Rapid enhancement of biodiversity occurrence records using unconventional specimen data

Katelin D. Pearson¹

Received: 31 January 2018 / Revised: 3 May 2018 / Accepted: 12 June 2018 / Published online: 18 June 2018
© The Author(s) 2018

Abstract

Distributions of taxa across time and space are central to understanding biodiversity and biotic change, yet currently available occurrence data, drawn from biodiversity specimen records and observational datasets, are often insufficient to answer many driving questions. Records of “associated taxa,” taxa co-occurring with a specimen at the time and place of collection, have the potential to fill data gaps and expand the spatiotemporal scope of current occurrence records. I developed a method to extract associated taxon records from 84,328 digitized specimen records and examined the potential of these data to improve the quantity and quality of existing species occurrence data. Adding associated taxon records increased the size of the test dataset by 18.5%, spanned multiple decades (1937–2016), and potentially extended the known range of 217 taxa in Florida and up to 1500 taxa in the United States, demonstrating the capacity of these records to deepen our understanding of changes in the distributions of taxa on Earth. These results suggest that increased attention to documenting associated taxa could be a promising way to maximize the impact of every collecting event.

Keywords Species distributions · Biodiversity · Specimens · Herbarium · Biological collections

Introduction

In this era of anthropogenic influence, the need to understand past and present species distributions to track biotic change has never been greater. Understanding geographical and temporal distributions of species is central to biogeography (Brown et al. 1996; Lomolino et al. 2016), biodiversity research (Gaston 2000; Ricklefs 2004), evolution (Sexton et al. 2009), and ecology (Weins and Graham 2005; Parmesan 2006), among other disciplines, and is vital for biodiversity conservation and planning (Ferrier 2002; Mota-Vargas and

Communicated by David Hawksworth.

✉ Katelin D. Pearson
kpearson@bio.fsu.edu

¹ Department of Biological Sciences, Florida State University, 319 Stadium Dr, Tallahassee, FL 32304, USA

Rojas-Soto 2012), yet our knowledge of where and when species occur is incomplete. Biodiversity specimens, such as dried, pressed plants housed in herbaria, are a significant source of species distribution data (e.g., Otero-Ferrer et al. 2017), as each specimen represents an occurrence of a species at a certain place and time. Recent efforts to digitize biodiversity specimen data have made millions of specimen records and images publically available on online portals (e.g., idigbio.org). However, even en masse, specimen data can be incomplete and geographically, temporally, or taxonomically biased, especially in understudied regions (Tobler et al. 2007; Stropp et al. 2016; Daru et al. 2017). Observational occurrence datasets such as those aggregated by the Global Biodiversity Information Facility (gbif.org) and iNaturalist (inaturalist.org) are also rapidly expanding our knowledge of species distributions, but because historical records are often rare, observational datasets often cannot answer essential questions such as how species distributions may shift in time and space with changes in climate and land use.

One potentially transformative resource for obtaining reliable historical occurrence data remains relatively untapped: records of associated taxa. “Associated taxa,” taxa co-occurring with a biodiversity specimen at the time and place of collection, are often documented on specimen labels in addition to standard date, locality, and collector data (Anderson 1965; Radford et al. 1974), and these data can serve as occurrence records of the associated taxa (Fig. 1). Like biodiversity specimen records, these observational records have the advantage of traversing time and space, and because collectors are usually experienced professionals, associated taxon records are likely to be reliable. Associated taxon records represent what the collector did not collect, perhaps because of time, resource, or technical restraints such as collecting permits, and therefore, once aggregated, they may help fill the gaps left by collecting biases. Moreover, many more associated taxon records can be created in the time that it takes to collect one biodiversity specimen, which suggests that associated taxon data, if consistently recorded, can rapidly expand current occurrence data.

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

Materials and methods

Observational dataset generation

All 84,328 available digitized herbarium specimen records (as of September 13, 2017) of the Florida State University Robert K. Godfrey Herbarium (henceforth “FSU herbarium”) were downloaded using the data portal provided by iDigBio, the U.S. National Science Foundation’s National Resource for Advancing Digitization of Biodiversity Collections and a major aggregator of biodiversity specimen records. The FSU herbarium is a large (220,000+ specimen) herbarium located within the North American Coastal Plain biodiversity hotspot (Noss et al. 2015) in Tallahassee, Florida, USA. Digitization efforts as of September 2017 have primarily focused on the flora of Florida, though the downloaded dataset contained specimens from around the world. This dataset was chosen because associated taxon records are consistently stored in the “habitat” database field in accordance

Fig. 1 Example herbarium specimen with associated taxa noted on the label

with FSU databasing protocol; however, the method developed here can be applied to any database field or multiple fields. Duplicate specimen records, defined as records of the same species collected in the same county on the same date, were removed, reducing the dataset to 72,120 unique occurrence records.

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

The resulting associated taxon dataset was cleaned by removing duplicate records (as defined above), records that had been created from words that the GNRD API misinterpreted as taxonomic names (e.g., Apalachicola, Wakulla), and a handful (8) of records that included the word “no” in front of the associated taxon name. Another R script was developed to resolve the likely identity of observational records with abbreviated binomial names (1510 records) by matching the abbreviated genus letter to the genus of the original

specimen record or, if the genus letter did not match the genus of the original record, the first genus listed in the habitat field. This algorithm was able to correctly infer the binomial name of the associated taxon for 89% of the records. All records with inferred genera were hand-checked for accuracy.

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

Identification of range extensions

Potential extensions of known species distributions were identified using an R script that compared the counties in which associated species were found to known county-level species distributions according to each of three databases: the Atlas of Florida Plants (for Florida specimens only; Wunderlin et al. 2017), the United States Department of Agriculture PLANTS database (for U.S. specimens only; USDA 2018), and iDigBio specimen records using the iDigBio API via the *ridigbio* package. Purported range extensions according to the Atlas of Florida Plants were manually verified to ensure each was not an artifact of incongruent taxonomy or other errors. Because the purpose of this paper is to examine the potential for associated taxon data to expand known taxon distributions rather than produce a full report of new county records, only a subset (100) of non-Florida range extensions of both the USDA PLANTS-based new county records and iDigBio-based new county records were examined to estimate the number “true” new county records that were not the result of errors.

Comparison of specimen data and associated taxon data

The habits and native statuses of original specimen records and associated taxon records were compared to determine whether certain plant types are more frequently documented as associated taxa rather than collected as specimens or vice versa. Plant habit (herb/forb, tree, shrub, or graminoid) and native status (native or introduced) were assigned to each taxon using the USDA PLANTS database (USDA 2018), the Flora of North America (efloras.org; eFloras 2008), and the Atlas of Florida Plants (Wunderlin et al. 2017). For these comparisons, “original specimen records” are only those from which the R script recovered associated taxon records in their habitat fields, and “associated taxon records” are the recovered observational records after data cleaning—including primary duplicate removal—but prior to combination with original specimen records and final duplicate removal.

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

Results

After data cleaning and both duplicate removal steps, 13,372 associated taxon records were extracted from the initial dataset of 72,120 unique herbarium specimen records, representing an 18.5% increase in total occurrence records (Fig. 2). Nearly two-thirds of these records (61.1%) were identified at least to species, and all but two of the remaining records were identified to genus. Of the associated taxon dataset, 1262 records (8.6%) had abbreviated scientific names (e.g., *E. elatus*) that were inferred to species using specimen data and geographic context.

Associated taxon records consisted of 2973 taxa, 207 plant families, and one family of lichen, while the original specimen dataset contained 9685 taxa and 317 plant families. Occurrences of the sunflower family (Asteraceae) were most frequent in both the associated taxon dataset and the original specimen dataset; however, the top ten most occurrence-rich plant families differed between datasets (Fig. 3). Notably, families containing dominant canopy and shrub taxa in this region—the oaks (Fagaceae), pines (Pinaceae), magnolias (Magnoliaceae), and palms (Arecaceae)—comprised 4.9, 2.3, 1.9, and 1.7% of the associated taxon dataset, respectively, while only comprising 1.7, 0.3, 0.3, and 0.2% of original specimen records.

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

Temporal trends in associated taxon data did not closely follow specimen collecting trends (Fig. 4). Associated taxon records spanned a narrower range of time (1937–2016) compared to specimen records (1880–2016), with the majority of associated taxa recorded during the mid-1980s. In one year (1988), the number of associated taxon records exceeded the number of collected specimens. This peak could reflect changes in

Fig. 2 Increases in occurrence records due to extraction of associated taxon records. The 10 most specimen-rich families in the original dataset of digitized herbarium specimen records are shown. These 10 families account for nearly 60% of the total specimen records

Fig. 3 Comparison of relative family composition of the associated taxon dataset and the original specimen dataset. The 10 most occurrence-rich families in the associated taxon dataset are shown

Fig. 4 Histogram of original specimen records (gray) and associated taxon records (black) excluding duplicate records

cultural norms of collecting, perhaps facilitated by advances in technology (e.g., printed labels) or increased activity of a few collectors who regularly documented associated species.

Conversely, spatial density of associated taxon records did correspond with specimen collecting locations (Fig. 5). The areas of highest record frequency for both associated taxon records and specimens were in Florida counties near the FSU herbarium: Leon,

Fig. 5 Heatmap comparison of record densities for the original dataset (left) and the associated taxon dataset (right). Colors indicate the density of records relative to each respective dataset independent of the other: red (darker) indicates higher record density and green (lighter) indicates lower record density. Black stars show the location of the FSU herbarium. Although the original dataset included all digitized specimen records from the FSU herbarium, which span the globe, only the state of Florida is shown in this figure since record density was highest in this region. The heatmap overlays were produced using identical settings for both datasets in the R packages *ggplot2* and *ggmap*, and the background map is courtesy of the Google API accessed using the same R packages. (Color figure online)

Franklin, Liberty, Wakulla, Gadsden, and Jackson. However, unlike the specimen dataset, the associated taxon dataset had an abundance of records from Escambia County that exceeded even those of Leon County, the location of FSU.

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

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

Identification of range extensions

The cleaned associated taxon dataset contained 247 new county records for 217 Florida plant species when compared to the Atlas of Florida Plants (Wunderlin et al. 2017). When compared both to the USDA PLANTS database and specimen records in the iDigBio portal, the associated taxon dataset produced 2371 and 1193 new county records, respectively. An estimated 66% of USDA PLANTS new county records and 75% of iDigBio new county

records could be confirmed as apparent range extensions rather than, for example, taxonomic inconsistencies. By these estimates, the newly generated observational dataset may provide as many as 894–1564 “true” new county records for these databases from the original 72,120 specimen dataset.

Discussion

Increasing our understanding of species distributions is crucial to many scientific aims, including assessing the impact of anthropogenic effects such as climate and land use changes. This analysis of FSU herbarium data demonstrates that accessing the relatively untapped resource of associated taxa noted on biodiversity specimen labels can significantly augment current distribution data. Extracting associated taxon data from 72,120 records resulted in 247 new county records for the state of Florida when compared to the Atlas of Florida Plants, 2371 (estimated 1564 true records) for the U.S. when compared to the USDA PLANTS database, and 1193 (estimated 894 true records) new county records for the U.S. compared to digitized herbarium records hosted on iDigBio. Furthermore, these records spanned multiple decades (1937–2016), providing an irreplaceable historical record of species’ past distributions, potentially in locations where the species can no longer be found. These data can be invaluable to, for example, conservation managers in determining pre-disturbance conditions or researchers seeking to understand spatiotemporal biodiversity change.

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

Imperiled species may also be under-collected due to their protected status (Daru et al. 2017), and their distributions may be poorly understood because they are rare. The associated taxon dataset contained 449 records of 161 state or federally threatened or endangered species and may therefore provide much-needed insight into the distributions of data-depauperate taxa of high conservation interest. Moreover, associated taxon records may provide a broader spatial and temporal range of data for these taxa, which is critical for species facing immediate anthropogenic threats.

On a more basic level, associated taxon records gleaned from biodiversity specimen records increase the quantity of data at hand, which is becoming increasingly important in an era of large-scale analytical methods. For instance, Environmental Niche Models have proven most effective with a high number of training points (i.e., large amount of starting data; Loiselle et al. 2008). Leveraging associated taxon records from digitized specimens

from the FSU herbarium increased the size of the usable dataset by 18.5% over a significant temporal and spatial distribution, demonstrating that this method can substantially boost species occurrence data across time and space.

Limitations

Associated taxon records may offer a new frontier for gaining valuable biodiversity data; however, like all datasets, they are subject to certain coverage, quality, and usage limitations. First, the spatiotemporal range of retrievable data from associated taxon records is limited by the coverage of specimen records. While these data may fill gaps in individual species distributions, they will not be able to address systematic temporal and spatial collecting biases such as lower data collection during World Wars (Delisle et al. 2003) and may instead introduce new biases such as increased occurrences in regions or time periods wherein collectors have been trained to record associated taxa (see Fig. 4). For this reason, associated taxon data are best combined with additional data sources to reduce spurious trends.

Second, associated taxa may be misidentified, and because associated taxon records are purely observational, they lack the verifiability of specimen records. Nevertheless, associated taxon identifications are expected to be reasonably accurate since collectors are often taxonomic experts and are likely to document associated taxa that they have confidently identified in the field. Misidentifications are not a new problem for users of specimen data (see Goodwin et al. 2015) and can be handled through outlier identification and other data quality control methods, or, in some cases, on-site verification. Further investigation on the reliability of associated taxon records and methods to overcome this potential limitation is needed.

Third, the methods developed in this study assume that the appropriate genus of abbreviated associated taxon names (e.g., *E. elatus*) could be found in the original specimen record or in the habitat field from which the associated taxon was gathered. This assumption appeared reasonable for 89% of records, and the remaining 11% could be corrected by hand using regional taxonomic knowledge. If employed on a large scale or without careful curation of the output, this method may be inefficient or cause data quality issues similar to those of misidentifications.

Future directions

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

The focus of this study was herbarium specimen label data, but other types of collections may offer similarly rich—or even greater—opportunities. For example, it is common

practice when collecting insects (Martin 1977) and fungi (Leonard 2010) to record the host plant or animal of the collected individual. Similarly, collectors of vertebrate specimens may record ecto- or endo-parasites or gut contents (RIC 1997; ISLES 2001). Thus, delving into the data of many types of biodiversity specimens may reveal additional, previously “hidden” occurrence data, even for taxonomically distant groups (e.g., insects and plants) and potentially for groups that are under-collected or difficult to preserve such as parasites.

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

The recent push for digitization of biodiversity specimens is making a vast amount of specimen data publically accessible, and we have the increasing opportunity to leverage these resources to produce new types of data. Extracting associated taxon data from existing specimen records may improve our knowledge of species and community distributions, as well as enable collectors and other biodiversity researchers to better identify data gaps, prioritize future collecting events, and optimize methods of data collection. Broadening our knowledge of species distributions and improving data- and specimen-collection practices may be as simple as examining the data we already have.

Acknowledgements Special thanks to my advisor, Austin Mast, for proposing the original idea of this project and encouraging me to submit for publication. I am also grateful for comments and suggestions on the manuscript from Gil Nelson, Brendan Scherer, and two anonymous reviewers. Thanks to Keith Bornhorst at the University of South Florida for access to species+county records and threatened/endangered status of species in the USF Atlas of Florida Plants. Thanks also to the United State Department of Agriculture Plant Data Team for assistance in accessing data from the USDA PLANTS database.

Funding This research was supported through iDigBio, which is funded by a grant from the National Science Foundation’s Advancing Digitization of Biodiversity Collections Program (Award number 1547229). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Anderson RM (1965) Methods of collecting and preserving vertebrate animals. National Museum of Canada no. 69 v. 18
- Boyle B, Hopkins N, Lu Z, Raygoza Garay JA, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, Mckay SJ, Lowry S, Freeland C, Peet RK, Enquist BJ (2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinform 14:16. <https://doi.org/10.1186/1471-2105-14-16>

- Brown JH, Stevens GC, Kaufman DM (1996) The geographic range: size, shape, boundaries, and internal structure. *Annu Rev Ecol Syst* 27:597–623
- Daru BH, Park DS, Primack RB, Willis CG, Barrington DS, Whitfield TJS, Seidler TG, Sweeney PW, Foster DR, Ellison AM, Davis CC (2017) Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytol*. <https://doi.org/10.1111/nph.14855>
- Delisle F, Lavoie C, Jean M, Lachance D (2003) Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. *J Biogeogr* 30:1033–1042
- eFloras (2008) Missouri Botanical Garden, St. Louis, MO & Harvard University Herbaria, Cambridge, MA. <http://www.efloras.org>. Accessed 28 September 2017
- Ferrier S (2002) Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Syst Biol* 51:331–363
- Garcillan PP, Ezcurra E, Vega E (2008) Guadalupe Island: lost paradise recovered? Overgrazing impact on extinction in a remote oceanic island as estimated through accumulation functions. *Biodivers Conserv* 17:1613–1625
- Gaston KJ (2000) Global patterns in biodiversity. *Nature* 405:220–227
- Goodwin ZA, Harris DJ, Filer D, Wood JRI, Scotland RW (2015) Widespread mistaken identity in tropical plant collections. *Curr Biol* 25(22):R1066–R1067
- Island Surveys to Learn about Endemic Species (ISLES) (2001) Instructions for the field collection and preservation of mammals. Museum of Southwestern Biology. <http://msb.unm.edu/isles/Instructions%20for%20the%20field%20collection%20%20and%20preservation%20of%20mammals.pdf>. Accessed 22 November 2017
- Leonard P (ed) (2010) A guide to collecting and preserving fungal specimens for the Queensland Herbarium. Queensland Herbarium, Department of Environment and Resource Management, Brisbane
- Loiselle BA, Jorgensen PM, Consiglio T, Jimenez I, Blake JG, Lohmann LG, Montiel OM (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *J Biogeogr* 35(1):105–116
- Lomolino MV, Riddle BR, Whittaker RJ (2016) Biogeography: biological diversity across space and time, 5th edn. Sinauer Associates, Sunderland
- Martin JEH (1977) The insects and arachnids of Canada, Part 1: collecting, preparing, and preserving insects, mites, and spiders. Biosystematics Research Institute, Ottawa, ON. http://esc-sec.ca/aafcm/onographs/insects_and_arachnids_part_1_eng.pdf. Accessed 22 November 2017
- Mota-Vargas C, Rojas-Soto OR (2012) The importance of defining the geographic distribution of species for conservation: the case of the bearded wood-partridge. *J Nat Conserv* 20(1):10–17
- Myltsev A, Mozzherin D (2016) Global Names Parser. <https://github.com/GlobalNamesArchitecture/gnparser>. Accessed 14 September 2017
- Noss RF, Platt WJ, Sorrie BA, Weakley AS, Means DB, Costanza J, Peet RK (2015) How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. *Divers Distrib* 21:236–244
- Otero-Ferrer F, González JA, Freitas M, Araújo R, Azevedo JMN, Holt WV, Tuya F, Haroun R (2017) When natural history collections reveal secrets on data deficient threatened species: atlantic seahorses as a case study. *Biodivers Conserv* 26:2791–2802
- Parmesan C (2006) Ecological and evolutionary responses to recent climate change. *Annu Rev Ecol Evol Syst* 37:637–669
- Radford AE, Dickison WC, Massey JR, Bell CR (1974) Vascular plant systematics. Harper & Row, New York
- Resources Inventory Committee (RIC) (1997) Fish collection methods and standards, Version 4.0 Ministry of Environment, Lands and Parks Resources Inventory Branch, Terrestrial Ecosystems Task Force, Resources Inventory Committee, The Province of British Columbia
- Ricklefs RE (2004) A comprehensive framework for global patterns in biodiversity. *Ecol Lett* 7:1–15
- Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009) Evolution and ecology of species range limits. *Annu Rev Ecol Evol Syst* 40:415–436
- Stropp J, Ladle RJ, Malhado ACM, Hortal J, Gaffuri J, Temperley WH, Skøien JO, Mayaux P (2016) Mapping ignorance: 300 years of collecting flowering plants in Africa. *Glob Ecol Biogeogr* 25:1085–1096
- R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> Version 3.3.2
- Tobler M, Honorio E, Janovec J, Reynel C (2007) Implications of collection patterns of botanical specimens on their usefulness for conservation planning: an example of two neotropical plant families (Moraceae and Myristicaceae) in Peru. *Biodivers Conserv* 16:659–677
- USDA, NRCS. The PLANTS Database. National Plant Data Team, Greensboro, NC, USA. <http://plant.s.usda.gov>. Accessed 25 January 2018

- Weins JJ, Graham CH (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annu Rev Ecol Evol Syst* 36:519–539
- Wunderlin RP, Hansen BF, Franck AR, Essig FB (2017) Atlas of Florida Plants. <http://florida.plantatlas.usf.edu/>. Accessed 14 September 2017