Diversity and Distributions, (Diversity Distrib.) (2016) 1–8

BIODIVERSITY RESEARCH

A Journal of Conservation Biogeography
Diversity and Distributions

Threatened or Data Deficient: assessing the conservation status of poorly known species

David L. Roberts[1], Lin Taylor[2] and Lucas N. Joppa[3]*

[1]*Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Marlowe Building, Canterbury, Kent CT2 7NR, UK,* [2]*Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK,* [3]*Microsoft Research, 14820 NE 36th St., Redmond, WA 98052, USA*

*Correspondence: Lucas N. Joppa, Microsoft Research, 14820 NE 36th St., Redmond WA 98052, USA.
E-mail: lujoppa@microsoft.com

ABSTRACT

Aim To determine whether extinction risk assessments based on biological collections and using Criterion B of the IUCN Red List Criteria reflect in part an accurate measure of species rarity and thus extinction risk.

Location Madagascar.

Methods We calculate the extent of occurrence (EOO) and area of occupancy (AOO) for orchids using herbarium specimen data. Correlations were made against range, occupancy, extinction risk, number of specimens and the date of description. We calculated the average increase in range per species specimen, correlated this against the date of description and determined significance of the observed EOO accumulation using randomization tests.

Results Significant negative correlations were found between date of description and all measures of range, occupancy and associated Red List Categories and number of specimens, as well as between the average range accumulation per specimen and date of description. Seventy-five percentage of species' observed EOO accumulations significantly differed from random. Maximum deviations between observed EOO accumulations and those derived from random sampling were always significantly positive. For most species, this occurred more frequently during the first half of the accumulation sequence.

Main conclusions Species described more recently have smaller ranges and occupancies, fewer specimens and greater perceived extinction risk status. Levels of geographic uniqueness of collections are higher in species described more recently. Awareness of a species range increased faster than random, particularly in the first half of the sampling process, suggesting that newly discovered or yet to be discovered species are rare and likely have a higher risk of extinction. For many species, biological collections represent the sum of our knowledge. While data may be limited, such species should be listed in an appropriate Red List Category in accordance with the IUCN Red List Guide lines rather than as Data Deficient.

Keywords

date of description, herbarium specimen, Orchidaceae, range, Red List, spatial sampling

INTRODUCTION

Extinction risk assessments, the International Union for Conservation of Nature's (IUCN) Red List being the most widely used, serve as an early warning system to the risk of extinction, enabling conservation planners and decision-makers to prioritize species with limited time and resources (Rodrigues *et al.*, 2006). Full Red List assessments can, however, be difficult to complete for the vast majority of species such as plants and insects (taxa frequently listed as Data Deficient in Red List assessments), as most assessment criteria require detailed information on population size, fragmentation or trends in abundance over time (IUCN, 2001). This is a problem as museum specimen data are often the only

DOI: 10.1111/ddi.12418
http://wileyonlinelibrary.com/journal/ddi

information that exists for many species (Ponder *et al.*, 2001).

Natural history collections, such as herbaria, are increasingly used as a basis for ecological and conservation studies (e.g. Nelson *et al.*, 1990; Rivers *et al.*, 2010; Robbirt *et al.*, 2011). Herbarium specimens are primary, verifiable observations, which provide information about the distribution of taxa through time and space (Ponder *et al.*, 2001). While herbarium and other museum datasets provide an invaluable source of information, they may carry problems of accuracy and interpretation, particularly in historic collections. In general, these problems include (a) incompleteness, information gaps in time and space (Kodric-Brown & Brown, 1993; Ponder *et al.*, 2001; Küper *et al.*, 2006); (b) geographical collection biases, notably due to accessibility (Nelson *et al.*, 1990; Reddy & Dávalos, 2003; Kadmon *et al.*, 2004); (c) spatial variations in collection effort (Freitag & Jaarsveld, 1998); (d) geographical imprecision due to post hoc georeferencing of specimens (Schatz, 2002; Golding, 2004); (e) subsequent taxonomic revisions (Schatz, 2002); (f) lack of reliable absence data (Elith *et al.*, 2006); and (g) collection biases due to variations in interspecific detectability (Sheth *et al.*, 2008).

Numerous studies have investigated the factors affecting the relative likelihood of species description (e.g. Gaston *et al.*, 1995; Allsopp, 1997; Collen *et al.*, 2004; Roberts & Marshall, 2009). A plant species is formally 'described' when it is assigned a Latin name and description in accordance with the International Code of Nomenclature for algae, fungi and plants (McNeill *et al.*, 2012). But few have studied how the accumulation of knowledge about a species post-description affects its subsequent Red List Assessment. In this study, we ask, does an extinction risk assessment based on herbarium specimens reflect an accurate measure of species extinction risk, or merely the extent of our knowledge at a specific point in time [a product of (geographical and temporal) sampling biases]?

We addressed this question by focussing on the sampling biases affecting the prediction of species' geographic ranges, which forms the basis of Criterion B under which many plant species are classified into IUCN Red List Categories (IUCN, 2001; Baillie *et al.*, 2008; Brummitt *et al.*, 2008). Specifically, we asked, (1) what is the relationship between species date of description, number of specimens and geographic range-based extinction risk assessment, (2) is perceived extinction risk based on spatial rarity a product of undersampling, and (3) does the accumulation of range for a species with increased sampling occur at random?

Understanding these issues will allow increased confidence in when, or whether, a species extinction risk can be inferred from incomplete data and can potentially affect Red List assessments by increasing the efficacy of the assessment process. Recent estimates suggest that 27–33% of the total flowering plant species are threatened with extinction (Joppa *et al.*, 2011). Given the current rate of species loss, estimated to be 100–1000 times the background rate (Pimm *et al.*, 2014), conservation planners cannot wait for the luxury of a more accurate and comprehensive dataset. Understanding and offering a way forward in terms of data reliability and associated biases is therefore of concern to those conducting and using the results of rapid extinction risk assessments (i.e. extinction risk assessments based the geographic range components of Criterion B).

METHODS

Study taxon

We investigated the plant family Orchidaceae on the island of Madagascar. With nearly 90% endemism in Madagascar (Hermans *et al.*, 2007; Cribb & Hermans, 2010), this is also a relatively well-studied family with good taxonomic resolution. High levels of species and generic endemism extend to other taxa on Madagascar, and with extensive habitat loss, it is a global biodiversity hotspot (Myers *et al.*, 2000).

Dataset

Using the Orchids of Madagascar checklist (Hermans *et al.*, 2007) as a stable taxonomy, the dataset was compiled from 21 herbaria, using all available herbarium specimens for the region (Rivers *et al.*, 2011), including all non-endemics. Each record was georeferenced; records with little or no location information were omitted. Records for the same species from the same locality (i.e. identical latitude and longitude) and date were treated as duplicates (i.e. non-independent) and therefore were only recognized once in subsequent analyses. Species with only one specimen were also removed, as a minimum of two records was required to predict a species' spatial occupancy (see Willis *et al.*, 2003; Moat, 2007). The final dataset represented 3950 records, of 491 species, from 53 genera, collected over a 174-year period (1830–2004). Early specimen records were not included for 28% of species in our dataset as they contained no georeferenceable information. This can be seen when the recorded date of description was earlier than the earliest date of collection for a georeferenced specimen; we termed this a 'negative lag time' (see Discussion).

Preliminary analysis

For each species, specimens were added in chronological order, each specimen resulting in a potential accumulation of the perceived range size of the species, as well as the corresponding change in estimated extinction risk level. This was calculated using the Conservation Assessment Tools software extension (Moat, 2007) for ArcView GIS 3.3 (ERSI). Preliminary extinction risk categories were determined based on the IUCN Red List Criterion B (IUCN, 2001), specifically extent of occurrence (EOO) and area of occurrence (AOO) following Brummitt *et al.* (2008). Ours is a preliminary effort because a formal Red List assessment would require the consideration of additional subcriteria after calculation of EOO and AOO. EOO and AOO are range and occupancy measures recognized by the IUCN Red List for assessing extinction risk status (see IUCN Criteria A, B and D; IUCN, 2001, Joppa *et al.*, 2015). EOO is defined as the minimum area encompassing all records. Here, this is calculated using a minimum convex polygon surrounding all sites of incidence. A minimum of three species records is required; species with fewer than three records were excluded. AOO is defined as the area within the EOO where a species is actually found to occur (IUCN, 2001). This is calculated by placing a grid of cells over the occurrence points. The total AOO is the product of the number of occupied cells and the area of a single cell. Two methods of defining grid cell size were used in this study. Firstly, we used a fixed cell size of 2 km × 2 km, as recommended by the IUCN (2001). Secondly, we used a shifting cell size, which is calculated as one-tenth of the longest axis between two occurrence points (Willis *et al.*, 2003; Moat, 2007). This second measure allows the cell size to be adjusted according to the unique spatial distribution of each species, reducing scaling problems associated with using a fixed cell size. It should be noted that the IUCN Guidelines for using the Red List Categories and Criteria do not endorse the use of the shifting cell size for calculating the AOO (IUCN Standards and Petitions Subcommittee, 2014). Finally, as a minimum of two occurrence points was required to calculate the shifting cell size, species with just one specimen were excluded from the AOO analyses.

Statistical tests

We determined the earliest date of description for each species using the World Checklist of Selected Plant Families (WCSP, 2009). We used Spearman's rank correlation to establish the relationship between species dates of description with (a) EOO, (b) EOO Red List rating, (c) AOO (shifting cell width), (d) AOO (shifting cell width) Red List rating, (e) AOO (fixed cell width), (f) AOO (fixed cell width and (g) number of specimens. AOO and EOO Red List assessments were ranked in order of extinction risk level, and each category was assigned a number from one to five: 1 = Critically Endangered (CR), 2 = Endangered (E), 3 = Vulnerable (V), 4 = Near Threatened (NT) and 5 = Least Concern (LC). The category NT does not have a recognized IUCN threshold (IUCN, 2001) and was thus assigned in keeping with the power relationship between the other EOO thresholds (Moat, 2007). LC is the default category for species that do not come under any of the range thresholds.

We also determined whether the number of specimens for a species was a function of the length of time that had elapsed since description, or whether it was independent of time, that is whether recently described species were truly less likely to be sampled (spatially rarer in the sense of AOO and EOO) or whether this was simply a sampling artefact. To do this, we calculated the average increase in range size (EOO) per specimen for a species, by dividing the total range size by the number of specimens, and then correlated this against the date of description for that species. We calculated this 'average range accumulation' for all three measures of species range size. We would expect species that are truly spatially rare, in the sense of AOO and EOO, to have smaller average range accumulations. Finally, we determined the significance of the observed EOO accumulation curves using randomization tests in a stepwise process:

1. Starting with the earliest three records, we calculated the area (EOO) of the minimum convex polygon (MCP). By iteratively adding observation records in order of date collected, we obtained an observed accumulation curve (OAC) from 3 to *n* records, where *n* equals the total number of observations of a given species.

2. We then created 100 random accumulation sequences by, for each new sequence, randomizing the original sequence of observations and calculating the randomized accumulation curve (RAC).

3. We then subtracted the OAC vector from each of the 100 RAC vectors ($D_{i...n}$), and for each comparison summed the absolute values of the observed differences (OD_{sum}), leaving us with a distribution of 100 OD_{sum} values.

4. We tested the significance of this distribution by creating another against which to compare it. This second distribution was the result of subtracting each of the 100 RACs from every other RAC, resulting in 4950 random D_{sum} (RD_{sum}) values.

5. This left two distributions. The first, OD_{sum} represented observed compared to random, and the second, RD_{sum} represented random compared to random. By comparing these two distributions via a simple Kolmogorov–Smirnov test, we were able to ask whether the OAC deviated significantly from random expectation.

We used an identical procedure as above to test whether the maximum deviation from random in the OAC deviated significantly from random expectation. In this instance, instead of summing the differences between the OAC and each RAC, we simply retained the value of the maximum deviation from zero (positive or negative). From this, we termed OD_{max}, and as before, we compared our 100 values of OD_{max} to 4950 values of RD_{max} via a Kolmogorov–Smirnov test. Finally, we asked whether or not the D_{max} values were more likely to fall in the first or second half of the species observation sequence using a Binomial sign test (i.e. do the largest deviations from random expectation occur early on in our knowledge of a species or towards the tail end?).

RESULTS

Significant negative correlations were found between date of description and all measures of species range, occupancy and associated IUCN Red List Categories, as well as number of specimens (Table 1). This suggests that species that are described more recently have smaller range sizes, smaller

Table 1 The correlation between range, occupancy, threat level and number of specimens for a species and the species date of description.

Variable	Date of description	
	r_s	d.f.
a) EOO	−0.47***	352
b) EOO Red List rating	−0.43***	352
c) AOO (shifting cell width)	−0.52***	473
d) AOO (shifting cell width) Red List rating	−0.44***	473
e) AOO (fixed cell width)	−0.53***	480
f) AOO (fixed cell width) Red List rating	−0.37***	480
g) Number of specimens	−0.51***	490

***Denotes P values < 0.001.

occupancies (by all measures) and fewer specimens, and as a result, their perceived extinction risk status is greater, based on range and occupancy under the IUCN Red List Criterion B.

A significant negative correlation was found between the average range accumulation per specimen for EOO ($r_s = -0.31$; $P < 0.001$; d.f. = 352) and AOO (sliding cell width) ($r_s = -0.41$; $P < 0.001$; d.f. = 473) range and occupancy measures, and the date of description. Interestingly, a significant positive correlation was found for the AOO (fixed cell width) ($r_s = 0.16$; $P < 0.01$; d.f. = 480), indicating that for species described recently, individual specimens tend to represent unique cells, whereas those described in the past tend to be represented by multiple collections per cell. In other words, levels of geographic uniqueness of collection localities are higher in species described more recently.

We found that 74.8% of species had an observed EOO accumulation (OAC) significantly different from random, while for 89.0% of species, the maximum deviation from random was significantly different from expected. Figure 1 shows examples for two representative species, one with an OAC significantly different from random expectation (*Vanilla madagascariensis* Rolfe) and the other (*Angraecum compactum* Schltr.) with no significant difference. The maximum deviation between the EOO derived from the observed accumulation of species observations and the EOO derived from random sampling was always significantly positive (Binomial sign test, P value < 0.05), meaning that at least in one point in time, observed sampling resulted in significantly greater awareness of a species EOO than random sampling would have. Moreover, for most species (57%) that greater awareness occurred (significantly – Binomial sign test, P value <0.05) more frequently in the first half of the accumulation sequence than in the second (for 17% of the species, it occurred significantly more frequently in the second half). Thus, the greatest gap in awareness between observed sampling and random occurs earliest on in the process of sampling species locations and results in significantly greater information than random sampling.

DISCUSSION

As expected, recently discovered species proved to have fewer specimens, smaller ranges and higher extinction risk status. This is in line with several other studies that have found geographic range size to play a part in the date of discovery of a species (Blackburn & Gaston, 1995; Cabrero-Sanudo & Lobo, 2003; Collen *et al.*, 2004). Identifying the causal relationship between these variables is an important distinction to make, as the validity of the use of herbarium specimens for the production of rapid Red List assessments relies on the assumption that species are not under-sampled.

It may seem contradictory that when controlling for number of specimens, one finds a significant negative correlation between date of description and average increase in EOO and AOO (sliding cell width) per specimen, but a corresponding significant positive correlation for AOO (fixed cell width). However, these results add support to the suggestion that recently discovered species are truly rare as opposed to under-sampled. To understand this, one must recognize the differences between EOO and AOO (sliding cell width) representing measures of total range size as determined by the distance between the furthest distribution points, and AOO (fixed cell width) as a measure of occurrence within this range.

Consider the initial phase of exploration in Madagascar during the 19th century, whereby the majority of specimens would have been collected from a few, intensively sampled locations, probably as a result of the limited points of entry to Madagascar (i.e. safe anchorage ports), with much of the interior inaccessible. In some cases, these initial collections may have been for commercial horticultural, rather than scientific, purpose, negating the need for random sampling. As these sites of anchorage and therefore collection would most likely have been far apart due to the island's topology, this would result in multiple specimens falling within the same 2 km × 2 km squares, thus low geographic uniqueness per specimen. This results in a large EOO and AOO (sliding cell width), due to the considerable distance between the anchorage and therefore collection points, but small AOO (fixed cell width) due to the limited number of sites. In contrast, recently discovered species have a smaller EOO and AOO (sliding cell width), suggesting that the entire specimen record is based on a smaller area. However, on average each specimen contributes more to the AOO (fixed cell width), that is high geographic uniqueness per specimen. This is likely to occur because the reason for specimen collection has changed to biodiversity surveys with the aim of maximizing coverage. Therefore, although sampling is more thorough, overall range sizes are still smaller, implying that species discovered more recently and have fewer specimens are highly localized (i.e. rarer) as opposed to being the products of a sampling artefact. This also implies that the date of description and number of specimens for a species could serve as initial indicators of the extinction risk status of a species.

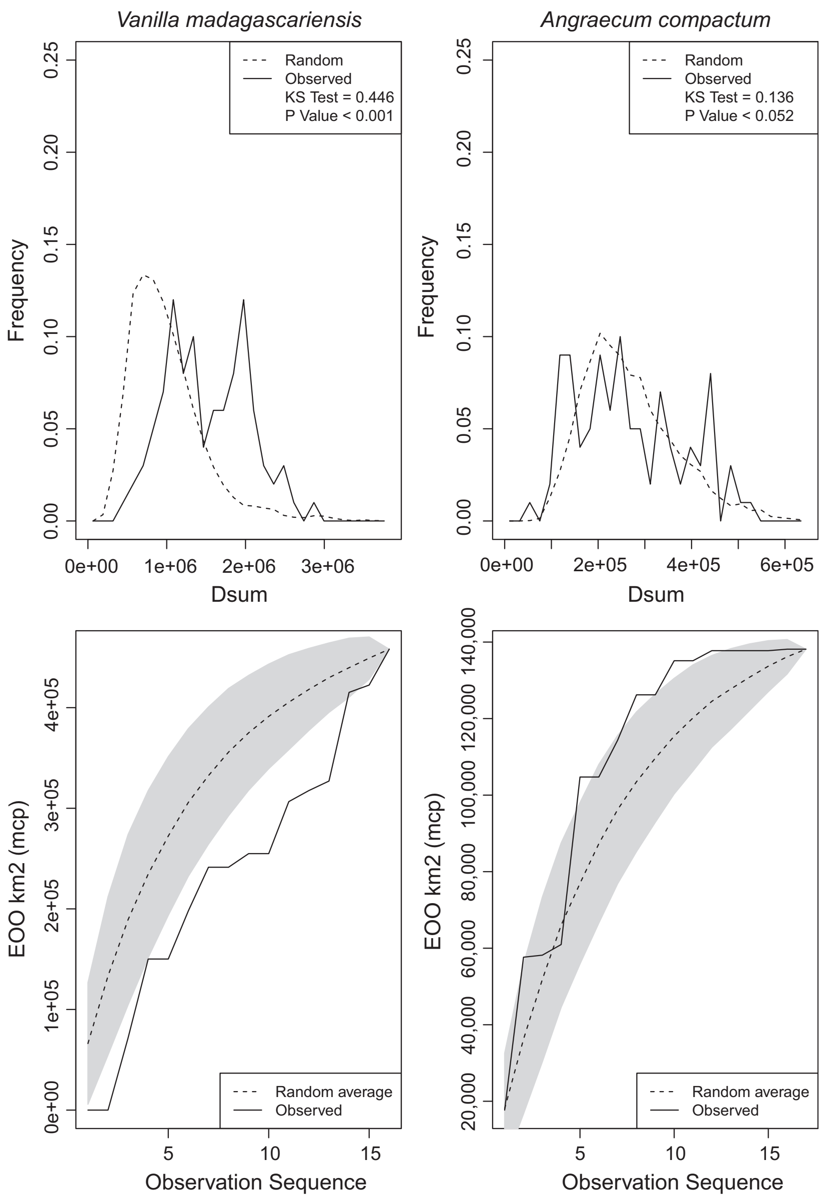

Figure 1 Examples of two species, one of which (*Vanilla madagascariensis*, left panel) has an accumulation pattern significantly different from random and the other (*Angraecum compactum*, right panel) shows no significant difference. Top: The distribution of OD_{sum} values (solid line) and RD_{sum} values (dashed line). Kolmogorov–Smirnov test and *P* values are reported in the legend for the comparison between the two distributions. Bottom: The calculated EOO (*y*-axis) with the inclusion of each additional species observation (*x*-axis) in solid line. The mean EOO at each iteration across all randomization attempts is shown by the dashed line, bounded by the mean ± 1 standard deviation in light grey.

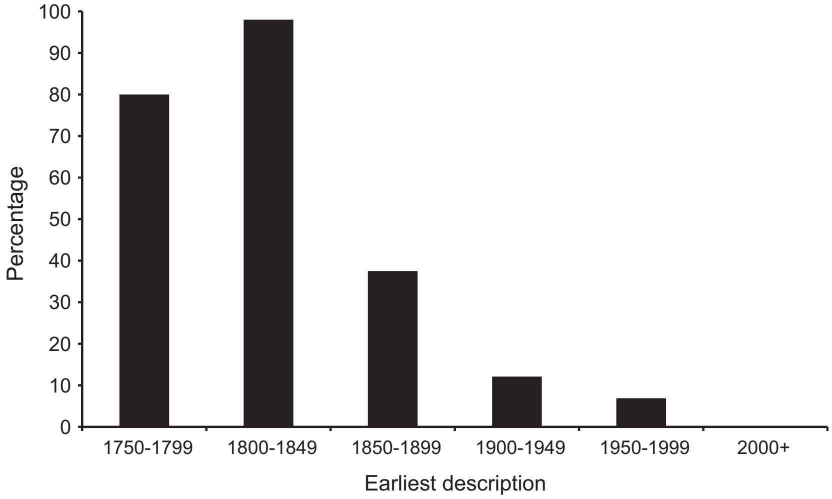

Figure 2 The percentages of negative lag times in the dataset (species with their earliest specimen after their earliest date of description) according to when the species was described. A marked decrease through time is observed.

The use of herbarium specimens has come under criticism due to the potential for sampling error in the dataset (e.g. Nelson *et al.*, 1990; Reddy & Dávalos, 2003; Kadmon *et al.*, 2004); however, these results show that while concerns may exist, herbarium specimens are a robust source of data for conservation decision-makers on which to determine the extinction risk status of a species. The proportion of the species in our dataset whose observed accumulation sequence differed significantly from random was high when compared with the distributions of all random-to-random D_{sum} and D_{max} values. The removal of early specimens due to the lack of georeferenceable information for a proportion of the species also appeared to have very little effect as they do not represent unique localities as seen in the differing results between EOO and AOO (sliding cell width) compared to AOO (fixed cell width) as discussed earlier.

Incomplete dataset

The comprehensiveness of the database fell short of the *c.* 1000 species and 60 genera of orchid known on Madagascar. We had to exclude 6% of records (28% species) in our original dataset as Data Deficient, as these belonged to species with only one specimen. Calculation of species EOO required a minimum of three specimens, removing a further 6% records (23% species) from this analysis. Subsequent analysis of the Data Deficient species revealed that they followed the trend highlighted by Spearman's rank correlation of having earlier dates of description. If our earlier inference is correct, this suggests the species that could not be assessed are in fact at high risk of extinction. IUCN guidelines (IUCN, 2001) acknowledge this, stating that the Data Deficient status should only be assigned as a last resort, due to the potential for these species to be overlooked. In addition, the dataset was missing entire specimen records for many species. The earliest record for a species in the database was often later than the earliest recorded date of description, as previously discussed, and many species had synonyms listed in the Orchids of Madagascar checklist (Hermans *et al.*, 2007) that had fewer specimens than synonyms. Possible reasons for an incomplete specimen record include (a) that many of the older species were poorly spatially resolved and could not be georeference and (b) that there was a greater chance that older specimens would have been lost or destroyed. These two points are supported by the fact that older specimens were more likely to have negative lag times, that is the older type specimens were more likely to be missing from the dataset (Fig. 2). This suggests an increased accuracy through time in terms of spatial resolution and reliability of specimen records. Further, older specimens are less likely to reflect the current day distribution of the species, making recent specimens more reliable and as a result knowledge of recently discovered species more reliable. It should be noted that the definition of EOO and AOO stipulate that they should be based on a species' current spatial distribution (IUCN 2001), rather than historic records that may not reflect a species' current EOO or AOO due to factors such as habitat clearance. One solution discussed by Rivers *et al.* (2010) is to work backwards in time, using the most recent specimens for extinction risk assessments. It would therefore be interesting to see whether the trend for recently discovered species to be more threatened still holds true if the extinction risk assessment is based solely on the five or 10 most recent specimens for a species.

CONCLUSIONS

With 10–20% of plant species still awaiting discovery (Joppa *et al.*, 2011), this study suggests that they are likely to be rare and at risk of extinction. For many species, not just plants, museum collections often represent the sum of our knowledge and therefore their conservation status. For many of these species, it is unlikely, given the current rate of bio diversity loss, that significant numbers of herbarium specimens and other data associated with more robust extinction risk assessments will be forthcoming in the near further. While data may be limited, reliable preliminary conservation assessments can be made, even for species that are only know from three specimens (i.e. the number required to calculate the EOO). Although those conducting extinction risk assessments may lean towards listing species as Data Deficient, the IUCN guidelines (IUCN, 2001) explicitly state that the Data Deficient status should only be assigned as a last resort. This study suggests that such species should be listed in accordance with the results of a preliminary assessment under Criterion B of the IUCN Red List (IUCN, 2001; Brummitt *et al.*, 2008).

ACKNOWLEDGEMENTS

The authors thank the following Phillip Cribb, Clare Hermans, Johan Hermans and David Houghton for compiling the dataset, Justin Moat, Landy Rajaovelona and Mijoro Rakotoarinivo for georeferencing the herbarium specimens records that made up the dataset and Stuart Pimm for comments on an earlier draft of the manuscript.

REFERENCES

Allsopp, P.G. (1997) Probability of describing an Australian scarab beetle: influence of body size and distribution. *Journal of Biogeography*, **24**, 717–724.

Baillie, J.E.M., Collen, B., Amin, R., Akcakaya, H.R., Butchart, S.H.M., Brummitt, N., Meagher, T.R., Ram, M., Hilton-Taylor, C. & Mace, G.M. (2008) Toward monitoring global biodiversity. *Conservation Letters*, **1**, 18–26.

Blackburn, T.M. & Gaston, K.J. (1995) What determines the probability of discovering a species?: a study of South American oscine passerine birds. *Journal of Biogeography*, **22**, 7–14.

Brummitt, N., Bachman, S. & Moat, J. (2008) Applications of the IUCN Red List: towards a global barometer for plant diversity. *Endangered Species Research*, **6**, 127–135.

Cabrero-Sanudo, F.J. & Lobo, J.M. (2003) Estimating the number of species not yet described and their characteristics: the case of Western Palaearctic dung beetle species (Coleoptera, Scarabaeoidea). *Biodiversity and Conservation*, **12**, 147–166.

Collen, B., Purvis, A. & Gittleman, J.L. (2004) Biological correlates of description date in carnivores and primates. *Global Ecology and Biogeography*, **13**, 459–467.

Cribb, P. & Hermans, J. (2010) *Field Guide to the Orchids of Madagascar*. Royal Botanic Gardens, Kew.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Freitag, S. & Jaarsveld, A.S.V. (1998) Sensitivity of selection procedures for priority conservation areas to survey extent, survey intensity and taxonomic knowledge. *Proceedings of the Royal Society B: Biological Sciences*, **265**, 1475–1482.

Gaston, K.J., Blackburn, T.M. & Loder, N. (1995) Which species are described first?: the case of North American butterflies. *Biodiversity and Conservation*, **4**, 119–127.

Golding, J. (2004) The use of specimen information influences the outcomes of Red List assessments: the case of southern African plant specimens. *Biodiversity and Conservation*, **13**, 773–780.

Hermans, J., Hermans, C., Du Puy, D., Cribb, P. & Bosser, J. (2007) *Orchids of Madagascar*, 2nd edn. Royal Botanic Gardens, Kew.

IUCN. (2001) *IUCN Red List Categories: Version 3.1*. IUCN, Gland, Switzerland.

IUCN Standards and Petitions Subcommittee. (2014) Guidelines for Using the IUCN Red List Categories and Criteria. Version 11. Prepared by the Standards and Petitions Subcommittee. http://www.iucnredlist.org/documents/RedListGuidelines.pdf (accessed 1 May 2015).

Joppa, L.N., Roberts, D.L. & Pimm, S.L. (2011) How many species of flowering plants are there? *Proceedings of the Royal Society B: Biological Sciences*, **278**, 554–559.

Joppa, L.N., Butchart, S., Hoffmann, M., Bachman, S., Akçakaya, R.H., Moat, J., Böhm, M., Holland, R.A., Newton, A. & Polidoro, B. (2015) The impact of alternative metrics on estimates of extent of occurrence for extinction risk assessment. *Conservation Biology*, doi:10.1111/cobi.12591.

Kadmon, R., Farber, O. & Danin, A. (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, **14**, 401–413.

Kodric-Brown, A. & Brown, J.H. (1993) Incomplete data sets in community ecology and biogeography: a cautionary tale. *Ecological Applications*, **3**, 736–742.

Küper, W., Sommer, J.H., Lovett, J.C. & Barthlott, W. (2006) Deficiency in African plant distribution data - missing pieces of the puzzle. *Botanical Journal of the Linnean Society*, **150**, 355–368.

McNeill, J., Barrie, F.R., Buck, W.R., Demoulin, V., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Marhold, K., Prado, J., vanPrud'homme Reine, W.F., Smith, G.F., Wiersema, J.H. & Turland, N.J. (2012) *International Code of Nomenclature for algae, fungi and plants (Melbourne Code) adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011*. Regnum Vegetabile, 154. Koeltz Scientific Books.

Moat, J. (2007) *Conservation assessment tools extension for ArcView 3. x, version 1.2*. GIS Unit, Royal Botanic Gardens, Kew.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nelson, B.W., Ferreira, C.A.C., da Silva, M.F. & Kawasaki, M.L. (1990) Endemism centres, refugia and botanical collection density in Brazilian Amazonia. *Nature*, **345**, 714–716.

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M. & Sexton, J.O. (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, **344**, 1246752.

Ponder, W.F., Carter, G.A., Flemons, P. & Chapman, R.R. (2001) Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology*, **15**, 648–657.

Reddy, S. & Dávalos, L.M. (2003) Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, **30**, 1719–1727.

Rivers, M., Bachman, S., Meagher, T., Nic Lughadha, E. & Brummitt, N. (2010) Subpopulations, locations and fragmentation: applying IUCN red list criteria to herbarium specimen data. *Biodiversity and Conservation*, **19**, 2071–2085.

Rivers, M.C., Taylor, L., Brummitt, N.A., Meagher, T.R., Roberts, D.L. & Lughadha, E.N. (2011) How many herbarium specimens are needed to detect threatened species? *Biological Conservation*, **144**, 2541–2547.

Robbirt, K.M., Davy, A.J., Hutchings, M.J. & Roberts, D.L. (2011) Validation of biological collections as a source of phenological data for use in climate change studies: a case study with the orchid *Ophrys sphegodes*. *Journal of Ecology*, **99**, 235–241.

Roberts, D.L. & Marshall, C.R. (2009) Are higher taxa described earlier or later than expected by chance? *Systematics and Biodiversity*, **7**, 243–247.

Rodrigues, A.S., Pilgrim, J.D., Lamoreux, J.F., Hoffmann, M. & Brooks, T.M. (2006) The value of the IUCN Red List for conservation. *Trends in Ecology & Evolution*, **21**, 71–76.

Schatz, G.E. (2002) Taxonomy and herbaria in service of plant conservation: lessons from Madagascar's endemic families. *Annals of the Missouri Botanical Garden*, **89**, 145–152.

Sheth, S.N., Lohmann, L.G., Consiglio, T. & Jiménez, I. (2008) Effects of detectability on estimates of geographic range size in Bignonieae. *Conservation Biology*, **22**, 200–211.

Willis, F., Moat, J. & Paton, A. (2003) Defining a role for herbarium data in Red List assessments: a case study of *Plectranthus* from eastern and southern tropical Africa. *Biodiversity & Conservation*, **12**, 1537–1552.

World Checklist of Selected Plant Families (WCSP) (2009) The Board of Trustees of the Royal Botanic Gardens, Kew. Published on the Internet; http://www.kew.org/wcsp/ accessed 2009.

BIOSKETCHES

David L. Roberts has a wide range of interests, from extinction modelling to the illegal wildlife trade, including the role of biological collection in conservation assessments. Central to this is the process of species discovery and description, and how this impacts on our understanding and perception of biodiversity.

Lin Taylor completed a PhD studying plant evolution and development, specifically the evolution of gene families that are important for flower development. Prior to graduate school, Lin undertook a placement at the Royal Botanic Gardens, Kew, where she worked on correlates of species discovery.

Lucas N. Joppa works to conserve ecological systems. His research ranges from quantifying the impacts of conservation actions to unravelling the complexities of species interactions and mapping where species are being discovered – and going extinct. He is most interested in predicting outcomes for ecological communities under an increasingly uncertain environment.

Author contributions: D.L.R conceived the ideas; L.T. collected the data; D.L.R., L.T. and L.N.J analysed the data; and D.L.R. and L.N.J. led the writing.

Editor: Jacqueline Beggs