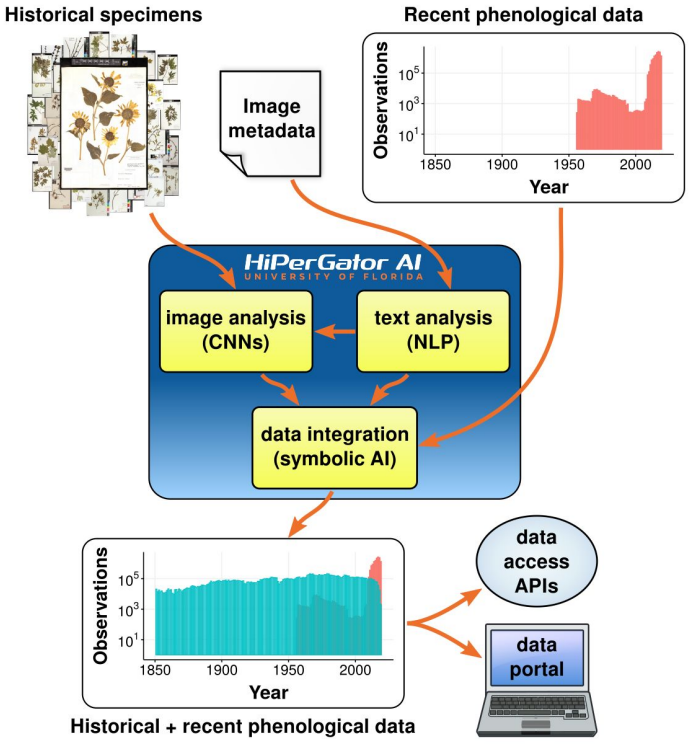
OR-DRD-AI2020: Using AI to uncover decades of global ecological change; Ecology and Biodiversity

Problem Statement

We propose to use a suite of statistical and symbolic AI techniques to deliver new, historical ecological data of unprecedented temporal and taxonomic scope by extracting and mobilizing *phenological information* (information about plant developmental events including flowering and fruiting) from the millions of historical plant specimen images currently available at UF through iDigBio. We will:

- Use natural language processing and computer vision techniques to generate new phenological data from the more than 28 million images of historical plant specimens available through the UF-led iDigBio project (some dating back more than 200 years).
- Use symbolic AI methods to integrate these new data with extant, recent phenological data.
- Deliver the data to researchers via a web portal and data access APIs (including an R package).
- Develop a module based on this work for a new *AI in Biology* course.

Significance of research. Plant phenological data are essential for understanding and forecasting ecosystem responses to climate and land use changes. Data from this project will provide vital historical baselines for forecasting long-term ecological and biodiversity trends, with impacts on both conservation and agriculture. This work will also establish a foundation for building new research infrastructure for generating, integrating, and delivering plant phenological data.



Team Members

- Brian Stucky (PI, FL Museum of Nat. Hist. [FLMNH])
- Rob Guralnick (Co-PI, FLMNH)
- Pam Soltis (Co-PI, FLMNH)
- Doug Soltis (Co-PI, FLMNH, Biology)
- José Fortes (Co-PI, Elec. and Computer Engineering)
- Matt Gitzendanner (Co-PI, Biology/UFIT Res. Comp.)
- Raphael LaFrance (software engineer, FLMNH)

Technical Process Description

Project timeline and key milestones.

	Q1	Q2	Q3	Q4
1. NLP analysis of free-text metadata.	■			
2. Single-pass image analysis.	■	■		
3. Windowed image analysis.		■	■	
4. Methods comparison, taxonomic scope.			■	■
5. Integrating new data with extant data.				■
6. New course module.		■	■	
7. Science applications, new proposals.		■	■	■

Required preparatory work. None.

Senior/key personnel roles. Stucky: project lead, 1-7; Guralnick: 5,7; P. Soltis: 4,7; D. Soltis: 4,7; Fortes:1-3,5,7; Gitzendanner: 1-3,6,7; LaFrance: 1-4

Sustainability of Research Program

Sponsors/programs targeted for extramural funding.

- NSF, Infrastructure Capacity for Biology (ICB) program
- NSF, Macrosystems Biology and NEON-Enabled Science Program (MSB-NES)
- NSF, Division of Environmental Biology Core Programs: Systematics and Biodiversity Science (SBS), Population and Community Ecology (PCE), Ecosystem Science (ES)
- USDA, Data Science for Food and Agricultural Systems (DSFAS)

Target dates for submission of proposals to sponsors.

- NSF ICB: February 2021
- NSF MSB-NES: November 2021
- NSF SBS, PCE, ES: early 2022
- USDA DSFAS: early 2022

Project Description

Problem Statement (498 words)

Human activities are pushing many of Earth's ecosystems to the brink of collapse, and the window of time for course correction continues to narrow [1–3]. Ensuring the future of our ecological heritage requires understanding its past, but our historical knowledge of most ecosystems remains frustratingly incomplete [4–6].

We will combine statistical and symbolic AI methods to deliver new, historical ecological data that will provide unprecedented temporal context for understanding the biological consequences of environmental change. Our project will focus on data about *plant phenology* – the timing of plant life-cycle events (e.g., leafout, flowering, fruiting). Plant phenology is a fundamental component of ecosystem function that often carries the earliest detectable ecological "fingerprints" of environmental change [7,8]. Historical phenological data are therefore crucial for analyzing and predicting the ecological impacts of ongoing climate and land use changes, not only to conserve natural ecosystems, but also to ensure the security of our agricultural systems [9–12].

We will exploit the historical record of plant phenology documented by *herbarium specimens* – preserved plant specimens that capture what a plant was doing (flowering, e.g.) at the time and place of collection [13,14]. Millions of herbarium specimens are housed in museums around the world, but the phenological information they contain remains largely inaccessible. Crucially, UF leads iDigBio (Integrated Digitized Biocollections), a national infrastructure for digitizing specimens that currently houses more than 28 million images of herbarium specimens ripe for extraction of phenological information (with hundreds more added daily).

Our project has four main objectives:

1. Use natural language processing methods to rapidly extract phenological information from the ~13% of iDigBio records with suitable free-text phenological metadata.
2. Use deep learning models to generate high-accuracy phenological information from iDigBio's herbarium specimen images that lack phenological metadata.
3. Integrate new data from objectives 1 and 2 with extant, recent phenological data using a scalable pipeline based on symbolic AI methods.
4. Develop a module for an *AI In Biology* course developed by PI Stucky and co-PI Gitzendanner (offered in Spring 2021), building on our use of natural history collections in the classroom [15–17].

This project will immediately impact our understanding of ecosystem responses to environmental change, with adjacent impacts on conservation and agriculture. Our work will also build a needed foundation of robust deep learning models for building infrastructure to automatically harvest and analyze new specimen images as they become available. Finally, we will establish a baseline for further innovation in phenological analysis (e.g., developing segmentation-based methods for producing quantitative phenological data).

Our extensive preliminary work in this area will allow us to immediately leverage UF's new computing resources. We have shown that for multiple, diverse plant genera, deep learning methods can generate research-grade phenological data from herbarium specimens with accuracy comparable or superior to expert human analysts [18,19]. We have also developed scalable software for phenological data integration and delivery, and we have used these tools to mobilize millions of recent phenological records [20–24]. The work proposed here will add millions of additional records with the potential to transform biodiversity research.

Technical Process Description (728 words)

<u>*Extracting phenological information from image metadata.*</u> We have determined that ~13% of all iDigBio images of herbarium specimens include metadata with "latent" phenological information in the form of unstructured, free text. We will use natural language processing (NLP) techniques to extract this latent phenological information. Our preliminary text mining analysis indicated that relatively simple, rule-based NLP methods will be suitable for rapidly extracting ~76% of the phenological information in these records. To get as much of the remaining information as possible, we will fine-tune BERT NLP models (Bidirectional Encoder Representations from Transformers) [25] to perform two tasks. First, we will identify plant structures of potential interest (e.g., floral parts, fruits), using named entity recognition. Then

we will determine the semantics and relationships of the structures to determine if they include qualitative phenological information. We will fine-tune freely available, pre-trained BERT models to reduce the implementation effort, and we will use human-generated datasets of phenological annotations from our prior work as training data.

Generating phenological information from herbarium specimen images. We will use convolutional neural networks (CNNs), which are currently the best algorithms known for image recognition tasks, to evaluate whether three key, qualitative phenological traits are present or absent on a herbarium specimen: 1) leaves; 2) flowers; and 3) fruits. These traits constitute the majority of phenological data used by researchers and are applicable to nearly all species of flowering plants (other phenological data are more limited in application). Furthermore, we know from our prior work that deep learning approaches can deliver research-grade presence/absence data with accuracy comparable to human experts [18,19].

We will build on the recently published "EfficientNet" architecture, which delivers state-of-the-art accuracy on the ImageNet challenge while providing highly efficient scaling of computational requirements with input image size [26]. We will pursue two approaches to image analysis: 1) CNNs that analyze entire herbarium specimen images in a single pass; and 2) CNNs that analyze images by decomposing them into a set of overlapping "chunks", each of which is analyzed separately and then recombined to provide a final evaluation of the entire image. We have previously found that the second approach works better for some plant species that require high-resolution images for accurate phenology assessment, while the first approach is better suited to species for which lower-resolution images are suitable (e.g., species with large, showy flowers). UF's AI resources will allow us to explore the tradeoffs between these techniques. We will also investigate whether models spanning many plant families are superior to more specialized models with narrower taxonomic focus.

Training data will come from our prior work and the results of the NLP analyses described above. We will employ our undergraduate research team at the FLMNH to provide additional human-generated training data as needed.

Integrating phenological data with symbolic AI. Phenological data from different sources are often highly heterogeneous (e.g., differences between quantitative and qualitative observations) and not amenable to simplistic data integration techniques [20]. To solve this problem, we have developed an ontology and custom data integration pipeline that uses symbolic AI techniques to efficiently integrate heterogeneous phenological data without information loss [20,23,24]. We have used these tools to build a knowledge base of more than 20 million recent phenological observations, including a prototype global plant phenological data portal and data access APIs (including an R package) [21,22]. We will use these tools to integrate our new, historical phenological data from herbarium specimens with the recent records already in the knowledge base, thus making the combined data available to researchers worldwide.

Computing requirements. The vast majority of the proposed work will only be feasible on a high-performance computing platform with GPU resources. For whole-image model development and training, we will use large CNNs to maximize input image size, and these models will require modern GPUs with large amounts of RAM. Methods based on decomposing images into chunks are potentially even more computationally intensive since they require multiple CNN passes for each image. BERT methods for analyzing free-text phenological information are also best trained and run on GPUs. Furthermore, given that iDigBio currently houses more than 28 million herbarium specimen images (and growing), a multi-GPU, high-performance computing environment will be essential for this work.

Multidisciplinary aspects. This project requires multi-disciplinary expertise spanning deep learning and computer vision, NLP, symbolic AI and data integration, botany, herbaria and natural history collections, and biodiversity informatics.

Team qualifications (200 words)

The range of expertise required for this work necessitates a team approach. Our interdisciplinary team has a long history of successful collaborations (NSF and DOE) and encompasses the expertise needed.

- Brian Stucky led much of the preliminary work. He will lead the project and contribute to text mining, analyzing herbarium specimen images, data integration, and mobilizing data for research.
- Rob Guralnick has built phenological data resources to address questions about environmental change. He will coordinate data integration and delivery and develop next-step science applications (forecasting, e.g.).
- Doug Soltis has extensive experience in the study of plant diversity and use of herbarium specimens and iDigBio data. He will provide botanical expertise and help develop next-step science applications.
- Pam Soltis leads the research applications component of iDigBio, emphasizing herbarium specimen data. She will help develop research applications for the mobilized data.
- José Fortes leads iDigBio cyberinfrastructure and has expertise with using NLP and combined human-machine intelligence methods to analyze specimens. He will support interactions with iDigBio and applications of machine learning.
- Matt Gitzendanner has expertise in computational methods and HPC applications. Gitzendanner and Stucky will co-develop a learning module based on this project for their *AI in Biology* course (Spring 2021).

Budget

Item	Salary and fringe	Total cost
3 months salary support for Brian Stucky	$18,378 + $5,458 fringe (29.7%)	$23,836
2.5 months salary support for Raphael LaFrance	$16,096 + $6,294 fringe (39.1%)	$22,390
undergraduate research support, 100 hours ($12/hr)	$1,200 + $23 fringe (1.9%)	$1,223
HiPerGator: 24-48 GPU cards, 256 CPU cores, 30 TB /blue storage, 70 TB of /orange storage	N/A	N/A
	Total budget:	$47,449

Start date: November 2, 2020

End date: November 1, 2021

Budget justification and need for support from AI Research Catalyst Fund. Our budget provides partial salary support to PI Brian Stucky (at 25% FTE / 3 months) and software engineer Rafe LaFrance (at 20.83% FTE / 2.5 months), who will together complete the majority of the implementation work required for this project. Co-PIs not receiving direct salary support plan to contribute as follows: Rob Guralnick at 2% FTE (0.24 months), Pam Soltis at 2% FTE (0.24 months), Doug Soltis at 2% FTE (0.24 months), Jose Fortes at 1% FTE (0.12 months), and Matt Gitzendanner at 10% FTE (1.2 months). Undergraduate research funding will support undergraduate researchers at the FLMNH working with Stucky, Guralnick, P. Soltis, and D. Soltis; these researchers will help fill any gaps in training data.

Given our prior work (described above), a timely injection of computing resources and personnel funding will permit us to rapidly scale efforts, allowing us to deliver high-impact results on a relatively short time scale. Our project requires substantial computing resources that we cannot provide within our labs. To permit some flexibility with resource allocation, we provide a range of GPU quantities that would work well for this project. We also request substantial /orange storage to serve as holding space for large numbers of full-resolution images with /blue space used for image subsets for model training or analysis. Our project's computing needs and readiness are an excellent fit for the AI Research Catalyst Fund.

Plans for continued support (198 words)

PI Stucky is leading an in-development proposal for NSF's Infrastructure Capacity for Biology (ICB) solicitation for submission in early 2021. This proposal focuses on creating more robust infrastructure for generating and mobilizing plant phenological data, includes a large AI component that will substantially build on the work proposed here, and will also use HiPerGator AI resources.

The team will use the methods, infrastructure, and data mobilized through this Catalyst project to address the effect of climate change on plant phenology during the past century. Intersections of these data with the plant Tree of Life and with spatial and ecological variation will be appropriate for proposals for NSF's Macrosystems Biology and NEON-Enabled Science Program and at least three of NSF's Division of Environmental Biology core programs: Systematics and Biodiversity Science, Population and Community Ecology, and Ecosystem Science. We envision submission to several of these programs in late 2021 and early 2022.

The research to be enabled by this project has important consequences for agriculture and forestry. The USDA's Data Science for Food and Agricultural Systems (DSFAS) program supports data-driven research on the impacts of climate change on crops and forest ecosystems and is another potential target for future funding.

Biographical Sketch – Brian J. Stucky

A. Professional Preparation

Bethel College	North Newton, KS	Computer Science	B.A., 2002
University of Colorado	Boulder, CO	Ecology and Evolutionary Biology	Ph.D., 2015
University of Florida	Gainesville, FL	Informatics, Ecology and Evol. Bio.	Postdoc., 2016-2018

B. Appointments

Jan. 2019 - present	Assistant Scientist, Florida Museum of Natural History, Univerity of Florida
Jan. 2017 - Dec. 2018	NSF postdoctoral research fellow, Florida Museum of Natural History, University of Florida
Jan. 2016 - Dec. 2016	Postdoctoral researcher and UF Informatics Institute (UFII) postdoctoral fellow, Florida Museum of Natural History and UFII, University of Florida
Aug. 2009 - Dec. 2009, Jan. 2012 - May 2012, Aug. 2013 - Dec. 2013, Aug. 2014 - Dec. 2014	Graduate Teaching Assistant, Department of Ecology and Evolutionary Biology, University of Colorado
Jan. 2010 - May 2013, Jan. 2014 - May 2014, Jan. 2015 - Aug. 2015	Graduate Research Assistant, Department of Ecology and Evolutionary Biology, University of Colorado

C. Products

C. i. Five most relevant products

Pearson, K.D., G. Nelson, M.F.J. Aronson, P. Bonnet, L. Brenskelle, C.C. Davis, E.G. Denny, E.R. Ellwood, H. Goëau, J.M. Heberling, A. Joly, T. Lorieul, S.J. Mazer, E.K. Meineke, B.J. Stucky, P. Sweeney, A.E. White, and P.S. Soltis. (2020) Machine learning using digitized herbarium specimens to advance phenological research. *BioScience* 70: 610-620. doi: 10.1093/biosci/biaa044

Stucky, B.J. (2020) ImageAnt: Software for fast and accurate image and audio annotation. https://gitlab.com/stuckyb/imageant

Brenskelle, L., B.J. Stucky, J. Deck, R. Walls, R.P. Guralnick. (2019) Integrating herbarium specimen observations into global phenology data systems. *Applications in Plant Sciences* 7(3): e1231. doi: 10.1002/aps3.1231

Stucky, B.J., R. Guralnick, J. Deck, E.G. Denny, K. Bolmgren, R. Walls. (2018) The Plant Phenology Ontology: A new informatics resource for large-scale integration of plant phenology data. *Frontiers in Plant Science* 9: 517. doi: 10.3389/fpls.2018.00517

Stucky, B.J., A. Luc. (2018) OntoPilot: Software for ontology development and deployment. https://github.com/stuckyb/ontopilot/

C. ii. Additional products

Li, D., B.J. Stucky, J. Deck, B. Baiser, and R. Guralnick. (2019) The effect of urbanization on plant phenology depends on regional temperature. *Nature Ecology and Evolution*. doi: 10.1038/s41559-019-1004-1

Stucky, B.J., J. Deck, T. Conlin, L. Ziemba, N. Cellinese, and R. Guralnick. (2014) The BiSciCol Triplifer: bringing biodiversity data to the Semantic Web. *BMC Bioinformatics* 15: 257. doi: 10.1186/1471-2105-15-257

Guralnick, R., T. Conlin, J. Deck, B.J. Stucky, N. Cellinese. (2014) The trouble with triplets in biodiversity informatics: a data-driven case against current identifier practices. *PLOS ONE* 9: e114069. doi: 10.1371/journal.pone.0114069

Stucky, B.J. (2013) Morphology, bioacoustics, and ecology of *Tibicen neomexicensis* sp. n., a new species of cicada from the Sacramento Mountains in New Mexico, U.S.A. (Hemiptera, Cicadidae, *Tibicen*). *ZooKeys* 337: 49-71. doi: 10.3897/zookeys.337.5950

Stucky, B.J. (2012) SeqTrace: a graphical tool for rapidly processing DNA sequencing chromatograms. *Journal of Biomolecular Techniques* 23: 90-93. doi: 10.7171/jbt.12-2303-004

D. Synergistic Activities

1. Open-source software. I have made substantial contributions to a variety of open-source software projects for scientific and general computing. As an example, I am the principal author and maintainer of OntoPilot (https://github.com/stuckyb/ontopilot/), which is a software program for developing and deploying ontologies.

2. Informatics education. Since 2016, I have served as an instructor or organizer for numerous Carpentries workshops at the University of Florida. These workshops teach undergraduates, graduate students, staff, and faculty basic principles of software development, data management, and data analysis.

3. Educational outreach. I have given interactive presentations about insects to students at two elementary schools in Gainesville, Florida, including a class of students for whom English is a secondary language.

4. Mentoring. Since 2016, I have regularly mentored undergraduate researchers in scientific software development and testing as part of various open-source software projects.

5. Professional service. I am a founding member of the Carpentries advisory board at the University of Florida (founded in 2017) and I have been an elected board member ever since then. The advisory board is responsible for supporting, managing, and directing Carpentries teaching activities at the University of Florida, including computer programming and data skills workshops and new instructor trainings.

Current and Pending Support

Investigator: Brian Stucky

Pending Support

Project/proposal title (this proposal): OR-DRD-AI2020: Using AI to uncover decades of global ecological change
Source of support: UF Research, Artificial Intelligence Research Catalyst Fund
Role: PI
Total amount requested: $47,449
Location of project: University of Florida
Award period covered: November 2, 2020 – November 1, 2021
Time: 3.0 person-months in year 1

Project/proposal title: OR-DRD-AI2020: Artificial intelligence for large-scale identification of insect pests and biodiversity discovery
Source of support: UF Research, Artificial Intelligence Research Catalyst Fund
Role: co-PI
Total amount requested: $49,979
Location of project: University of Florida
Award period covered: November 1, 2020 – October 31, 2021
Time: 1.8 person-months in year 1

Project/proposal title: Collaborative Research: Origins and drivers of extinction of Caribbean avifauna
Source of support: National Science Foundation
Role: co-PI
Total amount requested: $278,919 to UF
Location of project: University of Florida
Award period covered: January 1, 2021 – December 31, 2023
Time: 2.0 person-months per year in years 1 and 2

Current Support

Project/proposal title: Machine learning for high-throughput analysis of bioacoustic data
Source of support: Florida Museum of Natural History Research and Collections Grants
Role: PI
Total amount: $11,916
Location of project: University of Florida
Award period covered: January 1, 2020 - December 31, 2020
Time: 2 person-months per year (16.7% in year 1)

Biographical Sketch -- Robert P. Guralnick

A. Professional Preparation

U.C. Berkeley	Berkeley, CA	Pyshcology BA with high honors	1992
U.C. Berkeley	Berkeley, CA	Integrative Biology Doctor of Philosophy	1999
U.C. Berkeley	Berkeley, CA	Postdoctoral Fellow	1999

B. Appointments

Curator	2019-pres.	Dept. of Nat. Hist.	University of Florida
Associate Curator	2015-pres.	Dept. of Nat. Hist.	University of Florida
Associate Professor	2007-2014	Ecol. & Evol. Biol.	University of Colorado
Assistant Professor	2000-2007	Ecol. & Evol. Biol.	University of Colorado
Curator of Zoology	2000-2014	CU Mus. Nat. Hist.	University of Colorado

C.i Five Most Relevant Products

Allen, J., R. Folk, P. Soltis. D. Soltis*, and **R. Guralnick***. 2019. Biodiversity synthesis across the green branches of the tree of life. *Nature Plants* **5**: 11–13

Kissling, D, R. Walls, A. Bowser, M. O. Jones, J. Kattge, D. Agosti, J. Amengual, A. Basset, P. M. van Bodegom , J. H. C. Cornelissen, E. G. Denny, S. Deudero, W. Egloff , S. C. Elmendorf, E. A. García, K. D. Jones, O. R. Jones , S. Lavorel, D. Lear, L. M. Navarro,, S. Pawar , R. Pirzl, N. Rüger, S. Sal, R. Salguero-Gómez, D. Schigel, K-S Schulz, A. Skidmore, & **R. P. Guralnick***. 2018. Towards global data products of Essential Biodiversity Variables (EBVs) on species traits. *Nature Ecology and Evolution* 2:1531–1540.

Li, D. B. Stucky, J. Deck, B. Baiser, and **R. P. Guralnick**. 2019. The effect of urbanization on plant phenology depends on regional temperature. *Nature Ecology and Evolution 3:1661–1667.*

Stucky, B., R. Guralnick, J. Deck, E. Denny, K. Bolmgren and R. Walls. 2018. The Plant Phenology Ontology: A new informatics resource for large-scale integration of plant phenology data. *Frontiers in Plant Sciences 9:517.*

Brenskelle, L., B. Stucky, J. Deck, R. Walls, **R. P. Guralnick**. 2019 . Integrating herbarium specimen observations into global phenology data systems. *Applications in Plant Sciences 7(3): e01231.*

C.ii Other Products

Folk, R. A, R. L. Stubbs, M. E. Mort, J. M. Allen, P. Soltis, D. E. Soltis* and **R. P. Guralnick***. 2019. Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. *Proceedings of the National Academy of Sciences* 116 (22) 10874-10882.

Sun, M., R. Folk, M. Gitzendanner , P. Soltis , Z. Chen , D. Soltis , **R. Guralnick**. 2020. Recent, accelerated diversification in rosids occurred outside the tropics. *Nature Communications* 11:3333.

Jetz W., M. McGeoch, R. Guralnick, S. Ferrier, J. Beck, M. Costello, M. Fernandez, G. Geller, P. Keil, C. Merow, C. Meyer, F. Muller-Karger, H. Pereira, E. Regan, D. Schmeller, E. Turak. 2019. Essential information for mapping and monitoring of species populations for policy, management, and education. *Nature Ecology and Evolution*.

Jetz, W, J. Cavender-Bares, R. Pavlick, D. Schimel, F. Davis, G. P. Asner, **R. Guralnick**, J. Kattge, A. Latimer, P. Moorcroft, M. E. Schaepman. M. P. Schildhauer, F. D. Schneider, F. Schrodt, U. Stahl, S. L. Ustin. 2016. A global remote sensing mission to detect and predict plant functional biodiversity change. *Nature Plants* 2(3):16024.

Barve V.V, L. Brenskelle, D. Li, B. J. Stucky , N, V. Barve, M. M. Hantak, B. S. McLean, D. J. Paluh,, J. A. Oswald, M. Belitz, R. Folk, **R. Guralnick***. 2020. Methods for broad-scale plant phenology assessments using citizen scientists' photographs: a case study in Yucca. *Applications in Plant Sciences 8(1): e11315.*

D. Synergistic Activities

1) Associate Editor and Editorial Board for the journal Biodiversity Informatics and PLOS ONE, both fit my ethical stance on open access and open data.

2). 3-years serving on the reformed NEON Science, Technology & Education Advisory Committee (STEAC) including co-Chair for the Broadening Participation Committee on STEAC (2016-2018) and currently Co-Chair of the STEAC (May 2019-present).

3) Current Invited Member of NASA's Biological Diversity and Ecological Forecasting Working Group, program committee member for the Biodiversity Next meeting upcoming in October 2019, Data Task Force Lead for the Group on Earth Observations – Biodiversity Observation Network and Steering Committee Member Alliance for Biodiversity Knowledge.

4) Past President and Vice President, and Board of Directors member of the Board of the JRS Biodiversity Foundation (2011 - 2017). This is a primarily grant-making foundation that works in Africa and Latin America. The President and VP roles required significant time during the year plus two in-person board meetings a year.

5) My lab group has always remained diverse along gender and cultural lines, and I actively strive to engage more underrepresented groups into my lab and Museum collections; this includes female, latino, and international PhD and postdoctoral researchers.

Current and Pending Funding: Robert Guralnick
PI except where listed as a Co-PI or Senior Personnel
Current funding except where listed as Pending or Recommended

Pending, This Proposal
Co-PI
OR-DRD-AI2020: Using AI to uncover decades of global ecological change
Source of Support: UF Research, Artificial Intelligence Research Catalyst Fund
Amount of Award: $47,449
Period covered: 11/02/20-11/01/2021
Location of Project: University of Florida
Person-months committed to the project: 0.24

Pending
Collaborative PI
Collaborative Research: North American plants past, present, and future: Spatial phylogenetics on a continental scale.
Source of Support: NSF DEB
Amount of Award: $ $263,777
Period covered: 1/1/2021-12/31/2024
Location of Project: University of Florida
Person-months committed to the project: 0

Pending,
Co-Investigator
Cross-realm contrasts of scale: an integrated analysis of rates, patterns, and drivers of biodiversity change across marine and terrestrial systems.
Source of Support: NASA ROSES
Amount of Award: $176,067
Period covered: March 5, 2021-March 4, 2022
Location of Project: University of Florida
Person-months committed to the project: 0

Pending
Collaborative PI.
Collaborative Research: Origins and drivers of extinction of Caribbean Avifauna
Amount of Award: $278,919
Source of Support: NSF DEB
Period covered: 01/01/2021-12/31/2023
Location of Project: University of Florida
Person-months committed to the project: 0

Current: Collaborative PI.
Collaborative Research: GEODE: Genealogy, Geography and Ecology of Odonata: the first resolved evolutionary history and global biogeography of an entire insect order
Amount of Award: $233,438
Source of Support: NSF DEB
Period covered: 4/1/2020-3/31/2023
Location of Project: University of Florida

Person-months committed to the project: .1 (summer)

Current, NASA AIST Analytics Center
Co-Investigator
An Analytic Center for Biodiversity and Remote Sensing Data Integration
Souce of Support: NASA AIST
Amount of Award: $79,628
Period covered: 10/1/2019-9/31/2021
Location of Project: University of Florida
Person-months committed to the project: .15 months (summer)

Current, Collaborative PI.
IIBR RoL: Collaborative Research: A Rules Of Life Engine (RoLE) Model to Uncover Fundamental Processes Governing Biodiversity
Source of Support: NSF IIBR
Amount of Award: $293,949
Period covered: 9/1/2019-8/31/2022
Location of Project: University of Florida
Person-months committed to the project: .15 months per year (summer)

Current. Collaborative PI.
Collaborative Research: ABI Innovation: FuTRES, an Ontology-Based Functional Trait Resource for Paleo- and Neo-biologists
Source of Support: NSF ABI, $ 286,504 to UF
Period covered: 09/01/2018-8/31/2021
Location of Project: University of Florida
Person-months committed to the project: .5 months summer 2019-2021

Current. Co-Investigator.
Activities to advance, build, and deliver remote-sensing supported species distribution and species abundance EBVs.
Source of support: NASA GEO. $600K total, $59,925 to UF
Period covered: 09/20/17-8/31/2020
Location of Project: University of Florida
Person-months committed to the project: .2 months

Current. Collaborative PI
Collaborative Research: VertLife Terrestrial: A complete, global assembly of phylogenetic, trait, spatial and environment characteristics for a model clade
Amount of award: $2.5 million
Source of Support: NSF GoLife
Period covered: 10/1/2014-9/30/2020
Location of Project: University of Florida
Person months Committed to Project: .4 months summer.

BIOGRAPHICAL SKETCH: PAMELA S. SOLTIS

Florida Museum of Natural History & UF Biodiversity Institute
University of Florida
Gainesville, FL 32611
e-mail: psoltis@flmnh.ufl.edu

(a) Professional Preparation:

Central College	Pella, IA	Biology	B.A., 1980
University of Kansas	Lawrence, KS	Botany	Ph.D., 1986

(b) Appointments:

Director, University of Florida Biodiversity Institute, 2016 – present
Curator, Florida Museum of Natural History, and Professor, Genetics Institute, University of Florida, October, 2000 – present
Distinguished Professor, University of Florida, 2007 – present
Fulbright Distinguished Scholar, Royal Botanic Gardens, Kew, England, and Imperial College, Silwood Park, England, 2000 – 2001
Assistant, Associate, and Full Professor, Department of Botany, Washington State Univ., 1986-2000
Mellon Senior Fellow, Smithsonian Institution, 1994-95

(c) Products:
Five publications most closely related to the proposed project:

Soltis, P. S., G. Nelson, A. Zare, and E. K. Meineke. 2020. Plants meet machines: Prospects in machine learning for plant biology. *Applications in Plant Sciences* 8(6): e11371. doi:10.1002/aps3.1137.

Pearson, K. D., G. Nelson, M. F. J. Aronson, P. Bonnet, L. Brenskelle, C. C. Davis, E. G. Denny, E. R. Ellwood, H. Goëau, J. M. Heberling, A. Joly, T. Lorieul, S. J. Mazer, E. K. Meineke, B. J. Stucky, P. Sweeney, A. E. White, and P. S. Soltis. 2020. Machine learning using digitized herbarium specimens to advance phenological research. *BioScience*: biaa044, https://doi.org/10.1093/biosci/biaa044.

Soltis, P. S., G. Nelson, A. Zare, and E. K. Meineke. 2020. Plants meet machines: Prospects in machine learning for plant biology. Applications in Plant Sciences 8(6): e11371.
doi:10.1002/aps3.11371

Soltis, P. S., G. Nelson, A. Zare, and E. K. Meineke. 2020. Plants meet machines: Prospects in machine learning for plant biology. Applications in Plant Sciences 8(6): e11371.
doi:10.1002/aps3.11371

Soltis, P. S., G. Nelson, A. Zare, and E. K. Meineke. 2020. Plants meet machines: Prospects in machine learning for plant biology. Applications in Plant Sciences 8(6): e11371.
doi:10.1002/aps3.11371

Soltis, P. S., G. Nelson, A. Zare, and E. K. Meineke. 2020. Plants meet machines: Prospects in machine learning for plant biology. Applications in Plant Sciences 8(6): e11371.
doi:10.1002/aps3.11371

Lorieul, T., K. D. Pearson, E. R. Ellwood, H. Goëau, J.-F. Molino, P. W. Sweeney, J. M. Yost, J. Sachs, E. Mata-Montero, G. Nelson, P. S. Soltis, P. Bonnet, and A. Joly. 2019. Toward a large-scale and deep phenological stage annotation of herbarium specimens: case studies from temperate, tropical, and equatorial floras. *Applications in Plant Sciences* 7(3): e01223.

Allen, J., Folk, R. A., P. S. Soltis, D. E. Soltis, and R. P. Guralnick. 2019. Big data and biodiversity: Big challenges and broad applications. *Nature Plants* 5: 11-13.

Willis, C. G., E. R. Ellwood, R. B. Primack, C. C. Davis, K. D. Pearson, A. S. Gallinat, J. M. Yost, G. Nelson, S. J. Mazer, N. L. Rossington, T. H. Sparks, and P. S. Soltis. 2017. Old plants, new tricks: Phenological research using herbarium specimens. *Trends in Ecology and Evolution* 32: 531–546.

Five additional relevant publications:

Ellwood, E. R., J. A. Sessa, J. K. Abraham, A. E. Budden, N. Douglas, R. Guralnick, E. Krimmel, T. Langen, D. Linton, M. Phillips, P. S. Soltis, M. Studer, L. D White, J. Williams, and A. K. Monfils. 2020. Biodiversity science and the twenty-first century workforce. *BioScience* 70: 119–121.

Lu, L-M., L-F. Mao, T. Yang, J-F. Ye, B. Liu, H-L. Li, M. Sun, J. T. Miller, S. Mathews, H-H. Hu, Y-T. Niu, D-X. Peng, Y-H. Chen, S. A. Smith, M. Chen, K-L. Xiang, C-T. Le, V-C. Dang, A-M. Lu, P. S. Soltis, D. E. Soltis, J-H. Li, and Z-D. Chen. 2018. Evolutionary history of the angiosperm flora of China—cradle vs. museum. *Nature*: doi:10.1038/nature25485.

Soltis, P. S. 2017. Digitization of herbaria enables novel research. *American Journal of Botany* 104: 1-4.

Lacey, E. A., T. T. Hammond, R. E. Walsh, K. C. Bell, S. V. Edwards, E. R. Ellwood, R. Guralnick, S. M. Ickert-Bond, A. R. Mast, J. E. McCormack, A. K. Monfils, P. S. Soltis, D. E. Soltis, and J. A. Cook. 2017. Climate change, collections and the classroom: using big data to tackle big problems. *Evolution: Education and Outreach* 10:2.

Cook, J. A., S. V. Edwards, E. Lacey, R. P. Guralnick, P. S. Soltis, D. E. Soltis, C. Welch, K. C. Bell, K. E. Galbreath, C. Himes, J. Allen, T. A. Heath, A. C. Carnaval, K. L. Cooper, M. Liu, and and J. Hanken. 2014. Aiming Up: Natural history collections as emerging resources for innovative undergraduate education in biology. *BioScience* 64:725-734.

(d) Synergistic Activities (recent representative activities):

iDigBio, Director for Research (2011-present): In this position, I have had the pleasure of working with the broad community of natural history collections professionals and have contributed to the following: developed scientific requirements for iDigBio portal, led development of US Virtual Herbarium portal (in progress), developed use cases for specimen data in research & education, served as an organizer/co-organizer of research-related workshops and symposia, including a Plant Phenology Workshop, March, 2016, and Phenology and Deep Learning Workshop, January, 2019, served as organizer/co-organizer of training workshops on data analysis, served on the Organizing Committee for the Phenome Conference, and currently serving on the National Academy of Sciences' Committee on Biological Collections.

Professional Service (past five years): I am active in professional service, including during the past five years, the Board of Reviewing Editors, *Science* (2018-present), the NCBI Board of Scientific Counselors (2018-present), President-Elect, President, Past President, American Society of Plant Taxonomists (2018-21), the Merit Awards Committee of the Botanical Society of America (2013-15), Representative to SACNAS (2018), Board, American Institute of Biological Sciences (2018-present).

Editorial Service (past five years): I also serve in several editorial positions, including Consulting Editor, *The Plant Cell* (2015-present), Senior Associate Editor, *Applications in Plant Sciences* (2013-present), Associate Editor, *PhytoKeys* (2010-present), Board of Reviewing Editors, *Science* (2018-present), and handling editor for the *Proceedings of the National Academy of Sciences, USA* (2016-present).

NSF Workshops (past five years): I am a frequent participant in workshops sponsored by NSF, such as the US-China Biodiversity Workshop in Raleigh, NC (2015) and Hangzhou, China (2016), the Dimensions of Biodiversity PIs Meeting in Arlington, VA (2016), the Dimensions of Biodiversity Data Management Workshop in Gainesville, FL (2017), the Biodiversity Collections Network RCN Data Citation and Attribution Workshop in Lawrence, KS (2018), and the Biodiversity Literacy in Undergraduate Education Workshop in Berkeley, CA (2018).

CURRENT AND PENDING SUPPORT

Investigator: Pamela S. Soltis

Pending Support

Project/Proposal Title (this proposal): OR-DRD-AI2020: Using AI to uncover decades of global ecological change (Co-PI)
Source of Support: UF Research, Artificial Intelligence Research Catalyst Fund
Total Amount: $47,449
Total Award Period: 11/2/2020-11/1/2021
Location of Project: University of Florida
Time: 2% (0.24 person-months)

Project/Proposal Title: Collaborative Research: North American plants past, present, and future: Spatial phylogenetics on a continental scale (Co-PI)
Source of Support: NSF
Total Amount: $263,777
Total Award Period: 1/1/2021-12/31/2023
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: OPUS: Genome doubling in an evolutionary model: Synthesis across biological and temporal scales (Co-PI)
Source of Support: NSF
Total Amount: $321,782
Total Award Period: 6/1/2021-5/31/2023
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: SABI: iDigBio Phase 3: Sustaining the Digitization, Mobilization, Accessibility, and Use of Biodiversity Specimen Data in U.S. Museum and Academic Collections (Co-PI)
Source of Support: NSF
Total Amount: $19,477,861
Total Award Period: 9/1/2021-8/31/2026
Location of Project: University of Florida
Time: 1.0 mo/yr

Current Support

Project/Proposal Title: Infrastructure for Predicting, Understanding, and Mitigating Zoonotic Disease Outbreaks (PI)
Source of Support: NSF
Total Amount: $77,128
Total Award Period: 9/1/2020-8/31/2022
Location of Project: University of Florida (off campus)
Time: 0.5 mo/yr

Project/Proposal Title: CIBR: Collaborative Research: Integrating data communities with BiotaPhy: A computational platform for data-intensive biodiversity research and training (Co-PI)
Source of Support: NSF
Total Amount: $880,172
Total Award Period: 8/1/2019-7/31/2022
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: EDGE CT: Improving and streamlining systems for functional studies of non-model plants (Co-PI)
Source of Support: NSF
Total Amount: $849,007

Total Award Period: 8/1/2019-7/31/2022
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: Origins and impacts of nitrogen-fixing symbioses in a major clade of flowering plants (Co-PI)
Source of Support: NSF
Total Amount: $749,003
Total Award Period: 1/1/2020-12/31/2022
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: Collaborative Research: Temperate radiations and tropical dominance: the diversification and evolution of the plant clade Ericales (Co-PI)
Source of Support: NSF
Total Amount: $679,310
Total Award Period: 8/1/2019-7/31/2022
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: Phylogenomic discovery and engineering of nitrogen fixation into the bioenergy woody crop poplar (Co-PI)
Source of Support: DOE Plant Systems Design for Bioenergy Program
Total Amount: $4.1 million to UF ($7.3 million total)
Total Award Period: 9/15/2017-9/14/2022
Location of Project: University of Florida
Time: 0.25 mo/yr in yrs 1-3; 0.5 mo/yr in yrs 4-5

Project/Proposal Title: Digitization: iDigBio: Integrated Digitized Biocollections Phase 2 (Co-PI)
Source of Support: NSF
Total Amount: $15,486,747
Total Award Period: 7/1/16-6/30/21
Location of Project: University of Florida
Time: 1.0 mo/yr

Project/Proposal Title: Collaborative Research: ABI Innovation: Connecting resources to enable large-scale biodiversity analyses (Co-PI)
Source of Support: NSF
Total Amount: $834,828
Total Award Period: 5/1/15-6/30/21
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: PGRP: Evolution of specialized metabolite biosynthetic pathways in the Lamiaceae: Sources of chemical diversity for molecules essential for human use and plant defense (Co-PI)
Source of Support: NSF
Total Amount: $810,477 to UF ($4.1 million total)
Total Award Period: 4/1/15-3/31/21
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: Dimensions US-China: Collaborative Research: How historical constraints, local adaptation, and species interactions shape biodiversity across an ancient floristic disjunction (PI)
Source of Support: NSF
Total Amount: $1,199,043
Total Award Period: 1/1/15-12/31/20
Location of Project: University of Florida
Time: 0.5 mo/yr

DOUGLAS E. SOLTIS

Florida Museum of Natural History, Department of Biology, & the Genetics Institute,
University of Florida
Gainesville, FL 32611
email: dsoltis@ufl.edu

a. Professional Preparation

College of William and Mary	Williamsburg, VA	Biology	B.S.,1975
Indiana University	Bloomington, IN	Biology	M.A.,1977
Indiana University	Bloomington, IN	Biology	Ph.D.,1980
University of British Columbia	Vancouver, BC	Biology	1981

b. Appointments

Co-Director Grad Program, Genetics and Genomics, University of Florida, 2016-present
Distinguished Professor, University of Florida, 2008-present
Chair, Department of Botany, University of Florida, 2006-2008
Professor, University of Florida, 2000-present
Professor, Washington State University, 1990-2000
Acting Director of the Ownbey Herbarium, 1990-91
Associate Professor, Washington State University, 1986-1990
Assistant Professor, Washington State University, 1983-1986
Assistant Professor, The University of North Carolina at Greensboro, 1980-1983
Postdoctoral experience, University of British Columbia, summer 1981
William R. Ogg Fellowship, Indiana University 1979-1980

c. Products

(i) Five publications most closely related to the proposed project

Soltis, D.E. and P. S. Soltis. 2017. Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity. *Plant Diversity*, 38 (2016) 264e270.

Lu, L. M., Mao L, Yang T, Ye JF, Liu B, Li HL, Sun M, Miller JT, Mathews S, Hu HH, Niu YT, Peng DX, Chen YH, Smith SA, Chen M, Xiang KL, Le CT, Dang VC, Lu AM, Soltis PS, Soltis DE, Li JH, Chen ZD. 2018. Evolutionary history of the angiosperm flora of China. *Nature* 554:234–238. doi:10.1038/nature25485.

Allen, J., Folk, R.A., P.S. Soltis, D.E. Soltis, R.P. Guralnick. 2018. Big data and biodiversity: Big challenges and broad applications. *Nature Plants*, 5:11-13.

Allen, J., C. Germain-Aubrey, N. Barve, K. M. Neubig, L. C. Majure, S. W. Laffan, B. D. Mishler, H. Owens, S. A. Smith, W. M. Whitten, J. R. Abbott, D. E. Soltis, R. Guralnick, P. S. Soltis. 2019. Spatial phylogenetics of Florida vascular plants: The effects of tree uncertainty and ultrametricity. *iScience*11: 57–70

Mishler, B., S W. Laffan, P. S. Soltis, S. A. Smith, D. E. Soltis, N. Barve, J. M. Allen and R. Guralnick. 2020. Spatial Phylogenetics of the North American Flora. Journal of Systematics and Evolution. 10.1111/jse.12590

(ii) Five additional relevant publications

Hinchliff, C.E., S. A. Smith, J. F. Allman, J. G. Burleigh, R. Chaudhary, L. M. Coghill, K. A. Crandall, J. Deng, B. T. Drew, R. Gazisg, K. Gude, D. S. Hibbettg, L. A. Katz, H. D. Laughinghouse, E. J. McTavish, P. E. Midford, C. L. Owen, R. H. Reed, J. A. Rees, D. E. Soltis, T. Williams, and K. A. Cranston. 2015. Synthesis of phylogeny and taxonomy into a Comprehensive tree of life. *Proc. National Acad. Sciences* 112: 12764–12769.

Soltis, D. E., P. S. Soltis, P.K. Endress, M. W. Chase, W. S. Judd, L. Majure, E. Mavrodiev, S. M. Manchester. 2018. *Phylogeny and evolution of the angiosperms*. New Edition, University of Chicago Press.

Folk, R.A., R. L. Stubbs, M. E. Mort, N. Cellinese, J. M. Allen, P. S. Soltis, D. E. Soltis, R. P. Guralnick. 2019. Uncoupled diversification, ecological niche and phenotype in a temperate radiation. *Proc. National Acad. Sciences* 116: 10874-10882.

Green Plant Consortium. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679–685

Sun, M., C. Germain-Aubrey, M. Gitzendanner, S. A. Smith, P. S. Soltis, R. A. Folk, R. Guralnick, Z. Chen, D. E. Soltis. 2020. Recent, accelerated diversification in rosids occurred outside the tropics. *Nature Communications* 11: 3333.

d. Synergistic Activities (five recent representative activities):

Co-Organizer of PAG nitrogen-fixation symposium, 2018-present
Associate editor for *American Journal of Botany*, *Webbia*, *Phytokeys*, *Critical Reviews in Plant Science*
Advisory Committee Hardwood Genomics—Plant Genome Grant, 2011-2017
Advisory Committee, University of Vienna—2014-present
Chinese Academy of Sciences—Visiting Professorship for Senior Scientists 2011-2013

CURRENT AND PENDING SUPPORT

Investigator: Douglas E. Soltis

Pending Support

Project/Proposal Title (this proposal): OR-DRD-AI2020: Using AI to uncover decades of global ecological change (Co-PI)
Source of Support: UF Research, Artificial Intelligence Research Catalyst Fund
Total Amount: $47,449
Total Award Period: 11/2/2020-11/1/2021
Location of Project: University of Florida
Time: 2% (0.24 person-months)

Project/Proposal Title: Collaborative Research: North American plants past, present, and future: Spatial phylogenetics on a continental scale (Co-PI)
Source of Support: NSF
Total Amount: $263,777
Total Award Period: 1/1/2021-12/31/2023
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: QLCI-CI: The Quantum L.E.A.P. (Leveraging Enzymes and Advanced Phylogenetics)
Source of Support: NSF
Total Amount: $16,156,875
Total Award Period: 12/1/2020-11/31/2023
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: OPUS: Genome doubling in an evolutionary model: Synthesis across biological and temporal scales (PI)
Source of Support: NSF
Total Amount: $321,782
Total Award Period: 6/1/2021-5/31/2023
Location of Project: University of Florida
Time: 0.5 mo/yr

Current Support

Project/Proposal Title: CIBR: Collaborative Research: Integrating data communities with BiotaPhy: A computational platform for data-intensive biodiversity research and training (Co-PI)
Source of Support: NSF
Total Amount: $880,172
Total Award Period: 8/1/2019-7/31/2022
Location of Project: University of Florida
Time: 0.5

Project/Proposal Title: EDGE CT: Improving and streamlining systems for functional studies of non-model plants (PI)
Source of Support: NSF
Total Amount: $849,007
Total Award Period: 8/1/2019-7/31/2022
Location of Project: University of Florida
Time: 0.5

Project/Proposal Title: Origins and impacts of nitrogen-fixing symbioses in a major clade of flowering plants (Co-PI)
Source of Support: NSF
Total Amount: $749,003
Total Award Period: 1/1/2020-12/31/2022
Location of Project: University of Florida
Time: 0.5

Project/Proposal Title: Collaborative Research: Temperate radiations and tropical dominance: the diversification and evolution of the plant clade Ericales (Co-PI)
Source of Support: NSF
Total Amount: $679,310
Total Award Period: 8/1/2019-7/31/2022
Location of Project: University of Florida
Time: 0.5

Project/Proposal Title: Phylogenomic discovery and engineering of nitrogen fixation into the bioenergy woody crop poplar (Co-PI)
Source of Support: DOE Plant Systems Design for Bioenergy Program
Total Amount: $4.1 million to UF ($7.3 million total)
Total Award Period: 9/15/2017-9/14/2022
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: Collaborative Research: ABI Innovation: Connecting resources to enable large-scale biodiversity analyses (PI)
Source of Support: NSF
Total Amount: $834,828
Total Award Period: 5/1/15-6/30/21
Location of Project: University of Florida
Time: 0.5

Project/Proposal Title: PGRP: Evolution of specialized metabolite biosynthetic pathways in the Lamiaceae: Sources of chemical diversity for molecules essential for human use and plant defense (Co-PI)
Source of Support: NSF
Total Amount: $810,477 to UF ($4.1 million total)
Total Award Period: 4/1/15-3/31/21
Location of Project: University of Florida
Time: 0.5 mo/yr

Project/Proposal Title: Dimensions US-China: Collaborative Research: How historical constraints, local adaptation, and species interactions shape biodiversity across an ancient floristic disjunction (Co-PI)
Source of Support: NSF
Total Amount: $1,199,043
Total Award Period: 1/1/15-12/31/20
Location of Project: University of Florida
Time: 0.5

NSF BIOGRAPHICAL SKETCH

NAME: Fortes, José

ORCID: 0000-0001-8870-5205

POSITION TITLE & INSTITUTION: Professor and AT&T Eminent Scholar, University of Florida

(a) PROFESSIONAL PREPARATION

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
Universidade de Angola	Luanda, Angola	Electrical Enginering	BSEE (Licenciatura in EE)	1978
Colorado State University	Fort Collins, CO	Electrical Engineering	MSEE	1981
University of Southern California	Los Angeles, CA	Electrical Engineering	Ph.D.	1984

(b) APPOINTMENTS

2001 - present	Professor and AT&T Eminent Scholar, University of Florida, Department of Computer and Information Science and Engineering, Gainesville, FL
2008 - 2015	Founding Director of the NSF Industry-University Cooperative Center for Cloud and Autonomic Computing, University of Florida, Gainesville, FL
1999 - 2001	Professor and Assistant Head of Education, Purdue University, School of Electrical and Computer Engineering, West Lafayette, IN
1995 - 2001	Professor, Purdue University, School of Electrical and Computer Engineering, West Lafayette, IN
1993 - 1994	Visiting Professor, Universitat Politécnica de Catalunya - Campus Terrassa, Departament d'Arquitectura de Computadors, Terrassa, Catalunya
1989 - 1995	Associate Professor, Purdue University, School of Electrical Engineering, West Lafayette, IN
1989 - 1990	Program Director, National Science Foundation, Microelectronics Information Processing Systems Division, Arlington, VA
1987 - 1989	Assistant Professor, Purdue University, School of Electrical Engineering, West Lafayette, IN
1984 - 1987	Visiting Assistant Professor, Purdue University, School of Electrical Engineering, West Layafette, IN

(c) PRODUCTS

Products Most Closely Related to the Proposed Project

1. SELFIE: Self-Aware Information Extraction from Digitized Biocollections. Icaro Alzuru, Andréa Matsunaga, Maurício Tsugawa, and José Fortes, IEEE 13th International Conference on e-Science, Auckland (New Zealand); ; c2017. DOI: doi: 10.1109/eScience.2017.19.
2. Quality-aware Human-Machine Text Extraction for Biocollections using Ensembles of OCRs. Icaro Alzuru, Rhiannon Stephens, Andréa Matsunaga, Maurice Tsugawa, Paul Flemons and José Fortes, IEEE 15th International Conference on eScience, San Diego (CA, USA); ; c2019.

3. Task Design and Crowd Sentiment in Biocollections Information Extraction. Icaro Alzuru, Andréa Matsunaga, Maurício Tsugawa, and José Fortes, The 3rd IEEE International Conference on Collaboration and Internet Computing; ; c2017.
4. Cooperative Human-Machine Data Extraction from Biological Collections. Icaro Alzuru, Andréa Matsunaga, Maurício Tsugawa and José A. B. Fortes,12th IEEE International Conference on e-Science, October 24, 2016, Baltimore (MD, USA); ; c2016.
5. Page L, MacFadden B, Fortes J, Soltis P, Riccardi G. Digitization of Biodiversity Collections Reveals Biggest Data on Biodiversity. BioScience. 2015 September 01; 65(9):841-842. Available from: http://academic.oup.com/bioscience/article/65/9/841/237621/Digitization-of-Biodiversity-Collections-Reveals DOI: 10.1093/biosci/biv104

Other Significant Products, Whether or Not Related to the Proposed Project

1. Lee G, Fortes J. Improving Data-Analytics Performance Via Autonomic Control of Concurrency and Resource Units. ACM Transactions on Autonomous and Adaptive Systems. 2019 March 28; 13(3):1-25. Available from: https://dl.acm.org/doi/10.1145/3309539 DOI: 10.1145/3309539
2. Reaching Consensus in Crowdsourced Transcription of Biocollections Information. Andréa Matsunaga, Austin Mast and José Fortes 10th IEEE International Conference on e-Science, Guarujá, Brazil, October 20-24, pp. 57 - 64; ; c2014. DOI: doi: 10.1109/eScience.2014.30
3. A Computational-and Storage-Cloud for Integration of Biodiversity Collections. Andréa Matsunaga, Alex Thompson, Renato J Figueiredo, Charlotte C Germain-Aubrey, Matthew Collins, Reed S Beaman, Bruce J MacFadden, Greg Riccardi, Pamela S Soltis, Lawrence M Page and José AB Fortes, IEEE 9th International Conference on eScience, 2013, Bejing, October 22, pp. 78-87; ; c2013. DOI: doi: 10.1109/eScience.2013.48
4. Elliott M, Poelen J, Fortes J. Toward reliable biodiversity dataset references. Ecological Informatics. 2020 September; 59:101132-. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1574954120300820 DOI: 10.1016/j.ecoinf.2020.101132
5. Matsunaga A, Mast A, Fortes J. Workforce-efficient consensus in crowdsourced transcription of biocollections information. Future Generation Computer Systems. 2016 March; 56:526-536. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0167739X15002277 DOI: 10.1016/j.future.2015.07.004

(d) SYNERGISTIC ACTIVITIES

1. (Current) Co-PI (PI is Larry Page) and Lead PI for cyberinfrastructure of iDigBio, a National Resource for Advancing Digitization of Biodiversity Collections (ADBC) funded by NSF. ADBC and iDigBio, make available data and images for millions of biological specimens for the research community, government agencies, students, educators, and the general public (http://www.idigbio.org).

Current and Pending Support

Investigator: Fortes, José

Pending Support

Project/proposal title: OR-DRD-AI2020: Using AI to uncover decades of global ecological change
Source of support: UF Research, Artificial Intelligence Research Catalyst Fund
Total amount requested: $47,449
Location of project: University of Florida
Award period covered: November 2, 2020 – November 1, 2021
Time: 0.12 person-months in year 1

Project/proposal title: SABI: iDigBio Phase 3: Sustaining the digitization, mobilization, accessibility, and use of biodiversity specimen data in U.S. museum and academic collections
Source of support: NSF
Total amount requested: $15,594,623
Location of project: University of Florida
Award period covered: September 2021 – August 2026
Time: 1.25 person-months in year 1, 1 person-month in years 2-4

Project/proposal title: READA: Using Multimodal Data Streams for Real-time Equitable and Agile Damage Assessment During and Post Severe and Tropical Cyclones
Source of support: NSF
Total amount requested: $398,848
Location of project: University of Florida
Award period covered: July 2021 – June 2024
Time: 0.2 person-months in year 1, 0.3 in year 2, 0.2 in year 3

Current Support

Project/proposal title: CIBR: Collaborative Research: Integrating data communities with BiotaPhy: a computational platform for data-intensive biodiversity research and training
Source of support: NSF
Total amount: $878,597
Location of project: University of Florida
Award period covered: August 2019- July 2022
Time: 0.38 person-months in year 1, 0.25 in year 2, 0.13 in year 3

Project/proposal title: EAGER: Towards the Web of Biodiversity Knowledge: Understanding Data Connectedness to Improve Identifier Practices
Source of support: NSF
Total amount: $299,973
Location of project: University of Florida
Award period covered: October 2018 – September 2021
Time: 0.13 person-months in year 1, 0.07 in year 2

Project/proposal title: Digitization: IDigBio: Integrated Digitized Biocollections Phase 2
Source of support: NSF
Total amount: $13,786,561
Location of project: University of Florida
Award period covered: September 2016 – August 2021
Time: 1.5 person-months in year 1, 1.0 person-months in years 2-4

Project/proposal title: US-EA CENTRA: US-East Asia Collaborations to Enable Transnational Cyberinfrastructure Applications
Source of support: NSF
Total amount: $519,575
Location of project: University of Florida
Award period covered: October 2015 – September 2021
Time: 0.5 person-months in year 1, 0.25 person-months in years 2-4

Project/proposal title: Collaborative Research: ABI Innovation: Connecting resources to enable large-scale biodiversity analyses
Source of support: NSF
Total amount: $834,828
Location of project: University of Florida
Award period covered: July 2015 – June 2021
Time: 1.5 person-months in year 1, 1.0 person-months in years 2-4

Project/proposal title: SAVI:PRAGMA: Enabling Scientific Expeditions and Infrastructure Experimentation for Pacific Rim Institutions and Researchers
Source of support: UC San Diego
Total amount: $1,226,020
Location of project: University of Florida
Award period covered: October 2012 – September 2020
Time: 0.2 person-months in year 1, 0.1 person-months in years 2-4

BIOGRAPHICAL SKETCH: MATTHEW A. GITZENDANNER

Department of Biology
University of Florida
Gainesville, FL 32611
e-mail: magitz@ufl.edu

(a) Professional Preparation:

Reed College	Portland, OR	Biology	B.A., 1994
Washington State University	Pullman, WA	Botany	Ph.D., 2000

(b) Appointments:

2017-present	Scientist, Department of Biology, University of Florida, Gainesville, FL
2011-present	Coordinator of User Training and Bioinformatics Specialist, UFIT Research Computing, University of Florida, Gainesville, FL
2008-2017	Associate Scientist, Dept. of Biology, University of Florida, Gainesville, FL
2001-2008	Assistant Scientist, Dept. of Biology, University of Florida, Gainesville, FL
2000-2001	Lecturer, School of Biological Sciences, Washington State Univ., Pullman, WA
1995-2000	Teaching Assistant, Washington State University, Pullman, WA

(c) Products:

(i) Five publications most closely related to the proposed project:

Stull, G. W., P. S. Soltis, D. E. Soltis, **M. A. Gitzendanner**, and S. A. Smith. 2020. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. American Journal of Botany 107:790–805.

Sun, M., R. A. Folk, **M. A. Gitzendanner**, P. S. Soltis, Z. Chen, D. E. Soltis, and R. P. Guralnick. 2020. Estimating rates and patterns of diversification with incomplete sampling: a case study in the rosids. American Journal of Botany 107:895–909.

Leebens-Mack, J. H., M. S. Barker, E. J. Carpenter, M. K. Deyholos, **M. A. Gitzendanner**, and the One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature, doi: 10.1038/s41586-019-1693-2.

Li, H.-T., T.-S. Yi, L.-M. Gao, P.-F. Ma, T. Zhang, J.-B. Yang, **M. A. Gitzendanner**, P. W. Fritsch, J. Cai, Y. Luo, H. Wang, M. van der Bank, S.-D. Zhang, Q.-F. Wang, J. Wang, Z.-R. Zhang, C.-N. Fu, J. Yang, P. M. Hollingsworth, M. W. Chase, D. E. Soltis, P. S. Soltis, and D.-Z. Li. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. Nature Plants 5:461-470.

Gitzendanner, M. A., P. S. Soltis, G. K.-S. Wong, B. R. Ruhfel, and D. E. Soltis. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. American Journal of Botany. 105: 291–301.

(ii) Five additional relevant publications:

Gitzendanner, M. A., P. S. Soltis, T.-S. Yi, D.-Z. Li, and D. E. Soltis. 2018. Plastome Phylogenetics: 30 Years of Inferences Into Plant Evolution. Pp. 293–313 *in* S.-M. Chaw and R. K. Jansen, eds. *Advances in Botanical Research*. Academic Press.

Carpenter, E. J., N. Matasci, S. Ayyampalayam, S. Wu, J. Sun, J. Yu, F. R. Jimenez Vieira, C. Bowler, R. G. Dorrell, **M. A. Gitzendanner**, L. Li, W. Du, K. K. Ullrich, N. J. Wickett, T. J. Barkmann, M. S. Barker, J. H. Leebens-Mack, and G. K.-S. Wong. 2019. Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *Gigascience* 8.

Wang, Y., W. Jiang, W. Ye, C. Fu, **M. A. Gitzendanner**, P. S. Soltis, D. E. Soltis, and Y. Qiu. 2018. Evolutionary insights from comparative transcriptome and transcriptome-wide coalescence analyses in *Tetrastigma hemsleyanum*. BMC Plant Biology 18:208.

Payton, A. C., A. A. Naranjo, W. Judd, **M. A. Gitzendanner**, P. S. Soltis, and D. E. Soltis. 2019. Population genetics, speciation, and hybridization in *Dicerandra* (Lamiaceae), a North American Coastal Plain endemic, and implications for conservation. *Conservation Genetics*, 20: 531-543. doi: 10.1007/s10592-019-01154-8.

N. El Rouby, C. W. McDonough, Y. Gong, L. A. McClure, B. D. Mitchell, R. B. Horenstein, R. L. Talbert, Y. Bradford, D. C. Crawford, M. D. Ritchie, **M. A. Gitzendanner**, A. Takahashi, T. Tanaka, M. Kubo, C. J. Pepine, R. M. Cooper-DeHoff, O. R. Benavente, A. R. Shuldiner, J. A. Johnson. 2018. Genome-wide association analysis of common genetic variants of resistant hypertension. *The Pharmacogenomics Journal*. 2018/09/20. PMID: 30237584

(d) Synergistic Activities (five recent representative activities):

UF Research Computing Training:

I organize and present the Research Computing training program, offering weekly sessions on topics from Linux command line to using the SLURM scheduler. I also present at lab meetings and student groups about the cluster and provide application-specific training on request. Many of these training topics are also available online as pre-recorded, sub-titled videos. In addition, I provide training for our secure data analysis environment ResVault. UF has recently partnered with NVIDIA to dramatically expand access to hardware for AI research and I am developing training and course curriculum modules in AI.

Editorial Service:

Senior Associate Editor, *Applications in Plant Sciences*, 2015-present

Associate Editor, *Frontier in Plant Sciences: Plant Systematics and Evolution*, August 2018 - 2020

Teaching:

BSC4452: Computational Tools for Research. 3-credits. I teach this University of Florida graduate/undergraduate course which provides an introduction to the Linux command-line, cluster computing, Bash and Python scripting and SQL databases. Fall 2018 had 35 students enrolled. Summer 2020 had 30 students.

Outreach and increasing diversity:

Coordination of and Python instructor for Gator Computer Program, a 2-week computer science program for 9^{th} and 10^{th} graders. July 2017, 2018, 2019.

Presentations for Howard Bishop Middle School and Skype-A-Scientist.

Outreach activities with various programs through the Florida Museum of Natural History, including Junior Science and Humanities Symposium and WiSEGirlz spring break camp.

Current and Pending Support

Investigator: Matthew Gitzendanner

Pending Support

Project/proposal title (this proposal): OR-DRD-AI2020: Using AI to uncover decades of global ecological change
Source of support: UF Research, Artificial Intelligence Research Catalyst Fund
Role: co-PI
Total amount requested: $47,449
Location of project: University of Florida
Award period covered: November 2, 2020 – November 1, 2021
Time: 1.2 person-months in year 1

Project/proposal title: Data Science Training for Future Leaders in Soil Health Research and Extension
Source of support: National Institute of Food and Agriculture
Role: co-PI
Total amount requested: $251,704
Location of project: University of Florida
Award period covered: January 1, 2021 – January 1, 2026
Time: 0.12 person-months per year

Project/proposal title: QLCI-CI: The Quantum L.E.A.P (Leveraging Enzymes and Advanced Phylogenetics) Institute
Source of support: NSF
Role: co-PI
Total amount: $18,964,251
Location of project: University of Florida
Award period covered: September, 2021 - August, 2026
Time: 0.12 person-months per year

References

1. WWF (2020) Living Planet Report 2020: Bending the Curve of Biodiversity Loss.
2. Díaz S, Settele J, Brondízio ES, et al. (2019) Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* 366: eaax3100.
3. IPBES (2019) Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services. https://zenodo.org/record/3553579
4. Soga M, Gaston KJ (2018) Shifting baseline syndrome: causes, consequences, and implications. *Front Ecol Environ* 16: 222–230.
5. Mihoub J-B, Henle K, Titeux N, et al. (2017) Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures. *Sci Rep* 7: 41591.
6. Bonebrake TC, Christensen J, Boggs CL, et al. (2010) Population decline assessment, historical baselines, and conservation: Inferring population declines. *Conservation Letters* 3: 371–378.
7. Menzel A, Sparks TH, Estrella N, et al. (2006) European phenological response to climate change matches the warming pattern. *Global Change Biology* 12: 1969–1976.
8. Cleland E, Chuine I, Menzel A, et al. (2007) Shifting plant phenology in response to global change. *Trends in Ecology & Evolution* 22: 357–365.
9. Reilly J, Baethgen W, Chege F, et al. (1996) Agriculture in a changing climate: impacts and adaptation, In: Watson R, Zinyowera M, Moss R (Eds.), *Climate Change 1995. Impacts, Adaptation and Mitigation of Climate Change: Scientific–Technical Analyses. Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, UK, Cambridge University Press, 427–467.
10. Visser ME, Both C (2005) Shifts in phenology due to global climate change: the need for a yardstick. *Proceedings of the Royal Society B: Biological Sciences* 272: 2561–2569.
11. Willis CG, Ruhfel B, Primack RB, et al. (2008) Phylogenetic patterns of species loss in Thoreau's woods are driven by climate change. *Proceedings of the National Academy of Sciences of the United States of America* 105: 17029–17033.
12. Enquist CAF, Kellermann JL, Gerst KL, et al. (2014) Phenology research for natural resource management in the United States. *International Journal of Biometeorology* 58: 579–589.
13. Willis CG, Ellwood ER, Primack RB, et al. (including **P. Soltis**) (2017) Old Plants, New Tricks: Phenological Research Using Herbarium Specimens. *Trends in Ecology & Evolution* 32: 531–546.
14. Davis CC, Willis CG, Connolly B, et al. (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: 1599–1609.
15. Cook JA, Edwards SV, Lacey EA, et al. (including **Guralnick**, **D. Soltis**, **P. Soltis**) (2014) Natural History Collections as Emerging Resources for Innovative Education. *BioScience* 64: 725–734.
16. Lacey EA, Hammond TT, Walsh RE, et al. (including **Guralnick**, **D. Soltis**, **P. Soltis**) (2017) Climate change, collections and the classroom: using big data to tackle big problems. *Evo Edu Outreach* 10: 2.
17. Ellwood ER, Sessa JA, Abraham JK, et al. (including **Guralnick**, **P. Soltis**) (2020) Biodiversity science and the twenty-first century workforce. *BioScience* 70: 119–121.
18. Pearson KD, Nelson G, Aronson MFJ, et al. (including **Stucky**, **P. Soltis**) (2020) Machine learning using digitized herbarium specimens to advance phenological research. *BioScience* 70: 610–620.
19. Lorieul T, et al. (including **P. Soltis**) (2019) Toward a large‐scale and deep phenological stage annotation of herbarium specimens: Case studies from temperate, tropical, and equatorial floras. *Appl Plant Sci* 7: e01233.
20. **Stucky BJ**, **Guralnick R**, Deck J, et al. (2018) The Plant Phenology Ontology: A new informatics resource for large-scale integration of plant phenology data. *Front Plant Sci* 9: 517.
21. Deck J, **Stucky BJ**, **Guralnick R**. (2020) Global Plant Phenological Data Portal. https://plantphenology.org/
22. Deck J, **Stucky BJ**, **Guralnick R**. (2018) rppo R package. https://github.com/ropensci/rppo
23. Brenskelle L, **Stucky BJ**, Deck J, et al. (including **Guralnick**) (2019) Integrating herbarium specimen observations into global phenology data systems. *Applications in Plant Sciences* 7: e01231.
24. **Stucky BJ**, Luc A (2018) OntoPilot: Software for ontology deployment. https://github.com/stuckyb/ontopilot/
25. Devlin J, Chang M-W, Lee K, et al. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:181004805*.
26. Tan M, Le QV (2020) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv:190511946*.