. Climate change and biosphere response: unlocking the collections vault. *Bioscience* 61: 147-153.

Kevan, P.G. and B.F. Viana. 2003. The global decline of pollination services. *Biodiversity* 4: 3–8.

Klein, A.M., B.E. Vaissiere, J.H. Cane, I. Steffan-Dewenter, S.A. Cunningham, C. Kremen, and T. Tscharntke. 2007. Importance of pollinators in changing landscapes for world crops. *Proc Biol Sci* 274: 303-313.

Kress, W.J., W.R. Heyer, P. Acevedo, J. Coddington, D. Cole, T.L. Erwin, B.J. Meggers, M. Pogue, R.W. Thorington, R.P. Vari, M.J. Weitzman, and S.H. Weitzman. 1998. Amazonian biodiversity: assessing conservation priorities with taxonomic data. *Biodivers Conserv* 7: 1577-1587.

LaBerge, W.E. 1986. The zoogeography of *Andrena* Fabricius (Hymenoptera, Andrenidae) of the Western Hemisphere. In *The Prairie: Past, Present and Future (Proceedings of the Ninth North American Prairie Conference)* (Vol. 110, p. 115).

Lawler, J.J., S.L. Shafer, D. White, P. Kareiva, E.P. Maurer, A.R. Blaustein, and P.J. Bartlein. 2009. Projected climate-induced faunal change in the Western Hemisphere. *Ecology* 90: 588-597.

Lavoie, C. and D. Lachance. 2006. A new herbarium-based method for reconstructing the phenology of plant species across large areas. *Am J Bot* 93: 512-516.

Lebuhn, G., Droege, S., Connor, E. F., Gemmill-Herren, B., Potts, S. G., Minckley, R. L., Griswold, T., Jean, R., Kula, E., Roubik, D. W., Cane, J., Wright, K. W., Frankie, G. and Parker, F. (2013), Detecting Insect Pollinator Declines on Regional and Global Scales. *Conserv Biol*, 27: 113–120.

Lees, D.C., H.W. Lack, R. Rougerie, A. Hernandez-Lopez, T. Raus, N.D. Avtzis, S. Augustin, and C. Lopez-Vaamonde. 2011. Tracking origins of invasive herbivores through herbaria and archival DNA: the case of the horse-chestnut leaf miner. *Front Ecol Environ* 9: 322-328.

Loarie, S.R., P.B. Duffy, H. Hamilton, G.P. Asner, C.B. Field, and D.D. Ackerly. 2009. The velocity of climate change. *Nature* 462: 1052-1055.

Luedeling, E., K.P. Steinmann, M. Zhang, P.H. Brown, J. Grant, and E.H. Girvetz. 2011. Climate change effects on walnut pests in California. *Glob Chang Biol* 17: 228-238.

Marks, R. 2005. Native Pollinators. *NRCS Fish and Wildlife Habitat Management Leaflet* 34. http://plants.usda.gov/pollinators/Native_Pollinators.pdf (accessed 18 October 2013).

Memmott, J., P.G. Craze, N.M. Waser, M.V. Price. 2007. Global warming and the disruption of plant-pollinator interactions. *Ecol Lett* 10: 710-717.

Morris, A. B., Graham, C. H., Soltis, D. E. and Soltis, P. S. 2010. Reassessment of phylogeographical structure in an eastern North American tree using Monmonier's algorithm and ecological niche modelling. *J Biogeog*, 37: 1657–1667.

Motten, A.F., Campbell, D.R., Alexander, D.E., Miller, H.L. 1981. Pollination effectiveness of specialist and generalist visitors to a North Carolina population of Claytonia virginica. *Ecology*, 1278-1287.

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

National Resource Council. 2007. Status of Pollinators in North America. Washington, D.C.: The National Academies Press.

Netherer, S. and A. Schopf .2010. Potential effects of climate change on insect herbivores in European forests—General aspects and the pine processionary moth as specific example. *For Ecol Manage* 259: 831-838.

Nieukerken, E. J. van, L. Kaila, I. J. Kitching, N. P. Kristensen, D. C. Lees, J. Minet, C. Mitter, M. Mutanen, J. C. Regier, T. J. Simonsen, N. Wahlberg, S.-H. Yen, R. Zahiri, D. Adamski, J. Baixeras, D. Bartsch, B. Å. Bengtsson, J. W. Brown, S. R. Bucheli, D. R. Davis, J. De Prins, W. De Prins, M. E. Epstein, P. Gentili-Poole, C. Gielis, P. Hät-tenschwiler, A. Hausmann, J. D. Holloway, A. Kallies, O. Karsholt, A. Kawahara, S. J. C. Koster, M. Kozlov, J. D. Lafontaine, G. Lamas, J.-F. Landry, S. Lee, M. Nuss, K. T. Park, C. Penz, J. Rota, B. C. Schmidt, A. Schintlmeister, J. C. Sohn, M. A. Solis, G. M. Tarmann, A. D. Warren, S. Weller, R. V. Yakovlev, V. V. Zolotuhin, and A. Zwick. 2011. Order Lepidoptera Linnaeus, 1758.

In Zhang, Z.Q. ed., Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa* 3148: 212-221.

Opler, P.A., K. Lotts, and T. Naberhaus, coordinators. 2013. Butterflies and Moths of North America. http://www.butterfliesandmoths.org (Version 100313).

Panchen, Z.A., R.B. Primack, T. Anisko, and R.E. Lyons. 2012. Herbarium specimens, photographs, and field observations show Philadelphia area plants are responding to climate change. *Am J Bot* 99: 751-756.

Pape, T. and Thompson, F.C. 2013. Systema Dipterorum, Version1.5. http://www.diptera.org.
Pellmyr, O. 2003. Yuccas, yucca moths, and coevolution: a review. *Ann Missouri Bot Gard*: 35-55.

Peterson, A.T., M.A. Ortega-Huerta, J. Bartley, V. Sánchez-Cordero, J. Soberón, R.H. Buddemeier, and D.R.B. Stockwell. 2002. Future projections for Mexican faunas under global climate change scenarios. *Nature* 416: 626-629.

Pimentel, D., T. Petrova, M. Riley, J. Jacquet, V. Ng, J. Honigman, and E. Valero. 2006. Conservation of biological diversity in agricultural, forestry, and marine systems. *In* Schwartz, J., ed., *Focus on Biodiversity Research* (pp. 1-25). Nova Science Publishers, New York, 320 pp.

Potts. S.G., J.C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger, and W.E. Kunin. 2010. Global pollinator declines: trends, impacts and drivers. *Trends Ecol Evol* 25: 345-353.
Primack, R.B. and A.J. Miller-Rushing. 2012. Uncovering, collecting, and analyzing records to investigate the ecological impacts of climate change: a template from Thoreau's Concord. *BioScience* 62: 170-181.

Robbirt, K.M., A.J. Davy, M.J. Hutchings, and D.L. Roberts. 2011. Validation of biological collections as a source of phenological data for use in climate change studies: a case study with the orchid *Ophrys sphegodes*. *J Ecol* 99: 235-241.

Schuh, R.T. 2012. Integrating specimen databases and revisionary systematics. *ZooKeys* 209: 255-267.

Soltis, D.E., P.S. Soltis, V.A. Albert, D.G. Oppenheimer, C.W. DePamphilis, H. Ma, M.W. Frohlich, and G. Theißen. 2002. Missing links: the genetic architecture of flower and floral diversification. *Trends Plant Sci* 7: 22-31.

Tulig, M., N. Tarnowsky, M. Bevans, A. Kirchgessner, and B.M. Thiers. 2012. Increasing the efficiency of digitization workflows for herbarium specimens. *ZooKeys* 209: 103-113.

Tylianakis, J.M. 2013. The global plight of pollinators. *Science* 339: 1532-1533.
Vamosi, J.C., T.M. Knight, J.A. Steets, S.J. Mazer, M. Burd, and T.L. Ashman. 2006. Pollination decays in biodiversity hotspots. *Proc Natl Acad Sci U S A* 103: 956-961.

Walter, R.P., J.D. Hogan, M.J. Blum, R.B. Gagne, E.F. Hain, J.F. Gilliam, and P.B. McIntyre. 2012. Climate change and conservation of endemic amphidromous fishes in Hawaiian streams. Endangered Species Research 16: 261-272.

Ward, P.S., ed. 2013. AntWeb: Ants of California. http://www.AntWeb.org/california.jsp (accessed 21 September 2013).

Weiner, C.N., M. Werner, K.E. Linsenmair, and N. Blüthgen. 2013. Land-use impacts on plant-pollinator networks: interaction strength and specialization predict pollinator declines. *Ecology.* http://dx.doi.org/10.1890/13-0436.1

Willis, C.G., B. Ruhfel, R.B. Primack, A.J. Miller-Rushing, and C.C. Davis. 2008. Phylogenetic patterns of species loss in Thoreau's woods are driven by climate change. *Proc Natl Acad Sci U S A* 105: 17029-17033.

Willis, C.G., B. Ruhfel, R.B. Primack, A.J. Miller-Rushing, J.B. Losos, and C.C. Davis. 2010. Favorable climate change response explains non-native species' success in Thoreau's woods. *PLoS One* 5: e8878.

Winfree, R., I. Bartomeus, and D.P. Cariveau. 2011. Native pollinators in anthropogenic habitats. *Ann Rev Ecol Evol Syst* 42: 1-22.

Wolkovich, E.M. and E.E. Cleland. 2010. The phenology of plant invasions: a community ecology perspective. *Front Ecol Environ* 9: 287-294.

Youngsteadt, E. A.G. Dale, A. J. Terando, R. R. Dunn, and S.D. Frank. 2014. Do cites simulate climate change? A comparison of herbivore response to urban and global warming. Global Change Biology doi: 10.1111/gcb.12692.