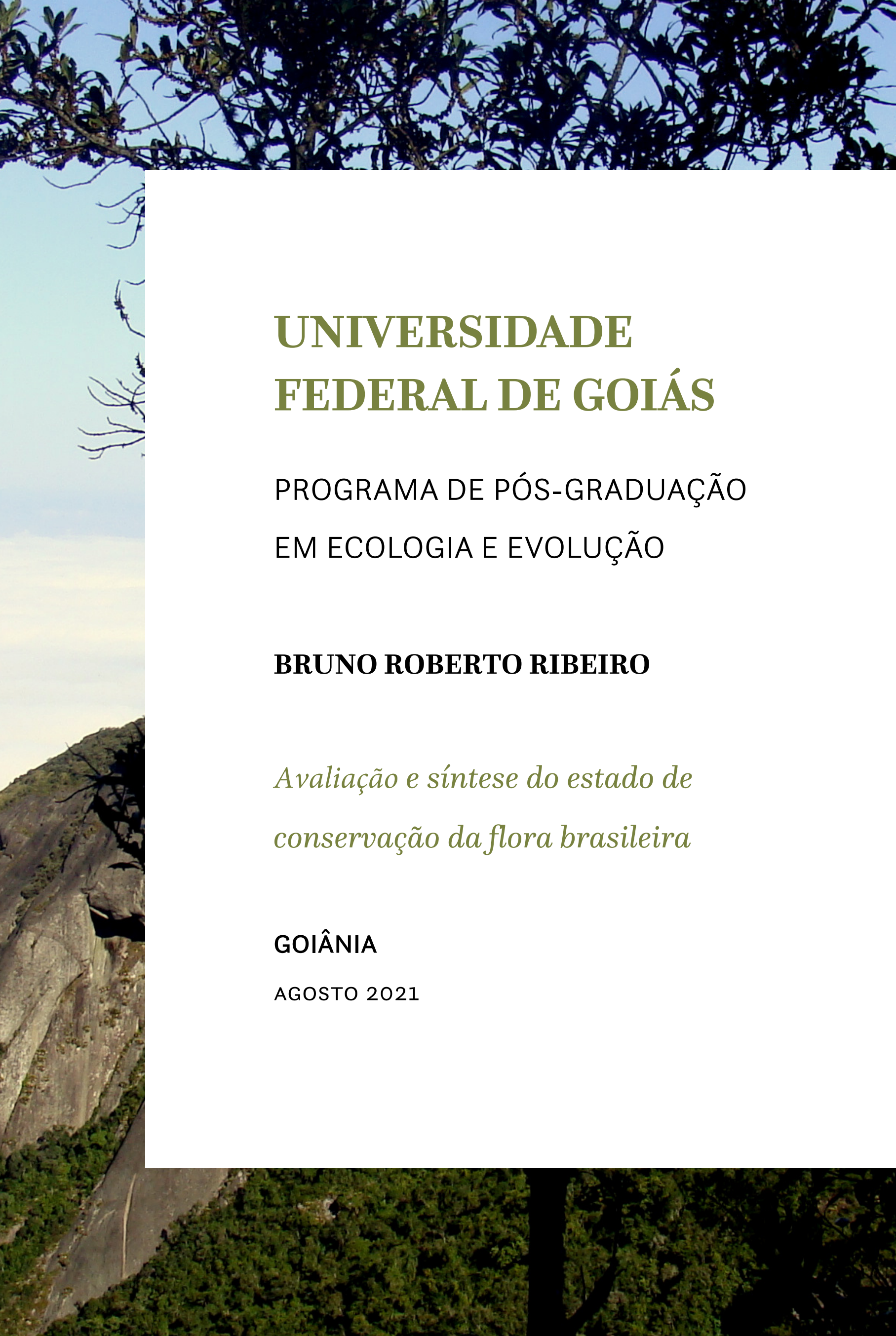


UNIVERSIDADE FEDERAL DE GOIÁS

PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E EVOLUÇÃO

BRUNO ROBERTO RIBEIRO

Avaliação e síntese do estado de conservação da flora brasileira

GOIÂNIA

AGOSTO 2021

Processo:
23070.038222/2021-18

Documento:
2324283

UFG

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[ ] Dissertação [x] Tese

2. Nome completo do autor

Bruno Roberto Ribeiro

3. Título do trabalho

Avaliação e síntese do estado de conservação da flora brasileira

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [x] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

---

sei! assinatura eletrônica

Documento assinado eletronicamente por **Rafael Dias Loyola**, **Professor do Magistério Superior**, em 02/09/2021, às 17:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

---

sei! assinatura eletrônica

Documento assinado eletronicamente por **BRUNO ROBERTO RIBEIRO**, **Discente**, em 17/09/2021, às 15:40, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Processo:
23070.038222/2021-18

Documento:
2324283

Referência: Processo nº 23070.038222/2021-18 SEI nº 2324283

BRUNO ROBERTO RIBEIRO

Avaliação e síntese do estado de conservação da flora brasileira

Tese apresentada ao Programa de Pós-Graduação em Ecologia e Evolução da Universidade Federal de Goiás (UFG) como requisito para obtenção do título de Doutor em Ecologia e Evolução.

Área de concentração: Ecologia e Evolução
Linha de pesquisa: ecologia e conservação

Orientador
Prof. Dr. Rafael Loyola

Goiânia
Agosto 2021

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Ribeiro, Bruno Roberto
Avaliação e síntese do estado de conservação da flora brasileira [manuscrito] / Bruno Roberto Ribeiro. - 2021.
CCLXVI, 266 f.

Orientador: Prof. Dr. Rafael Dias Loyola.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Ecologia e Evolução, Goiânia, 2021.

Apêndice.

1. Brasil. 2. Espécies ameaçadas. 3. Limpeza de dados. 4. plantas. 5.registros de ocorrência. I. Loyola, Rafael Dias, orient. II. Título.

CDU 574

UFG

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

ATA DE DEFESA DE TESE

Ata Nº **108** da sessão de Defesa de Tese de **Bruno Roberto Ribeiro** que confere o título de Doutor(a) em **Ecologia e Evolução**, na área de concentração em **Ecologia e Evolução.**

Aos **trinta e um dias do mês de agosto do ano de dois mil e vinte e um (31/08/2021)**, a partir das **quatorze horas (14h)**, por **webconferência, seguindo portaria CAPES no. 36 de 16 de março de 2020 e recomendação da UFG**, realizou-se a sessão pública de Defesa de Tese intitulada **"Avaliação e síntese do estado de conservação da flora brasileira"**. Os trabalhos foram instalados pelo Orientador, **Prof. Dr. Rafael Dias Loyola, (DECOL/ICB/UFG)**, com a participação dos demais membros da Banca Examinadora: **Profa. Dra. Mariana Pires de Campos Telles (DGEN/ICB/UFG)**, membro titular externo; **Dra. Rafaela Campostrini Forzza (DIPEC/Jardim Botânico do Rio de Janeiro)**, membro titular externo, **Prof. Dr. José Alexandre Felizola Diniz Filho (DECOL/ICB/UFG)**, membro titular interno; **Prof. Dr. Marcus Vinicius Cianciaruso (DECOL/ICB/UFG)**, membro titular interno. Durante a argüição os membros da banca **não fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo **Prof. Dr. Rafael Dias Loyola**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos **trinta e um dias do mês de agosto do ano de dois mil e vinte e um (31/08/2021)**.

TÍTULO SUGERIDO PELA BANCA

---

sei! assinatura eletrônica — Documento assinado eletronicamente por **José Alexandre Felizola Diniz Filho**, **Professor do Magistério Superior**, em 31/08/2021, às 17:39, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

---

sei! assinatura eletrônica — Documento assinado eletronicamente por **Marcus Vinicius Cianciaruso**, **Professor do Magistério Superior**, em 31/08/2021, às 17:40, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

---

sei! assinatura eletrônica — Documento assinado eletronicamente por **Rafael Dias Loyola**, **Professor do Magistério Superior**, em 31/08/2021, às 17:40, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

---

sei! assinatura eletrônica — Documento assinado eletronicamente por **Mariana Pires De Campos Telles**, **Professor do Magistério Superior**, em 31/08/2021, às 17:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

---

sei! assinatura eletrônica — Documento assinado eletronicamente por **Rafaela Campostrini Forzza**, **Usuário Externo**, em 01/09/2021, às 15:01, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Referência: Processo nº 23070.038222/2021-18 SEI nº 2313998

À minha família - meus pais

Rosângela e Carlos, minha esposa

Geiziane e meu filho Ian.

“[We should look] upon every living species of animal and plant now living as the individual letters which go to make up one of the volumes of our earth’s history; and, as a few lost letters may make a sentence unintelligible, so the extinction of the numerous forms of life ... will necessarily render obscure these invaluable records of the past.... future ages will certainly look back upon us as a people so immersed in the pursuit of wealth as to be blind to higher considerations. They will charge us with culpably allowing the destruction of [that] which we had it in our power to preserve; seeing many [species] perish irrecoverable from the face of the earth, uncared for and unknown”.

Alfred Russel Wallace,

Journal of the Royal Geographical Society (1863)

Agradecimentos

Gostaria, humildemente, de discordar do grande escritor mineiro Guimarães Rosa quando ele disse que “A colheita é comum, mas o capinar é sozinho”. O desenvolvimento dessa tese nunca foi um trabalho solitário. Pelo contrário, contou com ajuda de várias pessoas as quais, direta ou indiretamente, me levaram a finalizar este projeto. Esta tese é uma forma simplória de mostrar todo meu apreço e gratidão por todas essas pessoas.

Meu primeiro agradecimento vai para meu orientador Rafael Loyola, que com toda serenidade, respeito e paciência me guiou durante toda minha jornada de formação como cientista. Sem dúvida, aprendi muito com você não apenas sobre ciência, mas sobre vários outros temas fundamentais para minha formação enquanto cientista. Acredito que também me tornei um ser humano melhor convivendo com você todos estes anos. Guardo comigo seu exemplo de orientador e é muito bom tê-lo com um amigo, Rafa!

A minha família de Anápolis (GO), pelo pelos finais de semanas aconchegantes e reconfortantes tão necessários para dar continuidade aos trabalhos da tese. Agradeço ainda à paciência e à compreensão de meus pais, de minha irmã, e de toda minha família de Nova Resende (MG) que, mesmo que distante, sempre acreditaram em mim e me incentivaram a seguir firme no caminho que escolhi.

Aos meus amigos(as) da Laboratório de Biologia da Conservação (CB-Lab), com os quais tive a felicidade de conviver durante sete anos. Esta é uma parte delicada, pois corro o risco de esquecer de alguém importante. Se for o caso, por favor, me perdoe. Agradeço: a Lilian Sales, Fernando Resende, Raisa Vieira, Frederico Faleiro, Nathalia Machado, Fernanda Brum, Daniel Zacarias, Milena Diniz, Rafaela Silva, Nayara Rezende, Luisa Latorre, Rejane Santos-Silva, Fábio Borges, Thalline Rodrigues, Larissa Lemes, Dayany Joner, Maíra Sagnori e Lorena Ribeiro.

Agradeço ainda aos meus amigos(as) do programa de pós-graduação em Ecologia e Evolução (PPG-EcoEvol), em especial Marco Túlio Pacheco, Cristiele Valente, Margarita Florencio, Herlander Lima, Leila Meyer, Luciano Lima, Renato Dalla Corte, Fagner Oliveira, André Andrade, Daniel Plazas, Leandro Maracahipes, Mateus Atadeu, Lorena Simon, Fernando Landa, Davi Crescente, Fabrício Rodrigues, André Menegoto, Marcia Kurtz, e tantos outros.

Aos professores do Programa de Pós-Graduação em Ecologia e Evolução por seus ensinamentos e por me proporcionar uma formação de altíssima qualidade. Em especial, agradeço aos professores Paulo De Marco Jr., José Alexandre Diniz Filho, Marcus Cianciaruso, Adriano Melho e Thiago Rangel e por suas aulas inspiradoras e discussões sobre Ecologia.

Também não posso deixar de agradecer aos meus grandes amigos e coautores(as) dos capítulos dessa tese, Geiziane Tessarolo, Santiago Velazco, Lucas Jardim, Karlo Guidoni e Milena Diniz. Agradeço a paciência e o empenho de vocês. Que possamos colaborar em muitos outros trabalhos.

Ao Rhewter Nunes pela ajuda com a cluster, sem a qual seria impossível realizar as análises dessa tese.

Ao PPG-EcoEvol e à Universidade Federal de Goiás pelo apoio; à CAPES por financiar meus estudos durante estes mais de quatro anos.

Por fim, agradeço, com todo amor do mundo, à minha esposa e companheira Geiziane Tessarolo. Obrigado pelo seu amor, sua dedicação, seu carinho, sua paciência e apoio em todos momentos. Te amo! Ao meu filho Ian, fonte incessante de carinho, amor e beijinhos necessário para me manter atento e motivado.

Sumário

Resumo

O Brasil possui a flora mais rica e talvez mais ameaçada de extinção do planeta, e, por isso, sua conservação efetiva ainda é um grande desafio. No primeiro capítulo desta tese, nos debruçamos sobre o processo de avaliação de risco de extinção das espécies, destacando seus pontos fortes, suas deficiências assim como oportunidades para ampliar e agilizar as avaliações de risco. No segundo capítulo, desenvolvemos e agregamos uma série de ferramentas para integração e análise da qualidade de registros de ocorrência de espécies em um pacote desenvolvido no ambiente R de programação. Exemplificamos sua funcionalidade mostrando como erros e incertezas presentes nos registros de ocorrência alteram o padrão de riqueza de espécies de plantas no Brasil. No terceiro capítulo, avaliamos como erros presentes na informação taxonômica, espacial e temporal de registros de ocorrência podem gerar classificações equivocadas do risco de extinção de espécies de plantas no Brasil. Nossos resultados indicam que dados de alta qualidade são necessários para predição acurada de espécies ameaçadas, e que avaliações preliminares podem prever o risco de extinção de espécies não ameaçadas mesmo na presença de erros taxonômicos, geográficos e temporais. No quarto capítulo, avaliamos a eficácia das unidades de conservação

e das terras indígenas do Brasil em representar espécies de plantas atualmente ameaçadas de extinção. Descobrimos que de 10 a 33% das espécies encontram-se totalmente fora de áreas protegidas. Por fim, no quinto capítulo e no apêndice I apresentamos textos voltados às crianças que descrevem de forma simples e detalhada o processo de avaliação de risco e a situação atual de ameaça sofrida pelas espécies no Brasil. Os resultados dessa tese apontam caminhos em direção a uma maior compreensão e ampliação do processo de avaliação do risco de extinção, de modo a possibilitar a conservação efetiva da biodiversidade.

Palavras-chave: Brasil, espécies ameaçadas, limpeza de dados, lista vermelha, plantas, registros de ocorrência, risco de extinção.

Abstract

Brazil harbors the richest and perhaps the most endangered flora on the planet, and its effective conservation is still a significant challenge. In the first chapter of this Ph.D. dissertation, we thoroughly analyzed the extinction risk assessment process, highlighting its strengths, weaknesses, and opportunities to expand and streamline risk assessments. In the second chapter, we aggregated and developed a series of tools for integrating and analyzing the quality of species occurrence records in an R package. We exemplified its functionality by showing how errors and uncertainties in occurrence records change the species richness pattern of plants in Brazil. In the third chapter, we assessed how errors present in taxonomic, spatial, and temporal information from occurrence records can lead to erroneous classifications of the risk of extinction of plant species in Brazil. Our results indicate that high-quality data are needed for an accurate prediction of threatened species and that preliminary assessments can predict the extinction risk of non-threatened species even in the presence of taxonomic, geographic, and temporal issues. In the fourth chapter, we assessed the effectiveness of protected areas and indigenous lands in representing all known threatened plant species in Brazil. We found that between 10 and 33% of species are entirely outside protected areas.

Finally, in the fifth chapter and in Appendix I, we presented outreach texts aimed at children describing in a simple and detailed way the risk assessment process and the current situation of threat experienced by species in Brazil. The results of this dissertation point out avenues for a greater understanding and expansion of the extinction risk assessment process to enable the effective conservation of biodiversity.

Keywords: Brazil, threatened species, data cleaning, red list, plants, occurrence records, extinction risk.

INTRODUÇÃO GERAL

As plantas são fonte de recursos alimentares e de uma ampla gama de produtos e serviços fundamentais às pessoas, tais como madeira, purificação da água, controle de erosão do solo e regulação climática (Corlett 2016; Heywood 2017; IPBES 2019). Apesar da diversidade impressionante e do fato de serem essenciais para sobrevivência e o bem-estar humano, a conservação de plantas tem recebido muito menos atenção que a conservação de animais (Corlett 2016) e ainda é um grande desafio, principalmente em países megadiversos tais como o Brasil. Conhecer quais são e onde estão presentes as espécies ameaçadas de extinção é fundamental para definição de ações de conservação e manejo para frear o declínio da biodiversidade frente ao ritmo acelerado de alteração e destruição de habitats induzido pelas ações humanas (Vié et al. 2008; Betts et al. 2020).

O Brasil possui a maior flora do planeta, abrigando atualmente um total de 49.990 espécies das quais 2.113 encontram-se ameaçadas de extinção (Flora do Brasil 2020; Martinelli & Moraes 2013; MMA 2014; BFG 2021). Muitas outras espécies da flora brasileira podem estar em situação de ameaçada atualmente, visto que o risco de extinção é conhecido somente para duas a cada dez espécies da flora (Martins, E., Loyola, R., Martinelli 2017). A grande sub-representação de plantas avaliadas nas listas vermelhas (como é conhecido o catálogo

de plantas ameaçadas) dificulta o processo de decisão ambiental e o desenvolvimento de planos, estratégias e políticas públicas para frear o declínio da biodiversidade (Rodrigues et al. 2006; Bennun et al. 2018). Assim, a análise e síntese do estado de conservação da flora brasileira como um todo é fundamental e urgente.

A avaliação do risco de extinção de uma espécie é comumente realizada utilizando o sistema proposto pela União Internacional para a Conservação da Natureza (UICN; Rodrigues et al. 2006). No Brasil, o sistema da UICN é utilizado nas avaliações de risco de extinção da flora conduzidas pelo Centro Nacional de Conservação da Flora (CNCFlora). Esse sistema possui uma série de critérios baseados no número e na distribuição geográfica dos indivíduos de uma espécie para classificá-la em diferentes categorias de risco de extinção (Mace et al. 2008). Por serem quantitativos e explícitos, os critérios utilizados pela UICN podem ser empregados para avaliar do risco de extinção de qualquer espécie (exceto microrganismos; IUCN 2019).

Entretanto, a grande demanda de dados para avaliação e documentação do risco de extinção, somada à necessidade de avaliadores bem treinados, torna o processo de avaliação de risco de extinção demorado e dispendioso (Rondinini et al. 2014).

Apesar dos esforços recentes para tornar a lista vermelha mais representativa, o número de espécies avaliadas atualmente representa uma pequena amostra taxonomicamente e geograficamente enviesada da biodiversidade global (situação similar é encontrada nas avaliações da flora brasileira; Sousa-Baena et al. 2014; Martins, E., Loyola, R., Martinelli 2017).

Atualmente, avaliações do risco de extinção foram realizadas para apenas 5% das espécies descritas pela ciência, sendo que muitas das avaliações encontram-se desatualizadas (Rondinini et al. 2014; IUCN 2021). Para que a lista vermelha possa continuar atuando como um "termômetro" da saúde da biodiversidade (Stuart et al. 2010; Butchart et al. 2010), muitos desafios devem ser superados (Bachman et al. 2019). Nesse sentido, novas abordagens e ferramentas podem tornar o processo de avaliação de risco mais rápido e efetivo, e a lista vermelha de espécies mais abrangente e representativa (Nic Lughadha et al. 2019a; Zizka et al. 2021).

Melhorias contínuas nas tecnologias e ferramentas de computação voltadas à biodiversidade propulsionaram o aumento exponencial no número de registros de ocorrência de espécies disponibilizados em várias bases de dados online e de livre acesso (p. ex. GBIF, SpeciesLink, SiBBr; Soberón & Peterson 2004; Gadelha et al. 2021). Tais dados constituem evidências verificáveis e

citáveis da ocorrência de uma espécie e um recurso vital para avaliação do risco de extinção. Registros de ocorrência de espécies constituem o cerne de métodos e abordagens utilizadas para agilizar e reduzir os custos do processo de avaliação de risco de extinção (Raimondo et al. 2013).

Avaliações rápidas e geradas automaticamente utilizam de dados sobre a distribuição geográfica (critério B; para mais detalhes veja o Cap. 1 desta tese) e outros atributos que conferem maior vulnerabilidade à extinção (p. ex., tamanho populacional, informação taxonômica, nicho climático, perda de habitat) para gerar uma avaliação preliminar de risco de extinção (Darrah et al. 2017; Nic Lughadha et al. 2019a; Zizka et al. 2021). A avaliação é considerada preliminar pois outros critérios são necessários para realização de uma avaliação completa segundo o sistema da UICN (Nic Lughadha et al. 2019a).

Métodos de avaliação rápida têm mostrado alta acurácia (em torno de 90%) em classificar corretamente espécies como ameaçadas e não ameaçada, e podem ser particularmente úteis para informar espécies ou áreas necessitando de esforços de conservação (Nic Lughadha et al. 2019a; Zizka et al. 2021). Avaliações automatizadas agem como uma abordagem de triagem e podem facilitar a priozação de espécies para realização de avaliações completas (Raimondo et al. 2013; Bachman et al. 2019), além de fornecer a melhor estimativa do risco de extinção até que outras análises sejam realizadas.

Dados confiáveis e de alta qualidade são necessários para estimar com precisão a distribuição geográfica, um componente chave do processo de avaliação de risco de extinção (Rivers et al. 2011). Erros nas três dimensões básicas de registros de ocorrência de espécies (taxonômica, geográfica e temporal) podem levar a estimativas equivocadas do tamanho da área de distribuição de uma espécie e, por consequência, a classificações errôneas do risco de extinção (Nic Lughadha et al. 2019b; Panter et al. 2020).

Por exemplo, registros mal georreferenciados com coordenadas atribuídas ao centroide do Brasil podem inflar estimativas da área de distribuição. Como resultado, uma espécie ameaçada poderia permanecer desprovida de esforços de conservação por ter sido erroneamente avaliada como não ameaçada. Dessa forma, a realização de uma minuciosa inspeção da qualidade dos dados antes do uso torna-se fundamental. Este processo, também conhecido como "limpeza de dados" envolve a identificação, correção ou remoção de erros e inconsistências visando identificar qual fração dos registros que contém qualidade adequada

para um determinado fim (Veiga et al. 2017). Dados de baixa qualidade podem levar a resultados e conclusões equivocadas e inefetivas as quais, eventualmente, podem levar à perda de valiosos recursos para conservação da biodiversidade (Chapman 2005; Maldonado et al. 2015).

Como signatário da Estratégia Global para Conservação de Plantas (GSPC, CBD 2016), o Brasil se comprometeu a avaliações do risco de extinção de todas as plantas conhecidas (meta 2) além de garantir a proteção in *situ* 75% das espécies de plantas ameaçadas (meta 7). Entre as várias estratégias de conservação utilizadas para evitar a extinção de espécies, proteger as espécies em seus habitats naturais é a mais simples, barata e eficaz (Loucks et al. 2008).

Nesta tese, analisamos e sintetizamos vários aspectos relacionados à avaliação de risco de extinção, tais como uma maior compreensão do sistema utilizado nas avaliações, a importância de dados de alta qualidade para realização de avaliações de risco, e os desafios e oportunidades do uso de abordagens rápidas e automatizadas para predição do risco de extinção de espécies. Os resultados e conclusões apresentados também são um indicativo do nível de cumprimento do Brasil com as metas estabelecidas pela GPSC, e fornecem uma visão mais ampla e atualizada do estado de conservação das plantas brasileiras. Esta tese é composta de cinco capítulos:

No primeiro capítulo, apresentamos uma visão geral sobre o processo de avaliação de risco de extinção, bem como uma discussão sobre seus pontos fortes, equívocos e questões-chave, incluindo vieses e lacunas taxonômicas, escalas de implementação, incerteza de dados, o número de avaliações desatualizadas e o tempo e custos para realizar avaliações. Por fim, destacamos as oportunidades para ampliar e agilizar o processo de avaliação de risco de modo a fornecer uma visão mais completa do estado de conservação da biodiversidade.

No seguinte capítulo, introduzimos uma caixa de ferramentas chamada "*biodiversity data cleaning (bdc)*" (limpeza de dados de biodiversidade, em português), elaborada ao longo da tese e disponizada e desenvolvida no ambiente R de programação. O pacote contém uma série de ferramentas para facitar a integração e análise da qualidade dos dados de ocorrência de espécies, de modo a transformar dados brutos e potencialmente errôneos em informação de alta qualidade. Exemplificamos sua funcionalidade agrupando e analisando a qualidade de mais de 30 milhões de registros de ocorrência de espécies, e mostramos como dados errôneos podem alterar o padrão de riqueza de espécies de plantas no Brasil.

No terceiro capítulo, avaliamos como a incompletude de dados e erros presentes nas dimensões taxonômica, geográfica e temporal de registros de ocorrência de espécies afetam a acurácia de avaliações rápidas de risco de extinção. Mostramos que dados de alta qualidade são fundamentais para melhorar a sensibilidade da avaliação rápida em identificar espécies ameaçadas. Nossos resultados também indicam que avaliações rápidas possuem alta acurácia em identificar espécies não ameaçadas, mesmo na presença de erros taxonômicos, geográficos ou temporais.

No quarto capítulo, investigamos a eficácia das unidades de conservação e terras indígenas em representar espécies de plantas ameaçadas de extinção. Mostramos que entre 10 e 33% das espécies não estão presentes em áreas protegidas.

Acreditamos que conservação da biodiversidade é de responsabilidade de todas as pessoas. Nesse sentido, apresentamos no quinto capítulo e no apêndice I textos voltados às crianças que descrevem de forma simples e detalhada o processo de avaliação de risco e a situação atual de ameaça sofrida pelas espécies no Brasil.

Referências

Bachman SP, Field R, Reader T, Raimondo D, Donaldson J, Schatz GE, Lughadha EN. 2019. Progress, challenges and opportunities for Red Listing. Biological Conservation 234:45–55. Elsevier. Available from https://doi.org/10.1016/j.biocon.2019.03.002.

Bennun L et al. 2018. The Value of the IUCN Red List for Business Decision-Making. Conservation Letters 11:e12353. Available from http://doi.wiley.com/10.1111/conl.12353.

Betts J, Young RP, Hilton Taylor C, Hoffmann M, Rodríguez JP, Stuart SN, Milner Gulland EJ. 2020. A framework for evaluating the impact of the IUCN Red List of threatened species. Conservation Biology 0:cobi.13454. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13454.

BFG (The Brazil Flora Group). 2021. Flora do Brasil 2020. Page Jardim Botânico do Rio de Janeiro. Rio de Janeiro. Available from http://floradobrasil.jbrj.gov.br/reflora/listaBrasil/ConsultaPublicaUC/ResultadoDaConsultaNovaConsulta.do#CondicaoTaxonCP.

Butchart SHM et al. 2010. Global biodiversity: indicators of recent declines. Science 328:1164–8. Available from http://www.sciencemag.org/cgi/doi/10.1126/science.1187512.

CBD. 2016. Progress in Implementation of the Targets of the Global Strategy for Plant Conservation:23. Available from https://www.cbd.int/doc/meetings/sbi/sbi-01/information/sbi-01-inf-32-en.pdf.

Chapman AD. 2005. Principles of Data Quality. Global Biodiversity:58. Available from http://www2.gbif.org/DataQuality.pdf.

Corlett RT. 2016. Plant diversity in a changing world: Status, trends, and conservation needs. Plant Diversity 38:10–16. Available from http://linkinghub.elsevier.com/retrieve/pii/S2468265916300300.

Darrah SE, Bland LM, Bachman SP, Clubbe CP, Trias-Blasi A. 2017. Using coarse-scale species distribution data to predict extinction risk in plants. Diversity and Distributions 23:435–447.

Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. Available from http://floradobrasil.jbrj.gov.br/ (accessed August 7, 2021).

Gadelha LMR et al. 2021. A survey of biodiversity informatics: Concepts, practices, and challenges. WIREs Data Mining and Knowledge Discovery 11:1–41. Available from https://onlinelibrary.wiley.com/doi/10.1002/widm.1394.

Heywood VH. 2017. Plant conservation in the Anthropocene – Challenges and future prospects. Plant Diversity 39:314–330. Kunming Institute of Botany, Chinese Academy of Sciences. Available from http://linkinghub.elsevier.com/retrieve/pii/S2468265917300847.

IPBES. 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Page (S. Díaz, J. Settele, E. S. Brondízio E.S., H. T. Ngo, M. Guèze, J. Agard, A. Arneth, P. Balvanera, K. A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu, S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A and CNZ, editor). IPBES secretariat, Bonn, Germany. Available from https://ipbes.net/system/tdf/ipbes_global_assessment_report_summary_for_policymakers.pdf?file=1&type=node&id=35329.

IUCN. 2019. Guidelines for Using the IUCN Red List Categories and Criteria. Version 14. Prepared by the Standards and Petitions Committee. Available from http://intranet.iucn.org/webfiles/doc/SSC/RedList/RedListGuidelines.pdf.

IUCN. 2021. Table 1a: Number of species evaluated in relation to the overall number of described species, and numbers of threatened species by major groups of organisms. Available from www.iucnredlist.org/resources/summary-statistics (accessed August 1, 2021).

Loucks C, Ricketts TH, Naidoo R, Lamoreux J, Hoekstra J. 2008. Explaining the global pattern of protected area coverage: relative importance of vertebrate biodiversity, human activities and agricultural suitability. Journal of Biogeography 35:1337–1348. Available from http://doi.wiley.com/10.1111/j.1365-2699.2008.01899.x.

Mace GM, Collar NJ, Gaston KJ, Hilton-Taylor C, Akçakaya HR, Leader-Williams N, Milner-Gulland EJ, Stuart SN. 2008. Quantification of extinction risk: IUCN's system for classifying threatened species. Conservation biology : the journal of the Society for Conservation Biology 22:1424–42. Available from http://doi.wiley.com/10.1111/j.1523-1739.2008.01044.x.

Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Albán J, Chilquillo E, Rønsted N, Antonelli A. 2015. Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? Global Ecology and Biogeography 24:973–984.

Martinelli G, Moraes MÁ. 2013. Livro Vermelho da Flora do Brasil1 st. Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro.

Martins, E., Loyola, R., Martinelli G. 2017. Challenges and perspectives for achieving the Global Strategy for Plant Conservation targets in Brazil. Annals of the Missouri Botanical Garden in press:1–27.

MMA. 2014. Portaria no. 443, de 17 de dezembro de 2014.

Nic Lughadha EM et al. 2019a. The use and misuse of herbarium specimens in evaluating plant extinction risks. Philosophical Transactions of the Royal Society B: Biological Sciences 374:20170402. Available from http://www.royalsocietypublishing.org/doi/10.1098/rstb.2017.0402.

Nic Lughadha EM, Graziele Staggemeier V, Vasconcelos TNC, Walker BE, Canteiro C, Lucas EJ. 2019b. Harnessing the potential of integrated systematics for conservation of taxonomically complex, megadiverse plant groups. Conservation Biology 33:511–522. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13289.

Panter CT, Clegg RL, Moat J, Bachman SP, Klitgård BB, White RL. 2020. To clean or not to clean: Cleaning open source data improves extinction risk assessments for threatened plant species. Conservation Science and Practice 2. Available from https://onlinelibrary.wiley.com/doi/10.1111/csp2.311.

Raimondo DC, von Staden L, Donaldson JS. 2013. Lessons from the Conservation Assessment of the South African Megaflora. Annals of the Missouri Botanical Garden 99:221–230.

Rivers MC, Taylor L, Brummitt NA, Meagher TR, Roberts DL, Lughadha EN. 2011. How many herbarium specimens are needed to detect threatened species? Biological Conservation 144:2541–2547. Elsevier Ltd. Available from http://dx.doi.org/10.1016/j.biocon.2011.07.014.

Rodrigues ASL, Pilgrim JD, Lamoreux JF, Hoffmann M, Brooks TM. 2006. The value of the IUCN Red List for conservation. Trends in ecology & evolution 21:71–6. Available from http://www.ncbi.nlm.nih.gov/pubmed/16701477 (accessed March 3, 2012).

Rondinini C, Di Marco M, Visconti P, Butchart SHM, Boitani L. 2014. Update or Outdate: Long-Term Viability of the IUCN Red List. Conservation Letters 7:126–130.

Soberón J, Peterson AT. 2004. Biodiversity informatics: managing and applying primary biodiversity data. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 359:689–98. Available from http://www.ncbi.nlm.nih.gov/pubmed/15253354.

Sousa-Baena MS, Garcia LC, Peterson AT. 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Diversity and Distributions 20:369–381. Available from http://doi.wiley.com/10.1111/ddi.12136.

Stuart SN, Wilson EO, McNeely JA, Mittermeier RA, Rodriguez JP. 2010. The Barometer of Life. Science 328:177–177. Available from https://www.sciencemag.org/lookup/doi/10.1126/science.1188606.

Veiga AK, Saraiva AM, Chapman AD, Morris PJ, Gendreau C, Schigel D, Robertson TJ. 2017. A conceptual framework for quality assessment and management of biodiversity data. PLOS ONE 12:e0178731. Available from https://dx.plos.org/10.1371/journal.pone.0178731.

Vié J, Hilton-taylor C, Pollock C, Ragle J, Smart J, Stuart S, Tong R. 2008. THE IUCN Red List : A key conservation tools. Page in Hilton-Taylor and S.N. Stuart, editor. The 2008 Review of The IUCN Red List of Threatened Species. IUCN Gland, Switzerland.

Zizka A, Silvestro D, Vitt P, Knight TM. 2021. Automated conservation assessment of the orchid family with deep learning. Conservation Biology 35:897–908. Available from https://onlinelibrary.wiley.com/doi/10.1111/cobi.13616.

CAPÍTULO I

A panorama on the past, present, and future of extinction risk assessments.

Ribeiro, BR; Fiuza, M; Verdi, M; Fernandez, EP, Loyola, R.
A panorama on the past, present, and future of extinction risk assessments.

A panorama on the past, present, and future of extinction risk assessments

Bruno R. Ribeiro[1,2], Milena Fiuza[1,2], Marcio Verdi[3], Eduardo P. Fernandez[3], Rafael Loyola[1,4]

Author affiliations:

[1] Conservation Biogeography Lab, Department of Ecology, Universidade Federal de Goiás, Goiânia, Goiás, Brazil.

[2] Graduate Program in Ecology & Evolution, Universidade Federal de Goiás, Goiânia, Goiás, Brazil.

[3] Brazilian National Centre for Flora Conservation, Rio de Janeiro Botanical Garden, Rio de Janeiro, Brazil

[4] Brazilian Foundation for Sustainable Development, Rio de Janeiro, Brazil

The multiple paths towards extinction

Nearly 99% of species that have evolved on Earth over the 3.5 billion years are currently extinct (Novacek 2001). Although extinction is a natural process, the growing human population and its increasing demands for natural resources have accelerated extinction to 100 to 1000 times higher than the so-called background rate observed throughout geological history (Pimm et al. 2014). As a result of human pressure, several species have been pushed toward extinction, and our planet may already be experiencing the effects of an ongoing sixth mass extinction event (Ceballos et al. 2015).

The extinction process is simple to define – it occurs when the last individual of the last remaining population of a species dies – but it is hard to predict. Because extinction is often determined by the interplay of varying interacting species' traits and threats, multiple possible paths can lead species to disappear (but see, e.g.,. Chichorro et al., 2019; Davidson et al., 2009). To counter the current human-induced extinction crisis with sparse conservation funds, it is necessary to determine which, where, and why some species are more vulnerable to extinction than others, a pattern known as extinction selectivity or relative vulnerability (McKinney 1997; Humphreys et al. 2019; Chichorro et al. 2019).

Predicting species vulnerability to risk is only possible because extinction is a non-random process since certain biological traits predispose species to an elevated risk of extinction (Purvis et al. 2000; Mace et al. 2008; Collen et al. 2011). The list of extinction-promoting traits is extensive (e.g., small geographic range, low population density, slow life history, and large body size, to cite some (Chichorro et al., 2019; Collen et al., 2011; Mckinney, 1997). The importance of each trait on extinction risk varies among different taxonomic groups and biogeographic regions (Mckinney 1997; Cardillo et al. 2008). For example, in mammals, species with larger body sizes are at higher risk, but only in the tropics (Fritz et al., 2009 but see Davidson et al., 2017). Moreover, the influence of extinction-correlated traits varies depending on the threats species face (Isaac & Cowlishaw 2004).

Due to the idiosyncrasies and complexity involved in forecasting extinction risk based only on biological trai-ts, alternative systems for assessing species risk have been proposed in the last decades (De Grammont & Cuarón 2006). Among these systems, the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (herea-fter Red List) has been increasingly recognized and used as a global standard for assessing species extinction risk (Rodrigues et al. 2006).

Despite its crucial role in determining risk, there are several misconceptions regarding the criteria and categories used to assess extinction risk, primarily among young scientists and conservation practitioners (Collen et al. 2016). A well-trained group of people is crucial to avoid mis-classifying species risk and preclude the misapplication of scarce conservation resources.

Here, we provide a road map highlighting several facets of the Red List of threatened species aimed primarily at young scientists, conservation practitioners, or those with little experience in the red listing process. We present a short historical overview of risk classification systems and describe the characteristics of an ideal system before going deeper into details on the method most used for assessing extinction risk, i.e., the Red List. We also present an overview of the Red List process and discuss its strengths, mis-conceptions, and critical issues, including taxonomic biases and gaps, scales of implementation, data uncertainty, the number of outdated assessments, and the time and costs to carry out assessments. Finally, we focus on the oppor-tunities for broadening the Red List coverage and making it a more comprehensive "Barometer of Life" (Stuart et al. 2010) to provide a better picture of the state and distribu-tion of biodiversity worldwide.

Characteristics of an ideal system for assessing extinction risk

Although the Red List is currently the most widely recognized method for assessing species extinction risk, that was not always the case. Many countries worldwide had created their official lists of threatened species based on different systems (Andelman et al. 2004; De Grammont & Cuarón 2006). Early systems varied from qualitative assessments lacking a clear definition of categories and based on subjective criteria to a more objective method based on rules, point-scoring, and quantitative analyses (i.e., population viability analysis) (Master 1991; Mace & Lande 1991; Andelman et al. 2004; De Grammont & Cuarón 2006; Mace et al. 2008).

Ideally, a system assessing species conservation status should be based on the quantitative and explicit likelihood of extinction under current prevailing conditions over a precise time frame (Mace et al., 2008). Most systems used to assess extinction risk share a typical structure consisting of categories and criteria. Categories are names used to indicate extinction risk levels and are more easily understandable by the general public and decision-makers than systems that place extinction risk along a continuum (Mace & Lande 1991; De Grammont & Cuarón 2006). Categories are defined based on a set of criteria (i.e., risk indicators or symptoms of risk) which, in turn, are based on information on the species such as population size, geographic range, habitat availability, intrinsic characteristics related to their vulnerability to extinction (body size, generation length), and anthropogenic drivers threatening them (De Grammont & Cuarón 2006; Mace et al. 2008; Master et al. 2012).

There is a gamut of desirable characteristics which a threatened species categorization system must have (Mace & Lande 1991; De Grammont & Cuarón 2006). Most of all, it should be essentially simple, objective, scientifically sound, and easy to understand. The system should have well-defined and explicit categories expressing different extinction risk levels and a clear relationship to one another. It should have explicit – preferably quantitative and not ambiguous – criteria with which species are assigned into risk categories. It must be flexible in terms of data requirements and explicitly consider evaluations of levels of uncertainty of criteria. Ultimately, the system should apply to a wide range of taxonomic levels and groups (species and subspecies; from ants to elephants) and geographic scales (global and regional) and be updated regularly (Mace & Lande 1991; De Grammont & Cuarón 2006; Mace et al. 2008).

De Grammont and Cuarón (2006) evaluated the appropriateness of 25 categorization systems used in 36 threatened species lists in 20 countries against an ideal method's set of desirable characteristics. Their results showed that 76% of systems assessed had categories, 60% and 64% included explicit definitions of categories and criteria, respectively, and 72% included subjective elements in the description of categories such as "population reduced to a critical level". Most serious limitations were found in country categorization systems. They pointed out that the IUCN's system was the most suitable for assessing the extinction risk of species, followed by the Nature-Serve protocol. They stress the need to use a standardized categorization system to make it possible to compare different assessments and, thus, to provide a big picture on the status and trends of biodiversity.

Perhaps the significant strengths of the Red List are related to the quantitative nature of the criteria and the relatively lower number of categories and criteria compared with other systems. Conservation assessments based on a system lacking quantitative criteria are more prone to capricious categorization due to economic or political interests, which may hinder biodiversity conservation and scientific credibility (De Grammont & Cuarón 2006). Moreover, quantitative criteria are necessary to avoid ambiguities and subjective intrepretations of the criteria by leading to different outputs. Despite the plethora of drivers that can lead species to extinction, there seems to be little virtue in assessing threat levels based on numerous categories and criteria (Mace & Lande 1991). The Red List's lower number of categories and criteria makes it more objective and easily understandable by different users, such as the general public, decision-makers, and conservation practitioners.

The IUCN Red List of Threatened Species

Categories and Criteria of the Red List

The Red List is the most comprehensive and authoritative source of information on the global extinction risk of species (Rodrigues et al. 2006). It is based on data-driven and objective criteria used to assign species to nine categories of extinction risk. The Red List process commonly starts with a species list following by choice of which species will be prioritized for assessment (species not assessed are classified as Not Evaluated). Except for microorganisms, all described taxa (species, subspecies, varieties) can be evaluated (IUCN, 2019).

In the pre-assessment process, relevant information on species' biology (e.g., occurrence, population status and trends, and taxonomic data) are compiled from several sources, including natural history collections (e.g., herbarium and zoological collections often available in online databases), field observations, monographs, floras, articles, reports, and books, or provided by species experts. A wealth of supporting information is obtained at this stage, including threats to species, use and trade, habitat and ecology, and conservation efforts (e.g., presence of species in protected areas or national actions plans). Once compiled, data are analyzed to generate metrics (indicators of risk) underlying the Red List assessment (Bachman et al. 2019). If the assessor considers that sufficient data are lacking to assess, species are categorized as "Data Deficient". Suppose data are available to apply the criteria. In that case, species can be listed to a threatened category (Critically Endangered, Endangered, or Vulnerable), a not threatened category (Near Threatened or Least Concern), or as extinct (Extinct or Extinct in the Wild).

The classification is based on a set of five quantitative criteria of decreasing probability of extinction (A through E) associated with population size, geographic range, and rate of decline of both (IUCN, 2012; Mace and Land, 1991). If any criterion is met, species are assigned to a

threatened category. If any criterion is met, species are assigned to a threatened category. Most criteria include sub-criteria, such as the extent of occurrence (EOO) and area of occupancy (AOO), number of locations, habitat decline, fragmentation, and population size and trends. In this sense, sub-criteria can be understood as the evidence used to justify listing a taxon under a particular category (IUCN, 2019; Box 1).

Criteria A builds on the rate of population reduction (past, present, and/or projected), B on a geographic range size (either in the form of extent of occurrence or area of occupancy) in combination with fragmentation, decline, or extreme fluctuations, C on small population size that are currently declining or may decline shortly, and D on minimal population size. Criterion D2, a very restricted distribution, only applies to the Vulnerable category and is based on the size of the area of occupancy or in the number of locations. Criterion E regards quantitative analysis indicating a probability of a species going extinct over a given period (Mace et al. 2008; IUCN 2012a).

To qualify for listing in any category, a species needs to meet quantitative thresholds for at least one criterion and its respective sub-criteria (Mace et al. 2008). When more than one criterion is completed, the criterion that determines the conservation status is the one that returns the highest risk estimation (Mace et al. 2008; IUCN 2019). Threats to species are not explicitly considered in the assessment; instead, they are inferred by the assessors using criteria related to fragmentation level (population and distribution) and the number of locations (i.e., an indication that a threat event can vanish an entire population). Such information is appended to the documentation on the assessment.

The Red List assessment is made by assessors, who are experts with sufficient knowledge on a taxon that uses data to assign species to extinction risk categories. The evaluation can be done remotely or undertaken as part of a workshop where various people (assessors, species experts, and facilitators) gather together to process multiple assessments (Bachman et al. 2019). Each assessment is then reviewed by an appropriated Red List Authority (RLA) responsible for ensuring that the assessment is as accurate as possible. Assessments considered adequate and containing all required supporting information is then published on the IUCN Red List.

Box 1. Summary of key terms related to IUCN Red List

Threatened species: Species qualifying under the categories Critically Endangered, Endangered, and vulnerable facing an extremely high, very high, and high risk of extinction in the wild.

Non-Threatened species: a species that does not meet the quantitative thresholds based on population size, geographic range, and trends of both, to be considered a threatened species.

Subpopulations: Spatially distinction groups of the population between which there is little demographic or genetic exchange.

Populations: A taxon's total number of individuals (including mature and other life stages) across its entire distribution.

Population size: The total number of mature individuals within a population.

Mature individuals: individuals known, estimated, or inferred to be able to produce offspring.

Generation length: A measure of change in the population size over time defined as the average age of parents of the current cohort.

Reduction: Percentage of decline in population size over ten years or three generations used in criterion A (population size reduction).

Continuing decline: a recent, current, or future decline in a population measured in terms of population size (criterion C), or range size, habitat quality and availability, locations, subpopulations, and the number of mature individuals (criterion B).

Extreme fluctuation: functions in population size (criterion C) and range area (criterion B) occurring widely, rapidly, and frequently.

Geographic distribution: The spatial distribution of individuals of a population measure as extent of occurrence (EOO; an area encompassing all sites where a taxon is supposed to occur) or area of occupancy (AOO; the area within EOO with suitable habitat occupied by a taxon).

Severely Fragmented: when > 50% of individuals of a population are in small and isolated subpopulations (i.e., patches). A metric used in criterion B.

Location: a distinct area in which a single threatening event (e.g., fire, vegetation clearing) may affect all individuals.

Red List: key information for conservation action

The overall goal of the Red List is "to provide information on the status, trends, and threats to species to inform and catalyze conservation actions for biodiversity conservation" (Vié et al. 2008; IUCN 2019). The goal of any conservation assessment does not rely on red listing species but on preventing them from moving closer to extinction, i.e., to improve their conservation status. The results chain of this assessment included red listing a species (input) leads to outputs (enhanced scientific knowledge, raised awareness of stakeholders), desirable outcomes (conservation priorities better identified and understood, access to funding and resource availability, improved legislation, and policy), and ultimately impact (better-targeted conservation implementation and action needed for improving species conservation status) (Betts et al. 2020). Hence, assessing the threat status of species results in a piece of crucial information used to catalyze a chain of events that, ultimately, may lead to better-informed conservation actions and resource mobilization (Betts et al. 2020).

The information derived from Red List assessments constitutes a respected source of information that influences conservation in various ways (Rodrigues et al. 2006; Betts et al. 2020). The Red List provides information on the status, threats as well as a large number of data collected to support the assessments (e.g., distribution maps, population size, ecological requirements), which have been increasingly used in large-scale studies on different research areas, e.g., to assess the effectiveness of the network of protected areas (Rodrigues et al. 2004; Ribeiro et al. 2018), to identify correlates of extinction risk (Cardillo et al. 2008; Davidson et al. 2009), to measure the state of biodiversity and the effectiveness of conservation actions (Butchart et al. 2010; Hoffmann et al. 2010), to set priorities for conservation (Brum et al. 2017), to assess the effects of climate change on biodiversity (Faleiro et al. 2013), to determine the invasive potential of alien species (Sales et al. 2017), and many more.

The impact of the Red List in generating scientific knowledge is evidenced by the positive trend in references to "red list" and "red data book" in peer-reviewed articles and gray literature (Hoffmann et al. 2008; Betts et al. 2020). Threat status of species is also specific information to raise public awareness and interest in conservation, for example, the use of information on threat status in educational materials of zoos and botanic gardens (Betts et al. 2020). Besides, Red List is a global index of the state of change of biodiversity – a Barometer of Life – (Stuart et al. 2010; Butchart et al. 2010), used for monitoring trends in conservation status and to inform global and regional biodiversity targets such as Convention on Biological Diversity Aichi Target 12 ("By 2020 the extinction of known threatened species has been prevented and their conservation status, particularly of those most in decline, has been improved and sustained"; (Butchart et al., 2010; www.cbd.int/sp/targets) and Global Strategy for Plant Conservation Target 2 ("An assessment of the conservation status of all known plant species"; Sharrock et al., 2014; Martins et al., 2018).

Although not intended to act as a normative tool, the Red list frequently supports the national legislation and protection (for example, species protection, environmental law, protected areas, habitat protection) (Martinelli & Moraes 2013; Betts et al. 2020) and constitutes an essential instrument to the allocation of funding for species recovery plans and conservation planning (but see the section "A multitude of (mis)uses"; Bennun et al., 2018; Betts et al., 2020; Rodrigues et al., 2006).

Global versus regional assessments

Despite being primarily designed for application at global scales when the entire distribution and population of species (or subspecies) are considered, the Red List Categories and Criteria can be applied to any sub-global scale (regional, national; hereafter “regional assessments”), although the application within tiny areas is discouraged by the IUCN (IUCN 2019). Regional red lists serve two main purposes. First, to guide effective conservation actions and species/habitat management that hardly occur at the global and continental scales (Margules & Pressey 2000; De Grammont & Cuarón 2006). Second, they are a source of crucial information on the status of biodiversity within a region (i.e., country or state) that can be used to measure progress toward targets set on international policy agreements (e.g., Aichi Target 12; and GSPC Target 2) and, along with other information, to inform legislation and national strategies for biodiversity conservation, such as species recovery plans, restoration programs, monitoring centers, and environmental impact assessments (Gärdenfors 2001; Rodrigues et al. 2006; Azam et al. 2016). Currently, many countries have produced and published red lists based on the IUCN categories and criteria at the national (e.g., Martinelli and Moraes, 2013; Raimondo, 2011; see also National Red Lists available at www.nationalredlist.org) and regional level (Loyola et al. 2018).

Usually, the conservation status of a species assessed at the global level does not necessarily coincide with its regional status (Gärdenfors 2001; De Grammont & Cuarón 2006; Brito et al. 2010). Regional and global assessment should have equal conservation status for species restricted to one region (e.g., endemic). In such cases, regional assessments can be incorporated into the Red List (Rodríguez 2008). However, for non-endemics, it is expected that the threat category of a species at the global and regional levels differ, and such mismatches can result from genuine or artifactual differences (Vié et al. 2008; Brito et al. 2010). Genuine mismatches are expected because, in regional assessments, the criteria used apply to parts of species distribution/population arbitrarily delimited by geopolitical boundaries (Mace et al. 2008). In such cases, species may be threatened within a country because their population is minor and/or declining and not threatened at global levels where their populations are widespread and stable, and vice versa (Gärdenfors et al. 2001; Brito et al. 2010). Such mismatches, however, could also result from artifactual differences between regional and global red lists.

Artifactual mismatches may arise when different information regarding species taxonomy and distribution is used in regional and global assessments (Brito et al. 2010). One would expect more species to be threatened regionally than globally when assessments are carried out using the same species and with no errors (Brito et al. 2010). However, a comparison of classification outcomes of species listed at both scales highlighted that 14% of species are considered threatened only in the Red List (Brito et al. 2010). Such a difference probably results from artifactual rather than genuine mismatches, which in turn could undermine the credibility of red listing, make the comparison between regional and global status difficult, and ultimately, lead to extinction in cases where risk is underestimated (Brito et al. 2010).

Regional assessments are produced in a three-step process (Gärdenfors et al. 2001; IUCN 2012b). The first step consists of the selection of species and regional populations to be assessed. Next, a preliminary category is assigned by comparing the population within a region against the same criteria of the Red List. The preliminary category is then adjusted if appropriate (step three) by considering the likely "rescue effect" that the conspecific population in neighboring regions could exert on the likelihood of extinction of the population assessed (Gärdenfors et al. 2001). As a result, the preliminary category can be up- (e.g., EN to CR) or down (e.g., VU to NT) listed or remain the same depending on whether populations outside the region are judged to increase or decrease extinction risk of the population assessed (Gärdenfors et al. 2001).

The adjustment process is helpful to avoid over-or underestimating the risk, given the actual population size is generally greater than that defined by geopolitical borders (Gärdenfors et al. 2001). The possible modification of the

preliminary category is based on seven questions regarding whether the neighboring population can affect the extinction risk of the regional population (IUCN 2012b). This adjustment stage has been subject to criticism regarding its knowledge requirements, leading to arbitrary choices that can produce different conservation statuses (Eaton et al. 2005). Such arbitrary decisions, in turn, may also hinder the comparison of assessments made by other countries or between regional and global assessments (Eaton et al. 2005).

There are several benefits of regional red lists. Regional assessments provide valuable information on non--endemic species beneficial to the global Red List, further improving the consistency between regional and global assessments (Gärdenfors 2001; Rodríguez 2008). Moreover, regional assessment has the potential of increasing the comprehensiveness of the Red List as assessments of endemic species are equivalent to global assessments and can have their status information translated directly to the Red List (Rodríguez 2008). For example, nearly 60% of all plant species are endemic to a specific region, which highlights the great potential of regional assessments to be significant contributors to the Red List (Zamin et al. 2010; Bachman et al. 2018), especially given that now regional assessments can be submitted to the Red List in French, Spanish, Portuguese and English (Bachman et al. 2018).

Ultimately, regional assessment can catalyze conservation efforts at an appropriate scale where species are managed and organizations such as governments and NGOs act.

Dealing with uncertainties

The Red List system defines species conservation status by applying criteria based on data with inherent uncertainty associated. If not properly treated and acknowledged, such uncertainty could lead to undesirable outcomes and poor decision-making (Keith et al. 2004; Regan et al. 2005). Uncertainty can arise from one of these three factors: semantic (i.e., linguistic) uncertainty, natural variability, and measurement error (Akçakaya et al. 2000; Regan et al. 2002).

Semantic uncertainty arises from vagueness in the definitions of the criteria used in the system, which can result in inconsistencies of assessments performed by different assessors (Akçakaya et al., 2000; Keith et al., 2004; Regan et al., 2005). Indeed, inconsistencies of assessments carried out by different evaluators are expected (Keith et al., 2004). Some terms used to describe the criteria are intentionally left inexact to avoid some loss of generality (Akçakaya et al. 2000). For example, the interpretation of what consists of a geographical distribution "severally fragmented" (one sub-criterion on the criteria B) could vary markedly depending on the species' life-history characteristics (e.g., mobility) and could mean different things to different assessors (i.e., context-dependency). Such variation makes any attempt to make the definition of the criteria more exact almost impossible. Although some aspects of semantic uncertainty are irreducible, they should not be ignored (Akçakaya et al. 2000). Incorporating uncertainty in parameter estimates, training in applying the criteria, guidance on the use of the criteria for specific groups, and the use of consensus among multiple assessors are some possible solutions to deal with semantic uncertainty (Keith et al. 2004; IUCN 2019).

Natural variability occurs when an accurate value of a parameter (i.e., the number of adults in a population) varies in response to changes in independent variables (Regan et al. 2002). For example, population and distribution size are dynamic proprieties and change naturally in space and time due to multiple demographics, random, or contingent factors, making it extraordinarily difficult to measure or obtain precise information about them (Burgman et al. 1999). This type of uncertainty is attributable to the system's randomness and is not reduced by increased sampling effort (Akçakaya et al. 2000). For this reason, uncertainty in estimates of parameters can be better expressed in probabilistic terms, and these variations in parameter estimates can be compared with the criteria thresholds. The effects of natural variability on the variation of the criteria are expected to be low, given the values of parameters refer to a specific point in time or spatial scale (Akçakaya et al. 2000; IUCN 2019).

Measurement errors are the largest source of variability in conservation assessments. It arises from the lack of precise information on the parameters used in the criteria (e.g., number of mature individuals; population size) due to either inaccuracy in estimating the values or data availability and quality. For example, it is almost impossible to obtain the correct number of adult individuals of a population because we simply cannot count them all (Akçakaya et al. 2000). Alternatively, we can use the best estimates of the actual number of adult individuals and a range of plausible values defined based on intervals (e.g., probability or confidence) or expert opinion (Akçakaya et al. 2000). Different from semantic uncertainty, measurement errors can be reduced by acquiring additional data to estimate a range of plausible values for a parameter (Akçakaya et al. 2000)

When data necessary for carrying out conservation assessment contains low levels of uncertainty, the conservation status of species is supposed to remain the same. This occurs because the criteria used to define threat categories are flexible enough to handle uncertainty (Akçakaya et al. 2000). Therefore, precise figures on population size, reduction, and distribution are not required to undertake reliable assessments (IUCN 2019). For example, a species will qualify as Endangered under criterion C (i.e., small population size and decline) even if the number of mature individuals varies from 2449 to 250 individuals. Another feature of the Red List that allows classification even when data are highly uncertain includes projection and inference and the fact that only one out of the five criteria must be met (IUCN 2019).

On the other hand, highly uncertain data may lead to inconsistent classifications, making it even possible to categorize a species, for instance, as Endangered or Least Concern depending on the attitudes toward risk and uncertainty adopted by the assessors (IUCN 2019). To increase the consistency of assessments made by different assessors, at first, it is important to consider the full range of plausible values in assessments, except when extreme values are proven to be unrealistic (e.g., derived from biased expert opinions). In such a case, extreme values should be excluded from consideration (i.e., dispute tolerance; IUCN, 2019). Moreover, assessors should decide based on their values whether to adopt a precautionary or an evidentiary attitude to risk. The former (i.e., a risk-avert perspective) will classify a species as threatened unless it is unlikely that it is not threatened. An evidentiary attitude (i.e., risk-prone attitude), in contrast, would demand substantial evidence to support the classification of a species as threatened (Akçakaya et al. 2000; IUCN 2019).

The IUCN recommends that assessors adopt a "precautionary but realistic attitude" when applying the criteria. According to the precautionary approach and considering that a single category must be chosen, when uncertainty makes classifying species in more than one category possible, species should be assigned to a higher risk category. The other plausible categories should be explicitly documented as long as the assessor adopts the risk attitude (IUCN 2019). Finally, in cases where data is so uncertain, making all range of categories plausible, the category "Data Deficient" should be assigned (IUCN 2001), except when likely categories range from "Endangered" to "Data Deficient". In such a case, a more plausible category could be "Vulnerable" (see details in IUCN, 2019).

A multitude of (mis)uses

The main goals of the Red List are to assess the relative extinction risk of species and provide a global index on the state of degeneration of biodiversity; any other usage of the

Red List beyond its original intent could be improper (Possingham et al. 2002; Vié et al. 2008; Mace et al. 2008). Despite being a respected source of information that influences many aspects of conservation, the Red List – alone – may perform poorly to set priorities for conservation action, for developing and implementing recovery plans, to constrain development and exploitation, and to report on the state of the environment (Possingham et al. 2002).

Setting priorities for conservation action and assessing extinction risk is perhaps the most common misconception of the Red List, and various early categorization systems have confounded these processes (Master 1991; Eaton et al. 2005). To set conservation priorities, information on the extinction risk of species must be considered along with other biological, logistical, and socioeconomic variables such as costs, land use, habitats, and species distribution, ecosystems services, species' evolutionary and functional distinctiveness, existing legal frameworks for species protection (e.g., protected areas), and so forth (Box 2; Mace et al., 2007; Margules and Pressey, 2000). This is why Possingham et al. (2002) state that it is naïve and counterproductive to use the Red List as a single piece of information to guide resource allocation for recovery plans or constrain development projects. The Red List in isolation can perform poorly to guide the allocation of limited resources because often most funding is spending on few at imminent risk of extinction and with a generally low likelihood of success (Possingham et al. 2002; Bottrill et al. 2008). Efficient resource allocation can be improved by considering benefits (species persistence), the likelihood of success, and management costs (i.e., a triage approach; Bottrill et al., 2008). Accordingly, more species could be managed with a higher probability of success and relatively low recovery cost (McCarthy et al. 2008; Bottrill et al. 2008).

The Red List is often used as a normative tool by regulatory agencies to constrain development actions (Possingham et al. 2002). Again, the Red List in isolation could result in poor and inadequate conservation decisions. For example, the presence of threatened species may impede the impact development project. In contrast, a high-impact one to be implemented at a high species-rich site but with no threatened species could be approved with no mitigation requirements. A more effective decision could be achieved by considering additional criteria besides a list of threatened species such as other biodiversity and sociopolitical information (Possingham et al. 2002).

Box 2. Confusing concepts: assessing threat status vs setting conservation priorities

Setting conservation priorities and assessing extinction risk are related but different processes. Information on threat status simply indicates how quickly conservation action need to be taken to avoid species extinctions and, on its own, is not enough to set priorities for conservation action. To efficiently set priorities for conservation action, information on threat status must be considered along with other socioeconomic and biological factors. Below is a simplistic and hypothetical scenario in which five sites, each one holding one species, should be prioritized to receive conservation resources based on the information of conservation status alone or in combination with other variables.

I) It is not possible to distinguish which site (1, 2, or 3) is priority for conservation by using only information on threat status. Information on threat status in isolation could lead to poor conservation decision.

II) Note the changes in priority sites when the proportion of species geographic range is considered in the prioritization process (protection level?).

III) A cost-efficient prioritization could be achieved by considering management costs in the prioritization process. In this scenario, a species listed as Vulnerable ranks higher in priority than a Critically Endangered species.

Threat status				Protection level		Managment cost	
Site	Species	Risk	Rank	Protection	Rank	Cost	Rank
1	Sp1	CR (1)	1	1 (1x1=1)	1	100 (1x100=100)	1
2	Sp2	CR (1)	1	5 (1x5=5)	3	80 (5x80=400)	4
3	Sp3	CR (1)	1	20 (1x20=20)	4	20 (20x20=400)	4
4	Sp4	EN (2)	2	10 (2x10=20)	4	15 (10x15=300)	3
5	Sp5	VU (3)	3	1 (3x1=3)	2	50 (3x50=150)	2

Constraints for broadening the red list coverage

Despite the number of species assessed for the Red List keep increasing every year, several constraints influence the growth of the Red List and its ability to act as an up-to--date and comprehensive indicator of the world's biodiversity health (Stuart et al., 2010). These constraints include taxonomic gaps and biases, the time and costs needed to carry out assessments, and the number of outdated assessments. In this section, we present these issues and discuss the opportunities for broadening the coverage of the Red List assessments.

Gaps and biases in Red List assessments coverage

Despite providing the most complete picture of the extinction risk faced by species, the Red List still presents significant gaps. Currently, 112,432 species have been assessed for the Red List, representing only about 5% of the world's estimated ~2.1 million described species (Table 1; IUCN, 2020). Although the documented assessments constitute a small and biased sample of the total known species (Stuart et al. 2010), comprehensive Red List assessments have been achieved for several whole major taxonomic groups of vertebrates – birds (100% species evaluated), mammals (90%), amphibians (84%), reptiles (71%), and fishes (54%) – some groups of invertebrates – Horseshoe Crabs (100%), corals (40%) –, and also gymnosperms (91%) (IUCN 2021).

On the other hand, the vast majority of the less well-known and species-rich groups of invertebrates, fungi, and plant species are still under-assessed, despite recent efforts to increase their coverage (Bachman et al. 2019). In the face of the growing number of species that have been pushed towards extinction and the high number of threatened species (30,178 species equivalent to 27%; (IUCN 2021), to make the Red List more taxonomically representative is crucial to inform better conservation actions and policy decisions (Stuart et al. 2010; Bennun et al. 2018).

For groups such as plants, efforts in several countries (including megadiverse ones such as Brazil) have been mainly directed towards cataloging and organizing knowledge about their flora through the production of lists of all known plant species (GSPC Target 1 for 2010 and elaboration of an online monographed flora (GSPC Target 1 for 2020). Consolidating information about the flora's taxonomy and distribution is an important step towards more complete Red List assessments for plants. While taxonomic information of major vertebrate groups is relatively well-established and consolidated, it has been the opposite for some groups of plants (e.g., angiosperms, ferns).

We have experienced frequent changes in the classification and taxonomic treatment of plant species in the last decade with the advent and incorporation of molecular biology in classical plant taxonomy studies and due to the increase in the production of monographs. While these changes decrease taxonomic uncertainties, they also result in many species with outdated Red List assessments (see Martins et al. 2018). The impact of taxonomic revisions on Red List assessments will decrease as the botanical knowledge increases, and taxonomy becomes more consolidated and stabilized (Nic Lughadha et al. 2019b).

Financial costs

Although the Red List is currently sufficiently and sustainably financed, its long-term stability and expansion could be compromised if core operating costs are not met (Bachman et al. 2019). Assessing the extinction risk involves a pre-assessment stage in which all relevant data from various sources are analyzed to generate metrics that allow the application of the Red List criteria. This process is often undertaken as part of workshops where assessors, experts, and facilitators are gathered to process multiple assessments (Bachman et al. 2019). The cost of assessments in a workshop is US$333 per species on average (Rondinini et al. 2014).

Between 2011 and 2019, 7,817 extinction risk assessments were conducted for 7,520 Brazilian native plant species assessed and reassessed by the Brazilian Red List Authority (CNCFlora) at an average cost of USD 140,40 per species. To gain scale in the face of the challenge to evaluate the extinction risk of more than 36,000 species ("Flora do Brasil 2020" 2016), CNCFlora opted for investments in their extinction risk assessment system. The system has been identified as a crucial platform to document the whole process and enhance the engagement of remote collaborators, which are paramount to validate occurrence data and overall information associated with species' profiles. Remote assessments or desktop assessments have proven to be a cheaper and effective approach compared to assessments carried out in workshops (Rondinini et al. 2014).

Another factor limiting the expansion of the Red List is the cost to keep it up to date. After ten years, Red List assessments are declared outdated by IUCN as information underpinning assessments becomes less reliable over time (in some countries, this time is set to five years). The temporal expansion of the Red List, i.e., repeat assessment, is crucial for detecting trends in the overall extinction risk of species (i.e.,. the Red List index). The IUCN established an ambitious goal for broadening the taxonomic coverage of the Red List by assessing the conservation status of 160,000 species by 2020, with an estimated cost on the order of US$60 million (Stuart et al. 2010). However, such a cost could be underestimated if outdated assessments are updated (Rondinini et al. 2014). Assuming a mean cost of US$333 per species assessed in workshops, more than 9 million dollars will be needed to reassess the conservation status of 27,744 species currently outdated (25% of the total species assessed so far), almost twice the cost spent per year for growing and maintaining the Red List (US$4.7 million; Juffe-Bignoli et al., 2016).

Opportunities for broadening the Red List coverage

The Red List's taxonomic and geographic biases and gaps and the budget needed for expanding and keeping it up to date are primary limiting factors to make the Red List more complete. However, there are several opportunities based on new recent developments that could assist and expedite the growth of the Red List to make it more a comprehensive Barometer of Life (Stuart et al. 2010; Bachman et al. 2019). Some of these developments are summarized in Table 2 and briefly discussed in this section. A deeper discussion on opportunities for broadening the red list coverage can be found in (Bachman et al. 2019)

There are feasible alternatives that could both expedite and reduce the costs of the assessment process. South Africa, for example, assessed the conservation status of > 20,000 plant species with a total cost of US$593,291 (US$29 per species on average) within five years (Raimondo et al. 2013). Critical factors that allowed the assessment of such a high number of species within five years include, among other factors, the use of approaches for streamlining the assessment process (Raimondo et al. 2013).

South African flora assessment used automated tools to efficiently categorize non-threatened species (i.e., Least Concern; n = 9,387) and prioritize likely threatened ones for full Red List assessment (n = 6000). Rapid, automatically generated assessments produce a preliminary conservation assessment based primarily on the geographic range (criterion B), and digitally available species distribution data have proven to be highly accurate in predicting non-threatened and threatened plant species (Nic Lughadha et al. 2019a; Zizka et al. 2021). The increasing importance of tools to speed up the production of assessments (e.g., Bachman et al., 2011; Brummitt et al., 2008; Dauby et al., 2017) and documentation (Bachman et al. 2020) comes along with the recently reduced data requirements for publishing non-threatened (Least Concern) species on the Red List (Bachman et al. 2019). To facilitate this process, a new tool can assist and scale up the generation and documentation of Least Concern assessment (Bachman et al. 2020).

In addition, replace in-person workshops with web-based tools through which experts can contribute with data, review, and proposed changes to reassessments would be a cheaper and feasible alternative (Rondinini et al. 2014). Such a strategy has an estimated mean cost of US$68.5 per species (Rondinini et al. 2014) and an estimated cost of more than US$1.8 million to carry out all reassessments. As this

figure is for times higher than the budget currently spent by the IUCN in reassessments (US$400,000 per year), the strategy adopted by the IUCN is to keep expanding the Red List to achieve the goal of 160,000 species assessed by 2020 and carry out reassessments of selected species groups (IUCN List Committee 2017).

New avenues for broadening the Red List cover also include incorporating regional and national assessments into the Red List and training resources. Regional or national assessments of endemic species based on Red List criteria and categories are equivalent to global assessments and could readily be incorporated on the Red List. A set of recommendations can further increase the quality of regional assessments and their consistency with the Red List, which include the use of the standard taxonomic and distribution information in both regional and global assessments, to provide proper documentation and data as detailed as on the Red List (Brito et al. 2010), to follow the regional guidelines strictly and do not adapt the IUCN criteria to meet specific needs (IUCN, 2012b; Miller et al., 2007); and ensure that assessors receive proper training on how to apply the Red List categories and criteria at regional levels and on how to document an assessment (Miller et al. 2007; IUCN 2012b; Bachman et al. 2018). Some training resources are available online (www.conservationtraining.org).

Table 2 Opportunities for broadening the coverage of the Red List.

Opportunity	Description	Reference
Use of predictive models to assess the extinction risk of Data-Deficient and non-assessed species	Life-history traits and ecological variables of assessed species can be used to predict the conservation status of Data-Deficient and non-assessed species with high accuracy	Collen et al. (2011); Machado and Loyola (2013); Darrah et al. (2017); Jetz and Freckleton (2015); Nori and Loyola (2015); Lughadha et al. (2019)
Expedite the Red List process using automatically generated assessments	Preliminary conservation assessments based on criterion B (range size) have shown a high level of accuracy and can be used to identifying non-threatened species and prioritize the threatened ones.	Brummitt et al. (2008); Bystriakova et al. (2019); Lughadha et al. (2019); Raimondo et al. (2013)
Automatic assessment and documentation of Least Concern species	Take advance of reduced data requirements for publication of Least Concern species on the Red List, a new tool for scaling-up documentation of Least Concern assessments are available at spbachman.shinyapps.io/rapidLC	Bachman et al. (2020)
Reduced data requirement for publication of Least Concern species	Minimal data requirements are now necessary to support Least Concern (LC) assessments, which can expedite and facilitate the publication of LC species on the Red List	Bachman et al. (2019); IUCN List Committee (2017)
Batch transference of assessments to the Red List	The recently developed SIS Connect system allows the transference of multiple raw data that support assessments to the IUCN's data management system	Bachman et al. (2018)

Incorporation of regional assessments of endemic species into the Red List	Endemic species are equivalent to global assessments and can be incorporated directly into the Red List. To facilitate this process, now regional assessments can be submitted to the Red List in English, French, Spanish, and Portuguese	Bachman et al. (2019); Rodríguez (2008); Zamin et al. (2010)
Self-training resources on the application of the Red List Categories and Criteria	Trained assessors are crucial for executing consistent assessments, and self-training material on the application of the Red List guidelines are freely available online at www.conservationtraining.org	Bachman et al. (2019); Miller et al. (2007); Rodríguez (2008)
Replace workshop to carry (re)assessments by online meetings	Replace in-person workshops with online discussion tools can be a cost-effective strategy to carry out assessments and reassessments	Rondinini et al. (2014)

Conclusion

Over its more than fifty years of existence, the Red List has evolved to become the most comprehensive source on the health of biodiversity on Earth. Information on species threat status is fundamental for catalyzing action needed to save species from the brink of extinction and to halt the current biodiversity crisis. Despite recent progress in making the Red List more representative, the total number of species assessed so far constitutes a small and taxonomic biased sample of global biodiversity. To continue to act as an updated and comprehensive "barometer of life," many challenges must be overpassed. New strategic and cost-effective opportunities and tools available can help expand the Red List and make it more complete.

References

Akçakaya HR, Ferson S, Burgman MA, Keith DA, Mace GM, Todd CR. 2000. Making Consistent IUCN Classifications under Uncertainty. Conservation Biology 14:1001–1013. Available from http://doi.wiley.com/10.1046/j.1523-1739.2000.99125.x.

Andelman SJ, Groves C, Regan HM. 2004. A review of protocols for selecting species at risk in the context of US Forest Service viability assessments. Acta Oecologica 26:75–83. Available from https://linkinghub.elsevier.com/retrieve/pii/S1146609X04000578.

Azam CS, Gigot G, Witte I, Schatz B. 2016. National and subnational Red Lists in European and Mediterranean countries: Current state and use for conservation. Endangered Species Research 30:255–266.

Bachman S, Moat J, Hill AW, de laTorre J, Scott B. 2011. Supporting red list threat assessments with GeoCAT: Geospatial conservation assessment tool. ZooKeys 150:117–126.

Bachman S, Walker B, Barrios S, Copeland A, Moat J. 2020. Rapid Least Concern: towards automating Red List assessments. Biodiversity Data Journal 8. Available from https://bdj.pensoft.net/article/47018/.

Bachman SP, Field R, Reader T, Raimondo D, Donaldson J, Schatz GE, Lughadha EN. 2019. Progress, challenges and opportunities for Red Listing. Biological Conservation 234:45–55. Elsevier. Available from https://doi.org/10.1016/j.biocon.2019.03.002.

Bachman SP, Nic Lughadha EM, Rivers MC. 2018. Quantifying progress toward a conservation assessment for all plants. Conservation Biology 0:1–9. Available from http://doi.wiley.com/10.1111/cobi.13071.

Bennun L et al. 2018. The Value of the IUCN Red List for Business Decision-Making. Conservation Letters 11:e12353. Available from http://doi.wiley.com/10.1111/conl.12353.

Betts J, Young RP, Hilton Taylor C, Hoffmann M, Rodríguez JP, Stuart SN, Milner Gulland EJ. 2020. A framework for evaluating the impact of the IUCN Red List of threatened species. Conservation Biology 0:cobi.13454. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13454.

Bottrill MC et al. 2008. Is conservation triage just smart decision making? Trends in Ecology & Evolution 23:649–654. Available from http://www.ncbi.nlm.nih.gov/pubmed/18848367 (accessed March 20, 2014).

Brito D, Ambal RG, Brooks T, Silva N De, Foster M, Hao W, Hilton-Taylor C, Paglia A, Rodríguez JP, Rodríguez JV. 2010. How similar are national red lists and the IUCN Red List? Biological Conservation 143:1154–1158. Elsevier Ltd. Available from http://linkinghub.elsevier.com/retrieve/pii/S0006320710000546 (accessed March 20, 2014).

Brum FT, Graham CH, Costa GC, Hedges SB, Penone C, Radeloff VC, Rondinini C, Loyola R, Davidson AD. 2017. Global priorities for conservation across multiple dimensions of mammalian diversity. Proceedings of the National Academy of Sciences 114:7641–7646. Available from http://www.pnas.org/lookup/doi/10.1073/pnas.1706461114.

Brummitt N, Bachman SP, Moat J. 2008. Applications of the IUCN Red List: Towards a global barometer for plant diversity. Endangered Species Research 6:127–135.

Burgman MA, Keith DA, Walshe T V. 1999. Uncertainty in comparative risk analysis for threatened Australian plant species. Risk Analysis 19:585–598.

Butchart SHM et al. 2010. Global biodiversity: indicators of recent declines. Science 328:1164–8. Available from http://www.sciencemag.org/cgi/doi/10.1126/science.1187512.

Bystriakova N, Alves De Melo PH, Moat J, Lughadha EN, Monro AK. 2019. A Preliminary Evaluation of The Karst Flora of Brazil Using Collections Data. Scientific Reports 9:1–13. Springer US. Available from http://dx.doi.org/10.1038/s41598-019-53104-6.

Cardillo M, Mace GM, Gittleman JL, Jones KE, Bielby J, Purvis A. 2008. The predictability of extinction: biological and external correlates of decline in mammals. Proceedings of the Royal Society B: Biological Sciences 275:1441–1448. Available from https://royalsocietypublishing.org/doi/10.1098/rspb.2008.0179.

CBD. 2020. Strategic Plan for Biodiversity 2011–2020.

Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances 1:e1400253. Available from https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.1400253.

Chichorro F, Juslén A, Cardoso P. 2019. A review of the relation between species traits and extinction risk. Biological Conservation 237:220–229. Elsevier. Available from https://doi.org/10.1016/j.biocon.2019.07.001.

Collen B et al. 2016. Clarifying misconceptions of extinction risk assessment with the IUCN Red List. Biology Letters 12:20150843. Available from https://royalsocietypublishing.org/doi/10.1098/rsbl.2015.0843.

Collen B, Mcrae L, Deinet S, de Palma A, Carranza T, Cooper N, Loh J, Baillie JEM. 2011. Predicting how populations decline to extinction. Philosophical Transactions of the Royal Society B: Biological Sciences 366:2577–2586.

Darrah SE, Bland LM, Bachman SP, Clubbe CP, Trias-Blasi A. 2017. Using coarse-scale species distribution data to predict extinction risk in plants. Diversity and Distributions 23:435–447.

Dauby G et al. 2017. ConR: An R package to assist large-scale multispecies preliminary conservation assessments using distribution data. Ecology and Evolution 7:11292–11303.

Davidson AD, Hamilton MJ, Boyer AG, Brown JH, Ceballos G. 2009. Multiple ecological pathways to extinction in mammals. Proceedings of the National Academy of Sciences 106:10702–10705. Available from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2705575&tool=pmcentrez&rendertype=abstract.

Davidson AD, Shoemaker KT, Weinstein B, Costa GC, Brooks TM, Ceballos G, Radeloff VC, Rondinini C, Graham CH. 2017. Geography of current and future global mammal extinction risk. PLOS ONE 12:e0186934. Available from https://dx.plos.org/10.1371/journal.pone.0186934.

De Grammont PC, Cuarón AD. 2006. An Evaluation of Threatened Species Categorization Systems Used on the American Continent. Conservation Biology 20:14–27. Available from http://doi.wiley.com/10.1111/j.1523-1739.2006.00352.x.

Eaton MA, Gregory RD, Noble DG, Robinson JA, Hughes J, Procter D, Brown AF, Gibbons DW. 2005. Regional IUCN Red Listing: the Process as Applied to Birds in the United Kingdom. Conservation Biology 19:1557–1570. Available from http://doi.wiley.com/10.1111/j.1523-1739.2005.00213.x.

Faleiro F V., Machado RB, Loyola RD. 2013. Defining spatial conservation priorities in the face of land-use and climate change. Biological Conservation 158:248–257. Elsevier Ltd. Available from http://dx.doi.org/10.1016/j.biocon.2012.09.020.

Flora do Brasil 2020. 2016. Available from http://floradobrasil.jbrj.gov.br (accessed December 9, 2017).

Fritz S a, Bininda-Emonds ORP, Purvis A. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. Ecology letters 12:538–49. Available from http://www.ncbi.nlm.nih.gov/pubmed/19392714 (accessed March 20, 2014).

Gärdenfors U. 2001. Classifying threatened species at national versus global levels. Trends in Ecology and Evolution 16:511–516.

Gärdenfors U, Hilton-Taylor C, Mace GM, Rodríguez JP. 2001. The application of IUCN Red List criteria at regional levels. Conservation Biology 15:1206–1212.

Hoffmann M et al. 2008. Conservation planning and the IUCN Red List. Endangered Species Research 6:113–125. Available from http://www.int-res.com/abstracts/esr/v6/n2/p113-125/.

Hoffmann M, Hilton-Taylor C, Angulo A, Böhm M, Brooks TM, Et. al. 2010. The Impact of Conservation on the Status of the World's Vertebrates. Science 330:1503–1509. Available from http://www.sciencemag.org/content/330/6010/1503.abstract%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/20978281.

Humphreys AM, Govaerts R, Ficinski SZ, Nic Lughadha E, Vorontsova MS. 2019. Global dataset shows geography and life form predict modern plant extinction and rediscovery. Nature Ecology & Evolution 3:1043–1047. Available from http://www.nature.com/articles/s41559-019-0906-2.

Isaac NJB, Cowlishaw G. 2004. How species respond to multiple extinction threats. Proceedings of the Royal Society of London. Series B: Biological Sciences 271:1135–1141. Available from http://www.

pubmedcentral.nih.gov/articlerender.fcgi?artid=1691707&tool=pmcentrez&rendertype=abstract.

IUCN. 2012a. IUCN Red List Categories and Criteria: Version 3.1Second edi. Gland, Switzerland and Cambridge, UK.

IUCN. 2012b. Guidelines for Application of IUCN Red List Criteria at Regional and National Levels: Version 4.0. Gland.

IUCN. 2019. Guidelines for Using the IUCN Red List Categories and Criteria. Version 14. Prepared by the Standards and Petitions Committee. Available from http://intranet.iucn.org/webfiles/doc/SSC/RedList/RedListGuidelines.pdf.

IUCN. 2021. Table 1a: Number of species evaluated in relation to the overall number of described species, and numbers of threatened species by major groups of organisms. Available from www.iucnredlist.org/resources/summary-statistics (accessed August 1, 2021).

IUCN List Committe. 2017. The IUCN Red List of Threatened SpeciesTM Strategic Plan 2017 - 2020.:13.

Jetz W, Freckleton RP. 2015. -Towards a General Framework for Predicting Threat Status of Data-Deficient Species From Phylogenetic, Spatial and Environmental Information. Philosophical Transactions of the Royal Society B: Biological Sciences 370:1–10.

Juffe-Bignoli D et al. 2016. Assessing the cost of global biodiversity and conservation knowledge. PLoS ONE 11:1–22.

Keith DA et al. 2004. Protocol for listing threatened species can forecast extinction. Ecology Letters 7:1101–1108.

Loyola R, Machado N, Ribeiro BR, Martins E, Martinelli G. 2018. Áreas prioritárias para conservação da flora endêmica do estado do Rio de Janeiro. Graficci, Rio de Janeiro.

Mace GM, Collar NJ, Gaston KJ, Hilton-Taylor C, Akçakaya HR, Leader-Williams N, Milner-Gulland EJ, Stuart SN. 2008. Quantification of extinction risk: IUCN's system for classifying threatened species. Conservation biology : the journal of the Society for Conservation Biology 22:1424–42. Available from http://doi.wiley.com/10.1111/j.1523-1739.2008.01044.x.

Mace GM, Lande R. 1991. Assessing Extinction Threats: Toward a Reevaluation of IUCN Threatened Species Categories. Conservation Biology 5:148–157. Available from http://doi.wiley.com/10.1111/j.1523-1739.1991.tb00119.x.

Mace GM, Possingham HP, Leader-Williams N. 2007. Prioritizing choices in conservation. Pages 17–34 in D. W. Macdonald and K. Service, editors. Key Topics in Conservation Biology. Blackwell Oxford.

Margules CR, Pressey RL. 2000. Systematic conservation planning. Nature 405:243–53. Available from http://www.ncbi.nlm.nih.gov/pubmed/10821285.

Martinelli G, Moraes MÁ. 2013. Livro Vermelho da Flora do Brasil1 st. Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro.

Master LL. 1991. Threats and Setting Priorities for Conservation. Conservation Biology 5:559–563.

Master LL, Faber-Langendoen D, Bittman R, Hammerson GA, Heidel B, L.Ramsay, Snow K, Teucher A, Tomaino A. 2012. NatureServe Conservation Status Assessments: Factors for Assessing Extinction Risk. Arlington,VA.

McCarthy MA, Thompson CJ, Garnett ST. 2008. Optimal investment in conservation of species. Journal of Applied Ecology 45:1428–1435. Available from http://doi.wiley.com/10.1111/j.1365-2664.2008.01521.x.

Mckinney ML. 1997. EXTINCTION VULNERABILITY AND SELECTIVITY : Combining Ecological and Paleontological Views. Annual Review of Ecology and Systematics 28:495–516.

McKinney ML. 1997. EXTINCTION VULNERABILITY AND SELECTIVITY:Combining Ecological and Paleontological Views. Annual Review of Ecology and Systematics 28:495–516. Available from http://www.annualreviews.org/doi/10.1146/annurev.ecolsys.28.1.495.

Miller RM et al. 2007. National threatened species listing based on IUCN criteria and regional guidelines: Current status and future perspectives. Conservation Biology 21:684–696.

Nic Lughadha EM et al. 2019a. The use and misuse of herbarium specimens in evaluating plant extinction risks. Philosophical Transactions of the Royal Society B: Biological Sciences 374:20170402. Available from http://www.royalsocietypublishing.org/doi/10.1098/rstb.2017.0402.

Nic Lughadha EM, Graziele Staggemeier V, Vasconcelos TNC, Walker BE, Canteiro C, Lucas EJ. 2019b. Harnessing the potential of integrated systematics for conservation of taxonomically complex, megadiverse plant groups. Conservation Biology 33:511–522. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13289.

Novacek M. 2001. The Biodiversity Crisis: Losing What Counts. Page (Novacek, M J, editor). The New Press.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science 344:1246752–1246752. Available from http://www.ncbi.nlm.nih.gov/pubmed/24876501 (accessed July 9, 2014).

Possingham HP, Andelman SJ, Burgman MA, Medellın RA, Master LL, Keith DA. 2002. Limits to the use of threatened species lists. Trends in Ecology & Evolution 17:503–507. Available from https://linkinghub.elsevier.com/retrieve/pii/S0169534702026149.

Purvis A, Jones KE, Mace GM. 2000. Extinction. BioEssays 22:1123–1133. Available from https://royalsocietypublishing.org/doi/10.1098/rstb.2011.0015.

Raimondo DC. 2011. The Red List of South African plants: a global first. South African Journal of Science 107. Available from http://sajs--archive.ojs.knowledgearc.net/index.php/SAJS/article/view/653.

Raimondo DC, von Staden L, Donaldson JS. 2013. Lessons from the Conservation Assessment of the South African Megaflora. Annals of the Missouri Botanical Garden 99:221–230.

Regan HM, Colyvan M, Burgman MA. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications 12:618–628.

Regan TJ, Burgman MA, McCarthy M, Master LL, Keith DA, Mace GM, Andelman SJ. 2005. The consistency of extinction risk classification protocols. Conservation Biology 19:1969–1977.

Ribeiro BR, Martins E, Martinelli G, Loyola R. 2018. The effectiveness of protected areas and indigenous lands in representing threatened plant species in Brazil. Rodriguésia 69:1539–1546. Available from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2175-78602018000401539&lng=en&tlng=en.

Rodrigues ASL et al. 2004. Effectiveness of the global protected area network in representing species diversity. Nature 428:640–643.

Rodrigues ASL, Pilgrim JD, Lamoreux JF, Hoffmann M, Brooks TM. 2006. The value of the IUCN Red List for conservation. Trends in ecology & evolution 21:71–6. Available from http://www.ncbi.nlm.nih.gov/pubmed/16701477 (accessed March 3, 2012).

Rodríguez JP. 2008. National Red Lists: The largest global market for IUCN Red List Categories and Criteria. Endangered Species Research 6:193–198.

Rondinini C, Di Marco M, Visconti P, Butchart SHM, Boitani L. 2014. Update or Outdate: Long-Term Viability of the IUCN Red List. Conservation Letters 7:126–130.

Sales LP, Ribeiro BR, Hayward MW, Paglia A, Passamani M, Loyola R. 2017. Niche conservatism and the invasive potential of the wild boar. Journal of Animal Ecology 86:1214–1223. Available from http://doi.wiley.com/10.1111/1365-2656.12721.

Sharrock S, Oldfield S, Wilson O. 2014. Plant conservation report 2014: A review of progress in implementation of the Global Strategy for Plant Conservation 2011-2020. Page Technical SeriesTechnical. Secretariat of the Convention on Biological Diversity, Montréal, Canada and Botanic Gardens Conservation International, Richmond, UK.

Stuart SN, Wilson EO, McNeely JA, Mittermeier RA, Rodriguez JP. 2010. The Barometer of Life. Science 328:177–177. Available from https://www.sciencemag.org/lookup/doi/10.1126/science.1188606.

Vié J, Hilton-taylor C, Pollock C, Ragle J, Smart J, Stuart S, Tong R. 2008. THE IUCN Red List : A key conservation tools. Page in Hilton--Taylor and S.N. Stuart, editor. The 2008 Review of The IUCN Red List of Threatened Species. IUCN Gland, Switzerland.

Zamin TJ, Baillie JEM, Miller RM, Rodríguez JP, Ardid A, Collen B. 2010. National Red Listing Beyond the 2010 Target. Conservation Biology 24:1012–1020. Available from http://doi.wiley.com/10.1111/j.1523-1739.2010.01492.x.

Zizka A, Silvestro D, Vitt P, Knight TM. 2021. Automated conservation assessment of the orchid family with deep learning. Conservation Biology 35:897–908. Available from https://onlinelibrary.wiley.com/doi/10.1111/cobi.13616.

Supplementary material

IUCN versus NatureServe system

Although the Red List and NatureServe systems have similar data requirements, they differ in the way data are combined and weighted for assigning conservation categories. Moreover, different names are used to design conservation status (IUCN: categories; NatureServe: rank factors) and criteria (IUCN: criteria; NatureServe: factors). The NatureServe system uses conservation status rank raging from G1 (critically imperiled) to G5 (secure), which, in turn, is defined based on a set of 10 criteria (factors) organized in three broad groups –rarity, threats, and trends (Figure 1). A collection of rules specifies whether (and how) criteria should be used (Master et al. 2012).

Unlike the IUCN, the NatureServe system is based on a mixture of quantitative, qualitative, and subjective criteria (Keith et al. 2004; Master et al. 2012). The criteria are scaled and weighted according to their perceived impact on risk (i.e., an approach based on the weight-of--evidence; (Linkov et al. 2009). Each factor has a specific weight those in the rarity group (population size, range extent, area of occupancy, number of occurrences, number of occurrences or percent of area occupied with good viability/ecological integrity, and environmental specificity) receive the greatest weight (70%), thus, contributing more to the definition of risk status. Rarity factors are then combined with threat factors (overall threat impact and species' intrinsic vulnerability; accounting for 30% of the total weight) for creating an initial score, which is subsequently adjusted by the addition or subtraction of the trends score (long and short term trend in population size or area) to produce a final score that, eventually is translate in a conservation category (G1-G5) (Master et al., 2012; Figure S1). If trends are positive (i.e., the population is growing and range expanding), species are classified under a lower threat category. On the other hand, if population and distribution have been reduced and decreasing, respectively, the trend factor is used to qualify species under a higher threat category.

Unlike the IUCN system, NatureServe allows species to be classified in more than one risk category (e.g., G1-G3) to reveal the degree of uncertainty associated with the information available to assess species conservation status (Master et al. 2012).

Despite their structure and data requirement differences, NatureServe and IUCN systems tend to produce related but not identical conservation assessments (O'Grady et al. 2004). Generally, assessors using both methods concur in classify species in the highest and lowest categories but disagree in classifying species in intermediate categories. In its last revision, NatureServe sought to standardize the rating of information shared with the Red List, including range extent, area of occupancy, population size, and threats (Master et al. 2012). The standardization was intended to enhance the compatibility and exchange of information between the methodologies (Master et al. 2012). Whether such changes increase the consistency of extinction risk classification from both systems is not known yet.

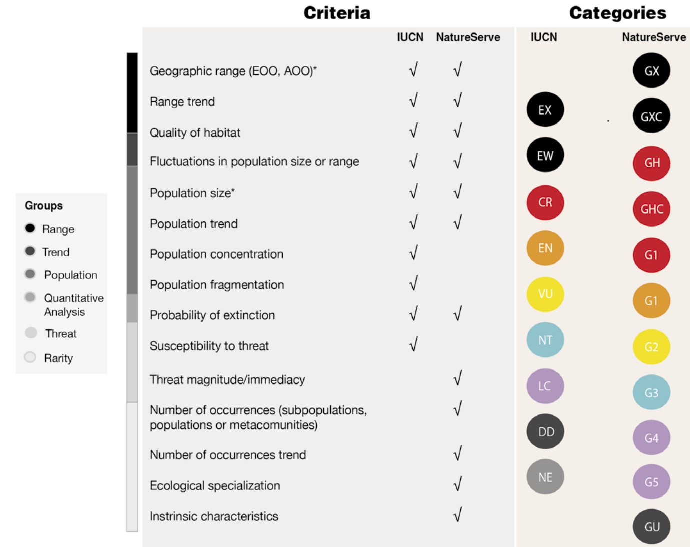


Fig. S1 Comparison between the categories and criteria of the IUCN's Red List and the NatureServe systems used for assessing extinction risk. The criteria used for classifying species under different threat categories are assembled in six major groups (greys scale). (*) criteria grouped in rarity according to the NatureServe classification. The correspondence between IUCN and NatureServe categories is shown by similar colors in the column "Categories". The list of criteria was based on O'Grady et al. (2004). Extinct (EX) Extinct in the Wild (EW), Critically Endangered (CR), Endangered (EN), Vulnerable (VU), Near Threatened (NT), Least Concern (LC), Data Deficient (DD), Not Evaluated (NE); Presumed Extinct (GX), Presumed Extinct in the Wild (GXC), Possibly Extinct (GH), Possibly Extinct in the Wild (GHC) Critically Imperiled (G1), Critically Imperiled (G2), Imperiled (G3), Vulnerable (G3), Apparently Secure (G4), Secure (G5), Unrankable (GU).

References

Keith DA et al. 2004. Protocol for listing threatened species can forecast extinction. Ecology Letters 7:1101–1108.

Linkov I, Loney D, Cormier S, Satterstrom FK, Bridges T. 2009. Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. Science of The Total Environment 407:5199–5205. Elsevier B.V. Available from http://dx.doi.org/10.1016/j.scitotenv.2009.05.004.

Master LL, Faber-Langendoen D, Bittman R, Hammerson GA, Heidel B, L.Ramsay, Snow K, Teucher A, Tomaino A. 2012. NatureServe Conservation Status Assessments: Factors for Assessing Extinction Risk. Arlington,VA.

O'Grady JJ, Burgman Ma, Keith Da, Master Ll, Andelman Sj, Brook Bw, Hammerson Ga, Regan T, Frankham R. 2004. Correlations among Extinction Risks Assessed by Different Systems of Threatened Species Categorization. Conservation Biology 18:1624–1635. Available from http://doi.wiley.com/10.1111/j.1523-1739.2004.00109.x.

CAPÍTULO II

bdc: An R toolkit for standardizing, integrating, and cleaning biodiversity data

Ribeiro, BR; Velazco, SJE; Guidoni-Martins, K; Tessarolo, G; Jardim, L; Bachman, SP; Loyola, R. ***bdc:* An R toolkit for standardizing, integrating, and cleaning biodiversity data**
Em revisão (Methods in Ecology and Evolution)

bdc: An R toolkit for standardizing, integrating, and cleaning biodiversity data

Bruno R. Ribeiro[1, *], Santigo José Elías Velazco[2,3,4], Karlo Guidoni-Martins[1], Geiziane Tessarolo[5], Lucas Jardim[6], Steven P. Bachman[7], Rafael Loyola[8,9]

Authors' affiliation

[1] Programa de Pós-graduação em Ecologia e Evolução, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[2] Department of Botany and Plant Sciences, University of California - Riverside, Riverside, California, USA

[3] Instituto de Biología Subtropical, Universidad Nacional de Misiones - CONICET, Puerto Iguazú, Misiones, Argentina

[4] Programa de Pós-Graduação em Biodiversidade Neotropical, Universidade Federal da Integração Latino-Americana (UNILA), Foz do Iguaçu, Paraná, Brazil

[5] Universidade Estadual de Goiás, UEG, Campus de Ciências Exatas e Tecnológicas - CCET, Anápolis, Goiás, Brazil

[6] Instituto Nacional de Ciência e Tecnologia em Ecologia, Evolução e Conservação da Biodiversidade, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[7] Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, U.K.

[8] Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[9] Fundação Brasileira para o Desenvolvimento Sustentável, Rio de Janeiro, Brazil

[*] Corresponding author's address

Programa de Pós-graduação em Ecologia e Evolução, Universidade Federal de Goiás, Avenida Esperança s/n, Campus Samambaia, CEP 74.690-900, Goiânia, Goiás, Brazil.

E-mail: ribeiro.brr@gmail.com

Abstract

The increase in online and openly accessible biodiversity databases provides a vast and invaluable resource to support research and policy. However, without scrutiny, errors in primary species-occurrence data can lead to erroneous results and misleading information. We present the Biodiversity Data Cleaning (bdc) R package to address quality issues and improve the fitness-for-use of biodiversity datasets. The *bdc* contains functions to harmonize and integrate data from different sources following common standards and implements various tests and tools to flag, report, clean, and correct taxonomic, spatial, and temporal biodiversity data. The package is accompanied by extensive documentation and tutorials. We demonstrated the *bdc* package's applicability in cleaning more than 30 million occurrence records for terrestrial plant species in Brazil. We found that around one-fifth of the original datasets hold the standard quality requirements. We hope the *bdc* package can facilitate the data-cleaning process and catalyze improvements to allow the wise and efficient use of primary biodiversity data.

Keywords: biodiversity, big data, data cleaning, data quality, fitness-for-use, GBIF, plants, taxonomy.

Resumo

O aumento de bancos de dados de biodiversidade online e de livre acesso fornece um recurso vasto e inestimável para apoiar o desenvolvimento de pesquisas e políticas públicas. No entanto, sem escrutínio, erros nos registros de ocorrência de espécies podem levar a resultados errôneos e informações equivocadas. Apresentamos o pacote *Biodiversity Data Cleaning* (bdc) criado e disponibilizado no ambiente R de programação para o tratamento e melhora da qualidade dados de biodiversidade de forma a torná-los apto ao uso. O pacote *bdc* contém funções para harmonizar e integrar dados de diferentes fontes seguindo padrões estabelecidos e implementa vários testes e ferramentas para identificar, relatar, limpar e corrigir dados taxonômicos, espaciais e temporais de biodiversidade. O pacote é acompanhado por uma extensa documentação e por tutoriais. Demonstramos sua aplicabilidade na limpeza de mais de 30 milhões de registros de ocorrência de espécies de plantas terrestres no Brasil. Descobrimos que cerca de um quinto do conjunto original de dados atende aos requisitos de qualidade estabelecidos. Esperamos que o pacote *bdc* possa facilitar o processo de limpeza de dados e catalisar melhorias para permitir o uso inteligente e eficiente de dados primários de biodiversidade.

Palavras-chave: biodiversidade, big data, limpeza de dados, qualidade de dados, adequação ao uso, GBIF, plantas, taxonomia.

Introduction

The development of biodiversity informatics tools and new computational platforms in recent decades has led to a significant increase in the online availability of primary species-occurrence data retrieved from natural history collections and citizen science observations (Bisby, 2000; Soberón & Peterson, 2004; Graham et al., 2004). Such openly accessible biodiversity databases provide a vast and invaluable resource to document species distributions through time and space for research, education, and environmental policy support (Chapman 2005a; Canhos et al. 2015; Ball-Damerow et al. 2019)

The importance of primary species-occurrence data for many biodiversity applications is evident, yet they have limitations, and their quality can vary substantially (Newbold, 2010; Meyer et al., 2016). Without scrutiny, issues related to difficulty standardizing data from different sources (Kissling et al., 2018), discrepancies and errors in taxonomic and nomenclatural data (e.g., Nic Lughadha et al., 2019a; Mesibov, 2013), and errors and inaccuracies in geographical and temporal information of primary species-occurrence data (e.g., Meyer et al. 2016; Daru et al. 2018) can lead to erroneous results and misleading information (Maldonado et al., 2015; Nic Lughadha et al., 2019a; Zizka et al., 2020). Although several efforts have already been made to develop tools for cleaning biodiversity data and improve fitness-for-use (e.g., functionalities found in GBIF (www.gbif.org), the Atlas of Living Australia (www.ala.org.au), SpeciesLink (splink.cria.org.br), and many R packages (details in Appendix 1; figshare.com/articles/thesis/Ap_ndice_1/15156342), significant challenges remain, especially when assembling large and heterogeneous databases from online aggregators (Chapman, 2005b; Kissling et al., 2018).

To the best of our knowledge, no existing R package handles most aspects needed to assess the quality of primary biodiversity data, including the standardization and integration of different datasets, taxonomic harmonization, and tests to detect and correct issues regarding the spatial and temporal information of data. In addition, features to visualize, document, and report data quality – which is essential for make data quality assessment transparent and reproducible – are rarely available in the existing packages (but see Gueta et al., 2018; a comparison between existing

R packages used to data cleaning and quality assessment can be found in Appendix 1; figshare.com/articles/thesis/Ap_ndice_1/15156342).

Here, we present the Biodiversity Data Cleaning (*bdc*) package to address quality issues and improve the fitness-for-use of a dataset. *bdc* contains functions to harmonize and integrate data from different sources following common standards and protocols and implements various tests and tools to flag, document, clean, and correct taxonomic, spatial, and temporal data. The package builds upon the integration and enhancement of cutting-edge functionalities (Carvalho 2017; Zizka et al. 2019; Norman et al. 2020) and on a series of new tests and tools developed for validating, documenting, and reporting data quality.

Description

Overview

The *bdc* package is organized in thematic modules related to different biodiversity dimensions (Fig. 1 and Table 1; Meyer et al., 2016), including: 1) Standardization and integration of different datasets; 2) Pre-filter: flagging and removal of invalid or non-interpretable information, followed by data amendments; 3) Taxonomy: cleaning, parsing, standardization and updated of scientific names against multiple taxonomic references; 4) Space: flagging of erroneous, suspect, and low-precision geographic coordinates; and 5) Time: flagging and, whenever possible, correction of inconsistent collection date (Fig. 1; Table 1). In addition, the package contains functions for documenting the results of the data-cleaning tests, including functions for saving i) records needing further inspection, ii) databases containing the results of each step, iii) figures, and iv) data-quality reports. These files facilitate the interpretation and visualization of the results by users and are automatically saved in a folder named "Outputs" (Fig. 1; Table 1). The modules illustrated, and functions within, can be linked to form a workflow, but can also be executed independently depending on user needs.

Each thematic modules (except the one used to merge datasets) is composed of tests to assert data quality (Table 1 and S1). When applying tests, the original data are retained, and the result of each test is added to the database as a separate field. Results are retrieved as TRUE or FALSE, in which the former indicates correct records and the latter potentially problematic or suspect records. This process follows data quality principles (Chapman, 2005b, 2005a) and allows users to track and validate all modifications in the original databases (Zizka et al. 2019).

In the *bdc* package, we sought to encompass a series of tests regarding the taxonomic, spatial, and temporal dimensions of data to resolve the most common data quality issues, some of those yet not available in online data aggregators, biological collections, or available software (e.g., R packages). Compared to other available R packages, the main strengths of the bdc package are that it brings together available tools – and a series of new ones – to assess the quality of different dimensions of biodiversity data into a single and flexible framework. The tools can be applied to a multitude of taxonomic groups, datasets (including regional or local repositories), countries, or worldwide.

The *bdc* package is implemented in R (R Core Team 2020) and it is based on standard tools for data quality assessments as well as data handling and visualization, including *taxadb* (Norman et al. 2020), *CoordinateCleaner* (Zizka et al. 2019), *flora* (Carvalho 2017), and *ggplo2* (Wickham 2016). We strongly encourage citing those packages when using the *bdc* package. We provide extensive documentation and tutorials on functions on the package website (github.com/brunobrr/bdc).

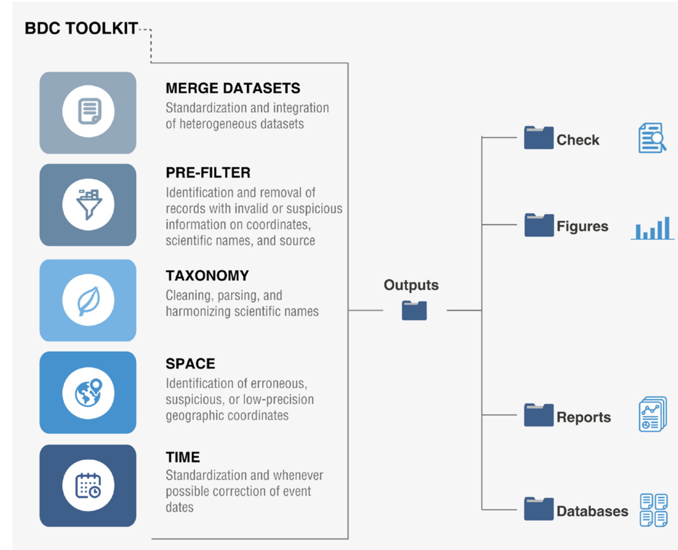


Fig. 1 The Biodiversity Data Cleaning (*bdc*) package contains functionalities for standardizing and integrating data from different sources and implement several tests for flag, document, clean and correct biodiversity data. The *bdc* package is organized in thematic modules (merge datasets, pre-filter, taxonomy, space, and time). Several outputs documenting the data-cleaning process can be saved, including files needing further inspections, figures, and reports.

Table 1 List and description of the main functions implemented in the bdc package (more details are presented in Table S1). Functions are grouped in thematic modules, namely Merge Datasets, Pre-filter, Taxonomy, Space, and Time.

Modules	Label	Description	Source
Merge databases	*bdc_standardize_datasets*	Harmonization and integration of different datasets into a standard database	bdc
Pre-filter	*bdc_scientificName_empty*	Identification of records lacking names or with names not interpretable	bdc
	bdc_coordinates_empty	Identification of records lacking information on latitude or longitude	bdc
	bdc_coordinates_outOfRange	Identification of records with out-of-range coordinates (latitude > 90 or -90; longitude >180 or -180)	bdc
	bdc_basisOfRecords_notStandard	Identification of records from doubtful sources (e.g., fossil or machine observation) impossible to interpret and not compatible with Darwin Core recommended vocabulary	bdc
	bdc_country_from_coordinates	Derive country name from valid geographic coordinates	bdc
	bdc_country_standardized	Standardization of country names and retrieve country code	bdc
	bdc_coordinates_transposed	Identification of records with potentially transposed latitude and longitude	bdc
	bdc_coordinates_country_inconsistent	Identification of coordinates in other countries or far from a specified distance from the coast of a reference country (i.e., in the ocean)	bdc
	bdc_coordinates_from_locality	Identification of records lacking coordinates but with a detailed description of the locality associate with records from which coordinates can be derived	bdc

Modules	Label	Description	Source
Taxonomy	*bdc_clean_names*	Name-checking routines to clean and split a taxonomic name into its binomial and authority components	bdc; rgnparser
	bdc_query_names_taxadb	Standardization of scientific names by correcting spelling errors and converting nomenclatural synonyms to currently accepted names. Names are standardized based on one out of ten taxonomic authorities (i.e., backbones) available in the taxadb package. Spelling errors are corrected using a fuzzy match algorithm based on a match distance defined by the user. The name standardization quality can be accessed in the column "notes" placed in the table resulting from the name standardization process	bdc; taxadb
	bdc_filter_out_names	This tool is used to filter out records according to their taxonomic status present in the column "notes". For example, to filter only valid accepted names categorized as "accepted"	bdc
Space	*clean_coordinates*	Identification of potentially problematic geographic coordinates based on geographic gazetteers and metadata. Include tests for flagging records: around country capitals or country or province centroids, duplicated, with equal coordinates, around biodiversity institutions, within urban areas, plain zeros in the coordinates, and suspect geographic outliers	CoordinateCleaner v2.0-18
	bdc_coordinates_precision	Identification of records with a coordinate precision below a specified number of decimal places	bdc
Time	*bdc_eventDate_empty*	Identification of records lacking information on event date (i.e., when a record was collected or observed)	bdc

Modules	Label	Description	Source
	bdc_year_outOfRange	Identification of records with illegitimate or potentially imprecise collecting year. The year provided can be out-of-range (e.g., in the future) or collected before a specified year supplied by the user (e.g., 1900)	bdc
	bdc_year_from_eventDate	This function extracts four-digit year from unambiguously interpretable collecting dates	bdc
All modules	*bdc_create_report*	Creation of data-quality reports documenting the results of data-quality tests and the taxonomic harmonization process	bdc
	bdc_create_figures	Creation of figures (i.e., bar plots and maps) reporting the results of data-quality tests	bdc
	bdc_filter_out_flags	Removal of columns containing the results of data quality tests (i.e., column starting with ".") or other columns specified	bdc
	bdc_quickmap	Creation of a map of points using ggplot2. Helpful in inspecting the results of data-cleaning tests	bdc
	bdc_summary_col	This function creates or updates the column summarizing the results of data quality tests (i.e.., the column ".summary")	bdc

Standardization and integration of heterogeneous datasets

Primary species-occurrence data found in online data aggregators or local repositories often exhibit different data and meta-data formats, thereby restricting comparison and integration (Seebens et al. 2020). The lack of terminology standardization makes the integration of large and heterogeneous datasets a challenge. To remedy this, the function *bdc_standardize_datasets* specifically handles the standardization of heterogeneous datasets. To do so, users must fill out a configuration table (see example in Appendix 2; figshare.com/articles/thesis/Ap_ndice_2/15156384) to indicate which field names (i.e., column headers) of each original dataset match a list of Darwin Core standard terms (Wieczorek et al. 2012). Once standardized, datasets are then integrated into a formatted database having a minimum set of terms required to share biodiversity data and metadata across a wide variety of biodiversity applications (Table S2; see also Simple Darwin Core standards at dwc.tdwg.org/simple).

Pre-filter

Large and heterogeneous datasets may contain thousands of records missing spatial or taxonomic information (partially or entirely) and several records outside a region of interest (Peterson et al. 2018; Jin & Yang 2020). Such lower quality data are not fit-for-use in many research applications without prior amendments. An advantage of removing suspect records with an initial first pass is a potential increase in processing speed with downstream analyses from the reduced number of records. Thus, the pre-filter module contains functions to flag and remove records i) missing species names, ii) missing partial or complete information on geographic coordinates, iii) out-of-range coordinates (latitude > 90 or -90; longitude >180 or -180), iv) records from doubtful sources (e.g., from drawings, photographs, or multimedia objects, among others), and v) records outside a region of interest, i.e., records in other countries or at an informed distance from the coast (e.g., in the ocean). This last step avoids falsely flagging records close to country limits as invalid (e.g., records of coast or marshland species; see Table S1 for additional details about this test). The pre-filter module also includes functions for data enhancement, such as deriving country names from valid geographic coordinates, standardizing country names, identification of records with potentially transposed geographic coordinates, and saving a table containing records with missing coordinates but with potentially useful locality information (Table 1 and S1; see Supplementary Material for more details).

Taxonomic harmonization

The combination of large datasets from several sources requires careful standardization of potentially thousands of taxonomic names. For instance, some data aggregators do not share a standard taxonomic reference and inevitably differ in the taxonomic concept applied (Boyle et al. 2013; Mesibov 2018; Bystriakova et al. 2019). The *bdc* package includes functions to help the names standardization by comparing scientific names against one of 10 taxonomic references. The taxonomic standardization uses *taxadb* package (Norman et al. (2020), which contains functions that allow querying millions of taxonomic names in a fast, automated, and consistent way using high-quality, locally stored, taxonomic databases. Querying names against these databases avoids significant drawbacks inherent to tools that implement queries using web APIs (Application Programming Interface), such as the need for internet access to perform queries, and limitations on number of names that can be queried at once; details in Norman et al. (2020).

A major limitation of *taxadb* package is that misspelled scientific names – commonly found in biodiversity databases – cannot be resolved by an exact matching algorithm, which may result in many unresolved names. To troubleshoot this, we developed functions (*bdc_clean_names* and *bdc_query_names_taxadb* functions) for i) cleaning and parsing scientific names; ii) resolving misspelled names or variant spellings using a fuzzy matching application, iii) converting nomenclatural synonyms to the currently accepted name, and iv) flagging ambiguous results. Scientific names improperly formatted usually cannot be resolved by exact or fuzzy matching applications. The *bdc_clean_names* is composed of several name-checking routines that optimize the taxonomic queries by increasing the probability of finding matching names (Table 2, Tables S1 and S3). After cleaned and parsed, names are then standardized based on one out of 10 taxonomic authorities (docs.ropensci.org/taxadb/articles/data-sources.html) available in the *taxadb* package. Even after running name-checking routines, a scientific name can remain unresolved because of typos or spelling variants. In such cases, a fuzzy matching algorithm processes name-matching queries to find a potential matching candidate from the specified taxonomic database based on a match distance defined by the user A detailed explanation of the taxonomic harmonization process can be found in the Supplementary Material and Table S1.

Table 2 Scientific names improperly formatted or containing authority or annotation information can constrain harmonizing names from heterogeneous sources. The *bdc* package uses a series of name-cleaning routines for cleaning (e.g., remove family names pre-pended to scientific names, terms denoting taxonomic uncertainty, infraspecific marks) and parsing names (i.e., separate taxon names from an authority, date, and annotations).

Original names	Names cleaned and parsed
LEGUMINOSAE Clitoria cf. guianensis (Aubl.) Benth.	*Clitoria guianensis*
DIOSCOREACEAE Dioscorea ? trifida L.f.	*Dioscorea trifida*
Agarista niederleinii (Sleumer) Judd var. niederleinii	*Agarista niederleinii niederleinii*
Erpodium coronatum (Hook. & Wilson) Mitt.	*Erpodium coronatum*
MELASTOMATACEAE Tibouchina aff. sellowiana Cogn.	*Tibouchina sellowiana*
Bauhinia forficata Link subsp. forficata	*Bauhinia forficata forficata*

Identification of errors in geographic coordinates

Flagging and cleaning potentially erroneous geographical coordinates is necessary to improve the quality of biological databases (Meyer et al., 2016; Jin & Yang, 2020). We used the *CoordinateCleaner*, an R package based on geographic

gazetteers, to flag potential erroneous coordinates (Zizka et al. 2019), which include records with i) zero coordinates in a radius around the point at zero latitudes and longitude; ii) equal latitude and longitude; iii) possible duplicate records with equal longitude, latitude, and accepted species name. Likewise, the package identifies records assigned to iv) country capitals; v) province centroids; vi) urban areas; vii) biodiversity institutions; and viii) geographic outliers (see Fig. 1, Table 1 and S1). Finally, we also used a tool for identifying records with low precision coordinates (Robertson et al., 2016). More details on each test can be found in Table S1 and Zizka et al. (2019).

Standardization and validation of temporal information

Temporal data are frequently present in biodiversity databases in the form of collection dates and represent a taxon's existence at a particular point in time (Kissling et al. 2018). To standardize and validate temporal data, *bdc* contain a function to extract the collection year whenever possible from complete and legitimate date information (Fig. 1, Table 1 and S1). Records with dubious collection year (e.g., 10/10/12) as well as with illegitimate (e.g., 1450, 2050) or no collection date supplied (e.g., 0 and NA) are flagged and can be subsequently removed.

Empirical example: the Brazilian Flora

We demonstrated the *bdc*'s applicability in cleaning >30 million occurrence records for terrestrial plant species in Brazil. All functions of the package were used to assess the quality of Brazilian flora data (the workflow used can also be checked on the package website) and the R scripts used in the analyses are available in Appendix 3 (figshare.com/articles/thesis/Script_R/15156396). More specifically, we aimed to assess the impact of data-cleaning on species richness (i.e., number of species before and after both the taxonomic harmonization and the application of spatial and temporal filters) and on the spatial pattern of species richness. Brazil harbors ~38,680 plant species (angiosperm, gymnosperm, ferns and lycophytes, and bryophyte, Flora do Brasil 2020 under construction), whose records are distributed in several heterogeneous online databases, making it an ideal case study. We assembled records for terrestrial plant species occurring in Brazil that could be accessed via nine public, freely, and openly available online databases (Table S4).

Results

From ~31 million records included in the original databases, only ~2.7 million records were considered high-quality data after the data cleaning and validation processes, representing a reduction of nearly 91% of the initial dataset (Fig. 2, 3; Table S5). Without removing records lacking information on collecting date, 13% of the original database were considered fit-for-use (Fig. 2 and 3; Table S5). The number of records flagged in each data-cleaning test can be found in Table S5.

Overall, 59% of the initial records (Fig. 1, 2), most corresponding to records lacking georeferencing (43%) or occurring outside Brazil (13%; Table S5; Fig S1), were excluded after applying filters of the pre-filter module. Data amendments were also performed by using functions available in the pre-filter module. Country names of 11% of records were derived from valid coordinates, and others 9% had country names standardized (Table S6). Finally, 0.14% records with transposed coordinates were corrected (Table S6; Fig. S2). The taxonomic step flagged 1.9% of records with names linked to non-accepted scientific names (Table S5). The most common spatial issues flagged corresponded to duplicate records (26%) followed by records within urban areas (2.2%; Table S5; Fig. S3). Around 19% of records lacked information on collecting date and 56 records with an out-of-range year (e.g., record collected before the year 1600 or in the future, e.g., 2030; Table S5). Records with valid date information had collecting years spanning from 1600 to 2020; most of them were recorded between 1980 and 2016 (Fig. S4).

The data-cleaning on richness maps changes the number of species with valid scientific names after both, the taxonomic harmonization, and the application of spatial and temporal filters (Fig. 3). In the pre-filter module, 213,303 specimens were recognized. This number was reduced to 38,783 species after taxonomic harmonization and to 38,207 and 36,540 species after applying space and temporal filters. The application of space and temporal filters led to a loss of 576 and 1,667 species, respectively, which had all records flagged as suspect or erroneous. While most records flagged in the space filters occurred within urban areas, around country or province centroids, or presented imprecise coordinates, most records removed in the time module were recorded before the year 1970.

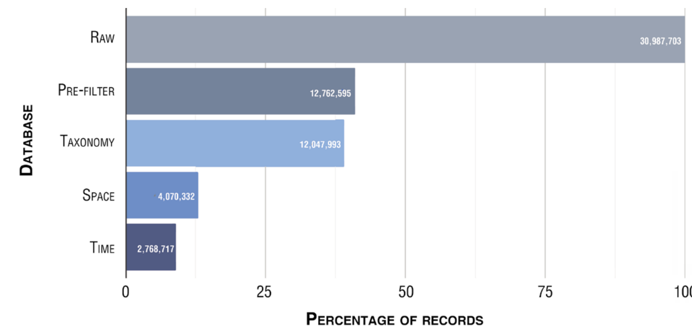


Fig. 2 Percentage of records of terrestrial plant species occurring in Brazil after applying functions of the *bdc* package to flag, correct, and eliminate suspect or erroneous records. The cleaning process comprises multiple tests for assessing data quality grouped in thematic modules namely pre-filter, taxonomy, space, and time. Absolute number of records is inside the bars.

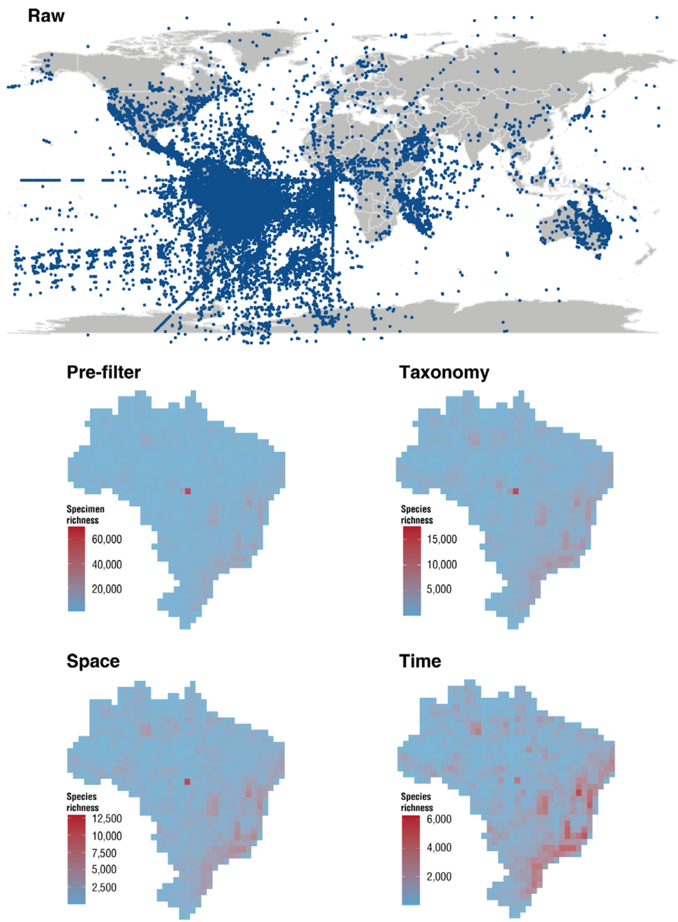


Fig. 3 Records of terrestrial plants occurring in Brazil in the raw database. Richness maps after the application of data-cleaning tests grouped in thematic modules named pre-filter, taxonomy, space, and time. Note legend has different scales.

Discussion

The *bdc* package is a toolkit that offers the means to convert raw data into high-quality information through a suite of core functions used to flag, clean, document, and enrich data quality. Such tools allow an automated and faster quality assessment of large datasets containing millions of records and thousands of species. The package contains flexible functions that can be applied to evaluate data quality. Nevertheless, there is no silver bullet to assess the quality of biodiversity data suitable for all purposes (Chapman, 2005a; Zizka et al., 2020). Biodiversity data have inherent limitations, and no hard rule exists to judge data quality needs to make data fit-for-use in all biodiversity applications; essentially, the data's adequacy depends mainly on the user's needs (Chapman, 2005b; Veiga et al., 2017). As demonstrated here the Brazilian flora as an example, the uncritical use of overly strict filters can result in loss of valid information; otherwise, errors can persist if no data cleaning is applied. In this sense, some researcher judgment will always be required to choose appropriate tools criteria to evaluate data quality and make data adequate for specific purposes (Zizka et al. 2019).

Perhaps the main novelty of the package is that scientific names of plant and animal species can be standardized by using taxonomic databases locally stored and through the application of exact and partial matching algorithms. Further, *bdc* processes are documented and auditable, making the management of biodiversity data more transparent and reproducible (a detailed comparison to available R packages can be found in Appendix 1). We hope the *bdc* package can facilitates and scales the data-cleaning process and catalyze improvements to allow the wise and efficient use of primary biodiversity data. We plan to add new functionalities in future versions of the package. In this sense, we encourage and welcome user's support to fixing issues and suggest new features.

Acknowledgments

We thank Eimear Nic Lughadha for her valuable comments and suggestions on this piece. We also thank researchers and citizens all over the world working to make knowledge on plants openly available online. BRR is supported by a CAPES scholarship. SJEV thanks the postdoctoral fellowships supported by the National Science Foundation (Award 1853697) and the Argentine National Council of Scientific and Technological Research received during this project. RL research is funded by CNPq (grant #306694/2018-2). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This paper is a contribution of the INCT in Ecology, Evolution and Biodiversity Conservation founded by MCTIC/CNPq (grant #465610/2014-5) and FAPEG (grant #201810267000023).

Data availability

The code of *bdc* R package is open and available at github.com/brunobrr/bdc. The scripts used in the analyses of the Brazilian flora are available in Appendix 3 (figshare.com/articles/thesis/Script_R/15156396). All data on the Brazilian flora were download from nine public, freely, and openly data sources (Table S4). An extensive documentation and tutorials on *bdc* package can be found at github.com/brunobrr/bdc.

References

Ball-Damerow JE, Brenskelle L, Barve N, Soltis PS, Sierwald P, Bieler R, LaFrance R, Ariño AH, Guralnick RP. 2019. Research applications of primary biodiversity databases in the digital age. PLoS ONE 14:1–26.

Bisby FA. 2000. The Quiet Revolution: Biodiversity Informatics and the Internet. Science 289:2309–2312. Available from https://www.sciencemag.org/lookup/doi/10.1126/science.289.5488.2309.

Boyle B et al. 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics 14:16. Available from https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-16.

Bystriakova N, Alves De Melo PH, Moat J, Lughadha EN, Monro AK. 2019. A Preliminary Evaluation of The Karst Flora of Brazil Using Collections Data. Scientific Reports 9:1–13. Springer US. Available from http://dx.doi.org/10.1038/s41598-019-53104-6.

Canhos DAL, Sousa-Baena MS, de Souza S, Maia LC, Stehmann JR, Canhos VP, De Giovanni R, Bonacelli MBM, Los W, Peterson AT. 2015. The Importance of Biodiversity E-infrastructures for Megadiverse Countries. PLOS Biology 13:e1002204. Available from https://dx.plos.org/10.1371/journal.pbio.1002204.

Carvalho G. 2017. flora: Tools for Interacting with the Brazilian Flora 2020. Available from http://www.github.com/gustavobio/flora.

Chapman AD. 2005a. Uses of Primary Species-Occurrence Data. Australian Biodiversity Information Services, GBIF papers 6:22–36. Available from http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=23464342&site=ehost-live&scope=site.

Chapman AD. 2005b. Principles of Data Quality. Global Biodiversity:58. Available from http://www2.gbif.org/DataQuality.pdf.

Chapman AD. 2005c. Principles and Methods of Data Cleaning. Report for the Global Biodiversity Information Facility:1–72.

Daru BH et al. 2018. Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytologist 217:939–955.

Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. Available from http://floradobrasil.jbrj.gov.br/ (accessed August 7, 2021).

Gadelha LMR et al. 2021. A survey of biodiversity informatics: Concepts, practices, and challenges. WIREs Data Mining and Knowledge Discovery 11:1–41. Available from https://onlinelibrary.wiley.com/doi/10.1002/widm.1394.

Graham C, Ferrier S, Huettman F, Moritz C, Peterson A. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution 19:497–503. Available from http://www.ncbi.nlm.nih.gov/pubmed/16701313 (accessed March 3, 2012).

Gueta T, Barve V, Nagarajah T, Agrawal A, Carmel Y. 2018. Introducing bdclean: a user friendly biodiversity data cleaning pipeline. Biodiversity Information Science and Standards 2:e25564.

Jin J, Yang J. 2020. BDcleaner: A workflow for cleaning taxonomic and geographic errors in occurrence data archived in biodiversity databases. Global Ecology and Conservation 21:e00852. Elsevier Ltd. Available from https://doi.org/10.1016/j.gecco.2019.e00852.

Kissling WD et al. 2018. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biological Reviews 93:600–625.

Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Albán J, Chilquillo E, Rønsted N, Antonelli A. 2015. Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? Global Ecology and Biogeography 24:973–984.

Mesibov R. 2013. A specialist's audit of aggregated occurrence records. ZooKeys 293:1–18. Available from http://zookeys.pensoft.net/articles.php?id=3928.

Mesibov R. 2018. An audit of some processing effects in aggregated occurrence records. ZooKeys 751:129–146. Available from https://zookeys.pensoft.net/articles.php?id=24791.

Meyer C, Weigelt P, Kreft H. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology letters 19:992–1006.

Newbold T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Progress in Physical Geography 34:3–22.

Nic Lughadha EM et al. 2019a. The use and misuse of herbarium specimens in evaluating plant extinction risks. Philosophical Transactions of the Royal Society B: Biological Sciences 374:20170402. Available from http://www.royalsocietypublishing.org/doi/10.1098/rstb.2017.0402.

Nic Lughadha EM, Graziele Staggemeier V, Vasconcelos TNC, Walker BE, Canteiro C, Lucas EJ. 2019b. Harnessing the potential of integrated systematics for conservation of taxonomically complex, megadiverse plant groups. Conservation Biology 33:511–522. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13289.

Norman KEA, Chamberlain S, Boettiger C. 2020. taxadb: A high performance local taxonomic database interface. Methods in Ecology and Evolution 11:1153–1159. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13440.

Peterson AT, Asase A, Canhos D, de Souza S, Wieczorek J. 2018. Data Leakage and Loss in Biodiversity Informatics. Biodiversity Data Journal 6. Available from https://bdj.pensoft.net/articles.php?id=26826.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria. Available from https://www.r-project.org/.

Robertson MP, Visser V, Hui C. 2016. Biogeo: An R package for assessing and improving data quality of occurrence record datasets. Ecography 39:394–401.

Seebens H et al. 2020. A workflow for standardising and integrating alien species distribution data. NeoBiota 59:39–59.

Soberón J, Peterson AT. 2004. Biodiversity informatics: managing and applying primary biodiversity data. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 359:689–98. Available from http://www.ncbi.nlm.nih.gov/pubmed/15253354.

Sousa-Baena MS, Garcia LC, Peterson AT. 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Diversity and Distributions 20:369–381. Available from http://doi.wiley.com/10.1111/ddi.12136.

Veiga AK, Saraiva AM, Chapman AD, Morris PJ, Gendreau C, Schigel D, Robertson TJ. 2017. A conceptual framework for quality assessment and management of biodiversity data. PLOS ONE 12:e0178731. Available from https://dx.plos.org/10.1371/journal.pone.0178731.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Available from http://ggplot2.org.

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D. 2012. Darwin Core: An Evolving Community--Developed Biodiversity Data Standard. PLoS ONE 7:e29715. Available from https://dx.plos.org/10.1371/journal.pone.0029715.

Zizka A et al. 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods in Ecology and Evolution 10:744–751. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13152.

Zizka A et al. 2020. No one-size-fits-all solution to clean GBIF. PeerJ 8:e9916. Available from https://doi.org/10.1101/2020.03.12.974543.

Supplementary material

Standardization and integration of heterogeneous datasets

The standardized database contains information on species taxonomy (e.g., scientific names), geolocation (e.g., latitude, longitude, coordinate precision, country, county, location), date of collection, and other relevant contextual information (e.g., basis of records, unique identification of the records, dataset name; Table S2). It is possible to insert other fields in the configuration table according to the user's needs. In such cases, we strongly recommend that the added terms follow the Darwin Core standards. A detailed explanation of the standardization and integration process and a description of each field are provided in Table S2. The standardization process follows the currently available methods for integrating biodiversity data in a standard format (Robertson et al., 2014; Seebens et al., 2020).

Taxonomic harmonization

Scientific names improperly formatted usually cannot be resolved by exact or fuzzy matching applications. To solve this issue, we developed the *bdc_clean_names*. This tool is composed of several name-checking routines that optimize the taxonomic queries by increasing the probability of finding matching names (Table 1). The *bdc_clean_names* is used to remove: i) family names pre-pended to species names, ii) qualifiers denoting the uncertain or provisional status of taxonomic identification (e.g., confer, species, affinis, among others, Table S3), and iii) infraspecific terms (e.g., variety [var.], subspecies [subsp], forma [f.], and their spelling variations; Sigovini et al. (2016)). The bdc_clean_names also includes applications to iv) standardize names, i.e., capitalize only the first letter of the genus name and remove extra whitespaces. This procedure is essential because the next step is case-sensitive; and v) parse names, i.e., separate author, date, annotations from taxon names. As a result of that process, a name parsing index ranging from 0 to 4 is returned, 1 denotes no problem detected and 4 severe problems detected; 0 indicates no parsing was performed, and probably the name is not interpretable. The name parser is executed using the *rgnparser* package, an implementation of Global Names - a scientific names parser originally written in Go programming language and encapsulated in an R package (Mozzherin et al. 2017; Chamberlain 2020). A detailed explanation of the name parsing process can be found in (Boyle et al. 2013). The output of name-cleaning routines is a table containing original and clean names and two columns indicating the quality of the name parsing process. The resulting table is saved in the "Output" folder.

The taxonomic standardization starts with the creation of a local database after downloading, extracting, and importing the taxonomic database informed by users using *taxadb* package (Norman et al. 2020). The remaining process is divided into two distinct phases according to the matching type to be undertaken. Firstly, the algorithm attempts to find an exact match for each original scientific name supplied using the function *filter_name* from *taxadb*. If an exact match cannot be found, names are returned as Not Available (NA). Also, it is possible that a scientific name matched multiple accepted names. In such cases, a helper function is used to flag records as "multipleAcceptedNames". Information on higher taxa (e.g., kingdom or phylum) can be used to disambiguate names linked to multiple accepted names. For example, the genus *"Casearia"* is present in both the Animalia and Plantae kingdoms. When handling names of Plantae, it would be helpful to get rid of names belonging to the Animalia to avoid flagging *"Caseria"* as having multiple accepted names. Following Norman et al. (2020), such cases are left to be fixed by the user. Users can choose to save a database containing names linked to multiple accepted names in the "Output" folder to assist this process.

Even after running name-checking routines, a scientific name can remain unresolved because of typos or spelling variants. In such cases, a fuzzy matching algorithm processes name-matching queries to find a potential matching candidate from the specified taxonomic database. Fuzzy matching identifies probable names (here identified as suggested names) for original names via a measure of orthographic similarity (i.e., distance). Orthographic distance is calculated by optimal string alignment (restricted Damerau-Levenshtein distance) that counts the number of deletions, insertions, substitutions, and adjacent characters' transpositions. It ranges from 0 to 1, being 1 indicative of a perfect match. A threshold distance, i.e., the lower value of match acceptable, can be informed by user (in the "suggest_distance" argument). If the distance of a candidate name is equal to or higher than the distance informed by user, the candidate name is returned as a suggested name. Otherwise, names are returned as NA. To reduce the number of cases when names are matched to a wrong name (i.e., false positives) or when a candidate name cannot be found (i.e., false negative), the match distance is set as 0.9 by default. This match distance is a conservative figure used to reduce the number of false-positive cases (Bachman et al. 2018). Only names very similar to the original names are matched without incurring any false matches. When multiple candidate names have an identical matching distance for the original name, the name with the lowest alphabetical sort order is presented as the best match (Boyle et al. 2013).

Two steps are performed before running the fuzzy matching. First, if supplied, information on higher taxon (e.g., kingdom, family) is used to filter the taxonomic database. This step removes ambiguity in matching by avoiding matching names from unrelated taxonomic ranks and decreases the number of names in the taxonomic database used to calculate the matching distance. Then, the taxonomic database is filtered according to a set of first letters of all input names. This process reduces the number of names in the taxonomic database that each original name should be compared (Carvalho 2017).

The quality of the name standardization processes can be accessed in the column "notes" from the output table after running the standardization process. The column "notes" contains assertions on the name standardization process based on (Carvalho (2017). The notes can be grouped into accepted names and those with a taxonomic issue or warning, needing further inspections. Accepted names can be returned as "accepted" (valid accepted name), "replaceSynonym" (a synonym replaced by an accepted name), "wasMisspelled" (original name was misspelled), "wasMisspelled | replaceSynonym" (misspelled synonym replaced by an accepted name), and "synonym" (original names is a synonym without accepted names in the database). Similarly, the following notes are used to flag taxonomic issues: "notFound" (no matching name found), "multipleAccepted" (name with multiple accepted names), "noAcceptedName" (no accepted name found), and ambiguous synonyms such as "heterotypic synonym", "homotypic synonym", and "pro-parte synonym". Ambiguous synonyms, names that have been published more than once describing different species, have more than one accepted name and cannot be resolved. Such cases are flagged and left to be determined by the user.

Data amendments

The pre-filter module also includes functions for data enhancement. The mismatch between informed country and coordinates can be the result of negative or transposed coordinates. This issue can be corrected if country names are supplied standardized, which is not usual when managing heterogeneous datasets. For solving this issue, we developed a tool to derive country names from valid coordinates. Next, country names are standardized using an exact match against a list of country names in several languages retrieved from Wikipedia. If any unmatched country names remain, a fuzzy matching algorithm can find potential candidates for each misspelled country name. Standardized country names are used to identify mismatches between country and coordinates. When a mismatch is detected, different coordinate transformations are made to correct the country and coordinates mismatch (Robertson et al., 2016). Verbatim coordinates are then replaced by the rectified ones in the returned database (a database containing verbatim and corrected coordinates is also created in the “Output” folder). Finally, coordinates can be derived from a detailed description of the locality associated with records in a process called retrospective georeferencing (Murphey et al. 2004; Wieczorek et al. 2004). A table containing records with missing coordinates but with potentially useful locality information can be saved in the “Output” folder.

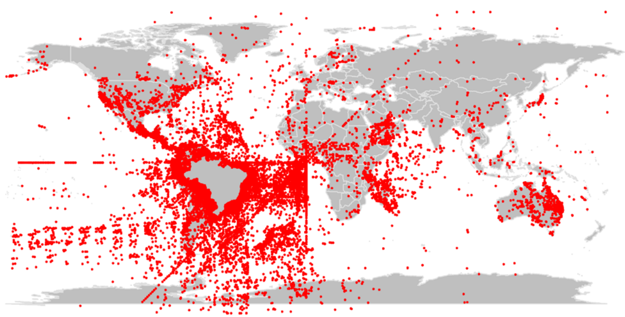


Fig. S1 Distribution of 3,924,464 records of land plants outside Brazil.

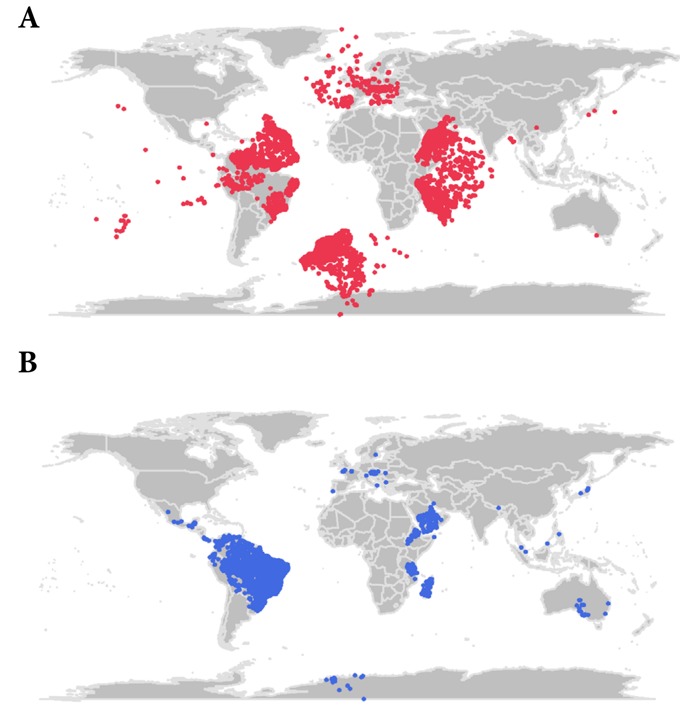


Fig. S2 Distribution of records of land plants supposed to occur in Brazil before (a) and after (b) correcting transposed geographic coordinates.

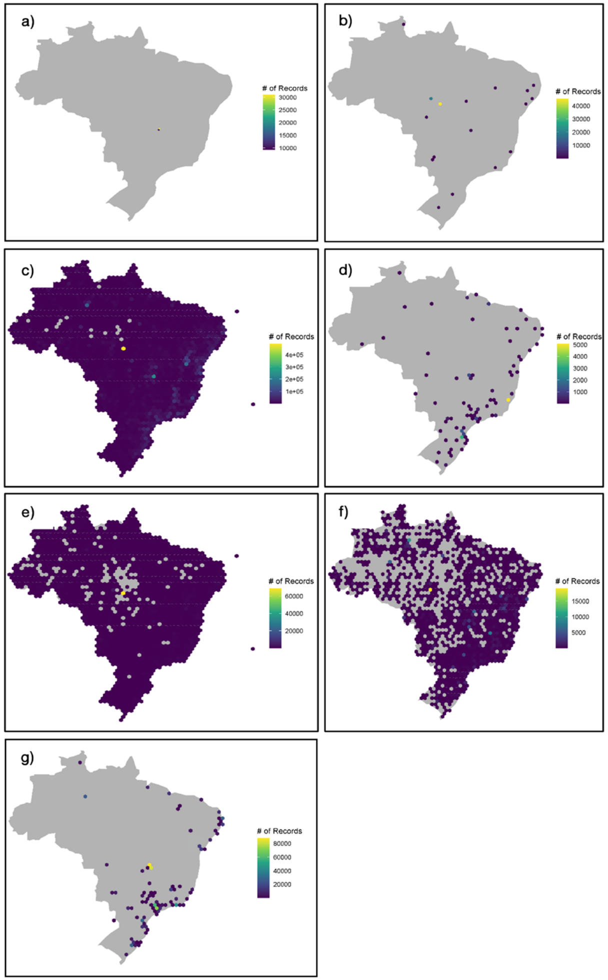


Fig. S3 Number of records of land plants occurring in Brazil with potentially erroneous or duplicated geographical coordinates. a) coordinates assigned to country capital; b) states centroids; c) duplicated; d) biodiversity institutions (e.g., herbarium), d) geographic outliers; e) imprecise coordinates, and f) urban areas. Values calculated after removing records from doubtful sources, outside Brazil, with empty coordinates, containing out-of-range coordinates (latitude >90 or < -90; longitude >180 or -180) or empty scientific names.

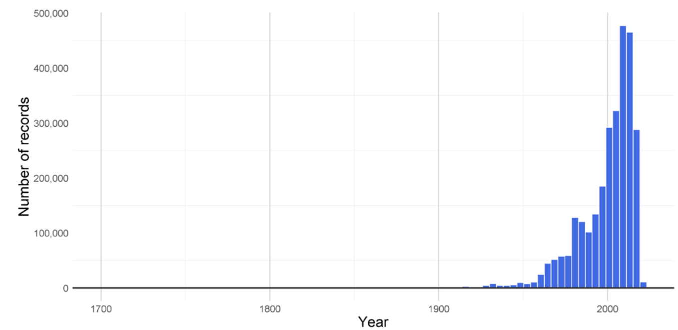


Fig. S4 Fig. S4 Number of records of land plants occurring in Brazil with valid information on the year of collection. Values calculated after removing records from doubtful sources, outside Brazil, with empty coordinates, containing out-of-range coordinates (latitude >90 or < -90; longitude >180 or -180) or empty scientific names.

Table S1 Description of data quality tests implemented in each thematic module of the bdc package. Tests are classified in the data quality framework (Veiga et al. 2017) as: *validation* (flags invalid or potentially problematic records), *completeness (flags records with non-interpretable or missing data),* and *amendment* (flags records amended to improve their quality). The information needed to run each test is based on metadata or external sources (geographical gazetteers). Darwin core formatted terms (i.e., columns headers) required to run each test are shown Information Elements field. Table S1 is available at "https://figshare.com/account/projects/120324/articles/15211257"

Table S2 Terms used to standardize and integrate different databases into a standard database containing information relevant for ecological and conservation studies. Terms were defined following Darwin Core Standards. Each field is classified in three categories according to its importance to run the data-cleaning tests: i) required (RQ), i.e., the minimum information necessary to run the tests, ii) recommended (RE), i.e., not mandatory but having important details on species records, and iii) additional (AD), i.e., information potentially useful for detailed data analyses.

Field	Category	Description of DarwinCore terms (if existing)	Type	Example of data
datasetName	RQ	A short name identifying the dataset	Character	GBIF
fileName	RQ	The path containing the dataset informed in "databaseName"	Character	C:/Users/Geizi/Projects/Flora/Input_files/GBIF.csv
scientificName	RQ	Name of the column containing scientific names with or without authorship information	Character	Myrcia acuminata
decimalLatitude	RQ	Name of the column containing information on geographic latitude in decimal degrees	Numeric	-6.370833
decimalLongitude	RQ	Name of the column containing information on geographic longitude in decimal degrees	Numeric	-3.25500
occurrenceID	RE	Name of the column presenting the unique identifiers in the original databases. If absent, NA	Numeric	1087566037

Field	Category	Description of DarwinCore terms (if existing)	Type	Example of data
basisOfRecord	RE	Name of the column in the original database containing the specific nature of the data record	Character	PreservedSpecimenHumanObservation
verbatimEventDate	RE	The name of the column containing information on verbatim date-time. Information only on collecting year can be provided. Recommended best practice is to use an encoding scheme, such as ISO 8601-1:201	date; numeric	2018-08-29T15:19 (3:19pm local time on 29 August 2018); 1970
country	RE	The name of the column containing information on the country or major administrative unit in which the data was recorded	Character	Brazil
stateProvince	RE	Name of the column presenting the name of the next smaller administrative region than country (state, province) where taxa were sampled or observed	Character	Rio de Janeiro
county	RE	Name of the column containing the full, unabbreviated name of the next smaller administrative region than “stateProvince” (county, shire, department, etc.) where taxon was sampled or observed	Character	Goiânia
locality	RE	Name of the column presenting the specific description of the place where taxon was sampled or observed	Character	Serra dos Pirineus
identifiedBy	AD	Name of the column presenting the list (concatenated and separated) of names of people, groups, or organizations who assigned the taxa to the subject	Character	M. Sobral
coordinateUncertaintyInMeters	AD	Name of the column presenting the horizontal distance (in meters) from the given “decimalLatitude” and “decimalLongitude”	Numeric	10
coordinatePrecision	AD	Name of the column containing the decimal representation of the precision of the coordinates given in the decimalLatitude and decimalLongitude	Numeric	0.0001
recordedBy	AD	Name of the column presenting the list (concatenated and separated) of names of people, groups, or organizations responsible for recording the original Occurrence	Character	Cervi, A.C

Table S3 Qualifiers applied to quantify taxonomic uncertainty or the provisional status of the taxonomic identification of specimens in the database of the Brazilian flora. See Sigovini et al. (2016) for further details on each qualifier.

Term	**Abbreviation**	**Meaning**
affinis	aff	Has affinity with. Indicates affinity of a specimen to a known species but they are not the same species.
Complex	complex	From the Latin complexus. Group of closely related species still waiting for a critical revision to clarify the taxa.
confer	cf, cfr, conf	Compare with. Provisional identification of a specimen that can be definitive after compared to reference material.
genus novum	gen nov, g nov	A new genus. This indicates that a species probably belongs to a new genus.
genus species	gen, g	Specimen not related to neither a known species nor genus.
Subspecies	ssp, subsp	It is used to ascertain uncertainty at the subspecies level.
Species	sp	The specimen has not been identified at the species level or is not related to any known species. Could also qualify species as non-indigenous or previously undescribed species.

Term	Abbreviation	Meaning
species (plural)	spp	Indicates the presence of more than one species not identified belonging to the same genus.
species incerta	?, inc	The question mark or the "inc" sign attached to a taxon indicates that the identification has been carried out correctly but is still uncertain. Further investigation is needed.
species inquirenda	inq	Species to be queried. Indicates a species of doubtful identity for which further investigation is needed.
species indeterminabilis (or species indeterminate)	indet, ind	Indeterminable or not determined species. Indicates specimen indeterminable beyond a particular taxonomic level.
species nova	nov, n	A new species. Indicates previously undescribed species.
species proxima	prox, nr	The specimen resembles a new known species but is not identical.
stetit	stet	Intentional absence of any qualifier. This indicates that further identification at a lower taxonomic level has not been attempted.

Table S4 Data sources used for compiling data on Brazilian flora. Due to each data source's particularities, different terms (i.e., filters) were used to download data manually.

Acronym	Data type	Term used	Website/citation	Link for download	Access
ATLANTIC	Occurrence and abundance		Ramos et al. (2019)	esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ecy.2541	2019/03/25
BIEN	Occurrence	Brazil	bien.nceas.ucsb.edu/bien	Package "BIEN"; function "BIEN_occurrence_country" 2019/03/25	2019/03/25
DRYFLOR	Occurrence	Brazil	www.dryflor.info	www.dryflor.info/data/datadownload	2020/10/08
GBIF	Occurrence	Tracheo-phyta; Brazil	www.gbif.org	api.gbif.org/v1/occurrence/download/re-quest/0069607-200613084148143.zip doi.org/10.15468/dl.v8yxab	2020/09/26
ICMBio	Occurrence	Plants	portaldabiodiversidade.icmbio.gov.br/portal	portaldabiodiversida-de.icmbio.gov.br/portal/search/downloadFile?fileName=/data/portal/tmp/portalbio_ex-port_09-04-2019-14-16-26.zip	2020/09/28
iDigBio	Occurrence	Plantae; Brazil; Brasil	www.idigbio.org	s.idigbio.org/idigbio-downloads/c7525055-fd26-4c13-9859-d-c298d57ff09.zip	2020/10/15
Neotropical TREE	Community (plot)	Brazil	www.neotroptree.info	www.neotroptree.info	2019/02/08
SiBBr	Occurrence	Plantae; Brazil; Brasil	sibbr.gov.br	Multiple files were downloaded because is it allowed to down-load 500,000 records per query	2020/09/01
*species*LINK	Occurrence	Plantae; Brazil	splink.cria.org.br	www.splink.org.br/downloads/speciesLink_all_97889_20200925235952.txt.zip	2020/09/25

Table S5 Absolute and relative number of suspect records flagged on each test of the *bdc* package used to assess data quality on Brazil's terrestrial plants. Individual records can be flagged by several tests; therefore, the sum of percentage from all tests can supersede the total percentage.

Modules	**Test function**	**Errors flagged**	**Total flagged**	**Fraction flagged (%)**
Pre-filter	bdc_scientificName_empty	Scientific name empty	970,214	3.13
	bdc_coordinates_empty	Coordinates empty	13,363,557	43.13
	bdc_coordinates_outOfRange	Coordinates out-of-range	211,983	0.68
	bdc_basisOfRrecords_notStandard	Records from uncertain source	23,313	0.08
	bdc_coordinates_country_inconsistent	Records outside the border of the reference country	3,924,464	12.66
Taxonomy	bdc_clean_names	Names containing terms denoting uncertain taxonomy	76,133	0.23
	bdc_query_names_taxadb	Names not found	536,199	1.73
		No accepted name	15,498	0.05
		Undefined status	22,763	0.07
		Multiple accepted names	8,802	0.02
Space	clean_coordinates*	Identical coordinates	0	0
		Zero latitude and longitude	3,411	0.01
		Coordinates around the country capital centroid	40,361	0.13
		Coordinates around the country or province centroids	68,428	0.22
		Coordinates within urban areas	751,936	2.43
		Geographical outliers	187,744	0.6
		Coordinates around the GBIF headquarters	0	0
		Coordinates around biodiversity institutions	18,656	0.06
		Duplicated records	8,206,238	26.48
Time	bdc_coordinates_precision	Coordinates imprecise	213,943	0.69
	bdc_eventDate_empty	Event date empty	5,842,295	18.85
	bdc_eventDate_outOfRange	Event date out-of-range	67,100	0.22

* Function of the *CoordinateCleaner* R package.

Table S6 Number of records amended to improve their quality in the pre--filter step of the *bdc* package.

Test function	Description	Total flagged	Fraction fragged (%)
bdc_country_from_coordinates	Extract of country names from valid coordinates	3,495,088	11.2
bdc_country_standardized	Standardization of country names against a list of country names in several languages retrieved from Wikipedia	2,791,217	9
bdc_coordinates_transposed	Correction of the mismatch between informed country and coordinates via several coordinates' transformations	44,417	0.14

References

Bachman SP, Nic Lughadha EM, Rivers MC. 2018. Quantifying progress toward a conservation assessment for all plants. Conservation Biology 0:1–9. Available from http://doi.wiley.com/10.1111/cobi.13071.

Boyle B et al. 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics 14:16. Available from https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-16.

Carvalho G. 2017. flora: Tools for Interacting with the Brazilian Flora 2020. Available from http://www.github.com/gustavobio/flora.

Chamberlain S. 2020. rgnparser: Parse Scientific Names. Available from https://github.com/ropensci/rgnparser (accessed March 15, 2021).

Mozzherin DY, Myltsev AA, Patterson DJ. 2017. "gnparser": a powerful parser for scientific names based on Parsing Expression Grammar. BMC Bioinformatics 18:279. Available from https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1663-3.

Murphey PC, Guralnick RP, Glaubitz R, Neufeld D, Ryan JA. 2004. Georeferencing of museum collections: A review of problems and automated tools, and the methodology developed by the Mountain and Plains Spatio-Temporal DatabaseInformatics Initiative (Mapstedi). PhyloInformatics 3:1–29.

Norman KEA, Chamberlain S, Boettiger C. 2020. taxadb: A high performance local taxonomic database interface. Methods in Ecology and Evolution 11:1153–1159. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13440.

Ramos FN et al. 2019. ATLANTIC EPIPHYTES: a data set of vascular and non vascular epiphyte plants and lichens from the Atlantic Forest. Ecology 100:e02541. Available from https://onlinelibrary.wiley.com/doi/abs/10.1002/ecy.2541.

Robertson MP, Visser V, Hui C. 2016. Biogeo: An R package for assessing and improving data quality of occurrence record datasets. Ecography 39:394–401.

Robertson T, Döring M, Guralnick R, Bloom D, Wieczorek J, Braak K, Otegui J, Russell L, Desmet P. 2014. The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. PLoS ONE 9:e102623. Available from https://dx.plos.org/10.1371/journal.pone.0102623.

CAPÍTULO III

Issues with species occurrence data and their impact on extinction risk assessments

Ribeiro, BR; Guidoni-Martins, K; Tessarolo, G; Velazco, SJE; Jardim, L; Bachman, SP; Loyola, R. **Issues with species occurrence data and their impact on extinction risk assessments.**
Em revisão (Biological Conservation)

Issues with species occurrence data and their impact on extinction risk assessments

Bruno R. Ribeiro[1], [*], Karlo Guidoni-Martins1, Geiziane Tessarolo[2], Lucas Jardim[3], Santigo José Elías Velazco[4],[5],[6], Steven P. Bachman[7], Rafael Loyola[8],[9]

Authors' affiliation

[1] Programa de Pós-graduação em Ecologia e Evolução, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[2] Universidade Estadual de Goiás, UEG, Campus de Ciências Exatas e Tecnológicas - CCET, Anápolis, Goiás, Brazil

[3] Instituto Nacional de Ciência e Tecnologia em Ecologia, Evolução e Conservação da Biodiversidade, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[4] Department of Botany and Plant Sciences, University of California - Riverside, Riverside, California, USA

[5] Instituto de Biología Subtropical, Universidad Nacional de Misiones - CONICET, Puerto Iguazú, Misiones, Argentina

[6] Programa de Pós-Graduação em Biodiversidade Neotropical, Universidade Federal da Integração Latino-Americana (UNILA), Foz do Iguaçu, Paraná, Brazil

[7] Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, U.K.

[8] Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[9] Fundação Brasileira para o Desenvolvimento Sustentável, Rio de Janeiro, Brazil

[*] Corresponding author's address

Programa de Pós-graduação em Ecologia e Evolução, Universidade Federal de Goiás, Avenida Esperança s/n, Campus Samambaia, CEP 74.690-900, Goiânia, Goiás, Brazil.

E-mail: ribeiro.brr@gmail.com

Abstract

Species extinction risk status is critical to support conservation actions. Species-occurrence data constitute the primary source of information underpinning risk assessments; however, the use of these data is hampered by uncertainties, errors, and gaps in their taxonomic, geographic, and temporal dimensions. Using a dataset of >30 million records of terrestrial plants occurring in Brazil, we quantified how incomplete data collation and errors in the taxonomic, geographic, and temporal dimensions of species-occurrence data translate into misclassifications of extinction risk. We conducted a preliminary risk assessment based on the extent of occurrence and assessed its accuracy in correctly predicting threatened and not threatened species. We found little impact of an incomplete data collation and issues in all three dimensions of species occurrence data on the overall accuracy and specificity of preliminary risk assessments. However, sensitivity rates (correct predicting threatened species) were strongly affected by incomplete data collation and spatial issues and marginally affected by temporal issues. Our results shed light on the crucial importance of higher quality data –derived from data cleaning– and a comprehensive collation of species-occurrence data for carrying out preliminary risk assessments.

Keywords: biodiversity data, fitness-for-use, data quality, GBIF, plants, rapid extinction risk assessment

Resumo

O risco de extinção de espécies é uma informação importante para apoiar a tomada de ações de conservação. Registros de ocorrência de espécies constituem a fonte primária de informações que sustentam avaliações de risco; entretanto, o uso desses dados é dificultado por incertezas, erros e lacunas presentes em suas dimensões taxonômica, geográfica e temporal. Usando um conjunto de dados de > 30 milhões de registros de plantas terrestres ocorrendo no Brasil, quantificamos como a coleta incompleta de registros e erros presentes em suas dimensões taxonômica, geográfica e temporal se traduzem em classificações incorretas de risco de extinção. Conduzimos uma avaliação preliminar de risco com base na informação sobre o tamanho da extensão da ocorrência das espécies e avaliamos sua precisão em identificar corretamente espécies ameaçadas e não ameaçadas. Encontramos pouco impacto de uma coleta incompleta de registros e problemas em todas suas três dimensões na precisão geral e especificidade das avaliações preliminares de risco de extinção. No entanto, as taxas de sensibilidade (previsão correta de espécies ameaçadas) das avaliações preliminares foram fortemente afetadas pela coleta incompleta de registros e por erros espaciais, e marginalmente afetadas por erros temporais. Nossos resultados destacam a importância de registros de ocorrência de espécies de alta qualidade - derivados da limpeza de dados - e uma ampla coleta dos registros disponíveis para realização de avaliações preliminares de risco de extinção.

Palavras-chave: dados de biodiversidade, adequação ao uso, qualidade de dados, GBIF, plantas, avaliação rápida de risco de extinção.

Introduction

Recent years have seen an explosion in the availability of species-occurrence data shared in online databases (Graham et al. 2004; Canhos et al. 2015). Such openly accessible biodiversity databases provide a vast and invaluable resource to document species distributions for many research uses (e.g., biogeographic studies, ecological applications, and conservation decision making; Ball-Damerow et al. 2019). However, the accumulation of large datasets may, without scrutiny, propagate errors that might influence outputs (Maldonado et al. 2015). An example of this is the evaluation of extinction risk as applied in systems such as the International Union for Conservation of Nature (IUCN) Red List of Threatened Species, hereafter the Red List (IUCN 2012).

The Red List is the most comprehensive and authoritative source of information on the global extinction risk of species (Rodrigues et al. 2006). It is used to assign species to categories of extinction risk based on a set of five quantitative criteria (symptoms of extinction risk) associated with population size, geographic distribution, and rates of decline of both (IUCN 2012). The extinction risk of most species is usually assessed by the application of the Red List criterion B (geographical range size), mainly because it is the most readily available data and IUCN metrics such as the extent of occurrence (EOO) and area of occupancy (AOO) can easily be calculated from geo-referenced records of species occurrence (Collen et al. 2016; IUCN 2019; Le Breton et al. 2019). The availability and quality of data are crucial for conducting species extinction risk assessment, particularly preliminary automated estimates based on criterion B (Nic Lughadha et al. 2019a; Zizka et al. 2020, 2021; Panter et al. 2020). Issues related to difficulties in standardizing and integrating data from different sources (e.g., GBIF and SpeciesLink; Kissling et al. 2018), discrepancies and errors in taxonomic and nomenclatural data (Nic Lughadha et al. 2019b), and errors and inaccuracies in geographical and temporal information of primary species-occurrence data (e.g., Meyer et al. 2016) may lead to under-or over-estimation of species range size. This bias, in turn, may trigger inappropriate conservation responses by misclassifying species extinction risk (Brummitt et al. 2008; Nic Lughadha et al. 2019a; Zizka et al. 2020; Panter et al. 2020).

On the one hand, unchecked taxonomy (e.g., misspelled names and synonyms) and the incomplete collations of species records (e.g., data compiled from few data sources), may lead to an overestimation of species risk by omitting records which, after being standardized or compiled, would increase the geographic range of taxa. Such issues may falsely classify as threatened a non-threatened species. On the other hand, records with suspect or wrong coordinates (e.g., within urban areas or geographical outliers) or the use of early collected or legacy data (which shed doubt on the continuing presence of species in a location given to, for example, habitat conversion), may overestimate the size of the species range, which could be wrongly qualified to a lower risk category (i.e., falsely classify a non-threatened a threatened species). In terms of allocating resources for conservation, wrongfully assessing a threatened species as non-threatened could mean vital resources are deprioritized. In contrast, to consider a species as threatened when it is not represents wasted effort. The potential impact of taxonomic, spatial, and temporal issues and an incomplete collation of records on the Red List assessment is summarized in Box 1.

Previously studies evaluating the influence of species occurrence records on preliminary risk assessment based on criterion B found contrasting results. Panter et al (2020) findings showed a stronger influence of spatial errors on sensitivity rates (i.e., correctly predicting threatened categories) than on specificity rates (i.e., correctly predicting non-threatened categories). Such impact varied depending on the data sources (e.g., GBIF vs. BIEN). Other researches, however, found that preliminary risk

assessment is relatively robust to the presence of records with spatial and taxonomic issues (Nic Lughadha et al. 2019b; Zizka et al. 2020, 2021), mainly due to significant thresholds of criteria used to assessing species' extinction risk. Hence, a comprehensive assessment to disentangle the impact of issues related to all biodiversity dimensions (taxonomic, spatial, and temporal) and an incomplete data collation on risk assessment is yet to be examined.

Using a dataset of >30 million records of terrestrial plants occurring in Brazil, we quantified how incomplete data collation and errors in the taxonomic, spatial, and temporal dimensions of species-occurrence data translate into misclassifications of extinction risk. Aggregating large amounts of data from heterogeneous sources across Brazil, a megadiverse country, will highlight challenges with handling and processing these data. Specifically, we ask the following questions: 1) How similar are online databases regarding species-occurrence records they share? 2) What is the impact of gaps generated by an incomplete collation of occurrence data – e.g., use of data from a single database (e.g., GBIF) – on estimates of species ranges and threat status? 3) What is the proportion of species risk assessments that are potentially over-or -underestimated due to inaccuracies and errors in taxonomic, geographical, and temporal dimensions of biodiversity data? 4) Which of these issues contributes most to misclassifying species extinction risk?

Materials and Methods

Data compilation and cleaning

The dataset underlying this study's analyses was compiled in a previous study (Ribeiro et al. under review). The dataset includes >30 million records for terrestrial plant species in Brazil accessed via ten public, freely, and openly available online databases. The compilation included data from eight biodiversity repositories: Botanical Information and Ecological Network version 4.1 (BIEN, bien.nceas.ucsb.edu/bien); Global Biodiversity Information Facility (www.gbif.org), Integrated Digitized Biocollections (iDigBio, www.idigbio.org/), speciesLink network (SpeciesLink, splink.cria.org.br), Brazilian Biodiversity Information System (SiBBr, www.sibbr.gov.br), Tree flora of the Neotropical Region (NEOTROPTREE, www.neotroptree.info), The Latin American Seasonally Dry Tropical Forest Floristic Network (www.dryflor.info), Brazilian Biodiversity Portal (ICMBio, portaldabiodiversidade.icmbio.gov.br/portal). We also included Atlantic forest epiphytes (Ramos et al. 2019), a scientific paper collating grey literature data.

We then used Biodiversity Data Cleaner (bdc) package for assessing the quality of species-occurrence data, considering their taxonomic, spatial, and temporal dimensions (Ribeiro et al. under review; more details can be found

in brunobrr.github.io/bdc). The package contains functions to clean and assess the quality of biodiversity data grouped in the following thematic modules: 1) Standardization and integration of different databases in a standardized format 2) Pre-filter: flagging and removal of invalid or non-interpretable information 3) Taxonomy: cleaning, parsing, and standardizing scientific names; 4) Space: flagging of erroneous, suspect, and low-precision geographic coordinates; and 5) Time: flagging and, whenever possible, correcting inconsistent collection dates.

Each module contains a series of functions (i.e., tests) to assert data quality. By executing each test, original data are retained the result is appended in a different field as TRUE (accurate records) or FALSE (records flagged as erroneous or suspect). We created a "raw" dataset after excluding records flagged as incorrect in the "pre-filter" step, including records missing coordinates or species names, in the ocean, with out-of-range coordinates, outside Brazil, and from distrustful sources (e.g., from drawing, photographs, among others). These records are commonly not fit for the Red List assessment without prior amendments. We also excluded records of non-native species (cultivated and naturalized),algae or fungi species.

We used the "raw" dataset to generate four databases of species occurrence with different levels of data curation (i.e., three “uncleaned” and one “clean” dataset; Panter et al. (2020)). In each “uncleaned” dataset, all issues were corrected except the problem to be tested. Thus, the “taxonomic” dataset contains only taxonomic issues, i.e., synonyms, nomenclatural variants, and misspelled names. The “spatial” dataset contains only geographical issues (e.g., records assigned to country capital, in urban areas, with low-precision coordinates, geographical outliers, among others (Zizka et al. 2019). The “temporal” dataset contains only temporal issues, including records collected before 1970 or containing illegitimate information (e.g., collection date in the future). Occurrence data collected in the last 20-30 years are generally associated with a moderate to a high level of inaccuracies (Tessarolo et al. 2017); in comparison, records collected after the year 1970 are more likely to be geo-referenced using GPS, and therefore more accurate (Graham et al. 2004; Boitani et al. 2011). Finally, we also created a “clean,”, well-curated, and near-comprehensive dataset in which all issues were removed or corrected. The “taxonomic”, “spatial”, “temporal”, and “clean” datasets were used to perform the downstream analysis described below.

Database similarity

To answer the question about the similarity of online databases regarding the species-occurrence records they share (Question 1), we compared the proportion of redundant information shared between them. To do so, we built a similarity matrix with the number of unique records (i.e., those with equal name, latitude, and longitude data) shared between databases. Low similarity values indicate the uniqueness of a database, i.e., few records are shared with other databases.

We evaluated the similarity considering the “raw” and “clean” datasets separately. Before assessing the similarity between original databases, we removed records missing coordinates or scientific names and authority names and annotations from scientific names (but kept terms denoting taxonomic uncertain and intraspecific levels such as “cf.”, “var”). These procedures were necessary to avoid falsely considering records as duplicated. Since databases contain coordinates with different precision, we also evaluated the similarity after rounding the geographical coordinates to four decimals degrees. Our approach is based upon the amount of potentially duplicate records shared between databases since comprehensive and rich meta-data is needed to detect actual duplicate records, but such data is seldom available.

Accuracy of preliminary risk assessments

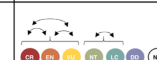
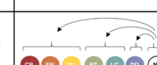
We used the R package *rCAT* (Moat 2017) to generate preliminary extinction risk assessments for each species based on Red List criteria B extent of occurrence (EOO), a range size metric commonly used for extinction risk assessment for plants due to scarcity of population data (Brummitt et al. 2008, 2015). The EOO, the minimum are encompassing all known records of species, was calculated as a minimum convex polygon (IUCN 2019). Based on EOO size estimates, *rCAT* classifies species in one of the following IUCN categories: critically endangered (CR; EOO < 10 Km2), endangered (EN; EOO > 10 and < 2,500 Km2), vulnerable (VU; EOO > 2,500 and < 10,000 Km2), near threatened (NT), or least concern (LC). Our assessment is preliminary because further subcriteria must be met to thoroughly justify a Red List rating (IUCN 2012).

To address question 2, we performed risk assessment using records from each database separately (e.g., only data from GBIF) to evaluate the impact of incomplete collation of occurrence data on risk assessment. To address questions 3 and 4, we used each "uncleaned" dataset separately to assess the effect of taxonomic, spatial, and temporal issues on preliminary extinction risk assessments. For example, to quantify the effects of taxonomic and nomenclatural errors, we used the "taxonomic" dataset to compare changes in the risk category due to non-standardized record names. Similarly, to quantify the impact of the spatial-related issues on risk assessment, we fixed all problems. Still, we kept records with one spatial issue (e.g., outliers) to quantify its relative impact on the risk category. This process was repeated to the nine spatial issues. Finally, to assess the impact of early collections or legacy data on risk assessments, we removed all records collected before 1970 or containing illegitimate information (e.g., collection date in the future). It is worth noting that changes in the risk category result from changes in the number of locality data available to each record due to spatial, taxonomic, or temporal issues not corrected or due to an incomplete data collation.

We evaluated our preliminary assessments' performance in correctly predicting the Red List categories of species with complete assessments carried out by the Brazilian Center for Flora Conservation (CNCFlora). We compared the performance of preliminary evaluations on both the detailed and binary levels (Zizka et al. 2021). At the detailed level, we compared the overall accuracy in assessing species in the same six categories as CNCFlora. At the binary level, we grouped species into two categories threatened (CR, EN, VU) and not threatened (NT, LC) and then evaluated the performance of preliminary assessments in terms of overall accuracy, sensitivity (correct prediction of threatened categories), and specificity (correct prediction of not threatened categories) (Nic Lughadha et al. 2019a). We repeated the accuracy tests with three additional reference datasets representing subsets of the CNCFlora data, including 1) species with up-to-date assessments carried out post 2018), 2) species with older assessments presented in the Red Book of the Brazilian Flora (Martinelli & Moraes 2013), and 3) species with > 15 occurrence records, a number suggested as the minimum for reliable automated assessments (Rivers et al. 2011).

We used R (version 4.02; R Core Team, 2020) to perform all analyses and *ggplot2* (Wickham 2016) to create all figures, except the Sankey diagram, which was generated using *networkD3* package (Allaire et al. 2017).

Box 1. Scenarios on how an incomplete data collation and errors, gaps, and inaccuracies in the taxonomic, spatial, and temporal information can influence preliminary Red List assessments. Incomplete data collation or issues in species-occurrence data may alter the number of records available to perform Red List assessment, which can change the estimates of Extent of Occurrence (a metric of range size used to assess species under the IUCN's criterion B). Minor issues in species-occurrence records may not change EOO estimates and species' threat category or lead to changes within categories (e.g., a threatened species continues to qualify as threatened but under a different category). On the other hand, insufficient data collation of major issues in species-occurrence records may lead to changes in species status (e.g., consider a threatened species as not threatened). Finally, certain species can only be assessed if new data is collated. This may occur, for example, when data from only one data aggregator (e.g., GBIF) is used in Red List assessments. According to the IUCN, species are classified as threatened (critically endangered [CR], endangered [EN], and vulnerable [VU]), not threatened (near threatened [NT] and least concern [LC]), data deficient (DD), or not evaluated (NE).

Scenario	Examples	Rationality	Implications
No change		• No errors found; • Minor errors in taxonomy, coordinates, or date **Errors causing minor changes in EOO**	Change in range size is not translated in changes in the threat category
Changes within categories		• Minor errors in coordinates or date of collection; • Taxonomy potentially outdated; • Enough data collation **Errors lead to changes in category but not between conservation status**	Change in range size; species category are up or down-graded. Implication for conservation priority
Changes the status		• Major errors and coordinates or date of collection; • Outdated taxonomy; • Insufficient data collation **Errors lead to changes in species conservation status**	Consider a threatened species as not threatened. The opposite is also true. In terms of conservation, the former is the worst
Not evaluated vs Evaluated		• New data collation **Assessment is only possible with new data collation**	More species assessed

Results

From > 30 million records presented in the original databases, only ~3.9 million records (13%) passed in all data-cleaning tests. The number of records detected and removed in each module of the data-cleaning workflow can be found in Table S1. As expected, the presence of taxonomic, spatial, and temporal change the number of both occurrence and species available for carrying out preliminary assessments. Without harmonizing species names (the "taxonomic" dataset), 63,112 specimens were recognized. After the taxonomic harmonization, the number of species was reduced to 38,690 in the "spatial" dataset and 38,207 in the "temporal" dataset. The "clean" dataset contains information from 37,519 species (Table S2). We found a similar pattern of species reduction in species' availability for carrying out a preliminary assessment after cross-referencing species from our dataset and CNCFlora data (Table S2).

Similarity between databases

Overall, we found a higher similarity of data shared between large data aggregators than with regional and group-specific databases (Fig. 1). Regarding the proportion of species-occurrence records shared between "raw" databases, we found higher similarity levels (varying from 75 to 95% of similarity) between GBIF, BIEN, iDigBio, and SpeciesLink databases after rounding coordinates to four decimals degrees (Fig. 1 and S1). Rounding coordinates to four decimals degrees and removing suspect or erroneous records (i.e., the similarity between "clean" databases) increased similarity levels (80-98%; Figs. S2-S3). SiBBr and NEOTROPTREE databases presented many exclusive records not shared with other databases. Considering the uniqueness of each database weighted by the number of records each database, databases containing fewer records (e.g., At_EPIPHYTES, DRYFLOR, and NEOTROPTREE) showed the highest number of unique records (Fig. 1). Most data from the DRYFLOR are available in NEOTROPTREE (but only after rounding coordinates), and species-occurrence data from both are not shared with other databases. Finally, around 25% of all records available in the nine databases can be obtained using data only from GBIF or SpeciesLink databases (Fig. 1).

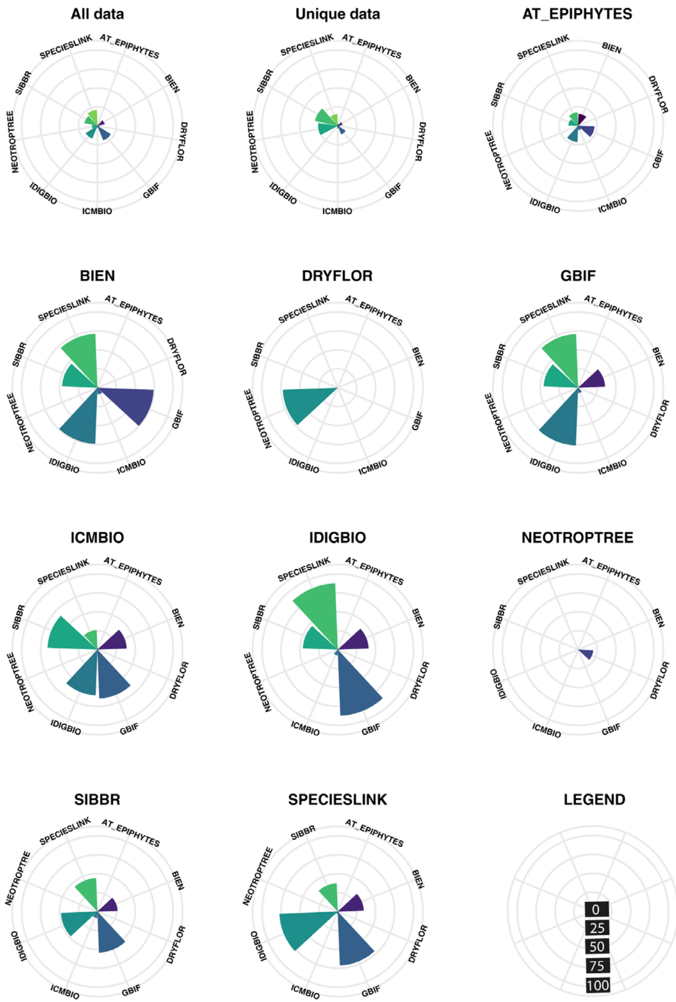


Fig. 1 Proportion of original data (n = 16,778,645 records) on land plants occurring in Brazil shared between nine heterogeneous online databases. The similarity among databases was measured through the number of unique records (i.e., with equal name, latitude, and longitude) shared between databases. For performing such comparisons, coordinates were rounded to four decimal degrees. The proportion of records and the number of records exclusive of each database are shown in figures "All data" and "Unique data".

Incomplete collation of occurrence data

We found a strong impact of an incomplete collation of occurrence data on the performance of preliminary risk assessments in correctly classifying species as threatened and not threatened. From the 5,524 species presented in the "clean" database (representing 94% of species officially assessed by CNCFlora), GBIF, BIEN, and SiBBr contained data for assessing the extinction risk of 80% (n = 4,625), 71% (n = 4,153), and 70% (n = 4,115) of species, respectively (Fig. 2, Table S3). The overall binary accuracy of preliminary assessments performed using data from only a single database was moderated and ranged from 61% (iDigBio) to 73% (DRYFLOR). Interesting, while the sensitivity rates were often low (13 to 48%), specificity rates ranged from 84% (ICMBio) to 94% (DRYFLOR), highlighting that data from single databases can be enough to estimate species' range and accurately identify not threatened species (Table S3). In this sense, the scenario "Changes the status" shown in Figure 2 highlights the wrong classification of threatened species mainly as data deficient.

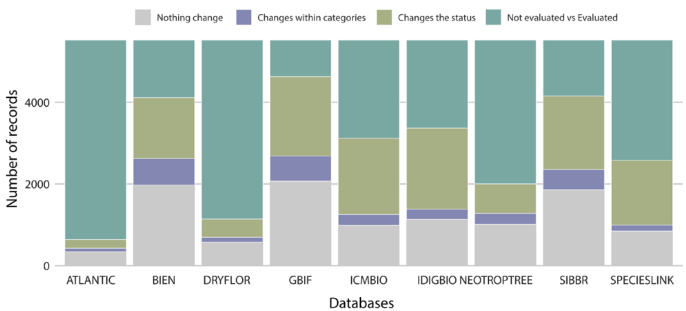


Fig. 2 Impact of an incomplete collation of species-occurrence records (i.e., the use of records from only a single database) on the performance of preliminary species extinction risk assessments. Changes in species category were evaluated by comparing the official category against the category derived from initial assessments using a single database after removing erroneous or suspect records. The impact of an incomplete records collation on the Red List assessment is summarized in four scenarios. *Nothing change:* no change in species risk category; *Change within categories:* changes in risk category occur within threat (CR, EN, VU) and not threat (NT and LC) status; *Change the status:* a threatened species becomes not threatened and vice-versa; Not evaluated vs. Evaluated: extinction risk of species cannot be evaluated because an incomplete collation of records.

Taxonomic, spatial, and temporal issues

The accuracy of preliminary risk assessments in correctly classifying species as threatened and not threatened are marginally affected by unchecked taxonomy, errors in geo--referenced information, and the use of early collected or legacy data (Fig. 3a, b). Accuracy was the highest (75%) by performing a preliminary assessment with a "clean" or "taxonomic" dataset and commonly more affected by spatial issues (66%; Fig. 4). Specificity rates were often high, ranging from 89% ("taxonomy" dataset) and to 92% ("spatial" dataset; Fig. 4). Sensitivity was the lowest due to the presence of spatial errors (38%), followed by temporal issues (48%), and no affected by the presence of taxonomic (61.5%) issues when compared with sensitivity (59.8%) after the removal of suspected or erroneous records (Fig. 4).

Using records with non-standardized taxonomy in risk assessments tended to overestimate the number of data deficient species and slightly underestimate the number of least concern species. In contrast, the presence of old records tended to overestimate the number of least concern species (Fig. 5). Compared to species correctly classified as not and not threatened using the "clean" dataset, geographic outliers and records in urban areas were the spatial issues that most contributed to the misclassification of species category (Fig. S4).

As expected, the detailed accuracy (i.e., correctly placing species into six IUCN categories) was often low (51-53%) and more affected by spatial issues (Figure S5). Relative to the assessment carried out using different reference datasets, the accuracy (64-78%) and specificity (78-99%) were the highest by using as reference the dataset containing only species with more than 15 records with no issue and with spatial issues, respectively (Fig. S6). Overall, the sensitivity rate was the lowest when the preliminary assessment was carried out based on a dataset containing both species with more than 15 records and spatial issues (Fig. S6). The sensitivity of preliminary risk assessment in predicting the category of species assessed after 2018 ("newer assessments") had the highest sensitivity rates (69%) when based on datasets with no issue or only taxonomic issues (Fig. S6).

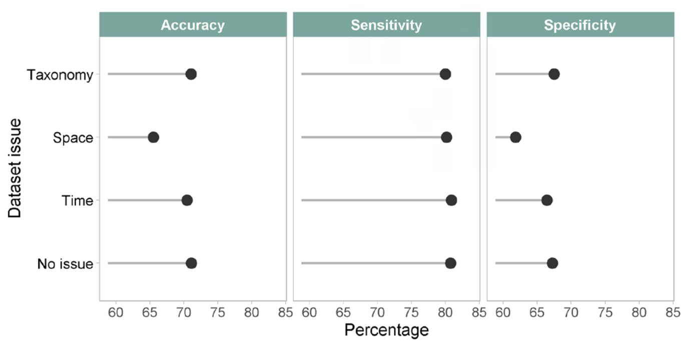


Fig. 3 The performance of preliminary extinction risk assessments carried out based on "uncleaned" datasets (containing taxonomic, spatial, and temporal issues) and on a "clean" dataset in correctly classifying species as threatened or not threatened compared with official assessments. The performance was measured in terms of overall accuracy, sensitivity (correct prediction of species as threatened), and specificity rates (correct prediction of species as not threatened).

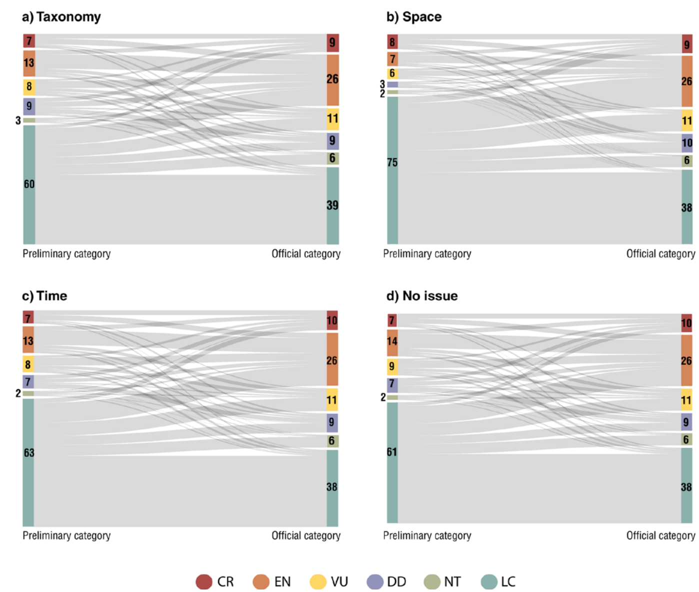


Fig. 4 Impact of taxonomic, spatial, and temporal issues on preliminary species extinction risk assessments. Changes in the preliminary assessment were evaluated by comparing species' official category against the category

derived from datasets with different data curation, including a) non-standardized taxonomy, b) spatial issues, c) temporal issues, d) no issue. Numbers within the bar represent the proportion of species in the category. Acronyms refer to IUCN Red List categories: CR, critically endangered; EN, endangered; VU, vulnerable; NT, near threatened; LC, least concern; DD, data deficient.

Discussion

We investigated the impact of both an incomplete data collation and issues in three dimensions of species-occurrence data—taxonomic, spatial, temporal—on preliminary species' risk assessments. We found little impact of an incomplete data collation and issues in all three dimensions of species occurrence data on the overall accuracy and specificity (correct predicting non-threatened species) of preliminary risk assessments. However, sensitivity rates (correct predicting threatened species) were strongly affected by incomplete data collation and spatial issues and marginally affected by temporal issues. Our results shed light on the crucial importance of higher quality data –derived from data cleaning– and a comprehensive collation of species-occurrence data for carrying out preliminary risk assessments.

Data from natural history collections have errors that could lead to misclassification of extinction risk and, consequently, undesirable conservation outputs (Maldonado et al. 2015; Panter et al. 2020). Our results showed that data cleaning is a fundamental process to improve the sensibility (correct prediction of threatened species) of preliminary risk assessment and that specificity (correct prediction of not threatened species) are marginally affected by both issues in occurrence records and incomplete collation of data (Panter et al. 2020). As the extent of occurrence (EOO) is purely a measure of range size, the presence of spatial issues (e.g., outliers and records in urban areas) often overestimates species range size, resulting in misclassification of many threatened species as not threatened. In contrast, the presence of spatial issues marginally improved specificity compared to specificity rates of preliminary assessment carried out using the "clean" dataset as reference. This may be due to the cleaning process that excluded records flagged as outliers or within urban areas.

Further, we found that all records of 1,171 species were flagged as suspect or erroneous, commonly including old or legacy records and those occurring in urban areas and country or province centroids. The removal of those records could lead to the loss of valuable data (Chapman 2005; Zizka et al. 2019). In such cases, further verification by the research or by an expert is highly recommended to determine whether the flagged records are erroneous to avoid information loss (Chapman 2005; Zizka et al. 2019).

A comprehensive collation of records is needed to improve the sensitivity of preliminary risk assessments (Rivers et al. 2011). In the era of Big Data, many biodiversity data today are increasingly digitalized and made available online through a wide range of heterogeneous databases. By assessing the commonalities of databases regarding the number of species-occurrence they share, we found an overall high similarity between large data aggregators (e.g., GBIF, BIEN, and SiBBr) and an often-lower similarity between such databases and regional and group-specific databases. A large number of records from small or local data sets that are not shared with large data aggregators highlights the importance of regional and local databases for estimating species' deriving a Red List category, especially for threatened species.

An incomplete collation of species records affects preliminary risk assessments' sensitivity than specificity rates (correct prediction of not threatened species). On the one hand, our results showed that preliminary assessments carried out based on data from single databases had moderate accuracy (61-72%) and high specificity (83-93%). These results highlight that even a reduced number of occurrence records are likely enough to represent the range of not threatened species, offering a helpful first step before investing in further collation of data (Rivers et al. 2011; Nic Lughadha et al. 2019a; Panter et al. 2020). Nevertheless, assessments based on records from single databases can only be performed by reducing species compared to the total species pool that could be assessed when compiling data from several sources.

On the other hand, our results showed that the sensitivity (correct prediction of threatened species) of species' preliminary risk assessment is strongly affected by an incomplete data collation and spatial issues and moderately affected by temporal issues. An insufficient collation of species-occurrence records is likely only to represent a fraction of species' range. Therefore, the preliminary risk assessment will underestimate the EOO, resulting in lower sensitivity rates. In contrast, the presence of spatial issues and older or legacy data often results in larger estimates of EOO, decreasing thus specificity rates.

Unchecked taxonomic and the removal of early collected or legacy records also led species to be classified as data deficient in a smaller proportion than an incomplete data collation and spatial issues. The negligible impact of unchecked taxonomy can indicate that synonyms and misspelled names had a small impact on the sensitivity of preliminary risk assessments or that databases had a relatively well-curated and updated taxonomy. Unchecked taxonomy most times are not enough to decrease the EOO to below the threshold to classifying a threatened species as not threatened (but see Nic Lughadha et al. 2019b). Outdated taxonomy, however, can result in species that cannot be assessed due to taxonomic remodeling (lumping and splitting; (Nic Lughadha et al. 2019b). Similarly, several species could not be assessed if older or legacy data were removed. These data constitute the best and unique information and support the Red List assessment (IUCN 2019).

Uncertainty in estimates of EOO and conservation status due to erroneous or suspect species-occurrence records should not be ignored; by the contrary, perhaps the best option relies on specifying the best estimate of EOO and risk category and document the range of plausible categories calculated using potentially problematic records (IUCN 2019). As shown by our results, drawbacks in taxonomic, spatial, and temporal dimensions of species-occurrence data can result in a range of possible estimates of EOO and conservation categories, particularly for rare species in which the exclusion of records could lead to extensive modification on the EOO estimates (Rivers et al. 2011; Zizka et al. 2020). When assessing the extinction risk of many species, it is challenging to determine whether a record assigned as suspicious is de facto erroneous. As a single category must be determined when assessing IUCN Red Lists status, it is

highly recommended to adopt a precautionary but realistic approach by informing the risk category from the best available information. (IUCN 2019).

Our results show that the accuracy of preliminary risk assessment was in the same range as previous studies (Nic Lughadha et al. 2019a; Panter et al. 2020; Zizka et al. 2021). Despite the need for preliminary risk assessment to meet further subcriteria to fully justify an IUCN Red List rating (Brummitt et al. 2008), these methods have shown high accuracy (89%) by correctly classifying as threatened or not threatened species of known extinction risk (Nic Lughadha et al. 2019a), and have been extensively used for assessing the extinction risk of species (Brummitt et al. 2015; Darrah et al. 2017; Bachman et al. 2018; Bystriakova et al. 2019; Nic Lughadha et al. 2019a). Preliminary risk assessments have been proven crucial to speed up the production of species extinction risk, mainly in megadiverse countries such as Brazil, where only 15% of species have been formally assessed so far (Martins et al. 2018).

There is a clear need for a drastic increase in the production rate of species risk assessment, especially in tropical countries, where species diversity and threats to plants are greatest. The robustness of preliminary risk, especially to identify non-threatened species, is encouraging, and existing tools are already available to speed up the generation and documentation of not threatened (i.e., least concern) species (Bachman et al. 2020). Our results showed that preliminary risk assessments are sensitive to incomplete data collation and spatial issues. Therefore, to achieve a more extensive and more balanced representation of species on the IUCN Red List, continuing efforts to the ensemble and generate high-quality primary biodiversity data in parallel with the increasing use and improvement of methods to facilitate and expedite the red listing process should be a priority.

Acknowledgments

BRR is supported by a CAPES scholarship. RL research is funded by CNPq (grant #306694/2018-2). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This paper is a contribution of the INCT in Ecology, Evolution and Biodiversity Conservation founded by MCTIC/CNPq (grant #465610/2014-5) and FAPEG (grant #201810267000023).

References

Allaire JJ, Gandrud C, Russell K, Yetman CJ. 2017. networkD3: D3 JavaScript Network Graphs from R. Available from https://cran.r-project.org/package=networkD3.

Bachman S, Walker B, Barrios S, Copeland A, Moat J. 2020. Rapid Least Concern: towards automating Red List assessments. Biodiversity Data Journal 8. Available from https://bdj.pensoft.net/article/47018/.

Bachman SP, Nic Lughadha EM, Rivers MC. 2018. Quantifying progress toward a conservation assessment for all plants. Conservation Biology 0:1–9. Available from http://doi.wiley.com/10.1111/cobi.13071.

Ball-Damerow JE, Brenskelle L, Barve N, Soltis PS, Sierwald P, Bieler R, LaFrance R, Ariño AH, Guralnick RP. 2019. Research applications of primary biodiversity databases in the digital age. PLoS ONE 14:1–26.

Boitani L, Maiorano L, Baisero D, Falcucci A, Visconti P, Rondinini C. 2011. What spatial data do we need to develop global mammal conservation strategies? Philosophical transactions of the Royal Society of London. Series B, Biological sciences 366:2623–32. Available from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3140738&tool=pmcentrez&rendertype=abstract (accessed March 22, 2012).

Brummitt N, Bachman SP, Moat J. 2008. Applications of the IUCN Red List: Towards a global barometer for plant diversity. Endangered Species Research 6:127–135.

Brummitt NA et al. 2015. Green plants in the red: A baseline global assessment for the IUCN Sampled Red List Index for Plants. PLoS ONE 10:1–22.

Bystriakova N, Alves De Melo PH, Moat J, Lughadha EN, Monro AK. 2019. A Preliminary Evaluation of The Karst Flora of Brazil Using Collections Data. Scientific Reports 9:1–13. Springer US. Available from http://dx.doi.org/10.1038/s41598-019-53104-6.

Canhos DAL, Sousa-Baena MS, de Souza S, Maia LC, Stehmann JR, Canhos VP, De Giovanni R, Bonacelli MBM, Los W, Peterson AT.

2015. The Importance of Biodiversity E-infrastructures for Megadiverse Countries. PLOS Biology 13:e1002204. Available from https://dx.plos.org/10.1371/journal.pbio.1002204.

Chapman AD. 2005. Principles of Data Quality. Global Biodiversity:58. Available from http://www2.gbif.org/DataQuality.pdf.

Collen B et al. 2016. Clarifying misconceptions of extinction risk assessment with the IUCN Red List. Biology Letters 12:20150843. Available from https://royalsocietypublishing.org/doi/10.1098/rsbl.2015.0843.

Darrah SE, Bland LM, Bachman SP, Clubbe CP, Trias-Blasi A. 2017. Using coarse-scale species distribution data to predict extinction risk in plants. Diversity and Distributions 23:435–447.

Graham C, Ferrier S, Huettman F, Moritz C, Peterson A. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution 19:497–503. Available from http://www.ncbi.nlm.nih.gov/pubmed/16701313 (accessed March 3, 2012).

IUCN. 2012. IUCN Red List Categories and Criteria: Version 3.1Second edi. Gland, Switzerland and Cambridge, UK.

IUCN. 2019. Guidelines for Using the IUCN Red List Categories and Criteria. Version 14. Prepared by the Standards and Petitions Committee. Available from http://intranet.iucn.org/webfiles/doc/SSC/RedList/RedListGuidelines.pdf.

Kissling WD et al. 2018. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biological Reviews 93:600–625.

Le Breton TD, Zimmer HC, Gallagher R V., Cox M, Allen S, Auld TD. 2019. Using IUCN criteria to perform rapid assessments of at-risk taxa. Biodiversity and Conservation. Springer Netherlands. Available from http://link.springer.com/10.1007/s10531-019-01697-9.

Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Albán J, Chilquillo E, Rønsted N, Antonelli A. 2015. Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? Global Ecology and Biogeography 24:973–984.

Martinelli G, Moraes MÁ. 2013. Livro Vermelho da Flora do Brasil1 st. Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro.

Martins E, Martinelli G, Loyola R. 2018. Brazilian efforts towards achieving a comprehensive extinction risk assessment for its known flora. Rodriguésia 69:1529–1537. Available from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2175-78602018000401529&lng=en&tlng=en.

Meyer C, Weigelt P, Kreft H. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology letters 19:992–1006.

Moat J. 2017. rCAT: Conservation Assessment Tools. Available from https://cran.r-project.org/package=rCAT.

Nic Lughadha EM et al. 2019a. The use and misuse of herbarium specimens in evaluating plant extinction risks. Philosophical Transactions of the Royal Society B: Biological Sciences 374:20170402. Available from http://www.royalsocietypublishing.org/doi/10.1098/rstb.2017.0402.

Nic Lughadha EM, Graziele Staggemeier V, Vasconcelos TNC, Walker BE, Canteiro C, Lucas EJ. 2019b. Harnessing the potential of integrated systematics for conservation of taxonomically complex, megadiverse plant groups. Conservation Biology 33:511–522. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13289.

Panter CT, Clegg RL, Moat J, Bachman SP, Klitgård BB, White RL. 2020. To clean or not to clean: Cleaning open source data improves extinction risk assessments for threatened plant species. Conservation Science and Practice 2. Available from https://onlinelibrary.wiley.com/doi/10.1111/csp2.311.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria. Available from https://www.r-project.org/.

Ramos FN et al. 2019. ATLANTIC EPIPHYTES: a data set of vascular and non vascular epiphyte plants and lichens from the Atlantic Forest. Ecology 100:e02541. Available from https://onlinelibrary.wiley.com/doi/abs/10.1002/ecy.2541.

Rivers MC, Taylor L, Brummitt NA, Meagher TR, Roberts DL, Lughadha EN. 2011. How many herbarium specimens are needed to detect threatened species? Biological Conservation 144:2541–2547. Elsevier Ltd. Available from http://dx.doi.org/10.1016/j.biocon.2011.07.014.

Rodrigues ASL, Pilgrim JD, Lamoreux JF, Hoffmann M, Brooks TM. 2006. The value of the IUCN Red List for conservation. Trends in ecology & evolution 21:71–6. Available from http://www.ncbi.nlm.nih.gov/pubmed/16701477 (accessed March 3, 2012).

Tessarolo G, Ladle R, Rangel T, Hortal J. 2017. Temporal degradation of data limits biodiversity research. Ecology and Evolution 7:6863–6870.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Available from http://ggplot2.org.

Zizka A et al. 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods in Ecology and Evolution 10:744–751. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13152.

Zizka A et al. 2020. No one-size-fits-all solution to clean GBIF. PeerJ 8:e9916. Available from https://doi.org/10.1101/2020.03.12.974543.

Zizka A, Silvestro D, Vitt P, Knight TM. 2021. Automated conservation assessment of the orchid family with deep learning. Conservation Biology 35:897–908. Available from https://onlinelibrary.wiley.com/doi/10.1111/cobi.13616.

Supplementary material

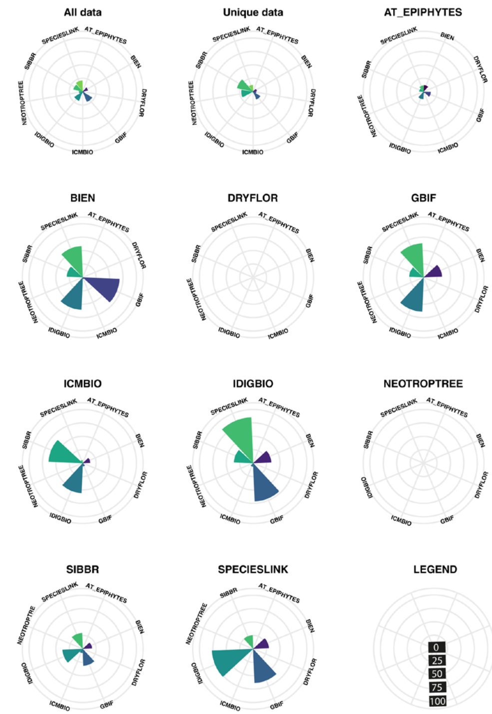


Fig. S1. Proportion of original data (n = 16,778,645 records) on land plants supposed to occur in Brazil shared among nine heterogeneous databases. The similarity among databases was measured through the number of unique records (i.e., with equal name, latitude, and longitude) shared between databases. Geographical coordinates were kept as in the original databases. The proportion of records and the number of records exclusive of each database are shown in figures "All data" and "Unique data", respectively.

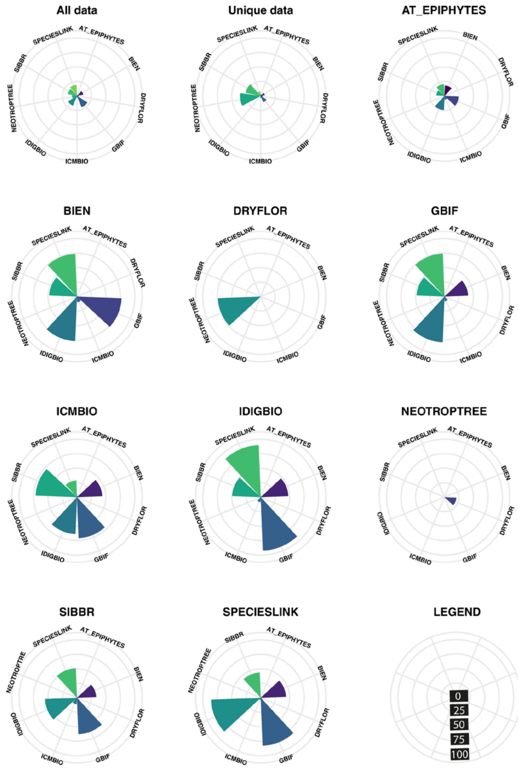


Fig. S2 Proportion of "clean" data (n = 3,887,565 records) on land plants supposed to occur in Brazil shared among nine heterogeneous databases. The similarity among databases was measured through the number of unique records (i.e., with equal name, latitude, and longitude) shared between databases. To do the comparison, geographical coordinates were rounded to four decimal degrees. The proportion of records and the number of records exclusive of each database are shown in figures "All data" and "Unique data", respectively.

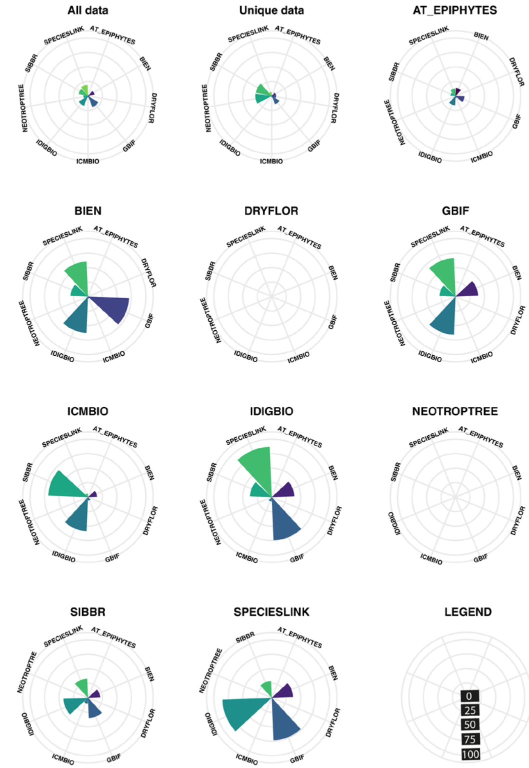


Fig. S3 Proportion "clean" data (n = 3,887,565 records) on land plants supposed to occur in Brazil shared among nine heterogeneous databases. The similarity among databases was measured through the number of unique records (i.e., with equal name, latitude, and longitude) shared between databases. Geographical coordinates were kept as in the original databases. The proportion of records and the number of records exclusive of each database are shown in figures "All data" and "Unique data", respectively.

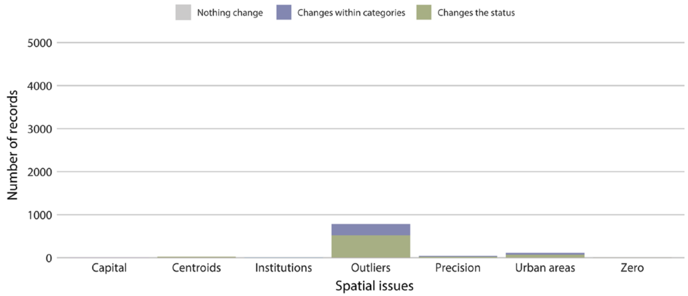


Fig S4. Impact spatial issues on preliminary species extinction risk assessments. Changes in species category were evaluated by comparing categories correctly classified as threatened or not threatened using as reference a "clean" dataset. The impact of spatial issues is summarized in three scenarios. *Nothing change*: no change in species risk category; *Change within categories*: changes in risk category occur within threat (CR, EN, VU) and not threat (NT and LC) status; *Change the status*: a threatened species becomes not threatened and vice-versa.

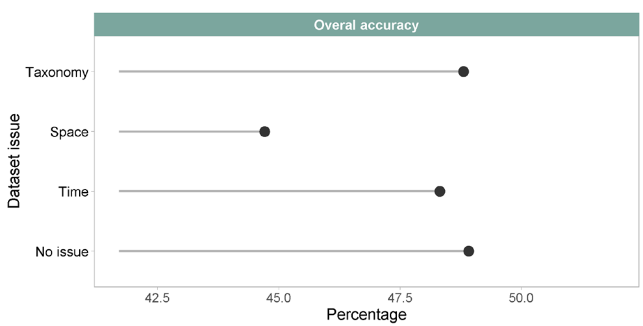


Fig S5. The accuracy of preliminary extinction risk assessments carried out based on "uncleaned" datasets (containing taxonomic, spatial, and temporal issues) and on a "clean" dataset in correctly placing species into six IUCN categories (critically endangered, endangered, vulnerable, near threatened, least concern, and data deficient). The accuracy was measured by comparing categories derived from preliminary assessment with categories from official Red List assessments carried out for plants of Brazil.

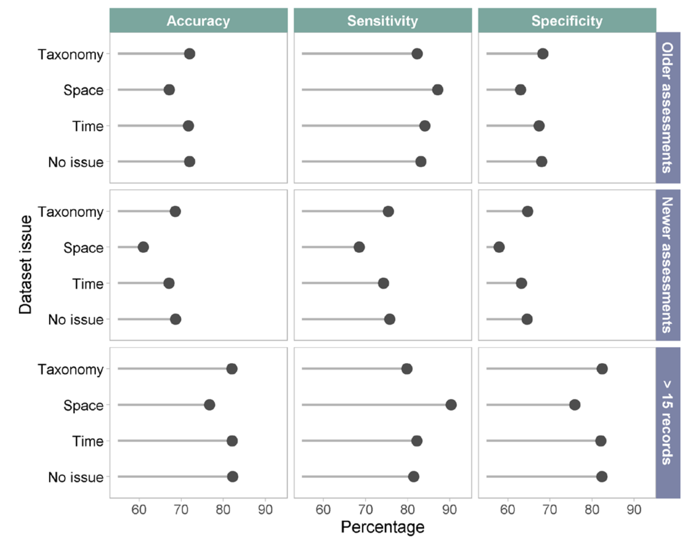


Fig S6. Impact of taxonomic, spatial, and temporal issues on preliminary species extinction risk assessments measured in terms of overall accuracy, sensitivity (correct prediction of species as threatened), and specificity (correct prediction of species as not threatened). Preliminary assessments were carried out using different reference datasets, including species assessed in the Brazilian Red Book published in 2013 ("Older assessments"), assessments carried out in 2018 ("Newer assessments"), and assessments performed considering only species with more than 15 records ("> 15 records").

Table S1 Absolute and relative number of suspect records flagged on each module of the data-cleaning workflow used to assess data quality of >30 million records of terrestrial plants supposed to occur in Brazil. Since individual records can be flagged by several tests, the sum of percentage from all tests can supersede the total percentage. Details about each test can be found in Ribeiro et al. (under review).

Workflow module	Errors flagged	Total flagged	Fraction flagged (%)
Pre-filter	Records lacking scientific name	970,214	3.13
	Records lacking latitude or longitude information	13,363,557	43.13
	Records with coordinates out-of-range	211,983	0.68
	Records from distrustful sources source	23,313	0.08
	Records outside Brazil (e.g., in other countries or in the ocean)	3,924,464	12.66
	Records of non-native (cultivated and naturalized), algae and fungi species.	66,534	0.2
Taxonomy	Names containing terms denoting uncertain taxonomy	76,133	0.23
	Records with no accepted names	583,262	1.87
Space	Identical coordinates	0	0
	Zero latitude and longitude	3,411	0.01
	Coordinates around the country capital centroid	40,361	0.13
	Coordinates around the country or province centroids	68,428	0.22
	Coordinates within urban areas	751,936	2.43
	Geographical outliers	187,744	0.6
	Coordinates around biodiversity institutions	18,656	0.06
	Duplicated records	8,206,238	26.48
	Coordinates imprecise	213,943	0.69
Time	Event date empty	5,842,295	18.85
	Event date out-of-range	67,100	0.22
	Event date older than 1970	601,327	1.93

Table S2 Number of occurrence records and taxa in all datasets used to assess the impact of taxonomic, spatial, and temporal issues on extinction risk assessments. The original dataset containing >30 million records of terrestrial plant occurring in Brazil was scrutinized to generate four datasets with different levels of data curation. In each dataset, all issues were corrected expect the issue to be tested. For example, the "taxonomic" database contains only taxonomic issues, including synonyms and misspelled names. All issues were corrected in the "clean" database. The number of taxa matching species officially assessed (n = 5,833) and used to carry out preliminary risk assessments are also shown.

Reference Dataset	# of occurrences	# of taxa	# of taxa used in preliminary assessment	# of records used in preliminary assessment
Taxonomic	3,973,371	63,112	4,902	522,897
Spatial	11,492,628	38,690	5,065	1,626,862
Temporal	4,070,313	38,207	5,568	604,474
Clean	3,887,565	37,519	5,025	574,168

Table S3 Impact of an incomplete collation of species-occurrence records (i.e., the use of records from only a single database) on the performance of preliminary species extinction risk assessments. Official species categories were compared to categories derived from preliminary assessments to generate three measures of performance: overall accuracy, sensitivity (correct prediction of species as threatened), and specificity rates (correct prediction of species as not threatened). Preliminary assessments were carried out using a "clean" database after the removal of suspect or erroneous records.

Database	Binary accuracy	Sensitivity	Specificity	Number of records
AT_EPIPHYTES	72.1	47.6	91.9	563
BIEN	68.3	47.3	88.5	3,806
DRYFLOR	72.1	25.4	93.3	1,095
GBIF	64.6	39.5	84.0	4,299
ICMBIO	61.0	26.3	83.7	2,970
IDIGBIO	59.7	19.5	83.4	3,178
NEOTROPTREE	71.1	38.1	92.7	1,857
SIBBR	66.1	38.4	85.4	3,903
SPECIESLINK	63.9	13.4	85.4	2.409

CAPÍTULO IV

The effectiveness of protected areas and indigenous lands in representing threatened plant species in Brazil

Ribeiro, BR; Martins, E; Martinelli, G; Loyola, R. **The effectiveness of protected areas and indigenous lands in representing threatened plant species in Brazil.** Rodriguésia, 69 (4) (2018), pp. 1539-1546, 10.1590/2175-7860201869404

The effectiveness of protected areas and indigenous lands in representing threatened plant species in Brazil

Bruno R. Ribeiro[1], [*], Eline Martins[2], Gustavo Martinelli[2], Rafael Loyola[1,3]

Authors' affiliation

[1] Universidade Federal de Goiás, Depto. Ecologia, Lab. Biogeografia da Conservação, Av. Esperança s/n, Campus Samambaia, 74690-900, Goiânia, GO, Brazil.

[2] Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Centro Nacional de Conservação da Flora, R. Pacheco Leão 915, 22460-030, Rio de Janeiro, RJ, Brazil

[3] Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

[*] Corresponding author's address

Programa de Pós-graduação em Ecologia e Evolução, Universidade Federal de Goiás, Avenida Esperança s/n, Campus Samambaia, CEP 74.690-900, Goiânia, Goiás, Brazil.

E-mail: ribeiro.brr@gmail.com

Abstract

Brazil is signatory of the Global Strategy for Plant Conservation (GSPC), which provides guidelines and directions to existing national policies. This strategy aims to halt the continuing loss of plant diversity through the achievement of 16 outcome-based targets set for 2020. One of these targets (target 7) states that at least 75% of known threatened plant species should be preserved in situ. Here, we assessed the effectiveness of the Brazilian current network of protected areas (PAs) and indigenous lands (ILs) in representing all known threatened plant species. We found that the number of species represented inside PAs and ILs varied according to data type. When using occurrence records, we found that 699 (33%) threatened plant species lie completely outside PAs (and/or ILs) and that 1,405 species (67%) have at least one record inside at least one PA (and/or IL). The number of species unrepresented decreased when we considered polygons of distribution. In this case, only 219 (10%) are supposedly unprotected. Although Brazil is almost reaching GSPC Target 7 in terms of absolute numbers, the government still needs to allocate resources for properly managing and improving the conservation status of its imperiled flora and expand the network of PAs.

Key words: Aichi Targets, conservation policy, gap analysis, GSPC, knowing-doing gap.

Resumo

O Brasil é signatário da Estratégia Global para Conservação de Plantas (GSPC), que fornece diretrizes e orientações para políticas públicas nacionais já existentes. Esta estratégia visa frear a perda contínua da diversidade de plantas por meio da implementação de 16 metas baseadas em resultados estabelecidas para 2020. Segundo uma dessas metas (meta 7), 75% das espécies de plantas ameaçadas conhecidas até então devem ser preservadas *in situ*. Neste trabalho, avaliamos a eficiência da atual rede brasileira de unidades de conservação (UCs) e terras indígenas (TIs) em representar todas as espécies de plantas ameaçadas. Descobrimos que o número de espécies presentes em UCs e TIs depende do tipo de dado utilizado. Ao utilizar registros de ocorrência, descobrimos que 699 (33%) espécies ameaçadas de plantas encontram-se completamente fora dessas áreas, e que 1.405 espécies (67%) possuem pelo menos um registro dentro de pelo menos uma UC ou TI. O número de espécies não representadas diminuiu quando consideramos polígonos de distribuição de espécies. Neste caso, apenas 219 (10%) das espécies encontram-se supostamente desprotegidas. Embora o Brasil esteja alcançando a meta 7 da GSPC em números absolutos, o governo ainda precisa alocar recursos para manejo adequado que leve a melhorias no estado de conservação da flora ameaçada e para expansão a rede de UCs.

Palavras-chave: Metas de Aichi, políticas públicas para conservação, análise de lacunas, GSPC, lacuna conhecimento-implementação.

Introduction

Plants are the backbone of life on Earth and are a source of food, timber, medicine, and a wide range of goods and services such as pure water, erosion control, and climate regulation; which are essential for human wellbeing (Kier et al. 2005; Corlett 2016). Nevertheless, flora conservation is still a big challenge to be faced in megadiverse countries.

Brazil, for example, harbors the richest flora on the planet, with at least 33,100 native terrestrial species described, being 53% endemics (BFG 2015; Flora do Brasil 2020 2017). This enormous biodiversity, however, is threatened by habitat loss and fragmentation, mining activities, infrastructure development, overexploitation of species of economic interest, invasive species, and climate change (Martinelli & Moraes 2013). As a consequence, 2,113 plant species now figure in the Brazilian official and most up-to-date list of threatened species (MMA 2014). Furthermore, much more species are supposed to be threatened, as just 17% of the Brazilian flora has been formally assessed so far by the National Centre for Flora Conservation (Martinelli & Moraes 2013; Martinelli et al. 2014; Martinelli et al. 2018).

As signatory of the Convention on Biological Diversity (CBD), Brazil has committed to protect 75% of its known threatened plant species inside the country's PAs.

This goal relates to target 7 of the Global Strategy for Plant Conservation (GSPC), a program agreed at the CBD meeting in Nagoya in 2010 (Convention on Biological Diversity 2016a). The GSPC aims to halt the continuing loss of plant diversity through the achievement of 16 outcome-based targets set for 2020 (Convention on Biological Diversity 2016a). So far, few countries have already assessed their progress towards Target 7. For example, 66% of South Africa and 44% of Spain threatened plant species have at least one population occurring in PAs (Convention on Biological Diversity 2016b; Muñoz-Rodríguez et al. 2016).

Among several conservation strategies designed to avoid species extinction, protecting species in their natural habitats is the simplest, cheapest, and most effective one (Loucks et al. 2008). Brazil has one of the largest PAs system in the world (MMA 2018; Pacheco et al. 2018). However, even with the expansion of its network of PAs in the last decades, PAs are not homogeneously distributed among Brazilian biomes (Pacheco et al. 2018). Further, information on species representation inside (or outside) PAs is barely available (but see Ferreira & Valdujo 2014; Oliveira et al. 2017).

Here, we assessed the effectiveness of the Brazilian current network of PAs in representing all known threatened plant species. We did a comprehensive gap analysis considering different types of species distribution data (occurrence records and polygons) to evaluate how close Brazil is to achieve GSPC's target 7.

Methods

Species and protected area data

Species records were compiled by CNCFlora (Centro Nacional de Conservação da Flora, National Center of Flora Conservation) using the following procedure: data on species records were retrieved from the Global Biodiversity Information Facility (www.gbif.org), SpeciesLink (splink.cria.org.br), Rio de Janeiro Botanic Garden herbarium (http://jabot. jbrj.gov.br), and botanic experts' databases. These data were cleaned and records without explicit geographic information were spatially projected using information from the herbaria vouchers, whenever possible.

Species distribution polygons were built by the CNCFlora using ArcGIS v1.0 (ESRI) following the precision of species occurrence records. All records had a geographic precision associated to nine classes: a) 0 to 250 m; b) 250 to 1,000 m; c) 1 to 5 km; d) 5 to 10 km; e) 10 to 50 km; f) 50 to 100 km; g) centroid of polygon; h) centroid of protected area; i) centroid of the municipality. The CNCFlora defines the likely distribution of a given species as the sum of area covered by these polygons (for details see also Martinelli & Moraes 2013 and Loyola et al. 2014a).

We obtained species distribution as both occurrence records and polygons for 2104 species listed in the Red Book of Brazilian Flora (Martinelli & Moraes 2013) from CNCFlora. CNCFlora is the Red List authority in Brazil (Martins et al. 2017) and this is the most complete and up-to-date dataset of Brazilian threatened flora distribution available. We obtained data on PAs and indigenous lands (ILs) from a range of sources (Tab. 1). We considered all classes of PAs, i.e., federal, state, and municipal, which include full protection and sustainable use PAs (IUCN categories I-III and IV-VII, respectively) and also the ILs (MMA 2016). Although not included in the Brazilian law that establishes the roles and categories of PAs (MMA 2000), ILs are natural areas that encompass nearly 13% of the Brazilian territory (Pacheco et al. 2018), playing a key role in preventing deforestation (Soares-Filho et al. 2010). We did not include in the analysis Permanent Preservation Areas (e.g., riverside and hilltops forest buffers) and Legal Reserves (i.e., part of private lands that must be set aside for conservation) because no spatial information on these areas was readily available.

Table 1 Source of data used in this study.

Data	Source
Species distribution data (records and polygons)	National Centre for Flora Conservation (www.cncflora.jbrj.gov.br)
Shapefile of Protected Areas (full protection and sustainable use)	Brazilian Ministry of the Environment (mapas.mma.gov.br/i3geo/datadownload.htm)
Shapefile of Indigenous Lands	National Indian Foundation (www.funai.gov.br/index.php/shape)
Shapefile of Brazilian biomes	Brazilian Ministry of the Environment (mapas.mma.gov.br/i3geo/datadownload.htm)

To evaluate the effectiveness of PAs in representing the threatened flora we superimpose species distribution data onto a map of Brazilian PAs. In the analysis, we considered a species as "represented" in a given PA if at least one record or any extent of its distribution polygon overlapped any PA; otherwise, the species was considered a "gap species". We used function over (package "sp") to overlap species records and PAs. To overlap species polygons and PAs, we created a grid of 2 km of resolution at the extent of each species polygon. Then, we use this grid to rasterizing species polygons and PAs which lied inside species polygons. All data sources and the R script used to run the analysis are available in the Appendix 1 (figshare.com/articles/dataset/Table_S1_xlsx_Appendix_1/7221971/1).

Results

Most records of threatened plant species were concentrated in the Atlantic rainforest and Caatinga biomes, whereas large areas in the Amazon remain poorly sampled or not sampled at all (Fig. 1). On the one hand, the Atlantic rainforest harbored the highest number of threatened plant species in Brazil (n = 1,507) although only ~11% of its extent is covered by PAs and ILs (Pacheco et al. 2018). On the other hand, the Brazilian Amazon had the largest coverage of PAs and ILs in Brazil (~53% of biome extent) but has only 85 threatened species (Fig. 1).

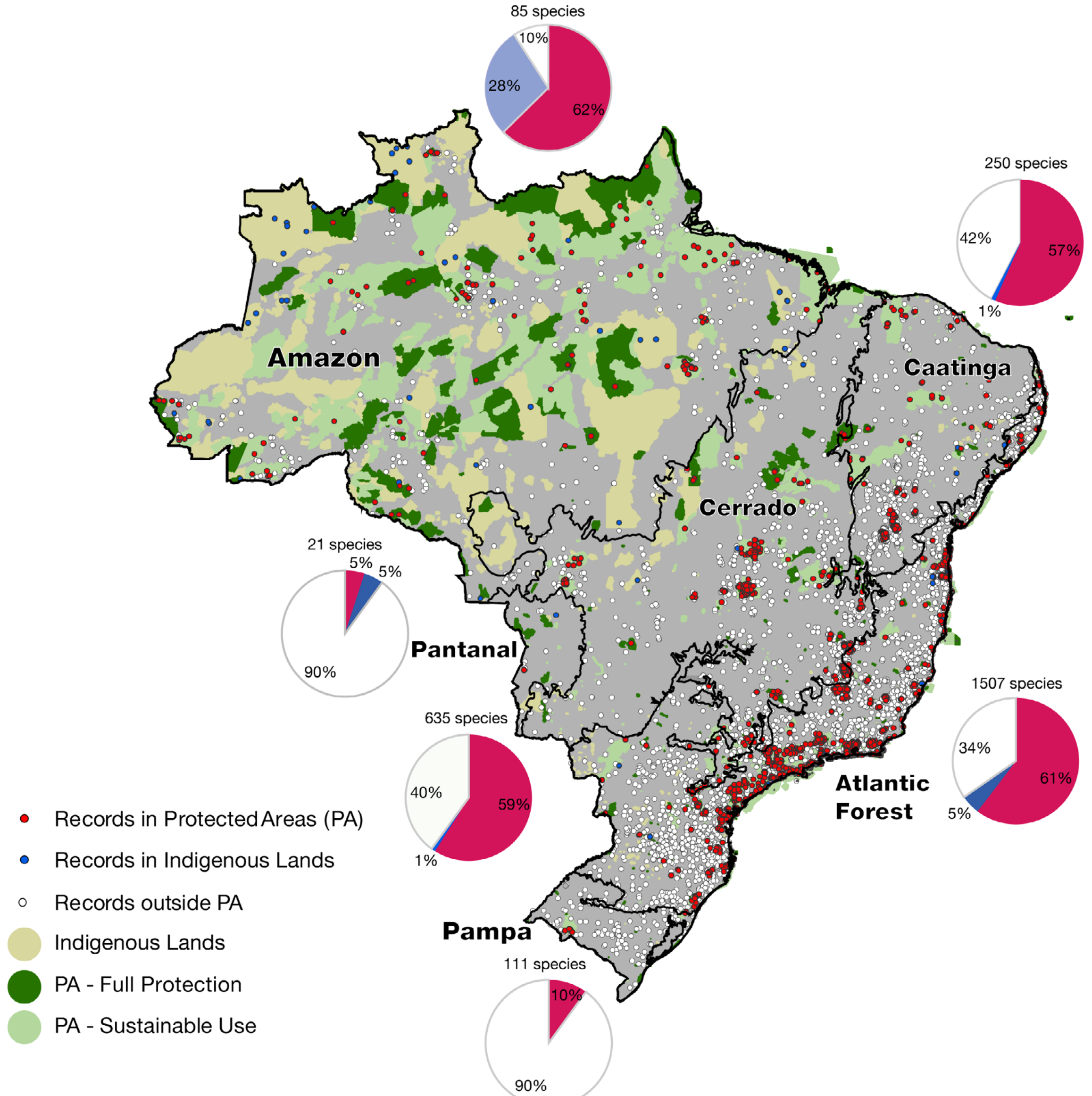

85 species
10%
28%
62%
250 species
42%
57%
1%
Amazon
Caatinga
Cerrado
21 species
5%
5%
90%
Pantanal
635 species
40%
59%
1%
1507 species
34%
61%
5%
Atlantic Forest
Pampa
111 species
10%
90%
Records in Protected Areas (PA)
Records in Indigenous Lands
Records outside PA
Indigenous Lands
PA - Full Protection
PA - Sustainable Use

Fig. 1 Distribution of Brazilian threatened plant species' records inside and outside protected areas (PAs) regarded as full protection, sustainable use PAs, and indigenous lands (ILs). The number and proportion of species representation inside each PA category and ILs are shown in the pie charts.

We found that the number of species occurring inside and outside PAs varied according to the type of distribution data used to do the analysis. When we used occurrence records, 699 species (33%) fell completely outside PAs or ILs (Fig. 2a). A total of 1,405 species (67%) had at least one record inside PAs and ILs; and 82% had up to 10 records inside PAs and ILs [Fig. 3; Tab. S1 (figshare.com/articles/dataset/Table_S1_xlsx_Appendix_1/7221971/1)].

When running the analysis with distribution polygons, we found that the number of gap species inside PAs or ILs reduced to 219 species (10%; Fig. 2b). Species fully covered (i.e., 100% of their distribution) by Brazilian PAs and ILs added to 29 species (1.3%); and 807 species (38%) had up to 20% of their range represented in these areas [Fig. 2b; Tab. S1 (figshare.com/articles/dataset/Table_S1_xlsx_Appendix_1/7221971/1)].

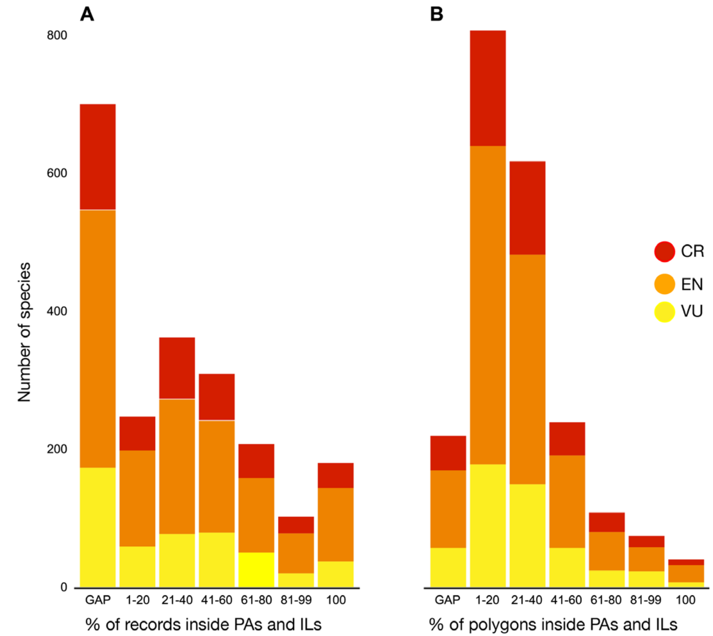


Fig. 2 Number of threatened plant species and their representation in the Brazilian network of protected areas and indigenous lands according to species' threat category and type of distribution data – a. species' records; b. distribution polygons. Species totally outside protected areas are shown as GAP. CR = Critically Endangered); EN = Endangered; and VU = Vulnerable.

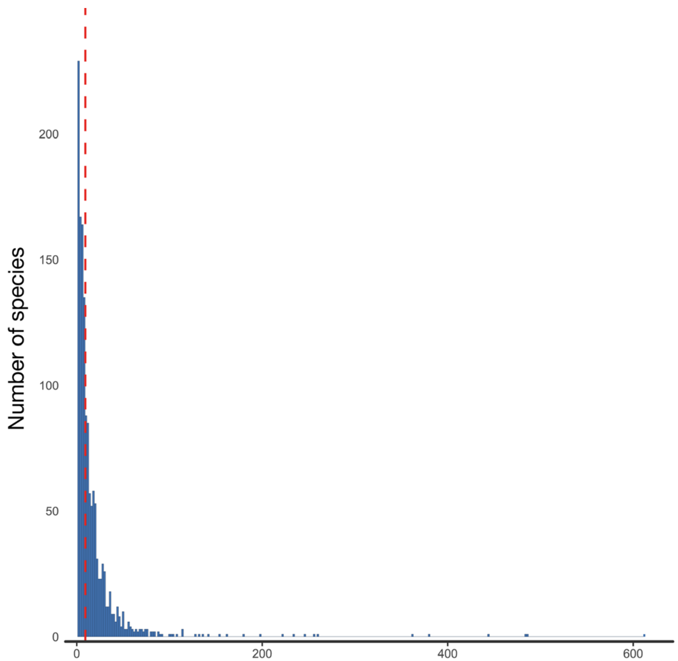


Fig. 3 Number of records of species inside protected areas (PAs) and indigenous lands (ILs). The red dashed line indicates that 82% species has less than 10 records inside PAs or ILs.

Species representation also varied between PAs and ILs. We found 81 (4%) threatened species occurring in ILs, 14 of them only occurred in these areas. In PAs, the number of species represented varied according to PA category: ca. 39% of species were found only inside full protection PAs, whereas 34% of species only occurred in sustainable use PAs, being one-third represented by Environmental Protection Areas, a category of PA in which a range of economic activities is allowed. Further, nearly 76% of species were found exclusively in terrestrial PAs and only 9% occurred in marine and coastal PAs; 15% of species were found in both.

On average, 55% of threatened species with at least one record inside PAs or ILs are currently listed as endangered (EN), 23% as vulnerable (VU), and 22% as critically endangered (CR) species. For gap species (i.e., those not represented in PAs or ILs), on average, 51%, 26% and 23% are EN, VU and CR, respectively (Fig. 2).

Discussion

How far is Brazil from achieving GSPC Target 7? Being pragmatic, we would state that Brazil is quite close to fulfilling its commitment, given that 67% of threatened flora is represented inside any given PA or IL. This figure is even higher when we analyzed data on species distribution (i.e., polygons), in which 90% of species occur in PAs or ILs. These results represent a great accomplishment for plant conservation in Brazil and we are optimistic that this target will be eventually achieved.

However, our results should be interpreted with caution given that the distribution of species records is heavily skewed towards the northeast and southeast Brazil (i.e., Atlantic Forest and Caatinga biomes), as also observed by (Schulman et al. 2007; Sousa-Baena et al. 2014a; Oliveira et al. 2017). Further, the occurrence of a species in a given PA or IL offers no guarantee to species' effective long-term conservation. Following the CBD (Convention on Biological Diversity 2016b), to be effective, *in situ* conservation must ensure viable populations of species in at least one PA. Therefore, a metric such as species' occurrence inside PAs could not be enough to ensure the conservation of threatened plant species given that 82% of them are represented in PAs by less than 10 records. Further, commission errors (*i.e.*, considering a species present in a given PA where it is absent) are also an inevitable consequence of using species polygon and can result in an overestimation of the representatives of species in PAs. Being aware of our current knowledge is a first step for supporting decision making regarding biodiversity conservation. There is a need to foster biodiversity inventories particularly in poorly sampled PAs (Sobral & Stehmann 2009; Martins et al. 2017; Loyola et al. 2018). To address this issue, more investments in PAs infrastructure, new surveys and inventories, trained personnel, management plans, institutional capacity, and public

engagement are urgently needed. Opening ILs to scientific expeditions is also paramount (Coelho et al. 2015).

We should also note that searching for an increase in the number of threatened species inside PAs or ILs might be the wrong focus (Heywood 2017); this number must actually reduce. The ultimate aim of GSPC is to halt the current and continuous loss of plant diversity by reducing species extinction risk and consequently reducing the number of threatened species (Heywood 2017). In situ conservation is indeed the first step required to avoiding habitat loss and to restore or sustain native plant populations (Convention on Biological Diversity 2016a; Heywood 2017). However, even inside PAs threats such as biological invasion (Nori et al. 2011; Loyola et al. 2012) habitat loss (Hansen et al. 2013), climate change (Lemes et al. 2013; Loyola et al. 2014b; Ferro et al. 2014), overexploitation (Castilho et al. 2017) still take place; and the very PAs could suffer from downgrading, downsizing and degazettement (Bernard et al. 2014; Oliver et al. 2016; Heywood 2017; Martins et al. 2017). Therefore, looking representation of threatened species inside PAs or ILs is not the big question; instead we should looking for how to guarantee the survival of these species once they are found inside these areas (Pressey et al. 2015).

We found that 33% of species records or 10% of species distribution polygons are fully outside any given PA or

IL. Gap species are the most in need of conservation actions owing to fast habitat loss, mainly in the Brazilian Cerrado, the most vulnerable savanna worldwide (Strassburg et al. 2017; Vieira et al. 2018). Conservation of gap species demands area-based conservation measures undertaken outside of and complementary to PAs (Heywood 2017) and ILs. For instance, the Brazilian Forest Code, if appropriately enforced, may offer a great opportunity for plant conservation in private properties since it requires that part of private area - ranging from 80% in the Amazon and 20% other regions of Brazil - should be set aside for conservation or restoration (Brancalion et al. 2016). Furthermore, priority areas for plant conservation have already been defined at the national (Loyola et al. 2014a) and regional (Loyola & Machado 2015a; Loyola et al. 2015b, Loyola et al. 2018) scale.

Beyond *in situ* conservation, Brazil and other countries also rely on species recovery plans as another tool for species conservation. These plans encompass habitat management and/or restoration, ex situ conservation, development of educational and outreach programs, environmental policy development, and continuing monitoring (Convention on Biological Diversity 2016; Heywood 2017). To date, 873 threatened plant species are included in species recovery plans in Brazil (Martins et al. 2014; Pougy et al. 2015a,b; Pougy et al. 2018). Surely, the main obstacles for improving species' threat status through the implementation of strategies set out in recovery plans are the inadequate funding for conservation and science in general. This situation is exacerbated by the interaction of inadequate funding with poor governance, internal corruption, lack of political will and the current political instability (Loyola 2014; Fearnside 2016; Fernandes et al. 2017).

Despite the encouraging progress made by some countries in pursuing GSPC Target 7, global progress remains difficult to be assessed because of slow progress with Target 2 (assessment of species conservation status; see Martins et al., this issue; Bachman et al. 2017) and the lack of distribution data for threatened species (Convention on Biological Diversity 2016b). The crucial importance of a global assessment is highlighted by the high number of species supposed to be threatened. An estimate made using nearly 16% of plants of the world showed that 43% of these plants are categorized as threatened with extinction (Sharrock et al. 2014).

Our quantitative analysis sheds light on the status quo of Brazilian responsibility and commitment regarding GSPC target 7. Great advances in plant conservation in Brazil have been achieved in the last decades (see Martinelli

& Moraes 2013; MMA 2018; Loyola et al. 2014a; BFG 2015; Martins et al. 2017). We conclude that although Brazil is almost reaching GSPC Target 7 in terms of absolute numbers, the government still needs to allocate resources for properly managing and improving the conservation status of its imperiled flora.

Acknowledgements

BRR thanks the CAPES for providing a PhD fellowship; RL research is funded by CNPq (grant 308532/2014-7); O Boticário Group Foundation for Nature Protection (grant PROG_0008_2013), and CNCFlora (grant 065/2016). This paper is a contribution of the INCT in Ecology, Evolution and Biodiversity Conservation founded by MCTIC/ CNPq/ FAPEG (grant 465610/2014-5).

References

Bachman, S. P.; Nic Lughadha, E. M.; & Rivers, M. C. 2017. Quantifying progress toward a conservation assessment for all plants. Conservation Biology, 0(0): 1–9.

Bernard, E.; Penna, L. A. O.; & Araújo, E. 2014. Downgrading, Downsizing, Degazettement, and Reclassification of Protected Areas in Brazil. Conservation Biology, 28(4): 939–950.

BFG. 2015. Growing knowledge: an overview of Seed Plant diversity in Brazil. Rodriguésia, 66(4): 1085–1113.

Brancalion, P. H. S.; Garcia, L. C.; Loyola, R.; Rodrigues, R. R.; Pillar, V. D.; & Lewinsohn, T. M. 2016. A critical analysis of the Native Vegetation Protection Law of Brazil (2012): updates and ongoing initiatives. Natureza & Conservação, 14 1–15.

Castilho, L. C.; De Vleeschouwer, K. M.; Milner-Gulland, E. J.; & Schiavetti, A. 2017. Hunting of mammal species in protected areas of the southern Bahian Atlantic Forest, Brazil. Oryx, 1–11.

Coelho, M. A. N.; Costa, D. P. da; Martinelli, G.; Moraes, M. Á.; Forzza, R. C.; & Fotografia: Azoury, R. 2015. Expedições Às Montanhas da Amazônia. Rio de Janeiro: Andrea Jakobsson Estúdio.

Convention on Biological Diversity. 2016a. Global Strategy for Plant Conservation: Introduction. Retrieved December 9, 2017, from https://www.cbd.int/gspc/intro.shtml

Convention on Biological Diversity. 2016b. Progress in Implementation of the Targets of the Global Strategy for Plant Conservation, 23. Retrieved from https://www.cbd.int/doc/meetings/sbi/sbi-01/information/sbi-01-inf-32-en.pdf

Corlett, R. T. 2016. Plant diversity in a changing world: Status, trends, and conservation needs. Plant Diversity, 38(1): 10–16.

Fearnside, P. M. 2016. Brazilian politics threaten environmental policies. Science, 353(6301): 746–748.

Fernandes, G. W.; Vale, M. M.; Overbeck, G. E.; Bustamante, M. M. C.; Grelle, C. E. V.; Bergallo, H. G.; ... Pillar, V. D. 2017. Dismantling Brazil's science threatens global biodiversity heritage. Perspectives in Ecology and Conservation, 15(3): 239–243.

Ferreira, M. N.; & Valdujo, P. H. 2014. Observatório de UCs: Biodiversidade em Unidades de Conservação. Brasília: WWF-Brasil.

Ferro, V. G.; Lemes, P.; Melo, A. S.; & Loyola, R. 2014. The Reduced Effectiveness of Protected Areas under Climate Change Threatens Atlantic Forest Tiger Moths. PLoS ONE, 9(9): e107792.

Flora do Brasil 2020. 2016. Retrieved December 9, 2017, from http://floradobrasil.jbrj.gov.br

Hansen, M. C.; Potapov, P. V.; Moore, R.; Hancher, M.; Turubanova, S. A.; Tyukavina, A.; ... Townshend, J. R. G. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342(6160): 850–853.

Heywood, V. H. 2017. Plant conservation in the Anthropocene – Challenges and future prospects. Plant Diversity, 39(6): 314–330.

Kier, G.; Mutke, J.; Dinerstein, E.; Ricketts, T. H.; Küper, W.; Kreft, H.; & Barthlott, W. 2005. Global patterns of plant diversity and floristic knowledge. Journal of Biogeography, 32(7): 1107–1116.

Lemes, P.; Melo, A. S.; & Loyola, R. D. 2013. Climate change threatens protected areas of the Atlantic Forest. Biodiversity and Conservation, 23(2): 357–368.

Loucks, C.; Ricketts, T. H.; Naidoo, R.; Lamoreux, J.; & Hoekstra, J. 2008. Explaining the global pattern of protected area coverage: relative importance of vertebrate biodiversity, human activities and agricultural suitability. Journal of Biogeography, 35(8): 1337–1348.

Loyola, R. D.; Nabout, J. C.; Trindade-Filho, Joaquim Lemes, P.; Urbina-Cardona, J. N.; Dobrovolski, R.; Sagnoli, M. D.; & Diniz-filho, J. A. F. 2012. Climate change might drive species into reserves: a case study of the American bullfrog in the Atlantic Forest Biodiversity Hotspot. Alytes, 29:61–74.

Loyola, R. 2014. Brazil cannot risk its environmental leadership. Diversity and Distributions, 20(12): 1365–1367.

Loyola, R.; Machado, N.; Vila-Nova, D. A.; Martins, E.; & Martinelli, G. 2014a. Áreas prioritárias para conservação e uso sustentável da flora brasileira ameaçada de extinção. Rio de Janeiro: Andrea Jakobsson Estúdio: Instituto de Pesquisas Jardim Botânico.

Loyola, R. D.; Lemes, P.; Brum, F. T.; Provete, D. B.; & Duarte, L. D. S. 2014b. Clade-specific consequences of climate change to amphibians in Atlantic Forest protected areas. Ecography, 37(1): 65–72.

Loyola, R.; & Machado, N. 2015a. Áreas prioritárias para a conservação da flora ameaçada de extinção da Serra do Espinhaço Meridional. In Plano de ação nacional para a conservação da flora ameaçada de extinção da serra do espinhaço meridional. Rio de Janeiro: Andrea Jakobsson Estúdio: Instituto de Pesquisas Jardim Botânico.

Loyola, R.; Machado, N.; & Vila-Nova, D. A. 2015b. Prioridades espaciais para a conservação da flora ameaçada de extinção de Grão Mogol - Francisco Sá. In Plano de ação nacional para a conservação da flora ameaçada de extinção de Grão Mogol - Francisco Sá. Rio de Janeiro: Andrea Jakobsson Estúdio: Instituto de Pesquisas Jardim Botânico.

Loyola, R.; Machado, N.; Ribeiro, B. R.; Martins, E.; & Martinelli, G. 2018. Áreas prioritárias para conservação da flora endêmica do Estado do Rio de Janeiro. Rio de Janeiro: Graficci.

Martinelli, G.; & Moraes, M. Á. 2013. Livro Vermelho da Flora do Brasil (1 st). Rio de Janeiro: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro.

Martinelli, G.; Messina, T.; & Filho, L. dos S. 2014. Livro Vermelho da Flora do Brasil: Plantas Raras do Cerrado. Rio de Janeiro: Andrea Jakobsson Estúdio: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro.

Martinelli, G.; Martins, E.; Moraes, M.; Loyola, R.; & Amaro, R. 2018. Livro Vermelho da Flora Endêmica do Estado do Rio de Janeiro. Rio de Janeiro: Andrea Jakobsson Estúdio.

Martins, E.; Fernandes, F. M.; Maurenza, D.; Pougy, N.; Loyola, R.; & Martinelli, G. 2014. Plano de Ação Nacional para a Conservação do

Faveiro-de-wilson (Dimorphandra wilsonii Rizzini). Rio de Janeiro: Andrea Jakobsson Estúdio: Instituto de Pesquisas Jardim Botânico.

Martins, E.; Loyola, R.; & Martinelli, G. 2017. Challenges and Perspectives for Achieving the Global Strategy for Plant Conservation Targets in Brazil. Annals of the Missouri Botanical Garden, 102(2): 347–356.

MMA. 2000. Sistema Nacional de Unidades de Conservação. Retrieved from http://www.mma.gov.br/areas-protegidas/sistema-nac

MMA. 2014. Portaria no. 443, de 17 de dezembro de 2014. Retrieved from http://www.mma.gov.br/biodiversidade/especies-ameacadas-de-extincao/atualizacao-das-listas-de-especies-ameacadas

MMA. 2018. Cadastro Nacional de Unidades de Conservação. Retrieved from http://www.mma.gov.br/areas-protegidas/cadastro-nacional-de-ucs/dados-consolidados

Muñoz-Rodríguez, P.; Draper Munt, D.; & Moreno Saiz, J. C. 2016. Global strategy for plant conservation: inadequate in situ conservation of threatened flora in Spain. Israel Journal of Plant Sciences, 63(4): 1–12.

Nori, J.; Urbina-Cardona, J. N.; Loyola, R. D.; Lescano, J. N.; & Leynaud, G. C. 2011. Climate Change and American Bullfrog Invasion: What Could We Expect in South America? PLoS ONE, 6(10): e25718.

Oliveira, U.; Soares-Filho, B. S.; Paglia, A. P.; Brescovit, A. D.; de Carvalho, C. J. B.; Silva, D. P.; ... Santos, A. J. 2017. Biodiversity conservation gaps in the Brazilian protected areas. Scientific Reports, 7(1): 9141.

Oliver, T. H.; Smithers, R. J.; Beale, C. M.; & Watts, K. 2016. Are existing biodiversity conservation strategies appropriate in a changing climate? Biological Conservation, 193 17–26.

Pacheco, A. A.; Neves, A. C. O.; & Fernandes, G. W. 2018. Uneven conservation efforts compromise Brazil to meet the Target 11 of Convention on Biological Diversity. Perspectives in Ecology and Conservation, 16(1): 43–48.

Pebesma, E.J., Bivand, R.S. 2005. Classes and methods for spatial data in R. R News 5 (2). Retrieved from https://cran.r-project.org/doc/Rnews/.

Pougy, N., Verdi, M., Martins, E., Loyola, R., & Martinelli, G. 2015a. Plano de Ação Nacional para a conservação da flora ameaçada de extinção da região de Grão Mogol-Francisco Sá. Rio de Janeiro: Andrea Jakobsson Estúdio: Instituto de Pesquisas Jardim Botânico.

Pougy, N., Verdi, M., Martins, E., Loyola, R., & Martinelli, G. 2015b. Plano de Ação Nacional para a conservação da flora ameaçada de extinção da Serra do Espinhaço Meridional. 1. ed. Rio de Janeiro: Andrea Jakobsson Estúdio: Instituto de Pesquisas Jardim Botânico.

Pougy, N.; Martins, E.; Verdi, M.; Fernandez, E.; Loyola, R.; Filho, T. B. S.; & Martinelli, G. 2018. Plano de ação nacional para a conservação da flora endêmica ameaçada de extinção do estado do Rio de Janeiro. Rio de Janeiro: Secretaria de Estado do Ambiente (SEA): Andrea Jakobsson Estúdio.

Pressey, R. L.; Visconti, P.; & Ferraro, P. J. 2015. Making parks make a difference: poor alignment of policy, planning and management with protected-area impact, and ways forward. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1681): 20140280.

Schulman, L.; Toivonen, T.; & Ruokolainen, K. 2007. Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. Journal of Biogeography, 34(8): 1388–1399.

Sharrock, S.; Oldfield, S.; & Wilson, O. 2014. Plant conservation report 2014: A review of progress in implementation of the Global Strategy for Plant Conservation 2011-2020. Technical Series (Technical). Secretariat of the Convention on Biological Diversity, Montréal, Canada and Botanic Gardens Conservation International, Richmond, UK.

Soares-Filho, B.; Moutinho, P.; Nepstad, D.; Anderson, A.; Rodrigues, H.; Garcia, R.; ... Maretti, C. 2010. Role of Brazilian Amazon protected areas in climate change mitigation. Proceedings of the National Academy of Sciences of the United States of America, 107(24): 10821–10826.

Sobral, M.; & Stehmann, J. R. 2009. An analysis of new angiosperm species discoveries in Brazil (1990-2006). Taxon, 58(1): 227–232.

Sousa-Baena, M. S.; Garcia, L. C.; & Peterson, A. T. 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Diversity and Distributions, 20(4): 369–381.

Strassburg, B. B. N.; Brooks, T.; Feltran-Barbieri, R.; Iribarrem, A.; Crouzeilles, R.; Loyola, R.; ... Balmford, A. 2017. Moment of truth for the Cerrado hotspot. Nature Ecology & Evolution, 1(4): 1–3.

Vieira, R. R. S.; Ribeiro, B. R.; Resende, F. M.; Brum, F. T.; Machado, N.; Sales, L. P.; ... Loyola, R. 2018. Compliance to Brazil's Forest Code will not protect biodiversity and ecosystem services. Diversity and Distributions, 24(4): 434–438.

CAPÍTULO V

Quem entra na lista vermelha?

Ribeiro, BR; Diniz, MF; Loyola, R. **À beira da extinção.**
Aceito para publicação na revista Ciência Hoje das Crianças

Quem entra na lista vermelha?

Bruno R. Ribeiro[1], Milena Fiuza Diniz[1] e Rafael Loyola[2,3]

Endereço

[1] Programa de pós-graduação em Ecologia e Evolução – Universidade Federal de Goiás

[2] Departamento de Ecologia – Universidade Federal de Goiás

[3] Fundação Brasileira para o Desenvolvimento Sustentável

Onça-pintada, ararinha-azul, mico-leão-dourado e tan-tas outras. Certamente, você já ouviu falar em espécies ameaçadas de extinção e em como o impacto humano têm feito a lista dessas espécies aumentar a cada ano. O que você provavelmente desconhece é como uma espécie ganha o tão famigerado rótulo de "ameaçada de extin-ção" e porque isso é tão importante para evitar que mais de 31 mil espécies de animais, plantas e outros organis-mos com esse título desapareçam em breve do planeta.

Se você já foi à uma unidade médica de pronto-socorro, deve ter percebido que os pacientes passam por um processo de triagem antes de serem avaliados por um médico. A triagem busca minimizar o risco de morte dos pacientes ao aplicar um sistema de classificação de risco para atribuir um grau de prioridade para o atendimento de acordo os sintomas iniciais apresentados. Agora, troque os pacientes por espécies, os médicos por cientistas, o risco de morte por risco de extinção e o protocolo médico pelo sistema usado pela União Internacional para Conservação da Natureza (UICN). Pronto! Você já tem o que é necessário para entender a essência da metodologia aplicada para avaliar o estado de conservação das espécies e a saúde da biodiversidade global. Mas, afinal, quais são os sintomas apresentados por uma espécie ameaçada de extinção? Quais as possíveis variações desse diagnóstico? Como o sistema de classificação de risco pode ajudar a salvar milhares de espécies da extinção no Brasil e no mundo?

A morte de uma espécie

Uma espécie é considerada extinta quando o último indivíduo, da última população restante da espécie, morre. A extinção é um processo natural e tem ocorrido ao longo de toda a história da vida no nosso planeta. Estima-se que cerca de 99% de todas as espécies que evoluíram na Terra já tenham se extinguido. Mas, afinal, se a extinção é um processo natural, por que deveríamos nos preocupar com as extinções de agora?

O grande problema do cenário atual se encontra na velocidade (ou seja, na taxa de extinção) com que o desparecimento de espécies tem ocorrido. Em apenas cinco eventos excepcionais em toda a história do planeta, conhecidos como extinções em massa, o desaparecimento de espécies ocorreu de forma intensa e rápida. Talvez o evento mais famoso entre eles seja a grande extinção do final do período Cretáceo que varreu da face da Terra cerca 75% de todas as espécies então existentes, incluindo os dinossauros não aviários.

Fora dos eventos de extinção em massa, e tirando o ser humano de cena, a perda de espécies ocorre em um ritmo muito lento. Mas o que acontece se adicionarmos o ser humano nesse cenário? Temos simplesmente um evento de extinção em massa! E é exatamente isso o que está acontecendo, pois a taxa com que espécies são sendo extintas atualmente é cerca de 100 vezes maior do que seria esperado em um processo natural de extinção.

O crescente número de pessoas no planeta associado com um padrão de consumo excessivo tem degradado profundamente os sistemas naturais e ameaçado inúmeras espécies. Somente nos últimos trinta anos, 75 espécies desapareceram de forma definitiva do planeta e outras 878 já se encontram extintas na natureza. O impacto humano sobre a sobrevivência das espécies é tão grande que muitos cientistas afirmam que nós já estamos em meio ao sexto evento de extinção em massa do planeta.

A Lista Vermelha de Espécies Ameaçadas

Como se já não fosse desafio suficiente lidar com a grande quantidade de espécies prestes a desaparecerem, os profissionais da conservação ainda enfrentam orçamentos restritos e tempo limitado. Portanto, uma etapa fundamental para frear a recente onda de extinção é identificar quais espécies correm um maior risco de serem extintas e desenvolver ações para protegê-las, tal como a criação de unidades de conservação e leis ambientais.

A UICN, desde 1964, produz a Lista Vermelha de Espécies Ameaçadas, o maior catálogo sobre o risco de extinção de espécie de plantas, fungos, protozoários e animais do planeta. A Lista Vermelha tem como objetivo fornecer informações científicas necessárias para embasar e facilitar ações de conservação. Ela também atua como um termômetro do estado de saúde da biodiversidade, pois indica se o estado de conservação das espécies tem melhorado ou piorado ao longo dos anos.

A Lista Vermelha é produzida e organizada pela UICN com a ajuda de milhares de pesquisadores e pesquisadoras voluntários, governos, e organizações não governamentais de todo mundo. O resultado da avaliação feita pela UICN pode ser acessado por qualquer pessoa através de sua plataforma (www.iucnredlist.org) e revela o estado de conservação das espécies em nível global.

Todas as espécies são avaliadas?

O primeiro passo para realizar uma avaliação do risco de extinção é compilar uma lista de espécies e decidir quais serão avaliadas. Uma espécie presente na lista, mas que, por algum motivo, não será avaliada nesse momento, é categorizada como Não Avaliada (NE, *Not Evaluated*, em inglês). Em seguida, é preciso reunir os dados necessários para realizar a avaliação, tais como: informações sobre a história de vida (por exemplo, a espécie é carnívora ou herbívora? Qual o seu tamanho?), sobre os fatores que podem levar a espécie à extinção (por exemplo, destruição de seu habitat

natural, poluição, presença de espécies invasoras, mudanças climáticas, caça e pesca), o habitat onde a espécie ocorre (floresta, campos, na Mata Atlântica, no Cerrado), dados sobre a população (tamanho total, número de indivíduos reprodutores) e, por fim, informações sobre sua distribuição geográfica (isto é, ela ocorre ao longo de todo Cerrado? Ocorre somente em parte da Serra do Espinhaço, em Minas Gerais?).

Na verdade, reunir os dados é uma das etapas mais demoradas e importantes do processo de avaliação de risco. Uma vez que isso tenha sido feito, é hora de realizar a avaliação. Caso os dados compilados não sejam bons o suficiente, a avaliação de risco não é realizada e a espécie será categorizada como Dados Insuficientes (DD, *Data Deficient*, em inglês). Vale ressaltar que a categoria DD não indica se uma espécie está ameaçada ou não. Indica apenas que não existem informações suficientes para avaliar o risco de extinção da espécie. Seguindo com a nossa analogia médica, não se pode dizer nada sobre o estado de saúde de um paciente cujos sintomas não puderam ser avaliados.

Quais são os critérios avaliados pela UICN?

As espécies que possuem dados suficientes seguem o processo de avaliação que resulta na classificação dessas em categorias de risco. As categorias são definidas com base em cinco critérios (ver Quadro) que incluem basicamente informações sobre o tamanho da população (isto é, o número de indivíduos existentes) e a área de distribuição geográfica da espécie. Os critérios foram elaborados de modo a permitir que, com exceção de microrganismos, todas as espécies possam ser avaliadas, mesmo aquelas muito diferentes quanto à sua biologia (por exemplo, formigas e elefantes).

O critério *A* avalia se houve uma redução no tamanho da população total das espécies. Para isso, o número atual de indivíduos de uma espécie é comparado com estimativas do número de indivíduos no passado ou no futuro. A redução no tamanho da população é medida em porcentagem e utilizada para atribuir à espécie uma categoria de risco.

O critério *B* avalia o tamanho da área de distribuição geográfica de uma espécie, juntamente com medidas que indicam se a espécie está ameaçada como, por exemplo, se o habitat da espécie dentro da área de distribuição se encontra dividido em pedaços pequenos e isolados. A

distribuição geográfica representa o conjunto de locais onde a espécie ocorre e é medida em quilômetros quadrados (km2).

O critério *C* avalia se o tamanho de uma população foi muito reduzido de modo que agora restam apenas poucos indivíduos na população capazes de se reproduzir.

O critério *D* é muito parecido com os critérios B e C, mas é especialmente utilizado para avaliar o risco de extinção de espécie com populações e com área de distribuição muito pequenas.

Por fim, o critério *E* utiliza vários dados sobre a espécie (por exemplo, número de filhotes, tamanho do corpo, entre outros) para calcular uma probabilidade de que uma espécie seja extinta dentro de alguns anos.

Categorias de ameaça da UICN

Os critérios acima descritos são utilizados para classificar as espécies em sete categorias de risco que podem ser reunidas em três grandes grupos.

O primeiro grupo inclui as espécies não ameaçadas de extinção presentes nas categorias Menos Preocupante (LC, *Least Concern*, em inglês) e Quase Ameaçada (NT, *Near Threatened*). Ambas as categorias incluem espécies abundantes e com ampla distribuição. Entretanto, espécies listadas na categoria NT, apesar de estarem seguras atualmente, podem se tornar ameaçadas num futuro próximo.

O segundo grupo é, de fato, a Lista Vermelha de Espécies Ameaçadas. Fazem parte desse grupo espécies com elevado risco de extinção na natureza que são classificadas em uma das seguintes categorias: risco de extinção extremamente alto (CR – Criticamente em Perigo, *Critically Endangered*), muito alto (EN – Em Perigo; *Endangered*) e alto (VU – Vulnerável; *Vulnerable*).

O terceiro grupo é formado pelas espécies Extintas (EX; *Extinct)* e Extintas na Natureza (EW; *Extinct in the Wild*). Uma espécie é considerada extinta quando não restam dúvidas de que o último indivíduo da espécie morreu. Já a categoria Extinta na Natureza é utilizada para classificar espécies que, apesar de muito esforço de pesquisa, nenhum indivíduo da espécie pôde ser encontrado na natureza. Espécies classificadas nessa categoria sobrevivem apenas em cativeiro ou em cultivo como uma população fora do seu habitat original.

Qual a situação das espécies brasileiras?

Os critérios e categorias da Lista Vermelha global podem ser utilizados para avaliar o estado de conservação das espécies presentes em um país. Atualmente, 2.104 espécies de plantas e 1.182 espécies de animais estão ameaçadas de extinção no Brasil, segundo as avaliações realizadas pelo Centro Nacional para Conservação da Flora do Instituto de Pesquisas Jardim Botânico do Rio de Janeiro (CNCFlora, no caso de plantas) e pelo Instituto Chico Mendes para Conservação da Natureza (ICMBio, no caso de animais).

Os resultados das avaliações de risco de extinção são publicados na forma de "livros vermelhos" e todas as espécies listadas nestes livros são protegidas por lei. Os livros são de livre acesso, e contém informações importantes sobre características das espécies, risco de extinção e fatores de ameaça. Tais informações são fundamentais para a criação de Planos de Ação Nacional (PANs), isto é, uma publicação que contém ase ações de conservação necessárias para que uma espécie deixe de estar ameaçada de extinção. Os PANs são organizados Ministério do Meio Ambiente e publicados pelo CNCFlora e pelo ICMBio.

Incompleta, mas imprescindível

É preciso lembrar que apenas as espécies identificadas como ameaçadas podem ser salvas, uma vez que os esforços de conservação são limitados e direcionados a elas. Apesar de ser o maior catálogo com informações sobre o estado de conservação das espécies, a Lista Vermelha, seja ela global ou nacional, está longe de estar completa. Apenas 5% das cerca de 2 milhões de espécies que ocorrem no planeta foram avaliadas até hoje. Embora ainda há muito pela frente para tornar a Lista Vermelha mais abrangente, mesmo incompleta, essa ferramenta é a informação mais confiável e importante sobre a saúde da biota do planeta.

Por fim, é importante destacar que a saúde da biodiversidade está diretamente associada à nossa própria saúde. A extinção de espécies não apenas torna as florestas, campos e mares menos exuberantes, como também coloca em risco o nosso bem-estar, assim como a segurança alimentar, hídrica e energética de toda a população mundial.

APÊNDICE I

Chá de sumiço

Diniz, MF; Ribeiro, BR; Loyola, R. **Chá de Sumiço.**
Publicado na revista Ciência Hoje das Crianças (2020).
Disponível em: http://chc.org.br/artigo/cha-de-sumico/

Chá de sumiço

Milena Fiuza Diniz[1], Bruno R. Ribeiro[1] e Rafael Loyola[2,3]

Endereço

[1] Programa de pós-graduação em Ecologia e Evolução - Universidade Federal de Goiás

[2] Departamento de Ecologia - Universidade Federal de Goiás

[3] Fundação Brasileira para o Desenvolvimento Sustentável

Um lobo capaz de carregar os filhotes em uma bolsa como os cangurus. Um peixe-boi de nove metros de comprimento. Uma revoada de pombos capaz de ofuscar até a luz do sol. Isso é fruto da imaginação? Que nada! Esses bichos existiram, mas, infelizmente, já não estão entre nós. Milhares de outras criaturas magníficas como essas podem desaparecer em breve. O que está acontecendo? Será que tomaram chá de sumiço?

A jandaia-amarela pode desaparecer devido à sua captura para o tráfico ilegal de animais silvestres.

Foto Takashi Hososhima/Wikimedia Commons

A Terra é repleta de animais e plantas. E eles são fantásticos! Muitos parecem ter saído de filmes de ficção científica, de tão curiosos. Mas são reais. Acontece que, a história no mundo real também possui vilões. E eles estão fazendo com que muitas espécies desapareçam para sempre. Quem são esses vilões? O que podemos fazer para combatê-los? Há tempo para salvar as demais espécies, que têm tanto direito ao planeta quanto nós? Eis a questão!

Extinção em massa

Quando uma espécie já não pode mais ser encontrada em nenhum canto da Terra, dizemos que ela foi extinta. Apesar de parecer trágico, o desaparecimento de espécies, assim como o surgimento delas, tem acontecido desde que os primeiros organismos apareceram na Terra, há mais de 3 bilhões de anos. Então, por que deveríamos nos preocupar com as extinções nos tempos de hoje?

O problema está na velocidade com que as extinções estão acontecendo atualmente. Durante toda a história da vida na Terra, em apenas cinco ocasiões muitas espécies foram extintas rapidamente. Esses eventos dramáticos ficaram conhecidos como extinção em massa.

Quando a humanidade ataca

Se você é um amante de dinossauros, assim como nós, já deve ter ouvido falar que a extinção inesperada dessas criaturas maravilhosas foi causada pela queda de um gigantesco meteoro e pela mudança no clima do planeta. Isso aconteceu há cerca de 66 milhões de anos e foi o último evento de extinção em massa da Terra que tivemos registro.

Fora de situações especiais como essas, as espécies se extinguem em um ritmo muito lento. Mas tudo isso começou a mudar muito nos últimos séculos com o aumento da população humana. Nós, humanos, somos hoje quase 8 bilhões de pessoas extraindo e consumindo intensamente os recursos fornecidos da natureza. Nossa marca no planeta é tão profunda que alguns cientistas acreditam que iniciamos uma nova época geológica, o Antropoceno.

Quem são os vilões?

Quando cidades, rodovias, fazendas, pastos, plantações e outras construções e atividades humanas avançam em direção às áreas naturais, os animais e plantas que ali viviam são mortos ou forçados a se mudar. O problema é que encontrar espaços apropriados na natureza está se tornado cada vez mais difícil, porque as atividades humanas têm rapidamente destruído ou degradado os ecossistemas. Essa é a triste realidade da maioria das espécies brasileiras ameaçadas de extinção, como o sapinho-admirável-da-barriga-vermelha, a jacutinga e a jararacuçu-de-murici.

A jacutinga é uma das muitas espécies brasileiras ameaçada pela destruição de seu ambiente natural.

Foto Milena Fiuza Diniz

A caça, a pesca ou a captura também são atividades humanas que têm ameaçado a existência de muitas espécies. Aparentemente, não há nada de errado com estas atividades. Aliás, elas são importantes para fornecer alimentos para muitas pessoas. Mas elas se tornam um problema quando a retirada de seres da natureza é maior que a velocidade com que eles conseguem se reproduzir. Aqui no Brasil, por exemplo, o queixada, a jandaia-amarela e o enorme peixe-mero podem desaparecer em breve devido à caça, captura e pesca, respectivamente.

O mero está ameaçado pelo excesso de pesca e pela degradação do seu ambiente natural.

Foto Albert Kok/Wikimedia Commons.

A caça e a perda do ambiente natural são as principais ameaças enfrentadas pelo queixada.

Foto Ana Cotta/Wikimedia Commons

A temida Lista Vermelha

Quando uma espécie corre o risco de desaparecer, os cientistas dão a elas o famoso título de 'espécie ameaçada de extinção' e as colocam na chamada Lista Vermelha de Espécies Ameaçadas.

Se você fosse um bicho ou uma planta certamente não gostaria de receber esse título e nem estar nessa lista, mas se estivesse sofrendo com as ações do ser humano, estar nessa lista poderia ser sua única salvação. Isso porque a Lista Vermelha aponta quais espécies precisam de cuidados urgentes para escaparem da extinção.

Cada espécie ameaçada é classificada em uma categoria que mostra qual a gravidade da sua situação. Uma espécie pode ser classificada como tendo risco de extinção extremamente alto (chamado de CR – Criticamente em Perigo), muito alto (EN – Em Perigo) ou alto (VU – Vulnerável).

Escolha preservar e proteger

Nós humanos também somos capazes de reduzir ou reverter o mal que causamos. A criação da Lista Vermelha é apenas o primeiro passo para ajudar uma espécie escapar da extinção. Conhecendo as principais ameaças às espécies,

os pesquisadores conseguem desenvolver estratégias para evitar que elas sejam extintas, como criar unidades de conservação e leis ambientais.

Mas ajudar nessa situação não é trabalho exclusivo dos pesquisadores. Nossas escolhas no dia a dia, como os produtos que compramos, o descarte correto do lixo que produzimos, entre outras ações, podem impactar diretamente as espécies na natureza. A informação e a adoção de hábitos conscientes são os principais elementos para impedir que mais espécies se aproximem do precipício da extinção e tomem o temido "chá de sumiço"!

Superpoderes

Como no mundo da fantasia, os animais do mundo real também têm poderes impressionantes. As abelhas, por exemplo, têm o poder da polinização, que é essencial para produção de frutos e sementes das plantas. Muitos morcegos, lagartos e aranhas são capazes de combater certos insetos que podem destruir as plantações e transmitir doenças, inclusive para os seres humanos.

Por isso, a extinção de animais e plantas não só torna a história da natureza um pouco mais triste, como pode tornar a nossa vida muito mais complicada. Se as abelhas desaparecerem, por exemplo, a quantidade e a qualidade dos alimentos que comemos diminuiriam, colocando em risco o abastecimento de comida para a população mundial, além de trazer um prejuízo financeiro enorme para os países.

CONCLUSÃO GERAL

Ao fim dessa tese obtivemos as seguintes conclusões:

Capítulo I

- O sistema proposto pela UICN, apesar de atualmente ser o mais utilizado na avaliação do risco de extinção das espécies, é complexo, demorado e custoso.

- O risco de extinção é conhecido apenas para 5% das espécies descritas pela ciência, o que representa uma parcela pequena taxonomicamente e geograficamente da biodiversidade global.

- Novas abordagens e ferramentas podem agilizar e reduzir o custo das avaliações de risco de forma que o risco de extinção de uma amostra mais representativa da biodiversidade possa ser avaliado.

Capítulo II

- Dos mais 30 milhões de registros de ocorrência de plantas brasileiras disponíveis em nove bases de dados online, apenas 13% são potencialmente aptos ao uso. Todos os registros de ocorrência de 1.667 espécies foram identificados como errôneos ou suspeitos.

- A limpeza de dados é um processo fundamental para garantir a qualidade dos dados de biodiversidade. Entretanto, a remoção acrítica desses registros pode levar a perda de valiosa informação.

- Nesse capítulo apresentamos o pacote "*bdc*", o qual contém meios para facilitar o processo de limpeza de dados e torná-lo mais estruturado, abrangente, auditável e reproduzível.

Capítulo III

- Uma baixa coleta de registros de ocorrência de espécies e erros taxonômicos, geográficos e espaciais podem impedir seu uso efetivo em avaliações rápidas de risco de extinção.

- Avaliações rápidas e automatizadas apresentam maior acurácia em identificar espécies não ameaçadas do que ameaçadas.

- Avaliações rápidas de risco de extinção podem identificar acuradamente espécies não ameaçadas, mesmo sem uma ampla coleta de registros de ocorrência e na presença de erros taxonômicos, geográficos e espaciais.

• A sensibilidade (capacidade de identificar corretamente espécies não ameaçadas) de avaliações rápidas é afetada pela baixa coleta de registros de ocorrência e por erros taxonômicos, geográficos e temporais.

Capítulo IV

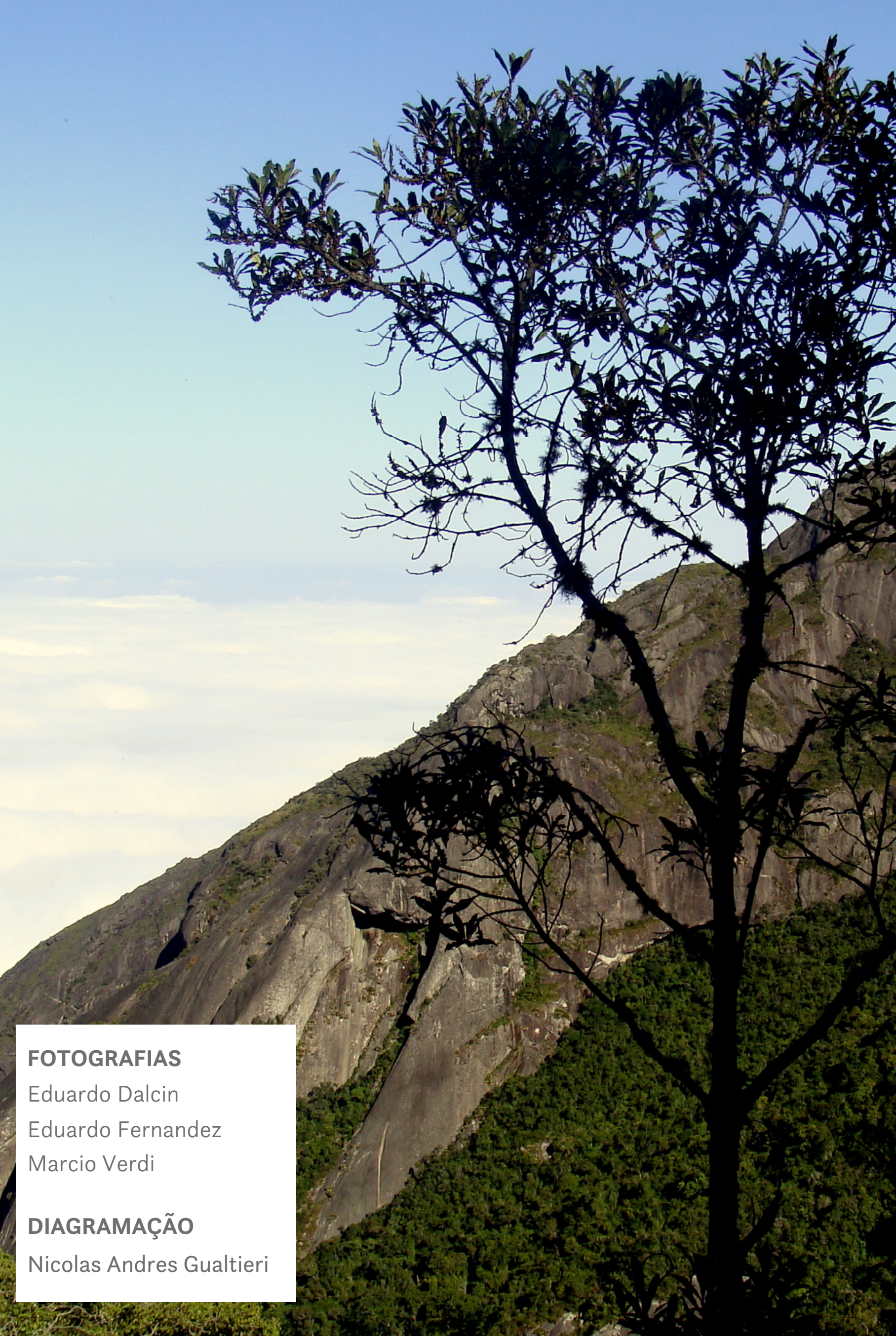
• Cerca de 10 a 33% das espécies de plantas oficialmente reconhecidas como ameaçadas de extinção no Brasil ocorrem fora de áreas protegidas.

• Em números absolutos, o Brasil está prestes a alcançar a meta 7 da Estratégia Global para Conservação de Plantas (garantir que 75% de todas as espécies atualmente ameaçadas estão representadas em áreas protegidas).

Capítulo V e Apêndice I

- Apresentamos textos detalhados e de fácil compreensão sobre o processo de avaliação de risco de extinção e a situação de ameaçada sofrida pelas espécies no Brasil. Estes textos são voltados ao público infantil.

Em suma, nesta tese analisamos e sintetizamos vários aspectos do processo de avaliação de risco e destacamos oportunidades para torná-lo mais representativo. Mostramos que métodos de avaliação rápida podem identificar acuradamente espécies não ameaçadas, e que dados de alta qualidade são necessários para aumentar a acurácia de tais métodos em identificar espécies ameaçadas. Nesse sentido, apresentamos o pacote “bdc” que contém uma série de funções para tratamento e limpeza de dados de modo a converter dados brutos de biodiversidade em informação de alta qualidade. Por fim, identificamos que cerca de 67% das espécies de plantas que se encontram ameaçadas atualmente no Brasil estão representadas em áreas protegidas. As ferramentas desenvolvidas e os resultados e conclusões dessa tese podem auxiliar a tornar o processo de avaliação de risco de extinção mais ágil e abrangente, e fornecem uma visão mais ampla e atualizada do estado de conservação das plantas no Brasil.



FOTOGRAFIAS

Eduardo Dalcin

Eduardo Fernandez

Marcio Verdi

DIAGRAMAÇÃO

Nicolas Andres Gualtieri