Biological Conservation xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Knowledge behind conservation status decisions: Data basis for "Data Deficient" Brazilian plant species

Mariane S. Sousa-Baena [a,*], Letícia Couto Garcia [a], A. Townsend Peterson [b]

[a] Centro de Referência em Informação Ambiental (CRIA), Av. Romeu Tórtima 388, 13084-791 Campinas, São Paulo, Brazil
[b] Biodiversity Institute,1345 Jayhawk Blvd., University of Kansas, Lawrence, KS 66045, USA

ARTICLE INFO

Article history:
Received 30 November 2012
Received in revised form 11 June 2013
Accepted 30 June 2013
Available online xxxx

Keywords:
Data Deficient species
Conservation status
Red List
Angiosperms
Digital Accessible Knowledge
Primary biodiversity data

ABSTRACT

Methods for evaluating risk of biodiversity loss are linked closely to decisions about species' conservation status, which in turn depend on data documenting species' distributions, population status, and natural history. In Brazil, the scientific community and government have differing points of view regarding which plant species have insufficient data to be accorded a formal threat category, with the official list of threatened flora published by the Brazilian Ministry of Environment listing many fewer species as Data Deficient than a broader list prepared by a large, knowledgeable group of taxonomists. This paper aims to evaluate, using diverse analyses, whether "Digital Accessible Knowledge" is genuinely lacking or insufficient for basic characterization of distributions for 934 angiosperm species classified as Data Deficient on Brazil's official list. Analyses were based on large-scale databases of information associated with herbarium specimens, as part of the *species*Link network. Evaluating these species in terms of completeness of geographic range knowledge accumulated through time, our results show that at least 40.9% of species listed as Data Deficient do not appear genuinely to be particularly lacking in *data*, but rather may be knowledge-deficient: data exist that can provide rich information about the species, but such data remain unanalyzed and dormant for conservation decision-making. Such approaches may be useful in identifying cases in which data are genuinely lacking regarding conservation status of species, as well as in moving species out of Data Deficient categories and into appropriate threat status classifications.

1. Introduction

Methods for evaluating biodiversity loss are linked closely to species conservation status decisions (Norris, 2012). To define conservation status, information about species' distributions, population status, and natural history is required. In Brazil, no consensus exists about optimal approaches to conservation status decisions for plant species. In particular, the scientific community and government have sharply contrasting points of view regarding which species are best classed as under some category of threat versus which should be considered Data Deficient.

A group of 300 scientists convened by the Biodiversitas Foundation (2005), in an effort to improve lists of threatened plant species (Scarano and Martinelli, 2010), analyzed 5212 species, classifying 1495 of them into five IUCN threat categories (IUCN Standards and Petitions Subcommittee, 2011) and 2513 as Data Deficient. However, after the Biodiversitas/IUCN list was submitted to the government, a substantially different list (the current "Official List of Threatened Brazilian Plants") was published by the Brazilian Ministry of Environment (MMA, 2008), which divided species among only two categories: Endangered or Data Deficient (Fig. 1). The Endangered list comprised 472 species, whereas the Data Deficient list included 1079 species, from which 934 were angiosperms. Most species on the Data Deficient list of the Ministry of Environment (hereafter referred to as "MMA") had been classed as Vulnerable, Endangered, or Critically Endangered on the Biodiversitas/IUCN list (Fig. 1).

According to IUCN guidelines, a species is designated as Data Deficient when data on its abundance and distribution are insufficient or lacking (IUCN Standards and Petitions Subcommittee, 2011): "a taxon is Data Deficient when there is inadequate information to make a direct, or indirect, assessment of its risk of extinction based on its distribution and/or population status." Hence, Data Deficient is not a category of threat; rather, it indicates that further research is necessary, not discarding the possibility that the species will turn out to be best considered as threatened (Butchart and Bird, 2010; Celep et al., 2010). Specific reasons for moving species classified as Vulnerable, Endangered, or Critically Endangered by Biodiversitas using IUCN criteria (Biodiversitas Foundation, 2005) to Data Deficient were not disclosed (MMA, 2008).

However, our preliminary review suggests that, in fact, considerable basic data exist for many angiosperms listed as Data

* Corresponding author. Tel.: +55 19 3288 0466; fax: +55 19 3249 0960.
E-mail address: baena@cria.org.br (M.S. Sousa-Baena).

0006-3207/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biocon.2013.06.034

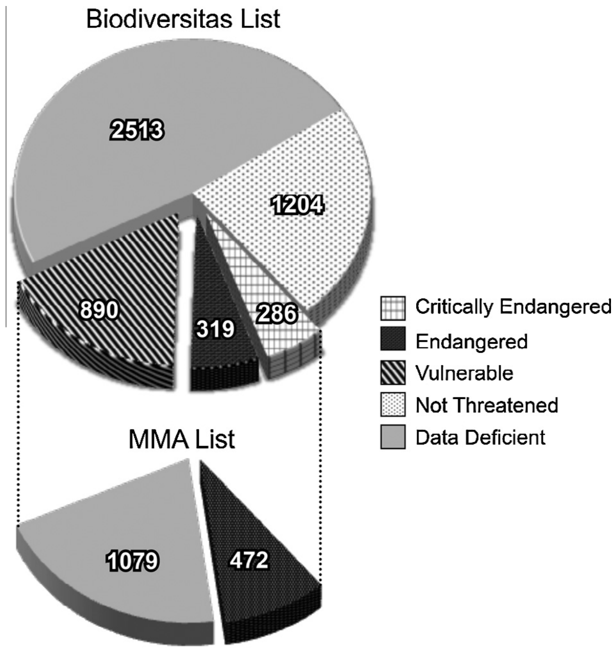

Fig. 1. Comparison of numbers of species considered as threatened according to Biodiversitas Foundation (2005) and the Brazilian Ministry of Environment (MMA, 2008). Virtually the same set of plant species is classified in very different ways in the two lists. Biodiversitas divides the species set in 890 Vulnerable, 319 Endangered and 286 Critically Endangered, whereas MMA divides the set in 1079 Data Deficient and 472 Endangered.

Deficient. This rapid review was achieved using *Lacunas* (Canhos et al., submitted), a data infrastructure developed by the Centro de Referência em Informação Ambiental (CRIA), fed with data from Brazil's *Virtual Herbarium*, which in turn is fed by the *species*Link network (CRIA, 2012). Currently, the *Virtual Herbarium* includes data on 91.4% of all Brazilian plant species, with 2,434,933 georeferenced records of angiosperms. Increasing amounts of primary biodiversity data are becoming available every year, such that what has been termed "Digital Accessible Knowledge" (DAK) is reaching critical mass for biodiversity in Brazil, making possible many novel, synthetic analyses that were heretofore impossible (Sousa-Baena et al., in press).

The use of diverse and novel analyses of primary biodiversity data to assess biodiversity threat and loss has considerable promise—for instance, linking primary biodiversity data with climate data and land-cover data can offer estimates of distributional area loss even in absence of actual monitoring data (Soberón and Peterson, 2009). Inventory statistics can offer useful information about data quality and status of knowledge (Colwell and Coddington, 1994). Ecological niche modeling can allow characterization of geographic distributions and evaluation of extinction risk for poorly-sampled species (e.g., Peterson et al., 2006; Siqueira et al., 2009). This technique can also be used for projecting potential population losses or gains though time and with environmental change (Peterson et al., 2006; Soberón and Peterson, 2009). As a consequence, opportunities for insightful views of biodiversity status are increasingly available.

The objective of this study, then, was to examine how much DAK exists for angiosperm species classified as Data Deficient in the Official List of Threatened Brazilian Plants, using openly available primary biodiversity data and diverse analytical approaches. Specifically, we base a subjective assessment of available knowledge on (1) a novel approach to inventory completeness statistics to assess completeness of knowledge of species' geographic distributions, (2) calculations of time since last record of each species, and (3) quality of preliminary ecological niche models based on known occurrences of each species. We do not set out to assign conservation status designations to species, nor do our assessments speak fully to all dimensions of Data Deficient designations, but rather we assess whether DAK exists and holds significant information for each species. The result is a view of the potential for improving knowledge about the conservation status of Brazilian plants via analysis of available data, balanced against the need for further study.

2. Methods

2.1. Input data

The analyses developed herein were based on large-scale databases of information associated with herbarium specimens as part of the *species*Link network (CRIA, 2012). We based our analyses on data available as of May 2012, at which point *species*Link provided access to data from 87 (presently 97) herbarium collections, including 83 from Brazil, plus the collections of the New York Botanical Garden, Muséum National d'Histoire Naturelle of Paris, U.S. National Museum of Natural History, and Missouri Botanical Garden (see list in Acknowledgements; data from Field Museum of Natural History are now also included in the network). We constrained queries to records with associated latitude-longitude coordinates (derived either from the original record or from centroids of the county) that were fully consistent (i.e., falling in the correct county, see below).

*Species*Link has incorporated advanced data-cleaning tools that examine data for likely erroneous records, helping to improve data quality by detecting and flagging errors, providing cautionary indicators for data users. These tools can be used to profile data sets as to a broad diversity of inconsistencies, ranging from nonstandard taxonomic names and coordinates in wrong administrative units. What is more, plant records in *species*Link have been harmonized with the *List of Species of the Brazilian Flora* (Forzza et al., 2012), a dynamic online platform that is updated regularly and scrutinized by a massive network of Brazilian plant taxonomists. Hence, species can be recognized and characterized via names that were either synonyms of the 1079 plant species classified as Data Deficient by MMA (2008) or phonetic homonyms that represent likely typographic errors (Table A1). We obtained records for 842 of the 934 angiosperm species listed as Data Deficient by MMA (2008), 384 Endangered species, and 24,795 additional angiosperm species for which conservation status has not been assessed.

2.2. Completeness of geographic distributional knowledge

We explored a novel extension of inventory completeness statistics (Colwell and Coddington, 1994; Peterson and Slade, 1998) that transposes the matrix of presences and absences of species at sites through time. That is, instead of examining the accumulation of species records at sites, we examined the accumulation of known occurrence sites for each individual species, following the example of Soberón et al. (2000). We processed initial *species*Link downloads as follows. We created a 'time' marker as the concatenation of year, month, and day (using only records for which day, month, and year were all available). Similarly, we created markers for 'place', in the form of a latitude–longitude combination; here, following lessons learned in our previous analysis (Sousa-Baena et al., in press), we rounded geographic coordinates to the nearest ½°. Finally, we used the binomial scientific name as an identifier for taxon.

In Microsoft Access 2010, we generated tables of unique combinations of taxon, place, and time. For each of (1) Data Deficient