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

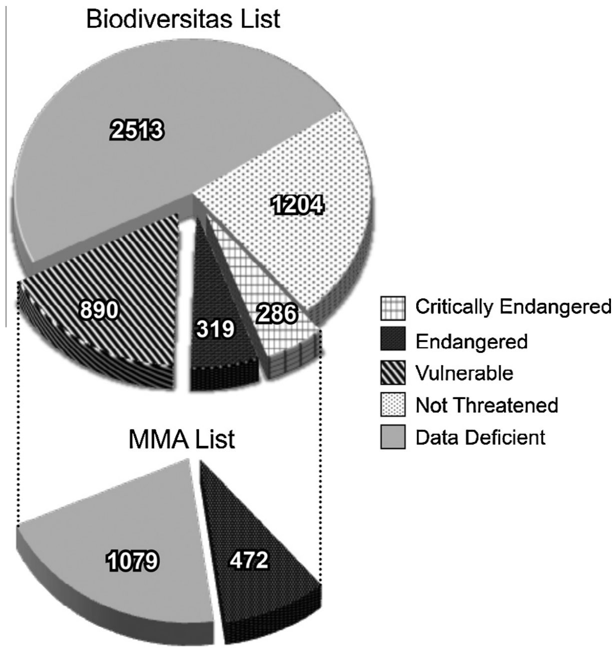

Fig. 1. Comparison of numbers of species considered as threatened according to Biodiversitas Foundation (2005) and the Brazilian Ministry of Environment (MMA, 2008). Virtually the same set of plant species is classified in very different ways in the two lists. Biodiversitas divides the species set in 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

Please cite this article in press as: Sousa-Baena, M.S., et al. Knowledge behind conservation status decisions: Data basis for "Data Deficient" Brazilian plant species. Biol. Conserv. (2013), http://dx.doi.org/10.1016/j.biocon.2013.06.034

species according to MMA (2008), (2) Endangered species according to MMA (2008), and (3) all remaining species, we calculated the observed number of ½° grid cells from which the species is known (*Lobs*), as well as the number of grid cells at which species had been collected once only (a) or twice only (b). From these three numbers, following equations presented by Chao (1987) that allow estimation of numbers of unsampled elements in an inventory, we calculated the expected number of grids cells for each species as $Lexp = Lobs + a^2/2b$. L_{exp} is a measure of likely overall range size, given the pattern of accumulation of knowledge of the species' range over time. Finally, we calculated geographic distribution knowledge completeness as $C = Lobs/Lexp$, following Colwell and Coddington (1994), Peterson and Slade (1998), and Soberón et al. (2000). To assess differences in C between MMA Endangered and Data Deficient species, we used a nonparametric Wilcoxon test, as implemented in Past 1.82b (Hammer et al., 2001).

2.3. Time since last record

Another criterion that can be used to clarify conservation status is the amount of time that has passed since the last record of a species (IUCN Standards and Petitions Subcommittee, 2011). Knowledge regarding whether specimens of a given species are all old, or whether recent records are available, may be useful in removing species from the Data Deficient category (Good et al., 2006). Hence, we identified the date of last collection for each species (focusing on records for which day, month, and year information were available), particularly for species with low range knowledge completeness indices ($C < 0.8$).

2.4. Niche modeling to delimit distributions

Species with numbers of records $N \geqslant 5$ and $C \geqslant 0.8$ were considered as likely well-documented and not Data Deficient. However, for the remaining species, which presented either or both of $N < 5$ and $C < 0.8$, we used available distributional data to develop a preliminary ecological niche model and model-based distributional estimate. Our aim was to obtain a first-pass assessment of likely quality of models that could be created for a species, based on a simple, default-parameter overview, but also based on years of experience with the technique (Peterson et al., 2011).

To create these models, we developed a principal component analysis of climates across Brazil, characterized via 19 ''bioclimatic'' environmental variables from Hijmans et al. (2005); we retained the 4 principal components that presented eigenvalues of $\geqslant 1$, and visualized variation in the first three components as a red–green–blue color space (see Fig. 2). We used Maxent on default settings, but with random seed and 5 bootstrapped replicate analyses per species, to generate preliminary niche models—we emphasize that a full, detailed analysis would customize many elements and spend far more time on model development for each species (Peterson et al., 2011). Our purpose herein, however, was simpler: we sought only an indication as to whether occurrence data available would be at all sufficient (in quantity and quality) for more in-depth informative niche models. Hence, for a random 10% of MMA's Data Deficient species, we inspected models visually, and categorized results as (1) no model possible (i.e., Maxent did not produce a prediction), (2) poor (Maxent output not in any way realistic), (3) good (Maxent output generally logical and consistent, but not precise), or (4) excellent (Maxent output identified a specific set of areas and conditions that likely constitute the species' distribution; Table A2). Although somewhat subjective, this classification serves to provide a general idea of model quality and quality of distributional knowledge.

3. Results

3.1. How comprehensive is DAK for Data Deficient species?

Records of Data Deficient plant species were mostly from southern and southeastern Brazil (Fig. 2). Most of these species were characterized by 5–50 records, with 316 species documented by $\geqslant$50 records (Fig. 3A); 148 (15.9%) species had no or single records. Regarding geographic distributions, 425 species (45.5%) were known from $\leqslant$5½° pixels; 207 species were known only from a single pixel (Fig. 3B). Smaller numbers of species have seen extensive geographic sampling, with 46 known from 20 pixels, 23 from 30 pixels, and 7 from 40 pixels (Fig. 3B).

Our analysis of range knowledge completeness (C) showed that many Data Deficient species' ranges remain poorly documented, with $C \leqslant 0.3$, whereas many other species had $C = 1$ (Fig. 3C). In the middle, 184 species (17.9%) had $0.35 < C < 0.75$ (Fig. 3C). Comparing completeness of range knowledge of MMA Endangered species with MMA Data Deficient species and plant species with no status, we found similar proportions of species across the C spectrum (Fig. 3D; $W = 144$; $P = 0.7368$). On the other hand, the major difference between 'no status' species (24,795 spp.) and those of the two status categories was the low proportion of species with $C = 1$ among 'no status' species (Fig. 3D). The vast majority of species, either considering only species with $C \geqslant 0.8$ (Fig. 4A) or the entire set of Data Deficient species (Fig. 4B), had been collected

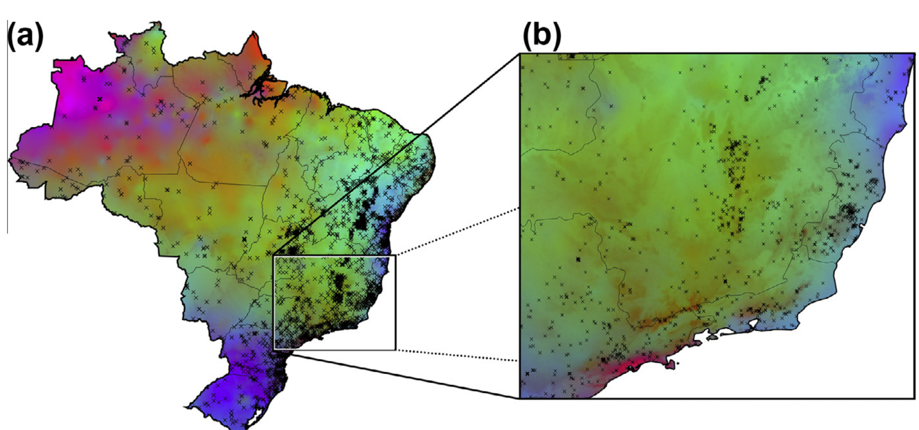

Fig. 2. (A) Primary biodiversity data for Data Deficient species (time-place unique records of Data Deficient species from MMA, 2008) overlaid on a principal components space based on 19 bioclimatic parameters across Brazil (Hijmans et al., 2005). (B) Inset: close-up of southeastern Brazil.

Please cite this article in press as: Sousa-Baena, M.S., et al. Knowledge behind conservation status decisions: Data basis for ''Data Deficient'' Brazilian plant species. Biol. Conserv. (2013), http://dx.doi.org/10.1016/j.biocon.2013.06.034

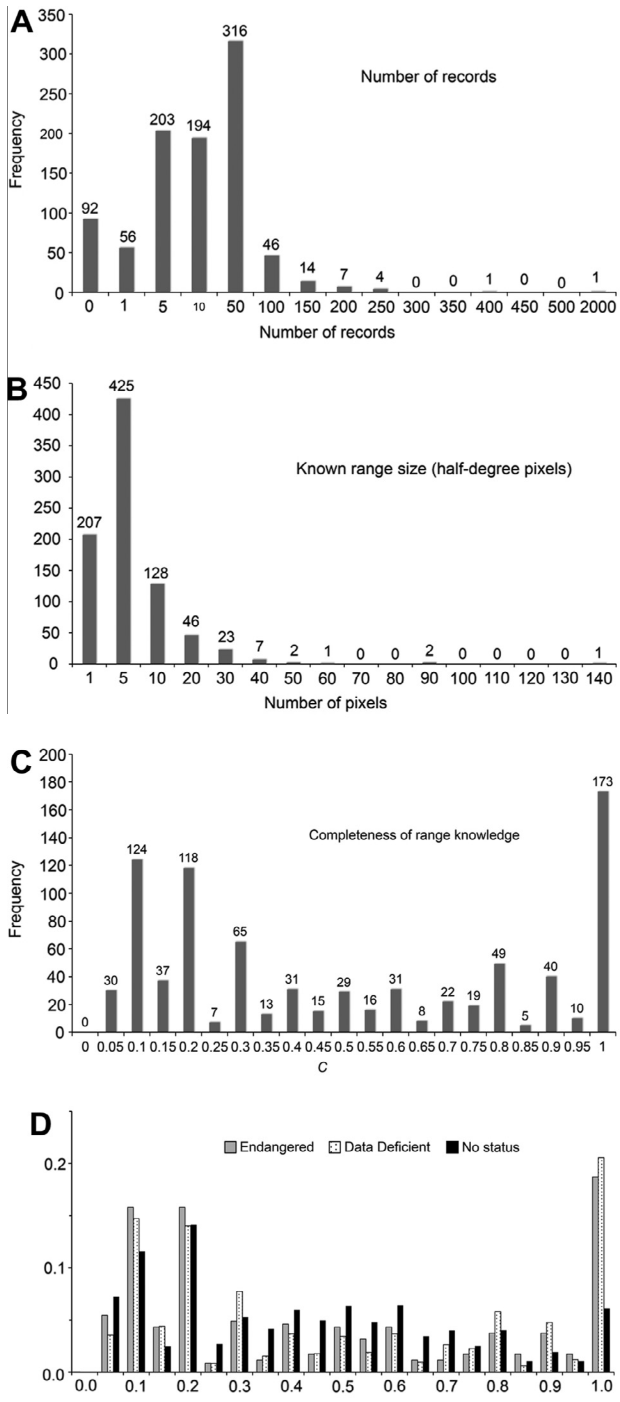

Fig. 3. Data Deficient species frequency according to (A) number of unique species-place-time records, (B) number of half-degree pixels, and (C) inventory completeness range. (D) Comparison among frequency distribution of species by inventory completeness (*C*) for Endangered (384 spp.), Data Deficient (842 spp.), and 'no status' (24,795 spp.) species according to MMA (2008).

within the last 10 years; specimens of 394 (42.2%) Data Deficient species had been collected in the past 5 years (Fig. 4).

3.2. Completeness of distributional knowledge

In all, 152 Data Deficient (16.3%) species met our criteria for consideration as well-documented, with $N \geqslant 5$ and $C \geqslant 0.8$ (Fig. 5). For the 690 remaining species, which presented either or both of $N < 5$ or $C < 0.8$, geographic distributional knowledge was less clear. We assessed niche model quality for a random selection of 10% of these species (Table A2): we found that excellent models could be calibrated for 33.3% of species, whereas for 17.4%, no model could be produced for lack of sufficiently rich data (Fig. 6A). We considered 30.4% of models as poor, and 18.8% as good (Fig. 6B).

For illustration, the model for *Heteropterys conformis* W.R. Anderson (Fig. 6B) was considered poor, because it predicted suitable conditions for occurrence of the species across areas with climatic conditions quite different from those along the Brazilian coast where the species is known to occur. The model for *Dyckia rariflora* Schult. & Schult. f. appeared to represent the potential distribution of the species reliably and was classified as good (Fig. 6C). As an example of an excellent model, the model for *Bactris pickelli* Burret. predicted suitable sites close to known occurrences, in good agreement with knowledge of the distribution of the species (Fig. 6D).

4. Discussion

4.1. Digital Accessible Knowledge and Brazilian plants

Our analyses were based on a large-scale effort to assemble information resources regarding herbarium-specimen documentation of the Brazilian flora. In line with recent global efforts (Canhos et al., 2004), Brazilian institutions have mobilized massive information resources, particularly as part of the *species*Link network (CRIA, 2012). At the time of our analyses (May 2012), the *species*Link network provided access to data from 87 (presently 97) herbarium collections, including 83 from Brazil, plus four major North American and European collections. *Species*Link also provides access to many zoological and microbiological biodiversity information resources as well.

This network has seen extensive and intensive attention to data quality and data fitness for use, including advanced data-cleaning tools and consistent treatment of taxonomic arrangements (Forzza et al., 2012). Indeed, an earlier attempt to develop such analyses (Peterson and Canhos, unpublished) failed precisely for lack of fitness for use, as the present tool set was then only incompletely implemented (in 2009); similarly, criticisms have been leveled at other such data networks based on data not being sufficiently 'fit for use' (Yesson et al., 2007). In this sense, Brazilian biodiversity information resources now present unique and rich opportunities by which to analyze the country's rich biodiversity, but which has heretofore been largely inaccessible or only accessible via laborious and time-consuming manual effort.

4.2. Conservation status and data sufficiency

Nine threat categories may be applied to taxa, including seven that are classifications of degree of threat, plus Data Deficient and Not Evaluated (IUCN Standards and Petitions Subcommittee, 2011). Not Evaluated and Data Deficient are not categories of threat. The former indicates only that the species in question has not been evaluated against Red List criteria, whereas the latter indicates taxa that were evaluated, but for which sufficient information appears to be lacking by which to estimate risk of extinction. Taxa in this category may have well-known biology, but insufficient data on their abundance and/or distribution (IUCN Standards and Petitions Subcommittee, 2011).

Via analysis of completeness of geographic range knowledge, time since last record, and quality of ecological niche models, we suggest that for at least 40.9% of MMA Data Deficient species considerable data and useful information exists (Fig. 7). This number

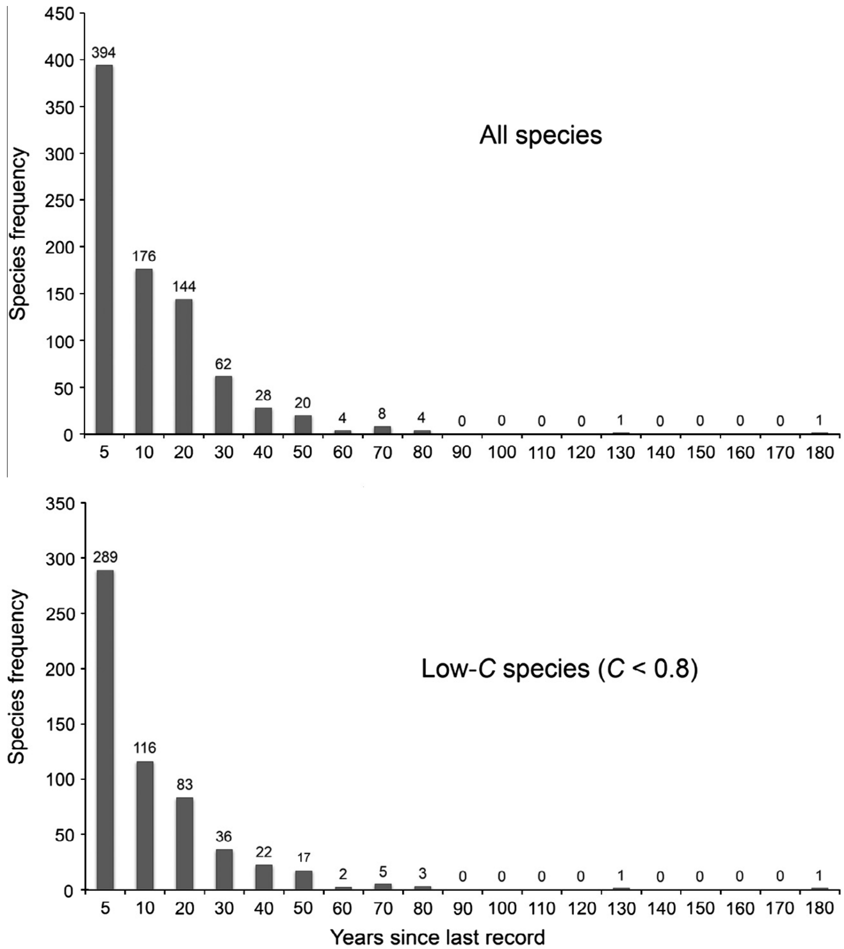

Fig. 4. Frequency of all Data Deficient and low-*C* Data Deficient species (MMA) in terms of years since the last record of the species.

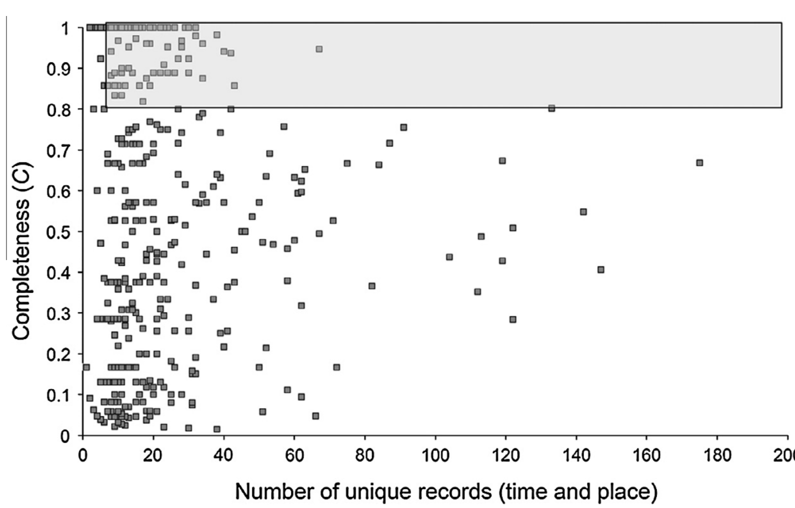

Fig. 5. Completeness of knowledge of the geographic distribution (*C*) for all Data Deficient species according to MMA (2008), showing numbers of unique species-place-time records; rectangle indicates species considered as well-documented ($N \geqslant 5$ and $C \geqslant 0.8$).

comprises species that are well-documented distributionally (i.e., exploratory criteria of $N \geqslant 5$ and $C \geqslant 0.8$), plus other species for which excellent niche models could easily be developed, even using very preliminary and cursory modeling approaches (Fig. 7). Some amount of DAK exists for still more species, but niche models generated for these species were not considered as excellent, at least based on our very preliminary efforts (Fig. 7). Hence, useful and informative DAK is available for at least 382 species, which corresponds to 40.9% of the angiosperm species presently considered as Data Deficient by MMA (2008; Fig. 7). On the other hand, combining 92 species, for which no records were available, with 330 species for which only poor niche models or no models at all

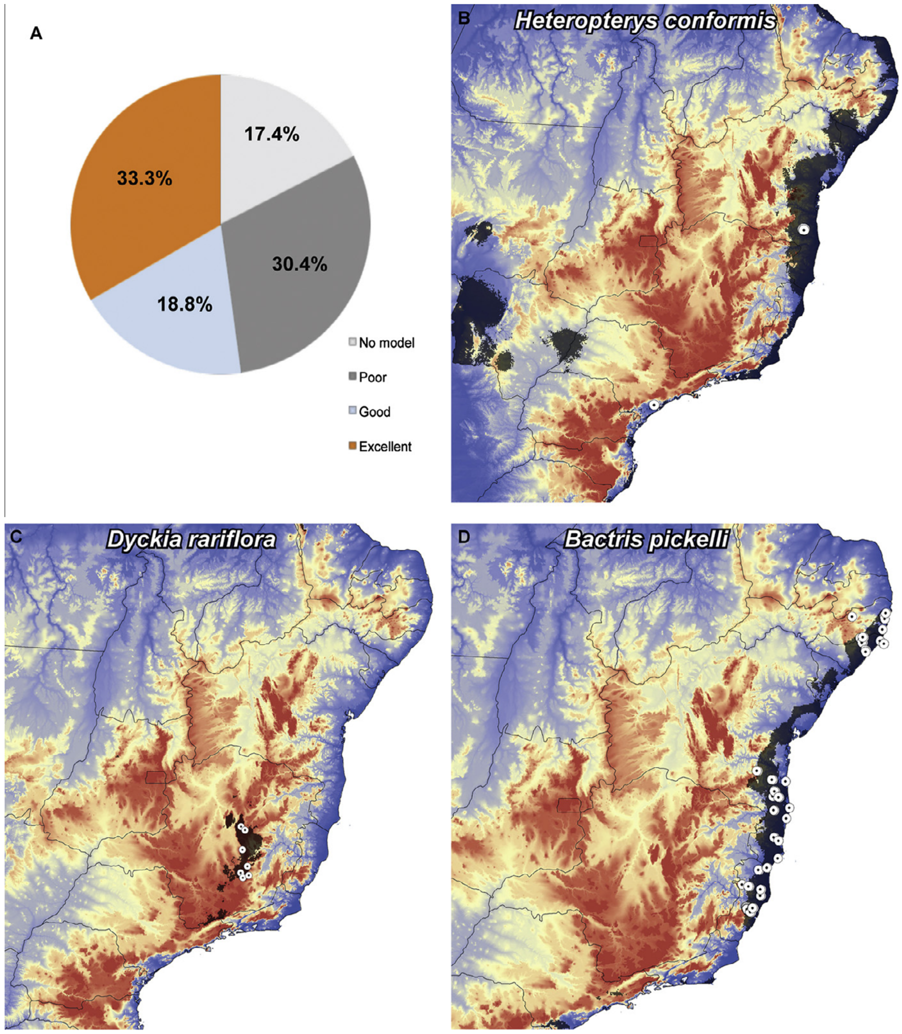

Fig. 6. Summary of ecological niche modeling results across species. (A) Pie chart showing percentage of models obtained according to subjective evaluation of quality. (B) Potential distribution model for *Heteropterys conformis* W.R. Anderson, which was considered a poor model. (C) Potential distribution model for *Dyckia rariflora* Schult. & Schult. f., which was considered a good model. (D) Potential distribution model for *Bactris pickelli* Burret., which was considered an excellent model.

could be generated, 422 (45.2%) "Data Deficient" species are defensibly lacking sufficient information on which to base decisions (Fig. 7).

4.3. Why so many species on the official Data Deficient list?

The official list published by MMA (2008) came after a much broader list elaborated years previously by Biodiversitas Foundation and collaborators (2005) using IUCN criteria, in which 1495 species were classified according to degree of threat (Critically Endangered, Endangered, Vulnerable; Fig. 1). However, in the MMA list, many species were left as Data Deficient, yet the details of which information was missing for each species were not provided by MMA (2008) in listing them as Data Deficient. Of 1079 total species, MMA's "Data Deficient" list includes 1054 species that were under threat according to Biodiversitas/IUCN (Fig. 1): 178 Critically Endangered, 209 Endangered, and 667 Vulnerable species (Fig. 1; Biodiversitas Foundation, 2005). Only two species are considered Data Deficient by both lists. The Data Deficient category of the Biodiversitas/IUCN summary, which included 2513 species out of a much-larger set of species considered (Biodiversitas Foundation, 2005), unfortunately, is not published or available online, and to our knowledge was not considered in the development of the MMA "official" lists.

To assign a conservation status designation to a given species, experts consider the entire spectrum of background knowledge and research about the species. Government decision-makers, however, may wish for greater certainty in such decisions, since, for every species added to the list of threatened species, they must create new policies and action plans towards its conservation (Scarano and Martinelli, 2010). Hence, it seems that a miscommunication occurred between scientists and decision-makers in Brazil, which in turn translates into disconnects in biodiversity conservation (Scarano and Martinelli, 2010). IUCN Red List criteria have become clearer and more objective over the years (Mace et al., 2008), but room still exists for subjective decisions in assigning conservation status. In fact, subjectivity and inconsistent use of IUCN criteria in national lists has already caused conflicts between global and regional lists, with the same species having been assigned different status ratings at different geographic extents (Brito et al., 2010).

IUCN recommends use of whatever data are available, highlighting that decisions between Data Deficient versus some threat status must be made carefully. In general, if the distribution of the poorly-known taxon in question is potentially restricted and if considerable time has passed since the last record, threatened status may well be justified. Indeed, sweeping, indiscriminate use of Data Deficient as a category is discouraged (IUCN Standards and Petitions

Please cite this article in press as: Sousa-Baena, M.S., et al. Knowledge behind conservation status decisions: Data basis for "Data Deficient" Brazilian plant species. Biol. Conserv. (2013), http://dx.doi.org/10.1016/j.biocon.2013.06.034

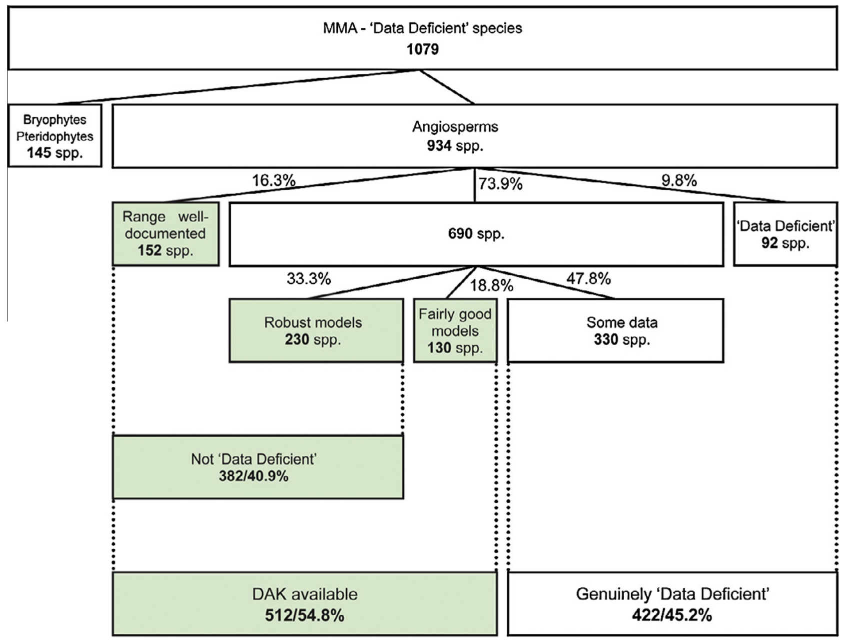

Fig. 7. Diagram summarizing Digital Accessible Knowledge (DAK) for Data Deficient species, and how we evaluated percentages of 'genuinely Data Deficient' species versus 'DAK available' species.

Subcommittee, 2011). Put another way, when information about taxa is genuinely lacking, Data Deficient species could be accorded the same degree of protection as threatened taxa (IUCN Standards and Petitions Subcommittee, 2011). Assignment of Data Deficient status should be accompanied by available supporting information, such as information regarding the condition of and threats to the habitats of the species (Butchart and Bird, 2010).

None of these suggestions appears to have been followed by MMA, and criteria used by MMA in deciding that particular species were Data Deficient were not disclosed. Because lack of knowledge of geographic distributions is one of the criteria for Data Deficient status, we expected to find more species with low *C*-values among Data Deficient species than among threatened species. Rather, we found similar distributions of *C* values among these two groups (Fig. 3D). In addition, although another criterion for Data Deficient status is time since the last collection, we found that 76.4% of MMA Data Deficient species have been collected within the last 20 years. Hence, in the metrics available to us, we find little evidence that species classified by MMA as Data Deficient are indeed less well-known than species that were accorded some threat category.

MMA indicated a requirement for specific action plans for each "official" Brazilian Endangered species. Specific action plans have been developed for 43 animal taxa by an environmental agency of the Brazilian government (Instituto Chico Mendes de Conservação da Biodiversidade, 2012), but the only plant groups currently with national action plans are species in the families Cactaceae and Eriocaulaceae (Silva et al., 2011; Instituto Chico Mendes de Conservação da Biodiversidade, 2011). Hence, conservation plans have not been developed even for most Endangered plant species.

4.4. Conclusions

Considering the huge numbers of species yet to be discovered and rapid potential rates of extinction even before formal scientific description (Costello et al., 2013), calls have been made for conventional academic paths to be reconsidered (Maddison et al., 2012). The situation is particularly critical in tropical countries, owing to more rapid rates of habitat destruction and possibly more vulnerable species (Vamosi and Vamosi, 2008). Although data from museum specimens have long proven key in assessing conservation status (Crandall et al., 2009), DAK derived from specimen data and now easily accessible can facilitate rapid assessments (van den Eynden et al., 2008).

Here, we have analyzed a case in which conservation status assignments by different actors with different priorities and different data sources ended in confusion (see parallel example in Pino del Carpio et al., 2011), which nonetheless affected the Official List of Brazilian Threatened Flora. Our analyses of DAK showed that for 40.9% of Data Deficient species (species with excellent models plus well-documented species with $N \geqslant 5$ and $C \geqslant 0.8$), data are sufficient to define geographic distributions in detail (Fig. 7). Even for plants for which occurrence data were sparse (~14% of species), records are frequently sufficient to generate reliable niche models. Although range descriptions are certainly not the only consideration in status assessments (IUCN Standards and Petitions Subcommittee, 2011), they are a critical starting point; combined with information on recency of records, this information may frequently be enough to avoid the Data Deficient designation.

A recent study focused on Brazilian Combretaceae used data from the literature and herbarium records, and reassessed conservation status of 11 species (Borges et al., 2012). This assessment did not agree completely with either the Biodiversitas/IUCN list or the MMA list. In addition, the authors concluded that 71% of species classified by MMA as Data Deficient were not actually lacking in sufficient data, and should rather be placed in different threat categories. Borges et al. (2012) emphasized that such category shifts would be a consequence of information access, rather than real changes in extinction risk.

In sum, we have explored the meaning of "Data Deficient" as applied to Brazilian plant species. Clearly, this label is complicated, since it is a function of many variables: detectability, gaps in taxonomic knowledge, unknown geographic distributions, vague existing information (e.g., on specimen labels), and the intensity of old collections (Golding, 2004; Callmander et al., 2005; Lõhmus, 2009; Butchart and Bird, 2010). In Brazil, however, we have documented large numbers of plant species currently indicated as Data

Please cite this article in press as: Sousa-Baena, M.S., et al. Knowledge behind conservation status decisions: Data basis for "Data Deficient" Brazilian plant species. Biol. Conserv. (2013), http://dx.doi.org/10.1016/j.biocon.2013.06.034

Table A1
Number of plant species from the Biodiversitas Foundation list (Biodiversitas Foundation, 2005), and Annexes I and II of Brazilian Ministry of Environment Official list that have been harmonized with the *List of Species of the Brazilian Flora* (LSBF).

		Biodiversitas	Annex I	Annex II
Species' status in the *List of Species of the Brazilian Flora* (LSBF)	Not found	42	13	45
	Found with incorrect spelling	56	22	43
	Found as synonym	20	5	14
	Total number of species validated by the LSBF	76	27	57

Deficient, but for which large amounts of information is available—while this information may or may not be sufficient to move the species to an official threat category (or to a no-threat status, in some cases), it clearly suggests that better use could be made of the DAK that exists and is readily available. Here, scientists have made massive data streams available, and yet the science-to-policy conduit appears to be broken. A detailed discussion of this disconnect and how it can be reconnected would be beneficial to all involved, and particularly to the plant species.

Acknowledgements

We are endlessly grateful to Sidnei de Souza for deriving the main datasets for analysis and helping in manifold ways, and to Dora A.L. Canhos for invaluable input and guidance. Principal thanks go to the herbaria in Brazil and elsewhere who have generously shared access to the unique and valuable datasets under their care: Herbário Alexandre Leal Costa da Universidade Federal da Bahia; Herbário da Universidade Federal de Sergipe; Herbário Antônio Nonato Marques; Xiloteca Calvino Mainieri; Herbário da Universidade Federal de Minas Gerais; Herbário Irina Delanova Gemtchújnicov da Universidade Estadual Paulista; Xiloteca Profa. Dra. Maria Aparecida Mourão Brasil da Universidade Estadual Paulista; Herbário do Centro de Pesquisas do Cacau; Herbário Centro Norte Mato Grossense da Universidade Federal do Mato Grosso; Coleção de Plantas Medicinais e Aromáticas da Universidade Estadual de Campinas; Herbário Pe. Dr. Raulino Reitz da Universidade do Extremo Sul Catarinense; Herbário da Reserva Natural Vale; Herbário Prisco Bezerra da Universidade Federal do Ceará; Herbário do Instituto Federal de Educação, Ciência e Tecnologia do Amazonas; Herbário Jaime Coelho de Moraes da Universidade Federal da Paraíba; Herbário da Escola Superior de Agricultura "Luiz de Queiroz"; Herbário do Departamento de Botânica da Universidade Federal de Santa Catarina; Fundación Puerto Rastrojo, Colombia; Herbário da Universidade Estadual de Londrina; Herbário da Fundação Universidade Federal de Mato Grosso do Sul; Herbário Dr. Roberto Miguel Klein da Universidade Regional de Blumenau; Herbário da Universidade Tecnológica Federal do Paraná Campus Campo Mourão; Herbário de Ilha Solteira da Universidade Estadual Paulista; Herbário do Jardim Botânico Plantarum; Herbário Rioclarense da Universidade Estadual Paulista; Herbário de São José do Rio Preto e Herbário de Algas de São José do Rio Preto da Universidade Estadual Paulista; Herbário Sérgio Tavares da Universidade Federal Rural de Pernambuco; Herbário do Trópico Semiárido; Herbário da Universidade Católica de Pernambuco; Herbário da Universidade Estadual de Feira de Santana; Herbário da Universidade Estadual de Goiás; Herbário Dárdano de Andrade Lima da Universidade Federal Rural do Semi-Árido; Herbário da Universidade Estadual de Maringá; Herbário da Universidade Estadual do Sudoeste da Bahia; Herbário da Universidade Estadual de Santa Cruz; Herbarium Uberlandense da Universidade Federal de Uberlândia; Herbário Mogiense da Universidade de Mogi das Cruzes; Herbário

Table A2
Species for which potential geographic distribution was modeled, with a subjective evaluation of the quality of the outcome.

Species	Model quality
Byrsonima cipoensis	Excellent
Odontocarya vitis	Excellent
Vriesea bituminosa	Excellent
Vriesea diamantinensis	Excellent
Vanhouttea leonii	Excellent
Encholirium heloisae	Excellent
Jacaranda ulei	Excellent
Vriesea sucrei	Excellent
Dyckia rariflora	Excellent
Bernardia similis	Excellent
Alophia coerulea	Excellent
Marsdenia queirozii	Excellent
Encholirium horridum	Excellent
Alstroemeria amabilis	Excellent
Mitracarpus anthospermoides	Excellent
Hemipogon hatschbachii	Excellent
Encholirium irwinii	Excellent
Hoffmannseggella ghillanyi	Excellent
Baccharis pseudoalpestris	Excellent
Mikania capricorni	Excellent
Piper laevicarpum	Excellent
Bactris pickelii	Excellent
Guatteria campestris	Excellent
Sellocharis paradoxa	Good
Microlicia psammophila	Good
Duguetia salicifolia	Good
Myrsine villosissima	Good
Camarea humifusa	Good
Eugenia oxyoentophylla	Good
Lychnophora brunioides	Good
Hyptidendron conspersum	Good
Tibouchina papyrus	Good
Byrsonima blanchetiana	Good
Agalinis ramulifera	Good
Smilax lutescens	Good
Byrsonima microphylla	Good
Richterago polyphylla	Poor
Cissus apendiculata	Poor
Constantia cristinae	Poor
Lychnophora crispa	Poor
Hippeastrum vittatum	Poor
Diplopterys amplectens	Poor
Dendrophorbium catharinense	Poor
Simaba floribunda	Poor
Cyrtopodium dusenii	Poor
Viguiera filifolia	Poor
Heteropterys admirabilis	Poor
Couepia monteclarensis	Poor
Staurogyne itatiaiae	Poor
Lychnophora blanchetii	Poor
Zephyranthes caerulea	Poor
Vriesea wawranea	Poor
Tillandsia heubergeri	Poor
Rudgea reflexa	Poor
Heterocoma albida	Poor
Heteropterys conformis	Poor
Alstroemeria malmeana	Poor
Lobelia santos-limae	No model
Microlicia flava	No model
Stenopadus aracaensis	No model
Paepalanthus grao-mogolensis	No model
Begonia crispula	No model
Sinningia carangolensis	No model
Trichocline incana	No model
Encholirium vogelii	No model
Encholirium longiflorum	No model
Lymania spiculata	No model
Encholirium biflorum	No model
Dyckia ursina	No model

da Universidade Estadual de Ponta Grossa; Herbário Vale do São Francisco da Universidade Federal do Vale do São Francisco; Herbário do Instituto Agronômico de Campinas; Carpoteca do Instituto Nacional de Pesquisas da Amazônia; Herbário do Instituto Nacional de Pesquisas da Amazônia; Coleção de Madeiras – Xiloteca do Instituto Nacional de Pesquisas da Amazônia; Herbário Dárdano de Andrade Lima do Instituto Agronômico de Pernambuco; Herbário do Parque da Ciência Newton Freire Maia; Herbário Dimitri Sucre Benjamin; Banco de DNA do Jardim Botânico do Rio de Janeiro; Xiloteca do Jardim Botânico do Rio de Janeiro; Herbário Lauro Pires Xavier da Universidade Federal da Paraíba; Laboratório de Botânica e Ecologia Vegetal da Universidade Federal do Acre; Herbário do Instituto do Meio Ambiente do Estado de Alagoas; Herbário do Museu Botânico Municipal de Curitiba; Herbário Mello Leitão do Museu de Biologia Professor Mello Leitão; Herbário do Museu Integrado de Roraima; Muséum National d'Histoire Naturelle – Herbier Virtuel Saint-Hilaire; Missouri Botanical Garden; Herbário do Museu da Pontifícia Universidade Católica do Rio Grande do Sul; U.S. National Museum of Natural History; The New York Botanical Garden; Sistema de Informações Biogeográficas dos Oceanos; Herbário Professor Vasconcelos Sobrinho da Universidade Federal Rural de Pernambuco; Herbário do Museu Nacional; Herbário do Museu Nacional – Tipos; Herbário Rondoniensis da Universidade Federal de Rondônia; Sistema de Informação Ambiental; Herbário Rosa Mochel da Universidade Estadual do Maranhão; Coleção de Fanerógamas do Herbário do Estado "Maria Eneyda P. Kaufmann Fidalgo" do Instituto de Botânica; Herbário do Departamento de Botânica da Universidade de São Paulo; Herbário Dom Bento Pickel do Instituto Florestal; Herbário Graziela Barroso da Universidade Federal do Piauí; Herbário da Universidade de Brasília; Herbário da Universidade Estadual de Campinas; Herbário da Universidade Federal do Acre; Herbário da Universidade Federal de Goiás; Herbário Geraldo Mariz da Universidade Federal de Pernambuco; Carpoteca da Universidade Federal de Pernambuco; Herbário da Universidade Federal do Rio Grande do Norte; Herbário do Departamento de Botânica da Universidade Federal do Paraná; and Herbário Central da Universidade Federal do Espírito Santo VIES. Support for M.S. Sousa-Baena and L.C. Garcia was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq grants573883/2008-4, 563342/2010-2, and CNPq/EuropeanComission co-funded project EUBrazilOpenBio FP7-ICT-2011/CNPq 590052/2011-0). A.T. Peterson's work in Brazil was supported by a grant from the Fulbright Program of the United States.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at See Tables A1 and A2.

References

Biodiversitas Foundation, 2005. Lista da flora brasileira ameaçada de extinção. Biodiversitas Foundation, Belo Horizonte <http://www.biodiversitas.org.br/floraBr/> (accessed 05.12).

Borges, R., Moraes, M.A., Monteiro, N.P., Bevacqua, A.M., Martinelli, G., Marquete, N., 2012. Available data and risk assessment of the Brazilian threatened species of Combretaceae. Rodriguesia 63, 31–38.

Brito, D., Ambal, R.G., Brooks, T., De Silva, N., Foster, M., Hao, W., Hilton-Taylor, C., Paglia, A., Rodríguez, J.P., Rodríguez, J.V., 2010. How similar are national red lists and the IUCN Red List? Biol. Conserv. 143, 1154–1158.

Butchart, S.H.M., Bird, J.P., 2010. Data Deficient birds on the IUCN Red List: what don't we know and why does it matter? Biol. Conserv. 143, 239–247.

Callmander, M.W., Schatz, G.E., Lowry, P.P., 2005. IUCN Red List assessment and the global strategy for plant conservation: taxonomists must act now. Taxon 54, 1047–1050.

Canhos, V.P., Souza, S., Giovanni, R., Canhos, D.A.L., 2004. Global biodiversity informatics: setting the scene for a "new world" of ecological forecasting. Biodiv. Inf. 1, 1–13.

Canhos, D.A.L., Sousa-Baena, M.S., Souza, S., Garcia, L.C., De Giovanni, R., Maia, L.C., Bonacelli, M.B.M., submitted for publication. Lacunas: a web interface to identify plant knowledge gaps to support informed decision-making. Biodivers. Conserv.

Celep, F., Dogan, M., Kahraman, A., 2010. Re-evaluated conservation status of *Salvia* (sage) in Turkey I: the Mediterranean and the Aegean geographic regions. Turk. J. Bot. 34, 201–214.

Chao, A., 1987. Estimating the population size for capture-recapture data with unequal catchability. Stat. Sinica 10, 227–246.

Colwell, R.K., Coddington, J.A., 1994. Estimating terrestrial biodiversity through extrapolation. Philos. Trans. Roy. Soc. B 335, 101–118.

Costello, M.J., May, R.M., Stork, N.E., 2013. Can we name Earth's species before they go extinct? Science 339, 413–416.

Crandall, K.A., Robison, H.W., Buhay, J.E., 2009. Avoidance of extinction through nonexistence: the use of museum specimens and molecular genetics to determine the taxonomic status of an endangered freshwater crayfish. Conserv. Gen. 10, 177–189.

CRIA, 2012. *species*Link. Centro de Referência em Informação Ambiental, Campinas <http://www.splink.org.br/index> (accessed 05.12).

Forzza, R.C., Leitman, P.M., Costa, A.F., Carvalho Jr., A.A., Peixoto, A.L., Walter, B.M.T., Bicudo, C., Zappi, D., Costa, D.P., Lleras, E., Martinelli, G., Lima, H.C., Prado, J., Stehmann, J.R., Baumgratz, J.F.A., Pirani, J.R., Sylvestre, L., Maia, L.C., Lohmann, L.G., Queiroz, L.P., Silveira, M., Coelho, M.N., Mamede, M.C., Bastos, M.N.C., Morim, M.P., Barbosa, M.R., Menezes, M., Hopkins, M., Secco, R., Cavalcanti, T.B., Souza, V.C., 2012. Lista de espécies da flora do Brasil. Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro <http://floradobrasil.jbrj.gov.br/2012> (accessed 05.12).

Golding, J.S., 2004. The use of specimen information influences the outcomes of Red List assessments: the case of southern African plant specimens. Biodivers. Conserv. 13, 773–780.

Good, T.C., Zjhra, M.L., Kremen, C., 2006. Addressing data deficiency in classifying extinction risk: a case study of a radiation of Bignoniaceae from Madagascar. Conserv. Biol. 20, 1099–1110.

Hammer, O., Harper, D.A.T., Ryan, P.D., 2001. Past: paleontological statistics software package for education and data analysis. Palaeontol. Electron. 4, 1–9.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

Instituto Chico Mendes de Conservação da Biodiversidade, 2011. Plano de ação nacional para conservação das sempre-vivas. Instituto Chico Mendes de Conservação da Biodiversidade, Ministério do Meio Ambiente, Brasília <http://www.icmbio.gov.br/portal/biodiversidade/fauna-brasileira/plano-de-acao/2050-plano-de-acao-nacional-para-a-conservacao-das-sempre-vivas.html>.

Instituto Chico Mendes de Conservação da Biodiversidade, 2012. Planos de ação nacionais. Instituto Chico Mendes de Conservação da Biodiversidade, Ministério do Meio Ambiente, Brasília <http://www.icmbio.gov.br/portal/biodiversidade/fauna-brasileira/planos-de-acao-nacionais.html>.

IUCN Standards and Petitions Subcommittee, 2011. Guidelines for using the IUCN Red List categories and criteria, version 9.0. International Union for the Conservation of Nature, Gland, Switzerland.

Lõhmus, A., 2009. Factors of species-specific detectability in conservation assessments of poorly studied taxa: the case of polypore fungi. Biol. Conserv. 142, 2792–2796.

Mace, G.M., Collar, N.J., Gaston, K.J., Hilton-Taylor, C., Akçakaya, H.R., Leader-Williams, N., Milner-Gulland, E.J., Stuart, S.N., 2008. Quantification of extinction risk: IUCN's system for classifying threatened species. Conserv. Biol. 22, 1424–1442.

Maddison, D.R., Guralnick, R., Hill, A., Reysenbach, A., McDade, L.A., 2012. Ramping up biodiversity discovery via online quantum contributions. Trends Ecol. Evol. 27, 72–77.

MMA, 2008. Normative Instruction n° 06, from September 23, 2008. Ministério do Meio Ambiente, Brasília <http://www.mma.gov.br/estruturas/179/_arquivos/179_05122008033615.pdf>.

Norris, K., 2012. Biodiversity in the context of ecosystem services: the applied need for systems approaches. Philos. Trans. Roy. Soc. B 367, 191–199.

Peterson, A.T., Slade, N.A., 1998. Extrapolating inventory results into biodiversity estimates and the importance of stopping rules. Divers. Distrib. 4, 95–105.

Peterson, A.T., Sánchez-Cordero, V., Martínez-Meyer, E., Navarro-Siqüenza, A.G., 2006. Tracking population extirpations via melding ecological niche modeling with land-cover information. Ecol. Model. 195, 229–236.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B., 2011. Ecological niches and geographic distributions. Princeton University Press, Princeton.

Pino del Carpio, A., Villarroya, A., Ariño, A.H., Puig, J., Miranda, R., 2011. Communication gaps in knowledge of freshwater fish biodiversity: implications for the management and conservation of Mexican biosphere reserves. J. Fish Biol. 79, 1563–1591.

Scarano, F.R., Martinelli, G., 2010. Brazilian list of threatened plant species: reconciling scientific uncertainty and political decision-making. Nat. Conserv. 8, 13–18.

Silva, S.R., Zappi, D.C., Taylor, N.P., Machado, M.C. (Eds.), 2011. Plano de ação nacional para conservação das cactáceas. Instituto Chico Mendes de Conservação da Biodiversidade, Ministério do Meio Ambiente, Brasília.

Siqueira, M.F., Durigan, G., de Marco Júnior, P., Peterson, A.T., 2009. Something for nothing: using landscape similarity and ecological niche modeling to find rare species. J. Nat. Conserv. 17, 25–32.

Soberón, J., Peterson, A.T., 2009. Monitoring biodiversity loss with primary species-occurrence data: toward national-level indicators for the 2010 target of the convention on biological diversity. AMBIO 38, 29–34.

Soberón, J.M., Llorente, J.B., Oñate, L., 2000. The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. Biodivers. Conserv. 9, 1441–1466.
Sousa-Baena, M.S., Garcia, L.C., Peterson, A.T., in press. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Divers. Distrib.
Vamosi, J., Vamosi, S.M., 2008. Extinction risk escalates in the tropics. PLoS ONE 3, 1–6.
van den Eynden, V., Oatham, M.P., Johnson, W., 2008. How free access internet resources benefit biodiversity and conservation research: trinidad and tobago's endemic plants and their conservation status. Oryx 42, 400–407.
Yesson, C., Brewer, P.W., Sutton, T., Caithness, N., Pahwa, J.S., Burgess, M., Gray, W.A., White, R.J., Jones, A.C., Bisby, F.A., Culham, A., 2007. How global is the global biodiversity information facility? PLoS ONE 2, e1124.

Glossary

Digital Accessible Knowledge (DAK): The set of primary biodiversity data that has been made both digital and accessible in standard formats
C: Completeness of knowledge of species' ranges, calculated as L_{obs}/L_{exp}.
L_{obs}: Number of ½ pixels from which a species has been recorded
L_{exp}: Number of ½ pixels in which a species is expected to occur