APPLICATION

Plant-O-Matic: a dynamic and mobile guide to all plants of the Americas

Gregory R. Goldsmith,1,2,3*, Naia Morueta-Holme4, Brody Sandel5, Eric D. Fitz3, Samuel D. Fitz3, Brad Boyle6, Nathan Casler6, Kristine Engemann5, Peter M. Jørgensen7, Nathan J. B. Kraft8, Brian McGill9, Robert K. Peet10, William H. Piel11,12, Nick Spencer13, Jens-Christian Svenning6, Barbara M. Thiers14, Cyrille Violle15, Susan K. Wiser13 and Brian J. Enquist6

1Ecosystem Fluxes Group, Paul Scherrer Institute, Villigen 5232, Switzerland; 2Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, OX13QY, UK; 3Ocotea Technologies LLC, Newton, MA 02458, USA; 4Integrative Biology, University of California Berkeley, Berkeley, CA 94720, USA; 5Section for Ecoinformatics & Biodiversity, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark; 6Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA; 7Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, USA; 8Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA 90025, USA; 9School of Biology and Ecology & Mitchell Center for Sustainability Solutions, University of Maine, Orono, ME 04469, USA; 10Department of Biology, University of North Carolina, Chapel Hill, NC 27599-3280, USA; 11Yale-NUS College, 16 College Avenue West, Singapore 138527, Singapore; 12Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore; 13Landcare Research, P.O. Box 69040, Lincoln, 7640 NZ, USA; 14The New York Botanical Garden, 2900 Southern Blvd., Bronx, NY 10348-5126, USA; and 15CEFDE UR 5175, CNRS - Université de Montpellier - Université Paul-Valéry Montpellier – EPHE, 1919 route de Mende, F-34293 Montpellier, CEDEX 5, France

Summary

1 Advances in both informatics and mobile technology are providing exciting new opportunities for generating, disseminating, and engaging with information in the biological sciences at unprecedented spatial scales, particularly in disentangling information on the distributions and natural history of hyperdiverse groups of organisms. 2 We describe an application serving as a mobile catalog of all of the plants of the Americas developed using species distribution models estimated from field observations of plant occurrences. The underlying data comprise over 3.5 million standardized observations of over 88 000 plant species. 3 Plant-O-Matic, a free iOS application, combines the species distribution models with the location services built into a mobile device to provide users with a list of all plant species expected to occur in the 100 × 100 km geographic grid cell corresponding to the user’s location. The application also provides ancillary information on species’ attributes (when available) including growth form, reproductive mode, flower color, and common name. Results can be searched and conditionally filtered based on these attributes. Links to externally sourced specimen images further aid in identification of species by the user. 4 The application’s ability to assemble locally relevant lists of plant species and their attributes on demand for anywhere in the Americas provides a powerful new tool for identifying, exploring, and understanding plant diversity. Mobile applications such as Plant-O-Matic can facilitate dynamic new approaches to science, conservation, and science education.

Key-words: botany, educational outreach, field guide, mobile app, natural history, smartphone, species identification, taxonomy

Introduction

Advances in mobile technology are facilitating new opportunities for generating, discovering, and accessing information while in the field. In particular, there has been rapid growth in mobile software applications for research and educational engagement in ecology (e.g., Farnsworth et al. 2013; Teacher et al. 2013). Many of these applications are focused on facilitating natural history observations, allowing the user to identify or record an observation of an organism (e.g., iNaturalist, LeafSnap, and Pi@ntNet). Such mobile applications have the advantage of making enormous quantities of information portable and accessible to large...
audiences. Moreover, they are frequently designed to leverage the capabilities of the mobile device, including access to a camera, location services (i.e., information from cellular, GPS, and Wi-fi services to determine your location), accelerometer, Bluetooth, and the Internet. This creates new opportunities for collecting, organizing, and sharing information (Goldsmith 2015). In short, mobile applications have the potential to revolutionize how we carry out our research, as well as how we communicate the results of that research.

We describe a new mobile application, Plant-O-Matic, which is a dynamic catalog of the plants of the Americas made possible by the unprecedented integration of mobile device technology with species range maps estimated from an extensive collection of plant species observations.

**Plant-O-Matic**

Plant-O-Matic is a free iOS application that generates a list of all the plant species expected to occur within the 100 × 100 km geographic grid cell corresponding to the user’s location anywhere in the Americas. Plant-O-Matic v1.0 was released on 14 August 2014 with a major revision (v1.2) released on 10 April 2015 (https://itunes.apple.com/us/app/plant-o-matic/id906932765).

**UNDERLYING DATA**

The first generation of Plant-O-Matic is based on integrating observations of plant species’ occurrences from a number of different sources. Led by the Botanical Information and Ecology Network (BIEN) working group (Enquist et al. 2009), observations from both herbarium specimens and vegetation plots were amalgamated through collaboration with >500 organizations (see http://bien.nceas.ucsb.edu/bien/). The BIEN data workflow integrates, standardizes, and identifies high-quality botanical observation records (Enquist et al. 2009, ms in prep). First, all botanical records were standardized using the Taxonomic Name Resolution Service (TNRS; Boyle et al. 2013). Next, all geographic coordinates were assessed and considered reliable when they fell within the hierarchy of associated place names (e.g., country, state, province) expected from the observation, validated with the Global Administrative Areas data set v2.0 (GADM, 2012). Finally, many known cultivated and non-native species were identified and removed. This final data set (BIEN2) contained 3 585 449 reliable plant observation records in the Americas. Of these observations, 95-7% (88 824 species) had sufficiently reliable geographic coordinates to support geographic range estimation (Morueta-Holme et al. 2013).

The species distributions used in Plant-O-Matic originate from the geographic range modeling workflow detailed in B. J. Enquist et al. (ms in prep). To summarize this methodology, the type of range map generated for a species depended on the number of occurrences available in the data base for that species. For example, for rare species with only one (20 926 species) or two occurrences (10 247 species), a species geographic range was defined as a square 75 000 km² area surrounding each of the occurrence points, similar to the definition of small-ranged species in Pitman et al. (1999). The range of a species with three or four occurrences (12 247 species) was defined using a convex hull. Finally, range maps for species with at least five occurrences (45 404 species) were produced using the Maxent species distribution modeling algorithm Maxent with default settings (Phillips, Anderson & Schapire 2006). As predictor variables in the models, B. J. Enquist et al. (ms in prep) used 19 layers representing current climate (average 1950–2000 conditions) from WorldClim 1.4 at 30-arc second resolution (Hijmans et al. 2005) and 19 spatial filters following Griffith & Peres-Neto (2006) and Blach-Overgaard et al. (2010). The spatial filters were geographic distance-based eigenvectors that capture the geometry of the study area and have previously been used to represent non-climatic range controls such as dispersal limitation and biotic interactions (Blach-Overgaard et al. 2010). By including them in the models, B. J. Enquist et al. (ms in prep) were able to constrain the species ranges and limit overpredictions across the study area that occurred when using only climate predictors (Griffith & Peres-Neto 2006; Blach-Overgaard et al. 2010; Kramer-Schadt et al. 2013). The resulting continuous suitability map for each species was converted to a presence–absence map by applying a 1% training presence threshold as described by Freeman & Moisen (2008). To further constrain potential overpredictions, any non-contiguous areas that were more than 1000 km away from the main range area of each species were removed. All range maps were produced on a 10 × 10 km Lambert equal-area projection and subsequently aggregated to 100 × 100 km for use in Plant-O-Matic. The species lists generated by Plant-O-Matic stem from the overlay of all geographic range maps for all plants in the Americas.

The coverage of BIEN data amalgamated from across the New World shows that large parts of the Americas are well sampled (B. J. Enquist et al. ms in prep). Nevertheless, there are some areas with limited sampling, particularly in the Tropics and at high latitudes. A forthcoming update (BIEN3; due early 2016) will include more detailed sampling in these areas where new data are available.

A key challenge for realizing the application involves merging species range maps (given in x and y coordinates on a grid) with the GPS location coordinates from the user’s mobile device (given based on the reference ellipsoid World Geodetic System 1984: WGS84). The grid is based on a Lambert azimuthal equal-area projection centered on North and South America. The projection also utilizes WGS84 as a reference ellipsoid to estimate Earth’s physical shape. Although both coordinates are based on the same ellipsoid, the projection transformation must still account for the ellipsoid. As iOS does not have a native library to convert coordinates between projections, it is necessary to convert latitude and longitude data to the azimuthal projection using the following algorithms in the PROJ.4 library (Warmerdam 2000).
First, to initiate the WGS84 coordinate system gathered from location services:

\[
\text{src}_\text{pj} = \text{pj}_\text{init}(_{+} \text{proj} = \text{longlat} + \text{ellps} = \text{WGS84} + \text{datum} = \text{WGS84} + \text{no}_\text{defs})
\]

eqn 1

where src_pj is the source projection. Next, to initiate the Lambert equal-area (destination) projection of the species range maps

\[
\text{dst}_\text{pj} = \text{pj}_\text{init}(_{+} \text{proj} = \text{laea} + \text{lat}_0 = 45 + \text{lon}_0 = -100 + x_0 = 0 + y_0 = 0 + \text{ellps} = \text{WGS84} + \text{units} = m + \text{no}_\text{defs})
\]

eqn 2

where dst_pj is the destination project and finally to transform the data

\[
\text{pj}_\text{transform}(\text{src}_\text{pj}, \text{dst}_\text{pj}, n^*, 1^*, x, y, z^*)
\]

eqn 3

where \(n^*\) is the number of points to be passed to the function each time it is called and could be passed as a variable, \(1^*\) is the number of steps in the array between coordinate values, and \(x\), \(y\), and \(z\) are geographic coordinate values that need to be converted to radians before being passed to \(\text{pj}_\text{transform}\). Here, \(z\), the elevation of the user, can be left null, as the projected species distribution maps were based on 2D \(x, y\) coordinates. Future work will assess whether the value of \(z\) will noticeably affect the accuracy of the results. The number of steps in the array is generally 1 for packed arrays, but could be set if the data comes in unpacked \(x/y/z\) values. The application assumes WGS84 latitude and longitude if no source projection is declared.

**APPLICATION FEATURES**

*Plant-O-Matic* uses the mobile device’s location services to determine the user’s location and then returns a list of all of the plant species that are expected to occur in the corresponding 100 × 100 km geographic grid cell. Species occurrence data, as well as ancillary information on each species, are stored locally on the device in a SQLite relational data base, and thus, only the device’s location services capabilities must be enabled to retrieve results. Version 1.2 of the application defaults to Tucson, Arizona, for users outside of the Americas. The results screen displays the list of taxa and ancillary information as described below (Fig. 1a). *Plant-O-Matic* also lists the total number of records associated with a given search or filtered search. For example, the number of species classified as trees that occur within the family ‘Fabaceae’. Each species record within the results is composed of the following information:

1 *Linnaean taxonomy*: Family, genus, and species are provided for each species, following standardization using the TNRS (see details above).

2 *Common name*: The common name is provided for each species, when available. Lists of common names are currently more complete for the United States (including Puerto Rico and Hawaii) and Canada (USDA NRCS 2015), but *Plant-O-Matic* does include common name compilations for the Neotropics (Dodson, Gentry & Valverde 1985; Nolazco & Vélez 1991; USDA NRCS 2015). Common names are currently available for 18% of all species.

3 *Growth form habit*: Each species was classified as an aquatic, bryophyte, cactus, epiphyte, fern, grass, herb, parasite, shrub, tree, or vine. Except for cactus, classifications follow Engemann *et al.* (2016). The cactus growth form was defined as all species in the Cactaceae. While recognizing that a species may have multiple growth forms, we list only the predominant form based on a consensus method across the different available data sets. Growth form is currently available for 49% of all species.

4 *Reproductive mode*: Each species is classified as a bryophyte, lichenid, pteridophyte (named fern for accessibility), gymnosperm, or angiosperm as defined by the major groups in *The Plant List* (www.theplantlist.org). This major group classification is available for 100% of the species. For angiosperms, flower color is also included when data are available (Dodson, Gentry & Valverde 1985; Nolazco & Vélez 1991; Quesada Quesada *et al.* 1997; Vindas 2003; Zamora-Villalobos, Jimenez-Madrigal & Poveda-Alvarez 2003; USDA NRCS 2015). Flower color is mainly represented for North American species and is currently available for 3% of angiosperm species.

5 *Images*: Each species is linked to a carousel of herbarium images of the species through the Tropicos data base (http://www.tropicos.org) established by the Missouri Botanical Garden (Fig. 1b). Images are available in low and high resolution. This feature is a result of a query through an application programming interface (API) and thus requires the device to have web access. Images are available for 32% of all species.

The records can be searched within the results screen or filtered from a separate screen (Fig. 1c). Conditional filtering allows the selection of one or more growth forms, flower colors, and plant families, as well as the ability to limit the results to only those species records with common names or images available (a ‘modular guide’ *sensu* Lawrence & Hawthorne 2006). Finally, a separate information screen provides a basic introduction to the data, links to further information on the Web, and the user’s grid cell and latitude/longitude.

**Discussion**

**ADVANTAGES AND LIMITATIONS OF MOBILE APPLICATIONS**

*Plant-O-Matic* delivers the botanical diversity of the Americas to any interested user with an iOS mobile device. It provides an easily accessible means by which scientists, naturalists, and the public can access and discover amalgamated, validated, and synthesized research-grade data on plant species occurrences. It is also a means of making public data collected from expeditions and ecological surveys and currently stored in herbaria.
museums, and on academic computers, available to broader public audiences. Our hope is that this new form of access will lead to novel uses and applications. More importantly, we designed this app in the hope of facilitating renewed engagement and excitement in botany. At a time when there are considerable threats to the sustainability of herbaria and long-term ecological data repositories (Kemp 2015), Plant-O-Matic is a particularly novel means of raising awareness. There is a critical need for understanding the importance of herbarium and museum collections, ecological surveys, and functional trait data as repositories of primary biodiversity data. These ‘primary biodiversity data’ (Costello et al. 2013) ultimately underlie the estimation of the geographic ranges of species, the definition of vegetation types, and are the basis of our knowledge of cladistic diversity and distribution (Viole et al. 2014). Together, these data form our only baseline for understanding species’ distribution and abundance.

A key strength of Plant-O-Matic is its ability to provide an on-demand list of only those species that are most likely to be present at the user’s location. Working with locally relevant subsets of species makes identification more manageable for all users, especially in regions of high species richness such as the humid tropics. Unlike typical guides, which feature only common species, Plant-O-Matic lists all or nearly all locally occurring species, thus aiding identification and discovery without sacrificing comprehensiveness. In addition, because Plant-O-Matic uses modeled ranges rather than just known occurrences, it can be used at locations where no local floras or checklists are available. The ability to generate predictive checklists will have many applications for biodiversity surveys and conservation assessment, in particular in poorly sampled tropical regions. In this sense, Plant-O-Matic differs from other research-grade mobile applications currently available. On-demand lists of plant species at large spatial scales are emerging from other applications (e.g., Map of Life, iNaturalist), but are not yet comprehensive. Similarly, comprehensive applications (e.g., PlantNet, Lucidmobile) are available, but are not yet available on demand at large spatial scales. Finally, a third and very different class of applications is primarily focused on novel approaches to identification (e.g., LeafSnap, FlowerChecker).

The accuracy and precision of the results are fundamentally limited by both the underlying species occurrence data and the subsequent derivation of species range maps (Feeley & Silman 2011). For instance, in remote locations where species observations are limited, the accuracy of species range models will be lower. Improving the results will require collecting additional data, improving species range modeling techniques, and reducing the grain of the maps. This last improvement must be weighed against the amount of information either stored locally on the application (currently 91 MB) or accessed by the application from a remote database through the Internet.

**FUTURE DIRECTIONS: INTEGRATING SCIENCE, SCIENCE EDUCATION, AND DISCOVERY**

The Plant-O-Matic initiative will continue to focus on providing an engaging mobile catalog of the plants of the Americas that facilitates the next generation of science and science education. The next generation of development will seek to improve the precision and accuracy of the results, as well as to exploit
the full capabilities of mobile devices to provide a more engaging user experience. For instance, we imagine facilitating (i) increased access to information and images for each species, (ii) the export of results through email, (iii) access to results outside of the user’s location through input of coordinates, and (iv) integration with citizen science tools that allow users to add their own geolocated, photographic observations of plants. In particular, providing combinations of results based on data generated by BIEN and data generated by users would improve both user engagement and the quality of the results. Moreover, subject to verification, user-generated observations can ultimately inform new species distribution models and trait data sets (Joly et al. 2014). We will also seek to develop the application on other operating systems (e.g., Android), as well as provide versions in additional languages. Such iterative improvements must be considered in the context of the users and their interests (Lawrence & Hawthorne 2006; McNerny et al. 2014). The application currently collects anonymous analytics on users to inform future application development. There is currently a lack of information regarding the efficacy of educational technology, particularly mobile applications, and their interests (Lawrence & Hawthorne 2006; McInerny et al. 2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics, 14, art11.

Acknowledgements

This work was conducted as a part of the Botanical Information and Ecology Network (BIEN) Working Group, 2008–2012 (Pis B.J.E., B.B., R.K.P., Steven Dolins, and Richard Condit) supported by the National Center for Ecological Analysis and Synthesis, a center funded by NSF (Grant #EF-0553768), the University of California, Santa Barbara, and the State of California. The BIEN Working Group was also supported by The iPlant Collaborative (NSF #DBI-0735191; URL: www.iplantcollaborative.org). We thank all of the scientists who have contributed to and supported various BIEN working group meetings since 2008. We thank the numerous individual data contributors and >500 herbaria who contributed their data to various data compiling organizations for the invaluable data and support provided to BIEN. We especially thank the following institutions for direct data access: Global Biodiversity Information Facility, World Information Network on Biodiversity, The speciesLink Project, New York Botanical Garden, Missouri Botanical Garden, Utrecht Herbarium, University of Arizona Herbarium, North Carolina State University Herbarium, and University of North Carolina Herbarium. Vegetation inventory data were provided by Amazon Forest Inventory Network, Carolina Vegetation Survey, Center for Tropical Forest Science, SALVias, Tropical Ecology and Assessment Monitoring Network, United States Forest Inventory and Analysis National Program, and VegBank. The staff at iPlant and the Texas Advanced Computing Center at the University of Texas provided critical computational assistance. N.M.-H. was supported by an EliteForsk Award, the Aarhus University Research Foundation, and the Villum Foundation. C.V. was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Program (DiversiTracts project, no. 220160) and by the European Research Council (ERC) Starting Grant Project ‘Ecophysiologically and biomechanical criteria for domestication in crop plants’ (Grant ERC-2014-639706-CONSTRAINTS). J.C.S. and B.J.E. were supported by the Center for Informatics Research on Complexity in Ecology (CIRCE), funded by the Aarhus University Research Foundation under the AU Ideas program, and the European Research Council (ERC Starting Grant: ERC-2012-SIG-310886-HISTFUNC). G.R.G. and E.H.F. were supported by a National Geographic Society Waitt Grant. G.R.G. was supported by funding from the European Community’s Seventh Framework Program (FP7-2007-2013) under grant agreement number 290605 (COFUND: PSI-FELLOW).

Data accessibility


References


Plant-O-Matic mobile guide 5


Received 5 October 2015; accepted 12 December 2015
Handling Editor: Robert Freckleton