



#### iDigBio Data Use: Enabling Research Through Training, Integration, and Tools

#### Pamela S. Soltis University of Florida







#### Overview

- Training workshops
- Sample publications
- Developing research programs: data integration
- Promoting research use through symposia, workshops, and special publications
- Enabling research through tool development
- Enabling research through tool availability
- Publication metrics





#### Overview

- Training workshops
- Sample publications
- Developing research programs: data integration
- **Promoting research use** through symposia, workshops, and special publications
- Enabling research through tool development
- Enabling research through tool availability
- Publication metrics





## **Training Workshops**

- Specific skills in data use iDigBio 2 (July 2016)
  - Soltis team
  - Georeferencing, ecological niche modeling, etc.
  - Venues
    - Botany 2016 (~50)
    - Denver (CU-Denver) (~25)
    - Botany 2017 (~45)
    - Applied to Ecological Society of America 2017
  - Webinar series planned for spring, 2018
  - Very positive reviews
  - Using the materials to train undergraduates, e.g.
     'mini-REU site program', summer 2017





## 'Mini-REU Site' Program - Summer 2017

- 5 students
  - 3 from UF, 1 from UCF, 1 from USC
  - 3 women, 2 men
  - Summer research
  - Final symposium
  - 2 presentations @

Botany 2017







## Training Workshops

- Data/Software Carpentry iDigBio 2 (July 2016)
  - M. Collins, D. Paul et al.
  - TDWG 2016
  - TDWG 2017
  - Florida State University 2017
  - ADBC Summit 2017
  - 8 workshops @ UF; UFBI/UFII
- Developing Data Carpentry lessons using iDigBio
- Plans:
  - Establishing group to share Carpentries membership
  - Instructor training at TDWG/SPNHC
  - Talk by Carpentries development manager at TDWG/SPNHC





#### **Sample Publications**

- Marchant et al. 2016 ecology of polyploid plants
- Willis et al. 2017 phenology review
- James 2017 new species

Northque	Rectar	Trends in Ecology & Exclusion	CelPress	EDENAUXER JOCENAL OF BOTANY. THE HA HA (RM) HE E Brann of Da Real Brann Cardin Lindwigh (2011)
Patterns of abiotic n progenitors	iche shifts in allopolyploids relative to their			dadi 10 cm 15 cm 4 pm 1 monork
		Heview		RHODODENDRON STANLEYI SJAMES & ARGENT:
Di Malan Marchandr <sup>14</sup> , Doughu K, Schöl <sup>3,2,4</sup> and Panada S, Schöl <sup>3,4,4</sup>		Old Plants, New Tricks: Phenological Research Using		A NEW RHODODENDRON SPECIES (ERICACEAE, SUBGENUS VIREYA) FROM PAPUA NEW GUINEA
	Lannary	Herbarium Specimens		S. A. JAMES' & G. ARDENT"
Kalife Interneting andread 1 March Marchan 1942 March 2010 1942 Marca Charles Andread 1943 Marca Charles Anton Acceleration (2010)	<ul> <li>Polytely has influence growth, physicity of a morphological and independent variabilities which the patients modeling the period in a constraintigati companion of polytechnik and being explaints which does there are noting on index well tangets advance.</li> <li>Post dark investment of all anguards provide the second polytect and these anguards are not an explaint of the second polytect and the second polytect and the anguards and the second polytect and the second anguards and the development of the endpolytect and the second anguards and advances of the endpolytect and index polytect and the second polytect and anguards and the second polytect and the second anguards and advances of the endpolytect and index polytect and advances of the endpolytect andvances of the endpolytect and advances of the endpolytect an</li></ul>	Charles G. Wile, <sup>1,1</sup> Basteth R. Bwood, <sup>1,4</sup> Richard B. Primeck, <sup>1</sup> Charles C. Davis, "Katelin D. Reeson, <sup>6</sup> Amanda S. Galfrat, <sup>2</sup> Jann M. Yost, <sup>1</sup> Gil Natson, <sup>2</sup> Busen J. Mazor, <sup>5</sup> Natalie L. Ressington, <sup>1</sup> Tim H. Sparka, <sup>4,7</sup> and Pamelia S. Soldi <sup>6</sup>		Bhulindowlow stanlays S.Janen B. August is described as a new species from Mount Wale, Central Province, Papus New Guman, Ils montphological positions in the webgems is discussed and the differences given from the most clearly similar species. A new on the habitat and conservation assessment is also provided. Keywood: Ericanome, new species, Papus Your Guman, Bhudidandow, subgema Thrys.
Ann Annunger (2010) Ann eine Ground ann an Annunger (2011) Annunger (2014), and Annung an Angel Annunger (2014), and Annung an Angel Annunger (2014) Annunger	• This devotes have parties on others levels, the para paragraphic Harden exploration, their sections of the proceed of the levels of the paragraphic devotes levels levels of the paragraphic devotes levels devotes and the paragraphic devotes levels devotes levels. The paragraphic angle asymptotic seque belows the devotes levels dev	The through of phonomeganal execute, south as leaf-out and Bewering, ethnological influence global executes and their study is visual to content leaderships from particle mesoned in leadership chicago, if thereas grant executes. Neverexis, is called to the horspace of phonomeganic, or phylogenetics to space of available data. Nevertherin and the horspace of phonomeganic, the phylogenetics to space of available data. Nevertherin and the horspace of phonomeganic phonomeganic to the phonomeganic phonomeganic phonomeganic phonomeganic phonomeganic phonomeganic phonomeganic explorations. Horizontow spacement is supported on characterize phonomeganic to characterize and there executed phonomeganic to show the state phonological phonomeganic and there executed phonomeganic to show the state phonomeganic to characterize to characterize the state of the phonomeganic to state phonological to state phonomeganic to characterize the state of the state state and phonological phonomeganic to characterize the state of the state of the state state and phonological transmission to characterize	Terration Terrating in a manual filtering factor of a minimum stress and a stress of terration stress and a stress of terration stress magnetic actions to report terration and stress terrations terration and stress terrations terration and stress terrations terration and stress terrations terration and stress terrations terrations and terrations terrat	INTERIOR/CTION New Guines already has more species of Pixys thededendrons than any of the other Malarian Education and approximately them times the number of species as the next most species-rich kindl, Bernon (Appen, 793). This reserve interest montantic building of the tistend has ind via a situation analogous to that of the Himaleys. In other genet area of diversification of Maladaindows, Green that many article of New Guines are tell
Howhichter Michiels Dahlegenens Agiliatist, ins fens of general materials in the mean of Appendix inspections, in the means of Appendix inspections, in the particular of particular inspections being and appendix inspections, and the Appendix and appendix in the Appendix and appendix in and oncoding (PA) the meaning appendix and oncoding (PA) the meaning and oncoding (PA) the meaning appendix appendix in a meaning appendix in an integra general applicity in a meaning appendix applications in a meaning appendix applications in a meaning appendix applications in a meaning appendix applications in a meaning appendix applications in a meaning appendix applications in a meaning appendix applications in a meaning appendix applications in a meaning applications appl	KGD, for some anisotic sectors of the sector of the sec	We sprace the convert status of behaviour based philosophic research, storethy painted the tases and therapic on the convertised status of the status based of spectrums, data, and introducts factors spectrameter for philosophic instances of spectrums, the based on the status based on the convertised based based of the convertised based on the status based on the convertised based based of the convertised based on the status based on the convertised based based on the convertised based on the status based on the convertised based based on the convertised based on the status based on the convertised based based on the convertised based on the status and the convertised based based on the convertised based on the status and the convertised based based on the convertised based on the status based on the convertised based based to the status based on the status of the status based on the convertised based based on the status based on the status based on the convertised based based based on the status based on the status based on the status based based based based on the status based on the status based on the status based based based based based based on the status based on the status based on the status based ba	And be and the addressing provides the basis. All the address is the address of the address is the address of the address of the address of the the address of the address of the address of the address	Birk known or addresses betasically, it is not surprising that any exploration is likely to provide investigation. <b>TAEUNCODE: TRAATMENT</b> <b>Biodecodem stability</b> 15. Same 4. Append was. Similar to Koludo-endows known politicalismus Shrumar the differing in hering contains in bases (not remained to invasidade), a smooth upper surface to the leaf after the sudie have goes (not permittenily nough-pupillose), long as without persistent to distance it is an advected with the subscription of the sudie that the distance it and a fine-aligned right as 2 mm in distance it to the leaf after the sudie have goes (not permittenily nough-pupillose), long without be subscription distance it and a fine-aligned right as 2 mm in distance it to the biologic distance distance it and a fine-aligned right as 2 mm in distance it to the biolic distance distance it and a fine-aligned right as 2 mm in distance it to the biolic distance distance it and a fine-aligned right as 2 mm in distance it to the biolic distance distance it and a fine-aligned right as 2 mm in the biolic distance it distance it and a fine-aligned right as 2 mm in the biolic distance it distance it and a fine-aligned right as 2 mm in the biolic distance it distance it and a fine-aligned right as 2 mm in the biolic distance it distance it and a fine-aligned right as 2 mm in the biolic distance it distance it and a fine-aligned right as 2 mm in the biolic distance it distance it and a distance it distance distance distance it distance it and a distance it dist
man of generating Mrs and Arabay. It and To Autom The Danigs of 100 City Paraget Trac.	White bank forms of a plotty, the "analogical spatial spatial and the polyhold plate the following (101) - see the polyhold plate as	the second as records of the phenodyna black of an inclusion, or gamma if an Next shares the second secon	A season of the second se	Descent from the concentration of the concentration of the Diff of DMD, suggests the Concentration of our section is





### Developing Research Programs: Data Integration - Genomics

- GenBank
  - Accession numbers added where available
  - Discussions with GenBank about changing their voucher requirements to include GUIDs rather than triplets
- Linking genomics with environmental data through specimen records: *Amborella*
- Spatial distribution of genome sizes in plants





#### Genome-scale SNPs & Environment

Selective sweeps and habitat variation in *Amborella* 





**Richie Hodel** 





#### Genome-scale SNPs & Environment

#### Selective sweeps and habitat variation in Amborella













#### Genome-scale SNPs & Environment

#### Selective sweeps and habitat variation in Amborella







### Developing Research Programs: Data Integration - Genomics

- GenBank
  - Accession numbers added where available
  - Discussions with GenBank about changing their voucher requirements to include GUIDs rather than triplets
- Linking genomics with environmental data through specimen records: *Amborella*
- Spatial distribution of genome sizes in plants



#### **Spatial Distribution of Genome Sizes**

RESEARCH

- Nucleic acids: large amounts of N and P; large genomes costly to build, replicate, transcribe, etc.
- Plants with large genomes should be selected against on N- and P-poor soil, favored on high-N/P soil
- Park Grass Experiment used to test this hypothesis:
  - GS of plants in high N+P plots higher than in control, N, P plots
- Continental scale, GS related to soil geochemistry?







#### **Spatial Distribution of Genome Sizes**

BOTANIC GARDEN



**GeoEcoEvo: USGS Powell Center Working Group** E. Bui, M. Goldhaber, PIs I. Pearse, J. Cartwright, I. Leitch, A. Leitch, D. Soltis



Atlas of the Biosphere Center for Sustainability and the Global Environ University of Whitpenlis - Made









## Developing Research Programs: Data Integration - Phenology

- Willis et al. (2017) TREE
- Plant Phenology Ontology
- Notes from Nature Guralnick, Mast, Allen, et al.
- Use of images machine learning
  - Stucky, Guralnick, Soltis et al.
  - Pl@ntNet Pierre Bonnet et al.; new collaboration with Nelson, Ellwood, Soltis et al.





## Willis et al. (2017) TREE

#### Review

Old Plants, New Tricks: Phenological Research Using Herbarium Specimens

Charles G. Willis,<sup>1,\*</sup> Elizabeth R. Ellwood,<sup>2,\*</sup> Richard B. Primack,<sup>3</sup> Charles C. Davis,<sup>1</sup> Katelin D. Pearson,<sup>2</sup> Amanda S. Gallinat,<sup>3</sup> Jenn M. Yost,<sup>4</sup> Gil Nelson,<sup>2</sup> Susan J. Mazer,<sup>5</sup> Natalie L. Rossington,<sup>5</sup> Tim H. Sparks,<sup>6,7</sup> and Pamela S. Soltis<sup>8</sup>





#### Willis et al. (2017) TREE







#### Plant Phenology Ontology: standards, terms, links



Figure I. Simplified Representation of Ontological Classes and Logical Structure. In a complete ontology, each term or 'class' has a specific definition and is linked to any and all related classes via 'relation terms' such as 'is\_a' or 'part\_of'. These structured linkages between classes allow integration among different methods of measuring a class (represented in blue), different subclasses within a class (white), and other types of data (yellow), which are subclasses of the general term 'quality' currently defined by the Phenotypic Quality Ontology.



#### Phenology: Engaging Citizen Scientists



#### **Rob Guralnick**





#### Julie Allen Austin Mast







#### Phenology: Data from Multiple Sources



20





## Pl@ntNet =

#### Free, web & mobile app dedicated to plant identification and gathering of botanical observations



Pl@ntNet is based on an innovative crowdsourcing workflow







## Promoting Research Use Workshops, Symposia, Special Pubs

- Phenology Workshop @ UC Berkeley, March 2016
   Willis et al. (2017) review; TREE
  - Yost et al. (in review) methods; APPS
- Annual Digital Data in Biodiversity Research Conference: U of Mich (2017), UC Berkeley (2018)
- Green Digitization Symposia @ Botany & IBC 2017
- Phenome 2018: promoting use of images









## Promoting Research Use Workshops, Symposia, Special Pubs

- Special issue of APPS (2018) based on symposia
   Nelson, James, Soltis (eds)
- Special issue of *Conservation Biology* (2018)
   Ellwood, Klein, Soltis (eds)
- Invited essay Amer. J. Bot. Soltis (2017)
  - On the Nature of Things Series
  - Digitization of herbaria and research
- iDigBio/ADBC grad students
  - paper on research use underway





#### Enabling Research Through Tool Development

- GUODA: Global Unified Open Data Access (ACIS)
- Jupyter: Python & R notebook infrastructure (ACIS)
- FreshData: Notifications of new data records (ACIS)
- Effechecka: Taxonomic Checklist Generator (ACIS)









#### Enabling Research Through Tool Development

- Interfaces with Open Tree of Life and Lifemapper (Soltis)
- USVH as a model linking resources (Soltis/Nelson)







## Enabling Research Through Tool Availability

- Coming soon!
  - Descriptions of and links to recommended and vetted software and tools for use with specimen data
  - Accessible via the Portal and the Research page





# Other Collaborations, Partnerships, Efforts

- Proposals to ABI for tool development
- Proposals to enhance research use through cyberinfrastructure development
- Proposals to enhance training in biodiversity and data science
- RCNs (e.g., Enhancing Participation; Biodiversity Literacy in Undergraduate Education)





#### **Publication Metrics**

- Compiled by Shari Ellis
- Number of publications
- Mentions of iDigBio vs. portal use in publications
- iDigBio vs. TCN portal use





#### **Publications by Year**







#### **Publications: Mentions vs Portal Use by Year**







#### Publications Using idigbio.org or TCN portals or resources





#### Thank you!



#### idigbio.org/wiki



facebook.com/iDigBio



twitter.com/iDigBio



vimeo.com/iDigBio



idigbio.org/rss-feed.xml



idigbio.org/events-calendar/export.ics







*iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.*