Citizen Science and Climate Change: Mapping the Range Expansions of Native and Exotic Plants with the Mobile App Leafsnap

W. JOHN KRESS, CARLOS GARCIA-ROBLEDO, JOÃO V. B. SOARES, DAVID JACOBS, KATHARINE WILSON, IDA C. LOPEZ, AND PETER N. BELHUMEUR

The mobile iPhone app Leafsnap, designed for the automatic identification of 220 tree species from the northeastern United States, was released to the public in 2011. In the first 3 years of its use, the app was downloaded by more than 1,500,000 users from five continents and 181 countries who recorded over 3,056,684 leaf images. The high levels of accuracy of Leafsnap identifications, as were confirmed by expert botanists, were used to map the geographic distribution of native and exotic species at a scale previously unachievable without this technology and without the aid of citizen scientists. Species demonstrated northerly migrations, southerly migrations, or little change from their estimated distributions in the 1950s. These results suggest that this tool carried on the phones of millions may potentially collect invaluable data that can be used to monitor the effects of climate change and exotic species on tree distributions at broad geographic scales.

Keywords: climate change, citizen science, identification, geographic distribution, deep learning

Citizen science, the participation of the public in scientific research, is becoming the cornerstone of many large-scale studies recording patterns, processes, and the distribution of biodiversity on Earth (Brossard et al. 2005, Bonney et al. 2009, Bonter and Cooper 2012, Snaddon et al. 2013). One of the challenges of citizen science is to recruit, train, and engage citizen scientists in the thinking process (Trumbull et al. 2000, Brossard et al. 2005). The ubiquity of smart phones is encouraging scientists and developers of mobile apps to combine educational tools with the engagement of citizen scientists in the collection of large data sets of species distributions at large geographical and temporal scales (Kumar et al. 2012, Goëau et al. 2013, Zilli et al. 2014, Goldsmith et al. 2016).

Leafsnap (figure 1; http://Leafsnap.com) is a free mobile app, currently available for the iPhone, which uses visual recognition software to help identify species of trees by taking a photograph of their leaves (Agarwal et al. 2006, Belhumeur et al. 2008, Kumar et al. 2012). Leafsnap, first released for popular use in 2011, currently includes 220 species of trees of northeastern North America (including most of the native and commonly cultivated trees of New York City and Washington, DC) and parts of eastern Canada (Kumar et al. 2012). Plans are to eventually expand the number of species to encompass the approximately 450 species of common trees of the entire continental United States and Canada. A version of Leafsnap has also been developed for the trees of the United Kingdom in collaboration with educators at The Natural History Museum in London (www.nhm.ac.uk/take-part/identify-nature/leafsnap-uk-app.html).

At the start, Leafsnap was originally conceived by the creators as a high-tech aid for taxonomic botanists searching for new species in poorly known habitats around the world. However, with the emergence of mobile phone devices, it rapidly became clear that this tree identification tool had the potential to be a new way for nonscientists as well as scientists to easily map the locations, geographic distributions, and diversity of trees in unprecedented detail and scope. However, to reach this potential, the app needed to be adopted by a large number of users across a broad geographic range. Fortunately, within a year after it was released, the media quickly recognized the utility and potential popularity of Leafsnap, which was recommended by both Scientific American and Wired magazine, received news coverage in newspapers and journals around the world (e.g., Science, The New York Times, The Guardian, and The Wall Street Journal).
Figure 1. Using Leafsnap, an iPhone-based identification tool for trees in the northeastern United States. (a) After a citizen scientist takes a picture of a leaf using Leafsnap, the identification algorithm compares the shape of the leaf with the shape of thousands of identified leaves in the previously compiled Leafsnap library of images. (b) A list of candidate species appears on the screen. For each candidate species, users can browse (c) short botanical descriptions and (d) images of reproductive and vegetative structures, both of which are useful for species identification. It is also possible to (e) browse through the trees included in Leafsnap looking for the correct species. Users then label their identifications and submit their image with locality data to the Leafsnap server; (f) specific localities of identified species are then mapped.
Leafsnap emphasizes interactivity (figure 1). A person uses the built-in camera of an iPhone to take a photograph of a single leaf; Leafsnap then compares the photograph with a central library of over 9000 images that have been collected and stored in a database on a remote server. Leafsnap’s algorithm establishes the contours of the leaf and uses visual recognition software to find an ordered set of possible matches for it in the database (Agarwal et al. 2006, Belhumeur et al. 2008). Leafsnap returns the top search results in 5–20 seconds, depending on the speed of the network connection, and brings up high-resolution images of the leaf along with highly detailed images of the species’ flower, fruit, seeds, and bark. The app also supplies background information on the species and its geographic distribution. Often, a leaf is so distinctive that the match to a species will be obvious. In other cases, when the identification is not so easy, Leafsnap users must scroll down the top-listed tree species and compare the informative images in the database for clues, such as pubescence on the petiole, the shape of the fruit, or the pattern of veins on the leaf. In the end, it is up to the user to make the determination of the correct species.

Once the Leafsnap user settles on a single species identification and “labels” the leaf, the identification along with the new photograph and location information taken from the device’s built-in global positioning system is automatically uploaded into Leafsnap’s database on the server. Leafsnap turns users into citizen scientists who can then share images, species identifications, and georeferenced species locations with their own social networks. Over time, it will become necessary for the community of scientists to stay engaged with the community of citizen scientists if they are to use this stream of distributional data to map and monitor how geographic ranges of different tree species are changing through time as a function of climatic change and habitat alterations. This ongoing engagement may be the biggest challenge of all in keeping this tool actively providing new data.

Here, we explore patterns in the number and distribution of users as well as the distribution of Leafsnap records to demonstrate a range of ways in which the data being collected can be used to support our understanding of species ranges in this age of global environmental change. After providing a summary of the magnitude and geographic distribution of users, we analyze the Leafsnap data to determine the accuracy of identification across the species included in the library. The georeferenced distributional data are then compared with the published distributions of selected native North American species to determine whether species migration as a result of climate change and habitat alterations can be detected with these citizen-scientist observations. Because Leafsnap includes a number of introduced and naturalized tree species not native to North America, we demonstrate that these georeferenced identifications can also be used to monitor the expanding radiations of exotic species across the landscape. We suggest that combined with other GIS technologies, Leafsnap records can assist ecologists, foresters, and arborists in locating specific individual trees for verification and investigation.

### A regionally oriented mobile app that has generated global interest

Although the target of the first version of Leafsnap was the identification of trees from the northeastern United States, this app has generated interest across the globe. During the first 3 years following its release in May of 2011, 480,561 users from all over the world downloaded Leafsnap and used the app to compare images of leaves with the then-current identification database (figures 2 and 3). During those 3 years, users recorded a total of 3,056,684 snaps (figure 2). Because users can choose not to share their locations, a little over half of the snaps (1,754,010 snaps) in the database were not georeferenced (figure 2). However, today, hundreds of images identified by citizen scientists with associated locations continue to be submitted to Leafsnap.

Leafsnap has been used in 181 countries (figure 3). Because of the lack of trees in both desert (e.g., Western Australia and the Saharan countries in Africa) and tundra (e.g., some parts of Canada, Greenland, and Russia), it appears that Leafsnap has been little used in these regions. Although the lack of Leafsnap identifications in the Amazon region of South America may in part be due to the scarcity
of mobile phone connectivity there, we find it encouraging that Leafsnap is used in other countries with the lowest Internet penetration rate (countries where 1.7% or less of the population has access to Internet), such as Burundi, the Democratic Republic of Congo, Ethiopia, Myanmar, Niger, and Sierra Leone, from which 413 snaps were submitted to our database (Chinn and Fairlie 2006). Most of the snaps submitted from outside the United States did not include verified identifications by the users. When citizen scientists from these regions submitted identifications, many were incorrect, in large part because this version of Leafsnap focuses primarily on trees of the northeast United States and eastern Canada.

The country with the highest number of users is the United States, with over 1.6 million snaps submitted to our database. The next four countries submitting the largest numbers of snaps are Canada, the United Kingdom, the Netherlands, and Italy. Inside the United States, Leafsnap is broadly used in all the states and the District of Columbia.

The submission of over three million snaps from citizen scientists worldwide suggests that Leafsnap has been successful in connecting a broad audience to nature, which is one of the main reasons it was developed. The interest generated across the globe by this version of Leafsnap, which was designed only for identifying the trees from a small region in North America, indicates that a similar tool may have application in other parts of the world.

It is clear that Leafsnap is an engaging tool for teaching users about trees. However, it also has the potential to provide valuable data for monitoring the distribution of trees and for plant conservation. In the following section, we focus on assessing the accuracy of identifications submitted by users in the focal states for which the app was designed. We also show how these records are being used to understand the extent of range expansions by native and exotic tree species.

Assessing the accuracy of tree identifications by citizen scientists using Leafsnap

The identifications of trees made by Leafsnap users depend on both the ability of the automated shape-matching algorithm provided by Leafsnap as well as the skills and knowledge of the user who must compare traits in the species descriptions with the actual tree (e.g., flowers, fruits, and bark). To assess the accuracy of these identifications by Leafsnap users, we tested our database of identifications by citizen scientists in the focal states of the northeast United States.

The submission of over three million snaps from citizen scientists worldwide suggests that Leafsnap has been successful in connecting a broad audience to nature, which is one of the main reasons it was developed. The interest generated across the globe by this version of Leafsnap, which was designed only for identifying the trees from a small region in North America, indicates that a similar tool may have application in other parts of the world.
were removed from consideration (indicated by an asterisk in figure 5). Most angiosperm trees can be distinguished by their leaf shapes, and these taxa accounted for most of the species tested. Twelve scientists in the Department of Botany at the Smithsonian's National Museum of Natural History participated in the verification of the identifications made by Leafsnap by comparing the leaf images with herbarium specimens and literature resources.

The accuracy of identifications made by citizen scientists using Leafsnap was highly variable among the tree species analyzed and ranged from 10% to 100% accuracy. For 38 species (19% of the total species tested), identification accuracy was 50% or less (figure 5). At the other end of the accuracy spectrum, 23 species (12% of the total species tested; 12 native and 10 exotic species indicated by green and red, respectively, in figure 5) were consistently identified with high accuracy, ranging between 90% and 100%. The accuracy of identification ranged between 50% and 90% for the remaining 137 species now present in Leafsnap (69% of total species tested). The accuracy of Leafsnap is now being improved using the latest advances in deep learning (Goodfellow et al. 2016; see below). In the analysis that follows, 22 of the species that showed the highest accuracy of identification were used as a proof of concept to determine the potential of Leafsnap as a tool to monitor changes in the geographic ranges of native trees and the extent of current distributions of exotic species of trees.

**Monitoring the migration of native and exotic trees in North America**

One of the original objectives in the development of Leafsnap was to use the potential power of nonscientists to collect data on the current location and distribution of both native and exotic tree species to compare with past known geographic ranges. To detect possible range expansions of native tree species in North America, we consulted the US Geological Survey's Tree Species Range Maps from "Atlas of United States Trees" by Elbert L. Little, Jr. (https://esp.cr.usgs.gov/data/little) to determine at least one estimate of the original distributions of 12 of the 14 native species that were identified with an accuracy of 90% or higher by Leafsnap (see previous section and figure 5). We omitted the native trees Abies concolor, because of the small number of snaps, and Amelanchier canadensis, because we could not obtain an accurate map of its original distribution.

We mapped the current localities of focal native and exotic tree species on the basis of the georeferenced snaps submitted by citizen scientists (figure 6). Four species, Acer rubrum, A. saccharinum, Quercus palustris, and Sassafras albidum (figure 6b, 6c, 6g, 6l), show no apparent range shifts. However, 8 of the 12 species show some shift in distributions compared with the US Geological Survey range maps. Three species (figure 6d, 6i, 6k), Catalpa bignonioides, Quercus shumardii, and Robinia pseudoacacia, show a distinct shift to the north. An additional three species (figure 6e, 6f, 6j), Cercis canadensis, Liriodendron tulipifera, and Quercus velutina, appear to be slightly shifting north. The final two species (figure 6a, 6h), Acer pensylvanicum and Quercus rubra, showed a distinct shift to the south. Leafsnap identifications can clearly add to our knowledge of the geographical ranges of our native trees.

We recognize that these results are by no means conclusive and unequivocal. Some of these records of range
extensions may represent human introductions outside the natural range of the species (e.g., *C. bignonioides*). Moreover, the absence of identification snaps in portions of the known distributions or outside the known distributions may also be an artifact of no Leafsnap users being active in the area. The dense distribution of Leafsnap users throughout the United States, as is indicated in figure 3, however, suggests that the app has been employed in these regions to identify at least some species of trees. The absence of some species formerly known from these regions in the Leafsnap identifications may therefore be a true reflection of their absence today. Our point is that these examples suggest that Leafsnap offers a significant tool to monitor shifts in native tree populations. As the users of Leafsnap continue to contribute to long-term geographic data sets (although it will clearly be a challenge to engage users of Leafsnap for decades to come), it will be

Figure 5. Assessing the accuracy of plant identifications by citizen scientists using Leafsnap. (a) Accuracy of citizen scientist identifications assessed by botanists at the Smithsonian Institution. Asterisks represent plant species that trained botanists have difficulty identifying from Leafsnap images only (mostly gymnosperms). Species displaying 90% or higher identification accuracy were used for further analyses of the distribution of native (highlighted in green) and exotic (highlighted in red) tree species (see figures 6 and 7). (b) The frequency of snaps by citizen scientists for each plant species in the overall Leafsnap database for the focal states.
Figure 6. Mapping range expansions of some species of native North American trees beyond their original distributions. Green areas represent the distribution of each plant species ascertained from US Geological Survey maps. Dots represent the location of native plant species identified by citizen scientists using Leafsnap. N\textsuperscript{Acer pensylvanicum} = 493, N\textsuperscript{Acer rubrum} = 5023, N\textsuperscript{Acer saccharinum} = 3064, N\textsuperscript{Catalpa bignonioides} = 590, N\textsuperscript{Cercis Canadensis} = 2465, N\textsuperscript{Liriodendron tulipifera} = 1908, N\textsuperscript{Quercus palustris} = 2179, N\textsuperscript{Quercus rubra} = 1680, N\textsuperscript{Quercus shumardii} = 452, N\textsuperscript{Quercus velutina} = 1011, N\textsuperscript{Robinia pseudoacacia} = 855, N\textsuperscript{Sassafras albidum} = 1041.
possible to monitor range shifts of tree species that are a result of human introductions or projected climate change.

We have also been able to explore the potential of Leafsnap as a tool to monitor invasions of nonnative, exotic species of trees in North America. Similar to the above, we selected 11 exotic tree species known to be in the United States that were consistently identified with an accuracy of 90% or higher by Leafsnap users and had a significant numbers of snaps in the database (figure 5). Citizen scientists submitted 7005 Leafsnap records of these 10 exotic plants from the continental United States, Canada, and Mexico. With the mapping these records, it can be seen that all 11 species are already distributed in various regions across North America (figure 7) and that at least a few may be potentially invasive in particular habitats where they are most common (e.g., *Acer platanoides* and *Ailanthus altissima*). As the Leafsnap database continues to grow, more accurate and extensive observations of range expansions will be possible.

**On-the-ground verification of novel localities of native and exotic trees of North America**

As the geographic ranges of native tree species are altered by climate change and as exotic species expand their distributions, the need to verify the identification of individual trees that have been recorded and submitted by citizen scientists using Leafsnap increases in importance. The geographic information system databases used by Leafsnap to determine new distributions of native and exotic trees in North America (figures 6 and 7) include all snaps, their locations, and links to photographs taken by citizen scientists (figure 8). These data sets are formatted as Keyhole Markup Language (kml) files, which can be opened in Google Earth (figure 8a). Users can access the information of each snap by clicking on the markers (figure 8b). This feature has the potential to allow further field verification of the snaps submitted by citizen scientists that can be tracked back to the original habitats and localities where each snap was collected. For example, an arborist can find the location of a tree reported by Leafsnap that was identified in an urban setting or a more natural forested area, such as a national park. It may even be possible for scientists, foresters, urban ecologists, and citizen scientists themselves to relocate the individual trees they earlier snapped to ensure correct identification and assess ecological correlates that may account for the locations of these specimens (figure 8C). The potential of this capability suggests that the long-term storage and archiving of this type of identification and distributional data will be critical, especially as our natural habitats and urban areas change.

**Next steps in the development of citizen-science apps for biodiversity monitoring**

Here, we have demonstrated that popular, easy-to-use, and effective tools used by citizen scientists for the collection of biodiversity data, such as Leafsnap, also allow ecologists to monitor the movement and migration of organisms inside and outside of their ranges. The rising redistribution of species across the planet, whether it is due to climate change (Davis and Shaw 2001) or human-caused biological invasions (Crall A et al. 2012), will have serious consequences for natural environments as well as our societies. Impacts on economic development, local livelihoods, health, and food production, which may be substantial (Pecl et al. 2017), call for increased monitoring of the movement of species both within their current ranges as well as migrations into new regions and habitats. The mobilization of millions of nonscientists as collectors of data employing accurate identification devices, such as Leafsnap, will be invaluable if we are to address these substantial changes in the distribution of nature on the planet.

Some citizen-science tools, such as Leafsnap, are quite effective in identification. However, the accuracy and ease of use of these handheld identification devices can be improved as technology advances. Since Leafsnap was released in 2011 (originally including only 160 native species of the northeast United States but soon expanding to all 450 species of common trees of North America), considerable progress has been made in the effectiveness of computational systems for visual recognition. Krizhevsky and colleagues (2012) demonstrated that systems trained using deep learning, a type of machine learning in artificial intelligence for unstructured data, could significantly outperform other approaches to visual recognition. This trend has continued, with virtually all state-of-the-art visual recognition systems now based on deep learning (Krizhevsky et al. 2012, Simonyan and Zisserman 2014, Szegedy et al. 2015, Goodfellow et al. 2016, He et al. 2016), including systems that demonstrate impressive performance on problems of species identification (Branson et al. 2014, Zhang et al. 2014, Cui et al. 2016). Advances such as these, which are now being incorporated into Leafsnap, promise that new electronic field guides will be more accurate and much easier to construct.

Prior to deep learning, recognition methods, which were applied to leaves (Abbasi et al. 1997, Kumar et al. 2012), flowers (Nilsson and Zisserman 2008), birds (Farrell et al. 2011, Berg and Belhumeur 2013, Berg et al. 2014), and other groups of organisms (Hemming and Rath 2001, Gamble et al. 2008, Crall JP et al. 2013), each used different approaches that were to some extent handcrafted for that particular domain. With deep learning, it is more likely that we can apply a single machine-learning architecture to recognize a wide range of both plant and animal species. The same computational approach should be applicable to beetles, trees, frogs, weeds, or birds. These improvements may make it possible to produce a “universal field guide” in which biologists need only collect and label appropriate training images to generate a new field guide for their domain. Moreover, the widespread use of digital photography and image-sharing platforms will allow biologists to collaborate with citizen scientists to acquire the appropriate images for these field guides. We predict that
Figure 7. Mapping distributions of some exotic tree species in North America. Dots represent the location of exotic trees identified by citizen scientists using Leafsnap. \( N_{\text{Acer palmatum}} = 2004, N_{\text{Acer platanoides}} = 2188, N_{\text{Acer pseudoplatanus}} = 1107, \)
\( N_{\text{Aesculus hippocastanum}} = 176, N_{\text{Ailanthus altissima}} = 375, N_{\text{Cornus mas}} = 175, N_{\text{Eucommia ulmoides}} = 280, N_{\text{Halesia tetraptera}} = 401, N_{\text{Ginkgo biloba}} = 1174, N_{\text{Salix babylonica}} = 393, N_{\text{Syringa reticulata}} = 262. \)
the construction of a unified core recognition technology will enable the creation of dozens of “Leafsnaps” for a wide range of plants, animals, fungi, and maybe even microorganisms. In the future, a single “Lifesnap” tool may allow citizen scientists to identify and record the presence of species across all domains of life.

Acknowledgments
We thank David Erickson, Yongli Fan, Masha Kuzmina, David Lopez, Serenity Montaño, Asmaa Muneer, Nancai Pei, and Karen Yu for their assistance in the Leafsnap identification accuracy tests. The creation of Leafsnap would not have been possible without the input and assistance of many colleagues and friends, including Finding Species, and funding support from the National Science Foundation (including grant no. 1526234), TD Bank, and the Smithsonian Institution.

Supplemental material
Supplementary data are available at BIOSCI online.

References cited

Figure 8. Exploring distributions of native and exotic tree species using Google Earth. (a) Snaps of native and exotic focal species (figures 6 and 7) can be uploaded to Google Earth using the kml files provided as supplemental materials (supplement S1 and S2). (b) Users can select a focal native or exotic species (in this example Acer rubrum, a native tree species) and zoom in to a particular geographic area (in this example Washington, DC). Users can click on each snap to display information such as common names, scientific names, and coordinates of a particular snap. This tab grants access to a link that shows the original image submitted to Leafsnap. By clicking on the directions option, users can travel to the locality where the snap was submitted. (c) Using the feature Street View in Google Earth, users can track individual trees when snaps were recorded near roads, in this example a red maple in an urban area in Washington, DC.


W. John Kress (kressj@si.edu) is distinguished scientist and curator of botany at the Smithsonian’s National Museum of Natural History, in Washington, DC. His research focuses on tropical biology and conservation, evolution, polination ecology, ethnobotany, and plant systematics. Carlos Garcia-Robledo is an evolutionary ecologist and assistant professor in the Department of Ecology and Evolution at the University of Connecticut, in Storrs. He is interested in the use of new technology to understand the ecology and evolution of plant–insect interactions in a changing world. João V. B. Soares is a research engineer at Yahoo!, in New York City. He has general interests in computer vision and machine learning. David Jacobs is a professor of computer science at the University of Maryland, in College Park. His research focuses on computer vision. He has also worked on problems in machine learning, computer graphics, human–computer interaction, and visual psychophysics. Katharine Wilson studied ecology, conservation, and biodiversity at Smith College. She is interested in ex situ plant conservation and invasive ecology. Currently, she is a curatorial scholar at the Jerusalem Botanical Gardens, in Israel. Idit C. Lopez is a museum specialist in the Botany Department at the Smithsonian’s National Museum of Natural History. She is the Leafsnap Project Coordinator at the Smithsonian. Peter Belhumeur is a professor of computer science at Columbia University, in New York City, and the technical lead for the Machine Learning Team at Dropbox. His research focuses on computer vision, machine learning, and artificial intelligence.