

Débora Christina Zuanny Martins da Silva

Grandes gêneros em Leguminosae: avanços taxonômicos e distribuição global



Salvador – Bahia
2023



Universidade Federal da Bahia
Instituto de Biologia
Programa de Pós-Graduação em Biodiversidade e Evolução

Grandes gêneros em Leguminosae: avanços taxonômicos e distribuição global

Débora Christina Zuanny Martins da Silva

Dissertação apresentada ao Instituto de Biologia da Universidade Federal da Bahia para a obtenção do Título de *Mestre em Biodiversidade e Evolução* pelo Programa de Pós-Graduação em Biodiversidade e Evolução.

Orientador: Dr. Domingos Benício Oliveira Silva Cardoso (UFBA)

Salvador – Bahia
2023

Banca Examinadora



Dr. Haroldo Cavalcante de Lima
Instituto de Pesquisas Jardim Botânico do Rio de Janeiro



Dra. Sara Ribeiro Mortara
Re.green



Dr. Domingos Benício Oliveira Silva Cardoso
Universidade Federal da Bahia
Orientador e Presidente da Banca



À minha família, dedico.

Agradecimentos

À Universidade Federal da Bahia e ao Instituto de Biologia pelo espaço físico que possibilitou meu desenvolvimento pessoal e profissional.

À Fundação de Amparo à Pesquisa do Estado da Bahia, pela bolsa de estudos de Mestrado.

Ao Programa de Pós-Graduação em Biodiversidade e Evolução, que contribuiu para que este trabalho fosse desenvolvido.

Ao Laboratório de Diversidade, Biogeografia e Sistemática do Instituto de Biologia, do qual faço parte desde 2018 e sempre me proporcionou grandiosas experiências.

Aos professores da Universidade Federal da Bahia, pelas discussões e ensinamentos, sobretudo ao meu orientador Prof. Dr. Domingos Cardoso, que sempre se dedicou a prover uma formação de qualidade e a guiar brilhantemente minha carreira acadêmica. Obrigada pelo suporte, atenção, pelas revisões minuciosas dos trabalhos e sugestões de leituras!

Ao Royal Botanic Garden Edinburgh e seus funcionários, que me acolheram muito bem durante minha estadia em 2022 e me permitiram participar das reuniões do grupo de pesquisa de Botânica da América Latina. Agradeço principalmente à Dra. Tiina Särkinen e ao Dr. Peter Moonlight pela recepção e ensino.

Aos amigos que me apoiam nesta jornada botânica de diversas formas, em especial a Catarina, o Gustavo, o Raphael, a Briggithe, a Bernarda, a Fernanda, o Raul, e a Bárbara. Meus agradecimentos aos novos amigos e àqueles de longa data!

À minha família.

Sumário

Resumo	i
Abstract	ii
Introdução Geral	1
Referências	4
Capítulo I: expowo: An R package for mining plant diversity and distribution data	7
Capítulo II: Big genera in Leguminosae: Overview of taxonomic changes and global distribution	36
Considerações finais	83
Apêndice	84

RESUMO

Grandes gêneros, também conhecidos como gêneros megadiversos, são em geral definidos como aqueles com mais de 500 espécies aceitas. O tamanho de muitos destes grandes gêneros mudaram significativamente através da descrição de novas espécies, divisão de gêneros não monofiléticos, ou por agrupamento ou *lumping* de gêneros. O número atual de grandes gêneros é desconhecido, e as estimativas de tamanho mais recentes desses gêneros estão significativamente desatualizadas. Assim, desenvolvemos um pacote em linguagem R para minerar o Plants of the World Online (POWO), um banco de dados global e taxonomicamente verificado do Royal Botanic Gardens, Kew, de modo a fornecer uma lista atualizada dos grandes gêneros da família Leguminosae e sumarizar: (i) o atual número de espécies; (ii) um resumo do histórico taxonômico sobre as principais mudanças no número de espécies reconhecidas em cada gênero; (iii) os mecanismos por trás das mudanças no tamanho dos gêneros a partir dos avanços filogenéticos moleculares nas últimas décadas; e (iv) fornecer mapas de distribuição dos gêneros nos diferentes países e regiões botânicas do mundo, como um primeiro passo para identificar possíveis padrões e determinantes da distribuição global das leguminosas megadiversas. Nós encontramos que os grandes gêneros de Leguminosae são *Astragalus*, *Acacia*, *Crotalaria*, *Indigofera*, *Lupinus*, *Mimosa* e *Oxytropis*. Adicionalmente, encontramos que o tamanho destes grandes gêneros variou principalmente devido às novas descrições ao longo dos anos. O número de espécies de *Astragalus* e *Acacia* reduziu desde a última revisão, *Cassia* deixou de fazer parte da lista, e *Lupinus*, *Mimosa* e *Oxytropis* agora são grandes gêneros. Em relação a sua distribuição, encontramos que nem todos os grandes gêneros têm sua riqueza de espécies associada à região tropical. Embora ricos em espécies, apenas uma parcela desproporcional da diversidade desses gêneros é utilizada atualmente e mostramos seu grande potencial para uso humano.

Palavras-chave: biogeografia; Fabaceae; grandes gêneros; sistemática; taxonomia.

ABSTRACT

Big genera, also known as megadiverse genera, are generally defined as those with more than 500 accepted species. The size of many of these big genera has changed significantly through (i) the description of new species; (ii) division of non-monophyletic genera; (iii) grouping or lumping of genera into a broader concept. The current number of large genera is unknown, and the most recent size estimates of these genera are significantly out of date. Therefore, we developed an R package to mine Kew's taxonomically verified global database Plants of the World Online (POWO) to provide an up-to-date list of megadiverse genera in the Leguminosae family and summarize: (i) the current number of species; (ii) a resume of the taxonomic history of major changes in the number of species recognized in each genus; (iii) the mechanisms behind changes in the size of genera from molecular phylogenetic advances in recent decades; and (iv) provide global distribution maps of species richness at country and botanical region levels, as a first step to identify possible patterns and determinants of the global distribution of megadiverse legumes. We found that the major genera in Leguminosae are *Astragalus*, *Acacia*, *Crotalaria*, *Indigofera*, *Lupinus*, *Mimosa* and *Oxytropis*, and not all of them are globally distributed in the tropical region. Additionally, we found that the greatest changes in the size of these big genera are due to new descriptions. The number of *Astragalus* and *Acacia* species has reduced since the last revision, *Cassia* is no longer listed, and *Lupinus*, *Mimosa* and *Oxytropis* are now included as big genera. Although rich in species, only a disproportionate portion of the diversity is currently under use and we show their great potential for human behoof.

Keywords: biogeography; big genera; Fabaceae; systematics; taxonomy.

Introdução Geral

Nas últimas décadas, estudos sobre a evolução, ecologia e distribuição da biodiversidade têm avançado de forma exponencial a partir do uso de ferramentas de bioinformática e com a digitalização massiva de dados de coleções de herbário (e.g. Park et al., 2021; Davis et al., 2022). Assim, tem sido possível investigar questões amplas em grandes famílias e gêneros de angiospermas, quando antes a dificuldade da sistematização de grandes conjuntos de dados era um fator limitante. Neste contexto, Leguminosae (ou Fabaceae), a terceira maior família de angiospermas, tem figurado como um grupo modelo em estudos que buscam entender a dinâmica evolutiva e biogeografia global de biomas tropicais (e.g. Lavin et al., 2004; Simon et al., 2009; Kursar et al., 2004). Nesta dissertação, trataremos da família cosmopolita Leguminosae, juntamente com as ferramentas trazidas pelos avanços na ciência de *Big Data* nos últimos anos, para revelar uma atualização sobre os avanços taxonômicos e distribuição global dos seus grandes gêneros.

A família Leguminosae compreende cerca de 765 gêneros e quase 19.500 espécies (LPWG, 2017), sendo classificada como a terceira família com maior riqueza e a segunda em importância econômica (LPWG, 2017). Segundo a classificação filogenética mais recente, as leguminosas estão divididas em seis subfamílias (LPWG, 2017): Duparquetioideae, Cercidoideae, Detarioideae, Dialioideae, Caesalpinioideae (recircunscrita de modo a incluir as mimosoides) e Papilionoideae. São plantas muito diversas em hábito e morfologia, porém marcantes pelo seu gineceu unicarpelar que dá origem ao legume, o fruto típico da família (Lewis et al., 2005). É nesta família que está não apenas *Astragalus* L., o maior gênero conhecido de angiospermas (Maassoumi & Ashouri, 2022), como também outros considerados megadiversos, a exemplo de *Acacia* Mill. e *Indigofera* L. (Frodin, 2004).

Estudar a distribuição dos grandes gêneros significa também estudar biogeografia, que é a área que busca reconstruir a distribuição geográfica dos organismos (Crisci, 2001) no espaço e no tempo e vem sendo estudada desde o século 19 por H. W. Bates e A. R. Wallace (Wallace, 1876). Os padrões biogeográficos globais podem ser determinados por processos evolutivos e fatores ecológicos em várias escalas espaciais e temporais (Wiens & Donoghue, 2004). Por exemplo, a distribuição das espécies pode ser limitada por fatores ambientais, que podem interferir na fisiologia do organismo sem precisar ser letal e limitam sua capacidade de

sobreviver e reproduzir de forma eficiente, tornando-o menos apto àquele ambiente. Possíveis fatores limitantes são os físicos, como a temperatura, precipitação e incidência solar, e bióticos, como a competição e predação. Podemos determinar e prever a distribuição geográfica de um clado a partir do nicho ecológico ancestral, da área inicial de dispersão, pela sua capacidade intrínseca de se dispersar, pela conservação filogenética de nicho, pela área fonte da dispersão, pelas oportunidades de evolução de nicho ecológico e pelo tempo total decorrido desde a cladogênese (Wiens & Donoghue, 2004). O fato de Leguminosae possuir amplo sucesso ecológico, além de possuir filogenias publicadas de muitos gêneros diversos (e.g. *Indigofera*, Schrire et al., 2009; *Swartzia*, Torque & Mansano, 2009; *Astragalus*, Su et al., 2021) a torna um grupo excelente para explorar os padrões de distribuição da biodiversidade em plantas.

Gêneros desproporcionalmente ricos na filogenia de plantas são convencionalmente aqueles com mais de 500 espécies (Frodin, 2004) e têm sido comumente chamados de grandes gêneros (Corner, 1961). No entanto, a ampla dimensão do tamanho de grandes gêneros era frequentemente visto como uma barreira ao seu estudo (Frodin, 2004). O número crescente de herbários e de espécimes oriundos de coleções tropicais têm sido apontados como fatores contribuintes para a taxa lenta de estudos em taxonomia (Goodwin et al., 2015), e esses fatores são particularmente agudos em grandes gêneros. Isso é problemático, pois os grandes gêneros incluem ca. 15% de todas as espécies de plantas do mundo (Frodin, 2004). Portanto, sua revisão é uma barreira significativa para a conclusão de tratamentos florísticos globais e regionais, como o World Flora Online (WFO, worldfloraonline.org/) e a Flora e Funga do Brasil (BFG, 2022). Tratamentos florísticos como esses são necessários para cumprir as metas da Estratégia Global de Conservação de Plantas (GSPC, 2011; Sharrock et al., 2018). A falta de conhecimento atualizado sobre o número de grandes gêneros, seu tamanho, e o estado de tais ferramentas dificultam o trabalho taxonômico, e este trabalho visa justamente combater essa lacuna de conhecimento.

A última revisão da história e tamanho de grandes gêneros foi publicada há mais de 15 anos (Frodin, 2004) e, desde essa revisão, o tamanho e a circunscrição de muitos grandes gêneros continuaram a mudar. Cerca de 2,064 espécies foram descritas entre 2005 e 2015 e muitas dessas novas descrições foram focadas em grandes gêneros como *Astragalus* L. (Leguminosae), *Begonia* L. (Begoniaceae) e *Bulbophyllum* Thouars (Orchidaceae) (Moonlight et al., 2017). A

descrição de novas espécies também resultou em novos gêneros atingindo o limiar de gêneros megadiversos. Por exemplo, *Mimosa* L. (Leguminosae) agora inclui c. 550 espécies (Simon et al., 2011; Borges et al., 2017), e não foi considerado um grande gênero na última revisão. Alguns gêneros também mudaram devido à divisão de gêneros polifiléticos. O exemplo mais emblemático é o de *Acacia* Mill. (Leguminosae), cuja informação filogenética mostrou ser polifilético e muitas espécies foram removidas (Miller & Sieglar, 2012). Processos semelhantes também resultaram na divisão de *Schefflera* J.R.Forst & G.Forst (Araliaceae) (e.g., Lowry et al., 2013, 2017, 2019) e *Polygala* Tourn. ex L. (Polygalaceae) em vários gêneros menores (Abbott et al., 2011; Pastore et al., 2010; Pastore, 2012; Pastore & Abbott, 2012).

O nível de conhecimento filogenético dos gêneros megadiversos também aumentou significativamente nos últimos anos. Filogenias bem amostradas têm sido publicadas em leguminosas (e.g. Schrire et al., 2009; Simon et al., 2011; Rockinger et al., 2017; Su et al., 2021) e as filogenias de outras famílias de plantas também têm sido acompanhadas de atualizações na classificação infragenérica (e.g., Moonlight et al., 2017). Tais sistemas estimulam a taxonomia, dividindo grandes gêneros em grupos menores, taxonomicamente gerenciáveis e filogeneticamente significativos. Assim, ao reunirmos as informações acerca do número e tamanho atual dos grandes gêneros, revelaremos qual o déficit Lineano e também contribuimos para a conclusão de tratamentos florísticos nacionais e globais. Além de sua importância para a taxonomia, os grandes gêneros também têm relevância para a compreensão da evolução e formação dos padrões atuais da diversidade de angiospermas, podendo se tratar de exemplos de irradiações adaptativas bem sucedidas. Adicionalmente, a avaliação da distribuição das espécies destes grandes gêneros vai nos permitir inferir onde tiveram maior sucesso evolutivo, bem como avaliar o nível do déficit Wallaceano.

Aqui, propusemos um estudo dos grandes gêneros de leguminosas utilizando uma base de dados global e estruturamos a dissertação em dois capítulos. No capítulo 1 **expowo: An R package for mining plant diversity and distribution data**, foi criado um pacote em linguagem de programação em R para extração dos dados do banco de dados Plants of the World Online. E, no capítulo 2 **Big genera in Leguminosae: taxonomic advances and global distribution**, foram determinados quais são, o tamanho, status e distribuição dos grandes gêneros de Leguminosae.

Referências

- Abbott, J.R. (2011). Notes on the disintegration of *Polygala* (Polygalaceae), with four new genera for the flora of North America. *Journal of the Botanical Research Institute of Texas*, 5: 125–137.
- Borges, L.M., Simon, M.F. & Pirani, J.R. (2017). Less is more: adjusting the taxonomy of the polytypic *Mimosa setosa* (Leguminosae, Mimosoid). *Rodriguesia*, 68: 515–540.
- Corner, E.J.H. (1961). *Evolution*. In McLeoud, A.M. & Cobley, L.S. (eds.) Contemporary botanical thought: Oliver & Boyd, Edinburgh.
- Crisci, J.V. (2001). The voice of historical biogeography. *Journal of Biogeography*, 28: 157–168.
- Davis, C.C., Lyra, G.M., Park, D.S., Asprino, R., Maruyama, R., Torquato, D., ... & Ellison, A.M. (2022). New directions in tropical phenology. *Trends in Ecology & Evolution*, 37: 683–693.
- Flora e Funga do Brasil*. Jardim Botânico do Rio de Janeiro. Disponível em: < <http://floradobrasil.jbrj.gov.br/> >. Acesso em: 25 set. 2022
- Frodin, D.G. (2004). History and concepts of big plant genera. *Taxon* 53: 753–776.
- Goodwin, Z.A., Harris, D.J., Filer, D., Wood, J.R. & Scotland, R.W. (2015). Widespread mistaken identity in tropical plant collections. *Current Biology*, 25: 16.
- Kursar, T.A., Dexter, K.G., Lokvam, J., Pennington, R.T., Richardson, J.E., Weber, M.G., ... & Coley, P.D. (2009). The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proceedings of the National Academy of Sciences*, 106: 18073–18078.
- Lavin, M., Schrire, B.P., Lewis, G., Pennington, R.T., Delgado-Salinas, A., Thulin, M., ... & Wojciechowski, M.F. (2004). Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359: 1509–1522.
- Lewis, G.P.; Schrire, B; Mackinder, B; Lock, M. (2005). *Legumes of the World*. Royal Botanic Gardens, Kew, 577p.
- Maassoumi, A.A., & Ashouri, P. (2022). The hotspots and conservation gaps of the mega genus *Astragalus* (Fabaceae) in the Old-World. *Biodiversity and Conservation*, 31: 1–21.

-
- Lowry, P.P., Plunkett, G.M. & Frodin, D.G. (2013). Revision of *Plerandra* (Araliaceae) I: a synopsis of the genus with an expanded circumscription and a new infrageneric classification. *Brittonia*, 65: 42–61.
- Lowry, P.P., Plunkett, G.M., Gostel, M.R. & Frodin, D.G. (2017). A synopsis of the Afro-Malagasy species previously included in *Schefflera* (Araliaceae): resurrection of the genera *Astropanax* and *Neocussonia*. *Candollea*, 72: 265–282.
- Lowry, P.P., Plunkett, G.M., Mora, M.M., Cano, A., Fiaschi, P. et al. (2019). Studies in Neotropical Araliaceae I: resurrection of the genus *Sciadaphyllum* P.Browne to accommodate most New World species previously included in *Schefflera* J.R.Forst & G. Forst. *Brittonia*, 72: 1–15.
- Miller, J.T. & Sieglar, D.S. (2012). Evolutionary and taxonomic relationships of *Acacia* s.l. *Australian Systematic Botany*, 25: 217–224.
- Moonlight, P.W., Ardi, W., Arroyo Padilla, A., Chung, K.F., Fuller, D. et al. (2017). Dividing and conquering the fastest growing genus: towards a natural sectional classification of the mega-diverse genus *Begonia* (Begoniaceae). *Taxon*, 67: 267–323.
- Park, D.S., Breckheimer, I.K., Ellison, A.M., Lyra, G.M., & Davis, C.C. (2021). Phenological displacement is uncommon among sympatric angiosperms. *New Phytologist*, 233: 1466–1478.
- Pastore, J.F.B. (2012). *Caamembeca*: generic status and a new name for *Polygala* subgenus *Ligustrina* (Polygalaceae). *Kew Bulletin*, 67: 435–442.
- Pastore, J.F.B. & Abbott, J.R. (2012). Taxonomic notes and new combinations for *Asemeia* (Polygalaceae). *Kew Bulletin*, 67: 801–813.
- Pastore, J.F.B., Cardoso, D.B.O.S. & Aymard, C. (2010). A synopsis, new combinations, and synonyms in *Acanthocladus* (Polygalaceae). *Novon* 20: 317–324.
- Rockinger, A., Flores, A. S. & Renner, S. S. (2017). Clock-dated phylogeny for 48% of the 700 species of *Crotalaria* (Fabaceae–Papilionoideae) resolves sections worldwide and implies conserved flower and leaf traits throughout its pantropical range. *BMC evolutionary biology*, 17: 1–13.
- Schrire, B. D., Lavin, M., Barker, N. P. & Forest, F. (2009). Phylogeny of the tribe Indigofereae (Leguminosae–Papilionoideae): Geographically structured more in succulent-rich and

-
- temperate settings than in grass-rich environments. *American Journal of Botany*, 96: 816–852.
- Sharrock, S., Hoft, R. & Ferreira de Souza Dias, B. (2018). An overview of recent progress in the implementation of the Global Strategy for Plant Conservation - a global perspective. *Rodriguesia*, 69: 1589–1511.
- Simon, M.F., Grether, R., Queiroz, L.P., Skema, C., Pennington, R.T., & Hughes, C.E. (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences*, 106: 20359–20364.
- Simon, M.F., Grether, R., de Queiroz, L.P., Sarkinen, T.E., Dutra, V.F. & Hughes, C.E. (2011). The evolutionary history of *Mimosa* (Leguminosae): toward a phylogeny of the sensitive plants. *American Journal of Botany*, 98: 1201–1221.
- Su, C., Duan, L., Liu, P., Liu, J., Chang, Z. & Wen, J. (2021). Chloroplast phylogenomics and character evolution of eastern Asian *Astragalus* (Leguminosae): Tackling the phylogenetic structure of the largest genus of flowering plants in Asia. *Molecular Phylogenetics and Evolution*, 156: 107025.
- Wallace, A. R. 1876. *The geographical distribution of animals; with a study of the relations of living and extinct faunas as elucidating the past changes of the Earth's surface*. New York: Harper & Brothers. Volume 1.
- Wiens, J.J.; Donoghue, M.J. (2004). Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, 19: 639–644.

Capítulo I

expowo: An R package for mining plant diversity and distribution data

*Manuscrito submetido para o periódico Methods in Ecology and Evolution
(ISSN: 2041210X)*



expowo: An R package for mining global plant diversity and distribution data

Débora C. Zuanny¹ | Bruno Vilela¹ | Peter W. Moonlight^{2,3} | Tiina E. Särkinen³ | Domingos Cardoso^{1,4}

¹Instituto de Biologia, Universidade Federal da Bahia, Rua Barão de Jeremoabo, s.n., Ondina, 40170-115, Salvador, BA, Brazil

²School of Botany, Trinity College Dublin, 1 College Green, Dublin 2, Ireland

³Royal Botanic Garden Edinburgh, 20a Inverleith Row, EH3 5LR, Edinburgh, Scotland, UK

⁴Instituto de Pesquisas Jardim Botânico do Rio de Janeiro (JBRJ), Rua Pacheco Leão, 915, 22460-030, Rio de Janeiro, RJ, Brazil

Correspondence

Débora C. Zuanny

Email: deborazuanny@gmail.com

Domingos Cardoso

Email: cardosobot@gmail.com

Funding information

Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB); Programa de Apoio à Pós-Graduação (PROAP) - Financial Code 001; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001; CNPq Research Productivity Fellowships (grant number 314187/2021-9)

Abstract

1. Data on plant distribution and diversity from natural history collections and taxonomic databases are increasingly becoming available online as exemplified by the RBG Kew's Plants of The World Online (POWO). This growing accumulation of biodiversity information requires an advance in bioinformatic tools for accessing and processing the massive data in downstream science.
2. We present herein the `expowo`, an open source package that facilitates extracting botanical data from POWO by reading the background HTML source code. We analysed this database and generated new functions to distil taxonomic and distributional information about all family, genera or species of flowering plants (Angiosperms). The `expowo`'s functions include the standardisation of the scientific plant family names as input, extraction of the global distribution at country and botanical subdivision levels, the number and complete list of species in each genus of any plant family, and the associated original publication. It is implemented in R and designed to handle the entire Angiosperm tree of life.
3. We demonstrate the utility of the package by mining lists of all plant species native to a set of biodiverse countries (Brazil, Indonesia, Madagascar, Malaysia, Mexico, Papua New Guinea and Thailand), and lists of all species in some of the most diverse and ecologically important families (Begoniaceae, Fabaceae and Poaceae).
4. By interacting with one of the world's most taxonomically comprehensive databases of plants that has been constantly updated by experts, `expowo` allows mining reliable data for downstream research. As such, the greatest strength of `expowo` is the extraction of curated data that can be readily exported in a spreadsheet and plotted in global distribution maps of species richness. Therefore, the `expowo` is a package that harnesses bioinformatic tools to bridge the gap between the most current botanical

online information and science user needs by providing global plant data in easy-to-handle outputs.

KEYWORDS

Angiosperm, biodiversity, collection, data mining, herbarium, POWO

1 | INTRODUCTION

Botanical collections provide extensive information to support science, education, and conservation efforts (Graham et al., 2004; Paton et al., 2020; Westwood et al., 2021; Walters & Pence, 2021). Open and publicly accessible botanical datasets have greatly improved our ability to produce comprehensive floristic checklists, taxonomic revisions, and phenology estimates, particularly for poorly studied tropical floras (Funk, 2003; Bebbier et al., 2010; Cardoso et al., 2017; Besnard et al., 2018; Sweeney et al., 2018; Park et al., 2023). To that end, a future goal is to create a “global metaherbarium” by digitising biodiversity data across the global botanical network of herbaria, which could become a key source of information on plant diversity and distribution (Davis, 2022). However, achieving this goal also requires developing bioinformatic tools to enable interactive exploration and easy download of these large botanical datasets, making the data available for downstream scientific fields, including taxonomy, conservation, ecology, evolution, and biogeography.

Several large botanical datasets have been developed to support research and conservation efforts. Current best examples of such large botanical datasets used by research and conservation include the Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>), which holds billions of specimen records from museums and herbaria worldwide. Another extensive database is the International Plant Names Index (IPNI,

<https://www.ipni.org/>), which provides a comprehensive list of all published Angiosperm names. The GenBank platform of the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>) holds publicly accessible DNA sequence data. More recently, global databases on all currently accepted plant species and their distribution have become available, including Kew's Plants of The World Online (POWO, <https://powo.science.kew.org/>). These datasets are supplied with information from the digitization processes of primary biodiversity data (Soberón & Peterson, 2004; Krishtalka & Humphrey, 2000), but not all information is readily available for use (Peterson *et al.*, 2018). Additionally, mistakes during data entry, translation between databases, and typographical errors can lead to misleading distribution and diversity patterns, causing severe inaccuracies in species lists (Maldonado *et al.*, 2015; Cardoso *et al.*, 2017). More recently, the issue of data standardisation has become increasingly important, especially because multiple databases are often required to address complex research questions (Willemse *et al.*, 2008).

Computational tools with capabilities to enable access, processing, cleaning, and analysing data are needed in order to tackle common issues involving biological information and to increase the value of botanical collections. To standardise the data available, some recently created R packages were developed to work with selected databases as taxonomic references (Kindt, 2020). For example, the `lcvplants` package is grounded on the Leipzig Catalog of Vascular Plants (LCVP; Freiberg *et al.*, 2020), and the `flora` package (Carvalho, 2020) interacts with the data available through the Flora e Funga do Brasil taxonomic database (<http://floradobrasil.jbrj.gov.br/>). Also, the use of GBIF data for understanding plant distribution and diversity patterns without expert checks, synonymy cleaning, and removal of cultivated records has been flagged up in several studies (Goodwin *et al.*, 2015; Cardoso *et al.*, 2017). Other packages, such as the `CoordinateCleaner` (Zizka *et al.*, 2019) and `rgbif` (Chamberlain *et al.*, 2021), have been extraordinarily useful tools for accessing and

cleaning plant biodiversity records available via GBIF (e.g., Moonlight et al., 2020; Maldonado et al., 2015; Guedes et al., 2018). On the other hand, the choice to use global taxonomically validated databases, such as POWO, is still with an issue of limited access, lacking a tool to download desired data.

Here, we introduce the new package `expowo` implemented in R (R Core Team, 2022) to perform automated plant taxonomic and geographical record extractions from the taxonomically verified POWO database. Unlike other R packages that deal with large botanical data by using APIs, `expowo` was designed to interact with POWO by reading the background HTML source pages for each plant taxa. We showcase the potential of `expowo` using two examples, where we produce a list of all accepted plant species for a set of countries, and lists of all accepted species for a set of economically and ecologically important families with associated distribution maps.

2 | PACKAGE DESCRIPTION

2.1 | Overview

The `expowo` R package is a comprehensive tool that comprises seven major functions, five auxiliary functions, and three associated data packages. It is specifically designed for extracting taxonomic and distribution data of accepted species and genera for any family (or a list of families) of flowering plants publicly available at RBG Kew's POWO taxonomic database. The latest package version is available on CRAN (<https://CRAN.R-project.org/package=expowo>) and GitHub (<https://github.com/DBOSlab/expowo/>) under the permissive MIT licence, making it easily accessible for use, copying, and modification.

The package workflow includes selecting the most appropriate function for the desired information, defining how the package will handle input data reception and check,

extracting and analysing the data, presenting the results either in a dataframe-formatted object or saving it as a comma-separated values (CSV) file in the designated working directory in a newly created folder. The results include the number of accepted species for any genus, as well as the full list of accepted species in any genus or family, their authorship, original publication, and global distribution at country and botanical country levels. Additionally, the `expowo` can easily generate global maps of species richness based on the queried taxonomic and distribution data. To further facilitate usage and comprehension, we offer complete tutorials for each function with examples and detailed vignettes, available on the package website (<https://dboslab.github.io/expowo/>).

3 | PACKAGE CONTENT

The package includes seven main functions (`powoFam`, `powoSpDist`, `powoGenera`, `powoSpecies`, `megaGen`, `topGen` and `powoMap`; Table 1) that requires only the name of the target family or a vector with multiple family names (Fig. 1). These functions work with three other main auxiliary functions (`getGenURI`, `getNumb` and `getDist`) to mine the plant data. The function `getGenURI` mines the URI for each genus, `getNumb` mines the total number of species within any genus, and `getDist` searches for native and introduced country-level distribution for any genus and species. Together, these functions can mine selected data and provide a full spreadsheet and global distribution maps of species richness as an output.

3.1 | Diversity data extraction

The lists can be extracted from R by providing a name of a plant family (or a vector of plant family names). Each major function automatically establishes a connection with the POWO

web page, by internally calling the respective URI of the queried family from the data package POWOCodes. Examples of custom scripts on how to use `expowo` functions are also given in the documentation. Because `expowo` uses the source code of each taxon page at POWO to scrape the data, the connection with the internet is mandatory.

Six of the seven main functions extract diversity and distribution data. The function `powoFam` was designed to count the total species number within any family and to save the queried results in a spreadsheet. The `megaGen` function uses a specified threshold number to return a list with all mega-diverse genera for any family of flowering plants and their associated number of accepted species. The default value of the threshold in the `megaGen` function is based on Frodin (2004), where it is defined that a genus must have more than five hundred species to be considered mega-diverse (or big). The function `topGen` is relatively similar to the previous one, while it selects all genera above a specific limit of species, but the `topGen` filters the most diverse genera within any family. The function `powoGenera` can generate a genus-level list with associated authorship, original publication of the protologue, the species number of each genus, the Kew ID and the URI of the web page, while the `powoSpecies` generates a species-level list with information about the associated authorship, original publication, the native distribution according to country and botanical country, the introduced distribution according to country and botanical country, the kew ID and the URI of the web page. The function `powoSpDist` generates the same type of spreadsheet, but it can mine the data for selected species by providing their names in the argument ‘species’. For those species missing information about the native or introduced distribution, the functions automatically assign as unknown.

3.2 | Extracting distribution lists

Data quality is a common issue when utilising plant distribution data. In our package, the process of extracting the distribution from each plant species was achieved via the auxiliary functions, and we were careful to choose an expert validated database and to internally standardise country names. Additionally, a major caveat within the POWO database is the original distribution information limited to just the TDWG’s botanical country level subdivision (level 3), without detail provided on political country level distribution. This may be somewhat problematic for comparative analyses in country-level conservation studies, because most effective conservation programs are often designed by national-level initiatives (Martins *et al.*, 2017; Ren *et al.*, 2019). To overcome this issue, we created the associated data package `botregions` that provides a dataframe with the global political country names and associated classification of botanical divisions. By doing so, all `expowo`’s functions that produce biodiversity lists are able to convert the botanical country distributions, as originally retrieved from POWO, into distributions across political country delimitation. As such, any list delivered by `expowo` will always include distribution at both country and botanical country levels.

3.3 | Plotting distributions

Our newly developed `powoMap` function uses the Natural Earth map data (<https://www.naturalearthdata.com/>) framework to plot global-scale maps of species richness at country and/or botanical country levels. Despite being originally designed to produce maps for any specific taxonomic level only extracted with `expowo`’s `powoSpecies` function, where the level is defined by the argument ‘`taxclas`’, `powoMap` now also accepts any dataframe-formatted input data that has at least a column with species names as binomial and the column(s) with corresponding distribution in the countries and/or botanical regions. A single map can be automatically produced for the entire input data or multiple richness maps

for any different taxonomic groups in a single run by just specifying a column name with the associated taxonomic classification. Additionally, this function can provide global maps according to botanical subdivision just by adjusting the arguments ‘botctrs’ and ‘distcol’. To generate publication-quality figures that best represent the species richness data graphically, we have designed the `powoMap` function using specific functions from `ggplot2` (Wickham, 2016). We also offer the possibility to display the colour scales of the maps based on colour palettes from both the packages `Viridis` and `RColorBrewer`, which provides visualisations that take into account colour blind users.

4 | EMPIRICAL EXAMPLES

We demonstrate `expowo`’s functionality with two sets of empirical examples: (1) Extracting species lists at country level for a set of biodiverse countries, and (2) Extracting species lists at family level for a set of economically and ecologically important plant families and mapping their diversity and distribution. See the Supporting Information for the scripts to generate the spreadsheets and maps with the full search results.

4.1 | Extracting species lists at country level

To show `expowo`’s power, we used `powoSpecies` to produce the entire list of accepted species of all angiosperm families available at POWO and generated global maps of species richness according to two different divisions of the world (Fig. 2). Overall, the output file resulted in a dataset of 334,132 rows and 4,347,161 cells of data. At country level, Brazil and Indonesia are revealed as the most speciose countries, while at the botanical country level, Colombia stands out as the most species-rich.

Additionally, we selected some countries that are renowned for their high species richness and endemism levels to produce their POWO-based list of accepted species. While several tropical countries are among the world's most biodiversity-rich, concerted efforts in summarising how diverse they are in plants have been made for only some of them (e.g., Brazil, BFG, 2021; Papua New Guinea, Cámara-Leret *et al.*, 2020; Mexico, Ulloa-Ulloa *et al.*, 2017; Madagascar, Antonelli *et al.*, 2022). Here, we have found the following numbers: Brazil (32,864 spp., Table S4), Papua New Guinea (12,306 spp., Table S5), Mexico (21,539 spp., Table S6), Malaysia (14,672 spp., Table S7), Madagascar (10,263 spp., Table S8), Indonesia (27,700 spp., Table S9) and Thailand (9,481 spp., Table S10). These species lists can serve as a starting point for directing actions in compiling a country checklist for those mega-diverse countries who still lack the documentation of their flora.

A recent expert-verified inventory showed more than one thousand species than the list retrieved from POWO and suggests that Papua New Guinea comprises the world's richest island flora, with 13,634 species identified. The list of plant names was compiled using additional databases than the one used here (see Cámara-Leret *et al.*, 2020). Additionally, a recently published review evidenced Madagascar's great endemism and species richness as a result of millions of years of evolution in isolation from Asia and Africa (Antonelli *et al.*, 2022). Furthermore, we compared the POWO results for Brazil with the data from Flora e Funga do Brasil (FFB) project, as a guide for future prospects. According to FFB database, there are 32,823 accepted native species in Brazil, which enables to retain its position as the country with the world's highest plant diversity (Forzza *et al.*, 2012), and indicates that it is home of even more species than previously shown (BFG, 2015). Finally, the data retrieved from POWO is slightly bigger than FBB, indicating areas to explore in both datasets to increase data quality.

4.2 | Extracting species lists at family level and generating global maps

In this empirical example, we use the `expowo` package to provide a global list of accepted species and genera in three economically and ecologically important plant families (Poaceae, Fabaceae and Begoniaceae). These examples show how the output files are produced from their queried lists of accepted species. We first queried the Kew's POWO database with the function `powoSpecies` in October (for Begoniaceae data) and November 2022 (for Poaceae and Fabaceae data). POWO currently accepts 11,811 species and 790 genera in Poaceae (Table S1), 22,186 species and 778 genera in Fabaceae (Table S2), and 1,994 species and 2 genera in Begoniaceae (Table S3). We also compared the total species number with two other databases. According to the Begonia Resource Centre (Hughes *et al.*, 2015-; <https://padme.rbge.org.uk/begonia/>), just the genus *Begonia* alone comprises 2,120 accepted species. Additionally, in the World Flora Online Plant List (WFO Plant List, <https://wfoplantlist.org/plant-list>, the old and superseded The Plant List), there are 12,648 species of Poaceae, 789 genera and 23,942 species of Fabaceae, and 2,196 species of Begoniaceae. These results demonstrate how `expowo` can be used to explore the dataset currently available online. While these figures show diverging species numbers across databases, we believe it is not related to the accuracy of the underlying data, since all of them are taxonomically validated. Rather, these differences are likely due to the frequency at which the databases are continuously maintained and updated. The `expowo` package mines what is presented online, performing web scraping to get the underlying data, and depends on maintenance efforts.

To represent the extracted data graphically, we generated global maps of species richness by country for those plant families (Fig. 3). The output shows that China and India have the most species for Poaceae, while Brazil, Australia and Tasmania are the most

country-rich for Fabaceae species, and Indonesia is the most species-rich country for Begoniaceae.

5 | CONCLUSIONS

Here, we have presented the new R package `expowo` by detailing all of its functionality and examples based on the data extraction of the authoritative and taxonomically verified RBG Kew's POWO database on global plant distribution and diversity. We further direct the package's users to utilise the vignettes in the website, which contains tutorials with examples. The package `expowo` represents another important step in facilitating and complementing the use of biological databases to foster further downstream botanical research. Specifically, the continued improvement of the package can increase the reproducibility in science and helps to flag types of data that need standardisation and refined curation while it generates spreadsheets and figures for publication. Now, all Angiosperm data can be automatically extracted from POWO, enabling to map their spatial distribution, which represents an important advance in documenting the world's flora. With the growing availability of data, it will be possible to take into account the distribution of endangered species and make better decisions about conservation strategies and delimitation of priority areas for funding. The native and introduced distribution areas produced for species/genera/families based on the `expowo` package are also of potential interest for many biogeographical and ecological research questions.

AUTHOR'S CONTRIBUTIONS

D.C.Z. and D.C. conceived the ideas and led the writing of the manuscript. D.C.Z. and D.C. wrote all R functions, the package, and the associated website. D.C. provided expertise on taxonomic aspects of the package. B.V. technically revised the R codes. D.C.Z. gathered the data and drafted the manuscript. D.C.Z., T.S. and P.M. designed the empirical examples. All authors revised the text and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We would like to thank the creators, maintainers, and everyone who contributed to these invaluable open databases, especially those working on the Kew's Plant of the World Online. We also thank Ludwig Baldaszti and Sara Mortara for the discussions about the R codes, and Rafael Pereira for the suggestions about the package's logo. This package is part of the first author's MSc Thesis developed at Programa de Pós-Graduação em Biodiversidade e Evolução and Universidade Federal da Bahia (UFBA), and was financially supported in part by the Programa de Apoio à Pós-Graduação (PROAP) - Financial Code 001 and the first author was supported by Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) scholarship. D.C.'s research in plant biodiversity is supported by grants from CNPq (Research Productivity Fellowship, grant 314187/2021-9), FAPESB (Edital Universal grant APP0037/2016) and Royal Society (Newton Advanced Fellowship grant NAF/R1/180331).

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

DATA ACCESSIBILITY

The package is available on CRAN (Appendix 1, <https://CRAN.R-project.org/package=expowo>) and on GitHub (developmental, <https://github.com/DBOSlab/expowo>). All scripts used to generate the examples are presented in the Supporting Information.

REFERENCES

- Antonelli, A., Smith, R. J., Perrigo, A. L., Crottini, A., Hackel, J., Testo, W., Farooq, H., Jiménez, M. F. T., Andela, N., Andermann, T., et al. (2022). Madagascar’s extraordinary biodiversity: Evolution, distribution, and use. *Science*, **378**, eabf0869. <https://www.science.org/doi/10.1126/science.abf0869>
- Bebber, D. P., Carine, M. A., Wood, J. R., Wortley, A. H., Harris, D. J., Prance, G. T., Davidse, G., Paige, J., Pennington, T. D., Robson, N. K. B., & Scotland, R. W. (2010). Herbaria are a major frontier for species discovery. *Proceedings of the National Academy of Sciences*, **107**, 22169–22171. <https://doi.org/10.1073/pnas.1011841108>
- Besnard, G., Gaudeul, M., Lavergne, S., Muller, S., Rouhan, G., Sukhorukov, A. P., Vanderpoorten, A., & Jabbour, F. (2018). Herbarium-based science in the twenty-first century. *Botany Letters*, **165**, 323–327. <https://doi.org/10.1080/23818107.2018.1482783>
- Cámara-Leret, R., Frodin, D. G., Adema, F., Anderson, C., Appelhans, M. S., Argent, G., Guerrero, S. A., Ashton, P., Baker, W. J., Barfod, A. S. et al. (2020). New Guinea has the world’s richest island flora. *Nature*, **584**, 579–583. <https://doi.org/10.1038/s41586-020-2549-5>
- Cardoso, D., Särkinen, T., Alexander, S., Amorim, A. M., Bittrich, V., Celis, M., Daly, D. C., Fiaschi, P., Funk, V. A., Giacomini, L. L., et al. (2017). Amazon plant diversity revealed

-
- by a taxonomically verified species list. *Proceedings of the National Academy of Sciences*, **114**, 10695–10700. <https://doi.org/10.1073/pnas.1706756114>
- Carvalho, G. (2022). *flora: Tools for Interacting with the Brazilian Flora 2020*. R package version 0.3.4, <URL: <https://CRAN.R-project.org/package=flora>>.
- Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., & Ram, K. (2022). *rgbif: Interface to the Global Biodiversity Information Facility API*. R package version 3.3.0, <URL: <https://CRAN.R-project.org/package=rgbif>>.
- Flora e Funga do Brasil. (2022). Jardim Botânico do Rio de Janeiro. <http://floradobrasil.jbrj.gov.br/>
- Forzza, R. C., Baumgratz, J. F. A., Bicudo, C. E. M., Canhos, D. A., Carvalho Jr, A. A., Coelho, M. A. N., Costa, A. F., Costa, D. P., Hopkins, M. G., Leitman, P. M. et al. (2012). New Brazilian floristic list highlights conservation challenges. *BioScience*, **62**, 39–45. <https://doi.org/10.1525/bio.2012.62.1.8>
- Freiberg, M., Winter, M., Gentile, A., Zizka, A., Muellner-Riehl, A. N., Weigelt, A., & Wirth, C. (2020). LCVP, The Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. *Sci Data*, **7**, 1–7. <https://doi.org/10.1038/s41597-020-00702-z>
- Frodin, D. G. (2004). History and concepts of big plant genera. *Taxon*, **53**, 753–776. <https://doi.org/10.2307/4135449>
- Funk, V. (2003). The importance of herbaria. *Plant Science Bulletin*, **49**, 94–95.
- Global Biodiversity Information Facility. (2017). Magnoliopsida. <https://doi.org/10.15468/dl.wquvxb>
- Goodwin, Z. A., Harris, D. J., Filer, D., Wood, J. R., & Scotland, R. W. (2015). Widespread mistaken identity in tropical plant collections. *Current Biology*, **25**, R1066–R1067. <https://doi.org/10.1016/j.cub.2015.10.002>

- 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. <https://doi.org/10.1016/j.tree.2004.07.006>
- Guedes, T. B., Sawaya, R. J., Zizka, A., Laffan, S., Faurby, S., Pyron, R. A., Bérnils, R. S., Jansen, M., Passos, P., Prudente, A. L. C., Cisneros-Heredia, D. F., Braz, H. B., Nogueira, C. C., Antonelli, A. (2018). Patterns, biases and prospects in the distribution and diversity of Neotropical snakes. *Global Ecology and Biogeography*, **27**, 14–21. <https://doi.org/10.1111/geb.12679>
- Hughes, M., Moonlight, P. W., Jara-Muñoz, A., Tebbitt, M. C., Wilson, H. P., Zhong, K. & Pullan, M. (2015–). Begonia Resource Centre. Online database available from <http://padme.rbge.org.uk/begonia/>. [Retrieved 06 December 2022].
- IPNI. (2022). International Plant Names Index. The Royal Botanic Gardens, Kew, Harvard University Herbaria & Libraries and Australian National Botanic Gardens. Published on the Internet <http://www.ipni.org>. [Retrieved 06 December 2022].
- Kindt, R. (2020). WorldFlora: An R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone data. *Applications in Plant Sciences*, **8**, e11388. <https://doi.org/10.1002/aps3.11388>
- Krishtalka, L., & Humphrey, P. S. (2000). Can natural history museums capture the future? *BioScience*, **50**, 611–617. [https://doi.org/10.1641/0006-3568\(2000\)050\[0611:CNHMCT\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2000)050[0611:CNHMCT]2.0.CO;2)
- Lima, R. A., Sánchez-Tapia, A., Mortara, S. R., ter Steege, H., & Siqueira, M. F. (2021). *plantR*: An R package and workflow for managing species records from biological collections. *Methods in Ecology and Evolution*, **00**, 1–8. <https://doi.org/10.1111/2041-210X.13779>

- Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., 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. <https://doi.org/10.1111/geb.12326>
- 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**, 347–356. <https://doi.org/10.3417/D-16-00009A>
- Moonlight, P. W., Silva de Miranda, P. L., Cardoso, D., Dexter, K. G., Oliveira-Filho, A. T.,
Pennington, R. T., Ramos, G., & Särkinen, T. E. (2020). The strengths and weaknesses
of species distribution models in biome delimitation. *Global Ecology and Biogeography*,
29, 1770–1784. <https://doi.org/10.1111/geb.13149>
- Park, D. S., Lyra, G. M., Ellison, A. M., Maruyama, R. K. B., Torquato, D. D. R., Asprino, R.
C., Cook, B. I., & Davis, C. C. (2023). Herbarium records provide reliable phenology
estimates in the understudied tropics. *Journal of Ecology*, **111**, 327–337.
<https://doi.org/10.1111/1365-2745.14047>
- Paton, A., Antonelli, A., Carine, M., Forzza, R. C., Davies, N., Demissew, S., Droge, G.,
Fulcher, T., Grall, A., Holstein, N. et al. (2020). Plant and fungal collections: Current
status, future perspectives. *Plants, People, Planet*, **2**, 499–514.
<https://doi.org/10.1002/ppp3.10141>
- Peterson, A. T., Asase, A., Canhos, D. A. L., Souza, S., & Wieczorek, J. (2018). Data leakage
and loss in biodiversity informatics. *Biodiversity Data Journal*, **6**, e26826.
<https://doi.org/10.3897/BDJ.6.e26826>
- Ren, H., Qin, H., Ouyang, Z., Wen, X., Jin, X., Liu, H., Lu, H., Liu, H., Zhou, J., Zeng, Y.,
Smith, P., Jackson, P. W., Gratzfeld, J., Sharrock, S., Xu, H., Zhang, Z., Guo, Q., Sun,
W., Ma, J., Hu, Y., Zhang, Q., & Zhao, L. (2019). Progress of implementation on the

-
- Global Strategy for Plant Conservation in (2011–2020) China. *Biological Conservation*, **230**, 169–178. <https://doi.org/10.1016/j.biocon.2018.12.030>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Soberón, J., & Peterson, T. (2004). Biodiversity informatics: Managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **359**, 689–698. <http://doi.org/10.1098/rstb.2003.1439>
- Sweeney, P. W., Starly, B., Morris, P. J., Xu, Y., Jones, A., Radhakrishnan, S., Grassa, C. J., & Davis, C. C. (2018). Large-scale digitization of herbarium specimens: Development and usage of an automated, high-throughput conveyor system. *Taxon*, **67**, 165–178. <https://doi.org/10.12705/671.9>
- The Brazil Flora Group. (2022). Brazilian Flora 2020: Leveraging the power of a collaborative scientific network. *Taxon*, **71**, 178–198. <https://doi.org/10.1002/tax.12640>
- Ulloa Ulloa, C., Acevedo-Rodríguez, P., Beck, S., Belgrano, M. J., Bernal, R., Berry, P. E., Brako, L., Celis, M., Davidse, G., Forzza, R. C., et al. (2017). An integrated assessment of the vascular plant species of the Americas. *Science*, **358**, 1614–1617. <https://doi.org/10.1126/science.aao0398>
- Walters, C., & Pence, V. C. (2021). The unique role of seed banking and cryobiotechnologies in plant conservation. *Plants, People, Planet*, **3**, 83–91. <https://doi.org/10.1002/ppp3.10121>
- Westwood, M., Cavender, N., Meyer, A., & Smith, P. (2021). Botanic garden solutions to the plant extinction crisis. *Plants, People, Planet*, **3**, 22–32. <https://doi.org/10.1002/ppp3.10134>

-
- 443 Wickham, H. (2016). *ggplot2: Create Elegant Data Visualisations Using the Grammar of*
444 *Graphics*. R package version 3.4.0, <URL: [https://CRAN.R-](https://CRAN.R-project.org/package=ggplot2)
445 [project.org/package=ggplot2](https://CRAN.R-project.org/package=ggplot2)>
- 446 Willemse, L. P., van Welzen, P. C., & Mols, J. B. (2008). Standardisation in data-entry across
447 databases: Avoiding Babylonian confusion. *Taxon*, **57**, 343–345.
448 <http://www.jstor.org/stable/25066008>
- 449 Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Ritter, C. D., Edler, D., Farooq, H.,
450 Herdean, A., Ariza, M., et al. (2019). *CoordinateCleaner*: Standardized cleaning of
451 occurrence records from biological collection databases. *Methods in Ecology and*
452 *Evolution*, **10**, 744–751. <https://doi.org/10.1111/2041-210X.13152>

TABLES

Table 1. All functions included in the `expowo` package and associated descriptions related to their specific functionalities when interacting with Kew’s Plants of the World Online (POWO).

Function	Description
megaGen	Extract mega-diverse genera from POWO
topGen	Extract topmost species-rich genera of any plant family at POWO
powoFam	Extract species number of any plant family from POWO
powoGenera	Extract list of genera of any family from POWO
powoSpecies	Extract list of species from POWO
powoSpDist	Extract list of species distributions from POWO
powoMap	Create global maps of species richness

FIGURES

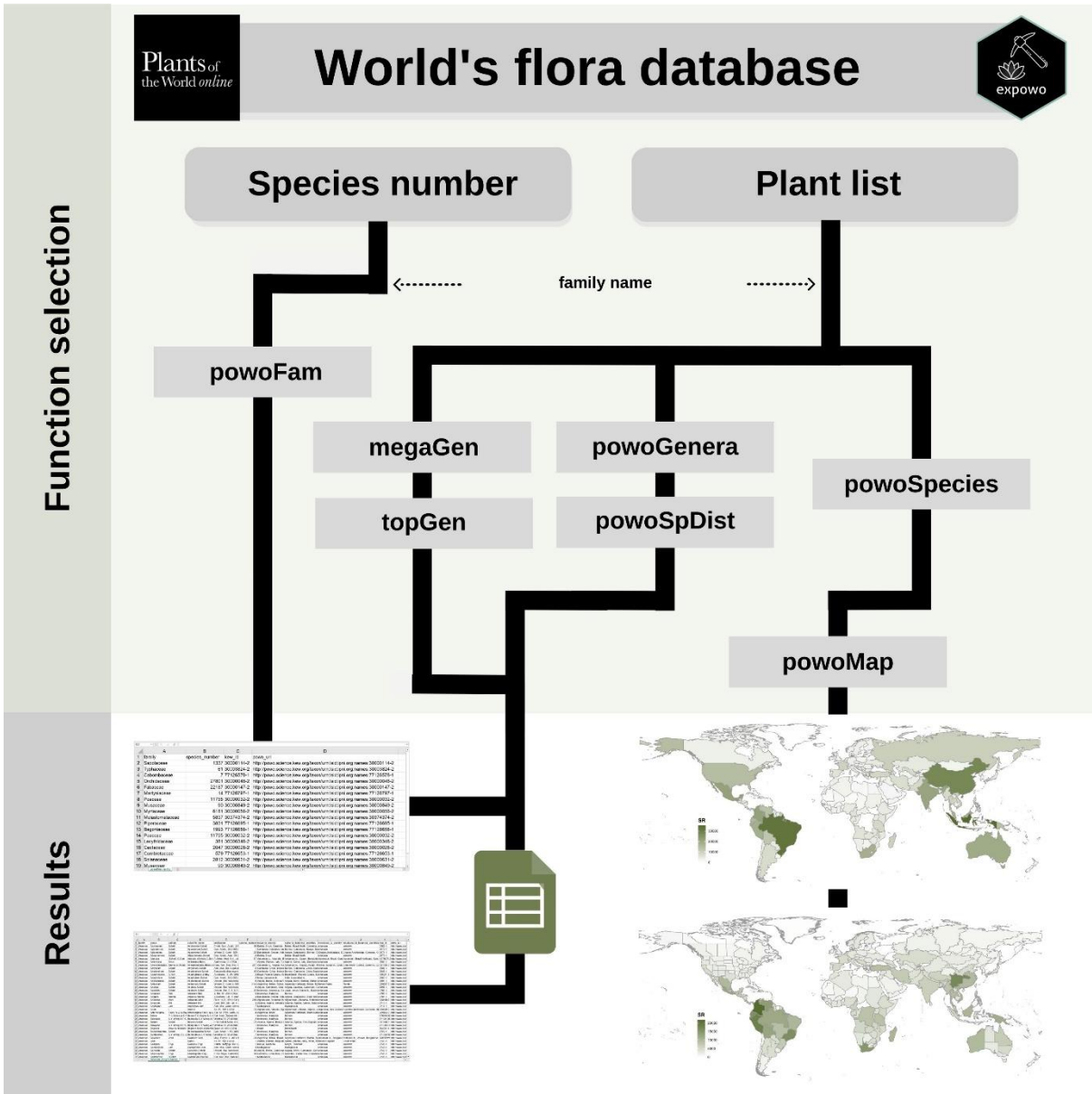


Figure 1 | Graphical representation of the structure and major functions of the `expowo` package for mining Kew’s Plants of the World Online database. The package is structured in two main objectives for the plant data extraction: providing species number for any angiosperm genus or family and a complete species list with associated global distribution. The arrows with dashed lines indicate the mandatory input information required for running the functions. The full lines in black connect the functions and the possible outputs: simple spreadsheets (top) and complete spreadsheets (bottom), and global distribution maps of species richness according to political country division (top) and botanical subdivision (bottom).

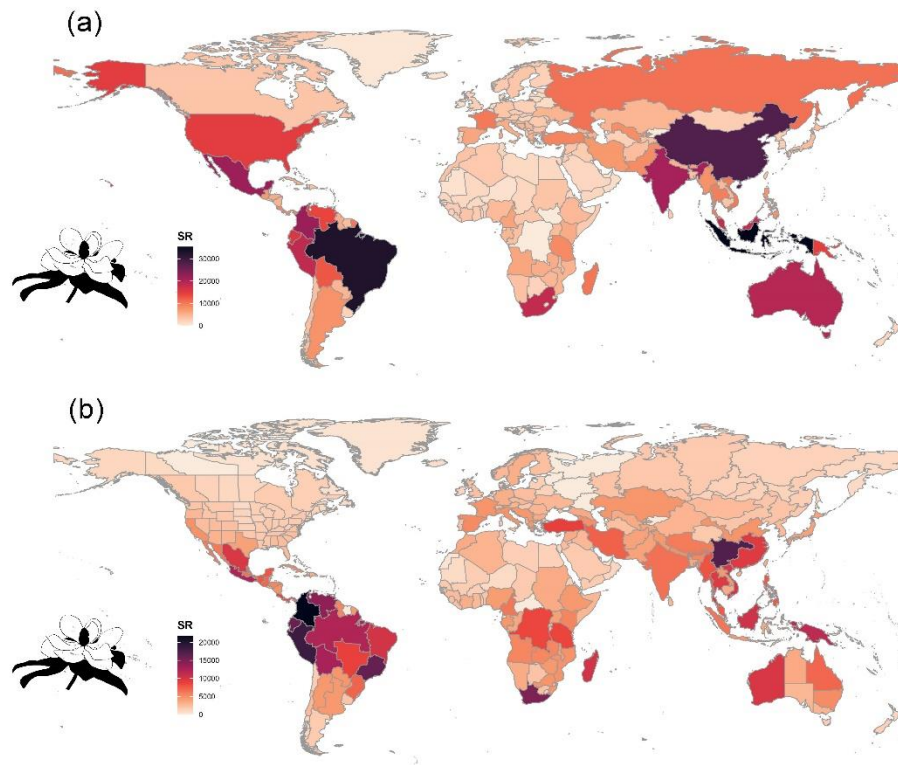


Figure 2 | Global native distribution of all Angiosperm species across (a) political country division of the world and (b) botanical subdivision. The data was extracted from Kew's Plants of the World Online using the function `powoSpecies` and plotted with `powoMap` of the `expowo` package. The scale bar and the global map is colored according to each number of species richness with the rocket color scale from `Viridis` package.

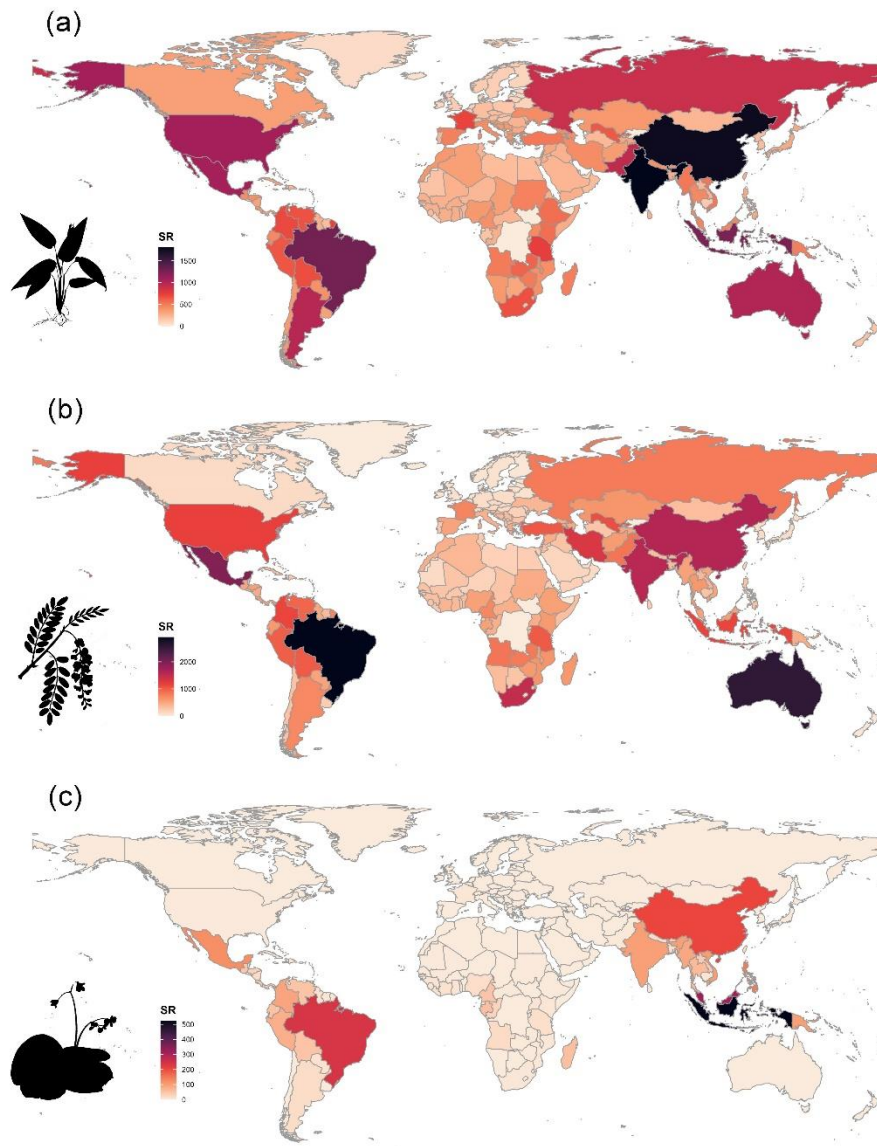


Figure 3 | Global native distribution of all accepted species across the political country division of the world within (a) Poaceae, (b) Fabaceae and (c) Begoniaceae. The data was extracted from Kew's Plants of the World Online using the function `powoSpecies` and plotted with `powoMap` of the `expowo` package. The scale bar and the global map is colored according to each number of species richness with the rocket color scale from `Viridis` package. The silhouettes of *Anomochloa marantoidea*, *Robinia pseudoacacia* (from PhyloPic, <http://phylopic.org/>) and *Begonia joshii* (by David A. Purvis) display the contour of their typical morphology as representatives of all species of the respective selected families included in the maps.

492 SUPPORTING INFORMATION

493 **Methods S1.** Custom R Scripts to extract the data used for each example and plot the maps
494 using our newly developed expowo package.

Supporting Information Methods S1

Article title: expowo: An R package for mining global plant diversity and distribution data

Authors: Débora C. Zuanny, Bruno Vilela, Peter W. Moonlight, Tiina E. Särkinen and Domingos Cardoso

Contents

Reproducing the examples

- Usage 1
- Extracting and plotting the data from all species of angiosperms 2
- Extracting and plotting the data for each selected plant family 3

Usage

We have created **expowo** to automatize the process of extracting data from Kew's, Plants of The World Online database using R. Our goal is to pull species lists with all associated data of any plant family. The empirical examples presented in the paper makes uses of this new R package. If you do not already have it, install **expowo** from CRAN and load it with:

```
install.packages("expowo")  
library(expowo)
```

Here we will show how to extract data of all plant species and for each selected plant family in the paper, and then use this data to plot global maps of species richness distribution.

Extracting and plotting the data from all species of angiosperms

To create the dataset with all angiosperms species available at Plants of the World Online, use the following script:

```
all_angio <- expowo::powoSpecies(family = POWOcodes$family,
                                hybridspp = FALSE,
                                country = NULL,
                                verbose = TRUE,
                                save = TRUE,
                                dir = "results_powoSpecies/",
                                filename = "all_plant_species")
```

With this dataset, it was possible to filter for the species occurrence in selected species-rich countries. To generate figures of the global native distribution according to political countries and botanical division, use the newly created object “all_angio” as input in powoMap. Both global maps generated in this step were included in Figure 2 of the present paper.

```
# To create one map according to political countries
expowo::powoMap(inputdf = all_angio,
                botctrs = FALSE,
                distcol = "native_to_country",
                taxclas = NULL,
                verbose = FALSE,
                save = TRUE,
                vir_color = "rocket",
                bre_color = NULL,
                leg_title = "SR",
                dpi = 600,
                dir = "results_powoMap/",
                filename = "Angio_global_richness_country_map",
                format = "jpg")

# To create one map according to botanical subdivision
expowo::powoMap(inputdf = all_angio,
                botctrs = TRUE,
                distcol = "native_to_botanical_countries",
                taxclas = NULL,
                verbose = FALSE,
                save = TRUE,
                vir_color = "rocket",
                bre_color = NULL,
                leg_title = "SR",
                dpi = 600,
                dir = "results_powoMap/",
                filename = "Angio_global_richness_botcountry_map",
                format = "jpg")
```

Extracting and plotting the data for each selected plant family

We can extract the data and generate individual lists with all POWO's accepted species within some of the most diverse and ecologically important families (Begoniaceae, Fabaceae and Poaceae) using a custom R script with the package expowo.

```
# Query for Poaceae using powoSpecies
poa_data <- expowo::powoSpecies(family = "Poaceae",
                                hybridspp = FALSE,
                                country = NULL,
                                verbose = TRUE,
                                save = TRUE,
                                dir = "results_powoSpecies/",
                                filename = "all_species_Poaceae")

# Query for Fabaceae using powoSpecies
fab_data <- expowo::powoSpecies(family = "Fabaceae",
                                hybridspp = FALSE,
                                country = NULL,
                                verbose = TRUE,
                                save = TRUE,
                                dir = "results_powoSpecies/",
                                filename = "all_species_Fabaceae")

# Query for Begoniaceae using powoSpecies
beg_data <- expowo::powoSpecies(family = "Begoniaceae",
                                hybridspp = FALSE,
                                country = NULL,
                                verbose = TRUE,
                                save = TRUE,
                                dir = "results_powoSpecies/",
                                filename = "all_species_Begoniaceae")
```

Then, the function creates species lists with all POWO's accepted species of the selected plant families Poaceae, Fabaceae, and Begoniaceae. This data can be used for mapping the global native distribution with powoMap. The output generated with the following R code composed the Fig. 3 of the present paper.

```
# To create one map for Poaceae according to political countries
expowo::powoMap(inputdf = poa_data,
                botctrs = FALSE,
                distcol = "native_to_country",
                taxclas = "family",
                verbose = FALSE,
                save = TRUE,
                vir_color = "rocket",
                bre_color = NULL,
                leg_title = "SR",
                dpi = 600,
                dir = "results_powoMap/",
                filename = "Poaceae_global_richness_country_map",
                format = "jpg")

# To create one map for Fabaceae according to political countries
expowo::powoMap(inputdf = fab_data,
```

```

    botctrs = FALSE,
    distcol = "native_to_country",
    taxclas = "family",
    verbose = FALSE,
    save = TRUE,
    vir_color = "rocket",
    bre_color = NULL,
    leg_title = "SR",
    dpi = 600,
    dir = "results_powoMap/",
    filename = "Fabaceae_global_richness_country_map",
    format = "jpg")

# To create one map for Begoniaceae according to political countries
expowo::powoMap(inputdf = beg_data,
    botctrs = FALSE,
    distcol = "native_to_country",
    taxclas = "family",
    verbose = FALSE,
    save = TRUE,
    vir_color = "rocket",
    bre_color = NULL,
    leg_title = "SR",
    dpi = 600,
    dir = "results_powoMap/",
    filename = "Begoniaceae_global_richness_country_map",
    format = "jpg")

```

Capítulo II

Big genera in Leguminosae: Overview of taxonomic changes and global distribution

*Manuscrito a ser submetido para o periódico Plants, People, Planet
(ISSN: 2572-2611)*



Big genera in Leguminosae: Overview of taxonomic changes and global distribution

Débora C. Zuanny^{1*} | Domingos Cardoso^{1,2*}

¹Instituto de Biologia, Universidade Federal da Bahia, Rua Barão de Jeremoabo, s.n., Ondina,
40170-115, Salvador, BA, Brazil

²Instituto de Pesquisas Jardim Botânico do Rio de Janeiro (JBRJ), Rua Pacheco Leão, 915,
22460-030, Rio de Janeiro, RJ, Brazil

*Corresponding authors: DCZ, deborazuanny@gmail.com; DC, cardosobot@gmail.com

Abstract

Societal Impact Statement

Big genera comprise ca. 15% of all angiosperm species and have a great ecological and economic importance. In the Leguminosae family, big genera have achieved a cosmopolitan distribution because of their incredible evolutionary success in different biomes. Here, we used an authoritatively updated checklist of big legume genera to explore their main changes in size over the years and global species richness distribution. The results presented here offer insights into the species assembly of big genera and lay the groundwork for further taxonomic and biogeographic studies.

Summary

- We used the taxonomically verified Kew's Plants of The World Online (POWO) to reveal an updated checklist of the largest genera in the ecologically and economically important legume family (Leguminosae or Fabaceae), to assess where they are distributed and how their circumscription changed over the years since the last review of big angiosperm genera in 2004. The global distribution patterns of the species diversity and nomenclatural history across the big legume genera were explored using the newly developed R package expowo.
- The biggest legume genera are currently comprised of a total of 7,411 species in the following seven genera: *Astragalus* (3,066 spp.), *Acacia* (1,086 spp.), *Indigofera* (731 spp.), *Crotalaria* (712 spp.), *Lupinus* (604 spp.), *Oxytropis* (612 spp.) and *Mimosa* (600 spp.). This order of most speciose genera and their associated species numbers contrast with previous estimates. While the species diversity of *Astragalus* and *Acacia* have decreased, *Astragalus* is still revealed as the biggest angiosperm genus, and

Lupinus, *Oxytropis*, and *Mimosa* stood out as the new big genera by reaching the threshold of more than 500 species.

- The outstanding species richness of the Leguminosae as a whole is repeatable in its big genera, which still continue increasingly growing, mostly due to the discovery and publication of many new species in recent years. Lastly, we found that species richness across the big genera is not equally predominantly associated with tropical countries.

1 INTRODUCTION

Exceptionally species-rich genera in the plant tree of life are those with more than 500 species (Frodin, 2004) and have been called as big genera (Corner, 1961) or mega-diverse genera (Moonlight *et al.*, 2017). Genera are the first nomenclatural category above the species level and where species are inserted (Turland *et al.*, 2018). There were periods in the history of taxonomy when some changes in genus delimitation prevailed, and the actual tendency is to maintain and create new big genera, which is related to large scale more than new molecular data (Humphreys & Linder, 2009). After more than one century of expansion, there were 50 genera recognized as mega-diverse (Mabberley, 1997), or 57 (Frodin, 2004). Because big genera have great representativeness across all flowering plants (angiosperms), comprising ca. 15% of all known species, it is fundamental to study them in order to summarize the global plant diversity in floristic databases, such as World Flora Online (WFO, 2022) and Plants of the World Online (POWO, 2022). While these databases are necessary to fulfill the goals of the Global Strategy for Plant Conservation (Sharrock *et al.*, 2018), the sheer size of big genera makes them unfeasible to monograph, reflecting in research focused mostly on geographically-restricted questions and encouraging the creation of infrageneric groups (e.g. Moonlight *et al.*, 2017).

Our still poor knowledge of the big plant genera, who and how big they are, and the mechanisms behind changes in their circumscriptions are reflected in our current underestimates of their global species richness. Despite the rates at which we have been losing biodiversity by deforestation (Giam, 2017; Antonelli, 2021; Ralimanana *et al.*, 2022), the ongoing discovery and description of new species are one of the main factors leading to increased genus size. While these top ten genera represent only 5% of all angiosperm species (Frodin, 2004; SOTWP, 2020), the new discoveries reported to them comprise c. 10% of the

angiosperm species described in the period (Bebber et al., 2010), suggesting a disproportionate number of new species descriptions that are still found in big genera.

Taxonomic lumping and splitting can also play a role in defining genus size. Lumping is related to an expanded generic concept, either by amalgamating paraphyletic genera or the inclusion of a satellite monophyletic genus (e.g., Forrest & Hollingsworth, 2003; Jacobs et al., 2008). On the other hand, splitting often involves a grade of a paraphyletic genus where the phylogenetic separation of multiple smaller clades is warranted for delimiting them into different genera (e.g., Miller & Siegler, 2012). As these mechanisms changed the genus size in the last fifteen years, the phylogenetic knowledge of the plant genera also increased over time, accompanied by updates in infrageneric classifications (Moonlight et al., 2017).

Here, we explore the Leguminosae family, also known as Fabaceae (Lewis & Schirire, 2003), as a model to show how its big genera changed over time, and how their species richness is globally distributed. The monophyly of the Leguminosae is well supported and it comprises ca. 765 genera and ca. 19.500 species, standing out as the third largest plant family in species richness, the second in economic importance (LPWG, 2017), and a good proxy to the overall angiosperm diversity (Yahara et al., 2013). Also, legumes are strategic for food and nutrition security, where well-known species of *Glycine* Willd. (soybean), *Lathyrus* L. (pea), *Medicago* L. (alfafa) and *Phaseolus* L. (common beans) are largely used. Yet, many other little known or geographically localized legume species have been explored for human and livestock consumption. The exportation of legumes is growing through the years (Food and Agriculture Organisation [FAO]: <http://www.fao.org/pulses-2016/en/>), as a response to its value as an inclusive source of protein and micronutrients (Yahara et al., 2013) for different types of diets. Many species are used in traditional medicine (Saslis-Lagoudakis et al., 2011), and general psychoactivity is found (Halse-Gramkow et al., 2016). The family also has

historical and cultural importance to some countries, such as Brazil, where the tree *Paubrasilia echinata* (Lam.) Gagnon, H.C.Lima & G.P.Lewis was exploited over centuries and gave rise to the country name. Additionally, Leguminosae is a cosmopolitan family, occurring in a high diversity of habitats and allowing us to explore global distribution patterns. The evolutionarily stable ecological association of legumes with various nitrogen-fixing bacteria (Faria *et al.*, 2022; Werner *et al.*, 2015) also explains why the family is ubiquitous and dominant across the world's major biomes (LPWG, 2013).

In the same way as it remains uncertain how much plant diversity we are losing (Rivers *et al.*, 2011), we do not know at what rate new species are being discovered. Also, it is difficult to estimate the actual species richness in different families due to distinct stages of taxonomic knowledge and associated uncertainty created by taxonomic lumping and splitting and reclassifications (Stropp *et al.*, 2022). Because the legume big genera are known to comprise a vast amount of the species within the family (Schrire *et al.*, 2005), investigating its global diversity and distribution may be helpful to understand how plant radiations have increased and successfully colonized worldwide. Grounded on our recent efforts towards assembling a global taxonomically verified checklist for the entire legume family (LPWG, 2022; POWO, 2022), in this study we address the following knowledge gaps: (i) to identify the big legume genera and to estimate the species number of each genus, (ii) to determine their global distribution and (iii) to summarize how their circumscription changed after the last review in 2004.

2 MATERIALS AND METHODS

2.1 Assembling the taxonomic and global distribution data

In setting the scope of this research, we defined the same absolute threshold of species number to be directly comparable to the last review (Frodin, 2004), which is a minimum of 500 species. For addressing the taxonomy, we chose the Kew's Plants of the World Online (POWO) database that is verified by taxonomic experts to create our dataset and explore the patterns within the big legume genera. We used the function `megaGen` of the newly developed package `expowo` (Zuanny *et al.*, submitted., <https://github.com/DBOSlab/expowo>) to download data and identify the big genera within Fabaceae. Following the data collection of big legume genera, we used the function `powoSpecies` of the same package to automatically extract the complete species list of each one of these genera. The extracted data using this function is according to the data publicly available at POWO in March 2023 and comes in a spreadsheet including: (i) family; (ii) genus; (iii) authors; (iv) species number; (v) scientific name; (vi) native and introduced distribution; (vii) the Kew ID; and the (viii) web page URI. We also extracted the list of all legume genera and their respective species number in December 2022 to compare and evaluate the threshold value to be considered big using the function `powoGenera` from the `expowo` package.

2.2 Bibliographic review

We used the data extracted from POWO and the list of big genera from Frodin (2004) as starting point to search for the species number in the protologue of each recognized big legume genera in order to access their original species number and compare the changes in size over time. We used the Tropicos (<https://tropicos.org>) database to search for the original publication and manually collected each species number. We also used important literature of Leguminosae and Angiosperm taxonomy (Brown, 1883; Mabberley, 1997; Frodin, 2004) to assess the species number in selected periods of time and we included the current species

number of the big legume genera, accessed by the data extracted from POWO in the previous step, to compare and evaluate the changes.

2.3 Nomenclatural data collection

In order to assess the major processes underlying the changes in generic size over the years, we compiled a list of nomenclatural data of the big legume genera also from Plants of the World Online, hosted by The Royal Botanic Gardens, Kew. We extracted the nomenclatural data using a custom R script. Then, we categorized the data according to: a) new births/species; b) new names due to taxonomic change. We included in the first category the species with absence of change in the year of valid publication and without synonyms, while all the species which the original publication was found in one of their synonyms were included in the second category. Hence, this step consisted in cleaning the data and comparing the year of publication of the current accepted name with their synonyms and selecting the year of their first valid publication. All data entries with unknown publication were excluded. The species with information about *orth. var.*, *nom. cons.*, *sensu auct.*, *nom. nud.*, *nom. provis.*, *nom. rej.*, *not. validly publ.*, *no type indicated*, *nom. rej. prop.*, *not accepted by the author*, *incomplete basionym ref.*, *pro syn.*, *no Latin descr. or type*, *no type*, *opus utique oppr.*, *contrary to Art. 37.6 (Vienna Code, 2006)*, *no latin descr.*, *tentatively listed as a synonym*, *with incorrect basionym ref.*, *contrary to Art. 37.5 ICBN (1988)*, *nom. illeg.*, *not validly publ.*, *nom. subnud.*, *combination is not actually made*, *nom. superfl.* were not included in the comparison, as we are only considering formal standing names and valid publications. Then, we used this data to create individual graphs of accumulation of species for each one of the seven big genera.

2.4 Geospatial analysis of global species richness

To examine the distribution patterns, we generated global maps of species richness with the native occurrence data extracted in the previous steps of all Leguminosae species and for each big genera. We used the function `powoMap` of `expowo` package based on two geographical divisions to plot the distribution of each genus: political country and botanical countries according to the level 3 of the World Geographical Scheme for Recording Plant Distributions (WGSRP, <https://www.tdwg.org/standards/wgsrpd/>). Despite that the data extracted from POWO also include the introduced distribution, we are only considering the native distribution for this study. To color the legend bar and the maps according to species richness, we used the `rocket` scale color from `Viridis` R package, which is fully implemented within the `expowo`'s function `powoMap`.

3 RESULTS

3.1 Taxonomic species diversity across the big legume genera

The list of Leguminosae extracted from POWO's database in March 2023 shows that 778 genera are currently accepted and includes seven big genera: *Astragalus* L. (3,066 spp.), *Acacia* Mill. (1,086 spp.), *Crotalaria* L. (712 spp.), *Indigofera* L. (731 spp.), *Lupinus* L. (604 spp.), *Mimosa* L. (600 spp.), and *Oxytropis* DC. (612 spp.). The taxonomic history and the estimation of species number of the seven big legume dates as further back as 1753 and are summarized in Table 1.

Overall, we extracted 7,411 species for the seven big genera and 118,576 entries of data. Also, those big genera represent ca. 33% of all legume species (22,186 spp.). For more information about the data mining, see the Supplementary Material S1. The remaining legume genera have between one and 366 species. In the extracted list, the top ten most species-rich genera include the seven big genera followed by *Chamaecrista* Moench (366

spp.), *Tephrosia* Pers. (358 spp.) and *Aspalathus* L. (291 spp.). This result reinforces our choice in maintaining the threshold in 500 species, because there is no genus near the threshold that was not included and, within these three remaining rich genera, only *Chamaecrista* were in the last review.

Regarding the changes in size at species rank for the currently accepted names, there were 3,762 new species described and 3,649 taxonomic changes for the seven big genera published between 1753 and 2022 (POWO, 2023): 1,523 new species and 1,543 taxonomic changes for *Astragalus*, 100 new descriptions and 986 taxonomic changes for *Acacia*, 499 new species and 213 taxonomic changes for *Crotalaria*, 381 new species and 350 taxonomic changes for *Indigofera*, 445 new species and 159 taxonomic changes for *Lupinus*, 376 new species and 236 taxonomic changes for *Oxytropis*, and 438 new species and 162 taxonomic changes for *Mimosa* (Figure 5). The genus *Astragalus*, formerly the first largest genus of this family and the biggest of all angiosperms in species richness according to some databases (e.g., WFO and POWO), continues increasing in size over time.

3.2 Global distribution of the species richness

The global distribution of all Leguminosae species is illustrated in Figure 1. The country-level botanical regions that are rich in legumes are found from Southwestern United States to South Brazil in the Neotropics; in Southern and Central China to Thailand, and from India to Turkey (highlight to Iran) and Kazakhstan in Asia; in the Democratic Republic of the Congo, Tanzania, South Africa, Cameroon, Angola, and Zambia in Africa; and mainly in Western Australia in Oceania. Overall, the most species-rich botanical divisions are Western Australia, Iran, and Brazil North. The representation of the geospatial data of Leguminosae species richness with a fully global coverage shows that they are mainly tropical or

subtropical, with exceptions being made to some extent of the prevalence in temperate regions for *Astragalus*, *Oxytropis*, and *Lupinus*.

Different patterns of distribution have emerged for each big genus (Figures 1–4). For *Astragalus*, the species richness is almost restricted to the northern hemisphere, exceptionally high in Iran (Figure 1c and 1d), whereas its more phylogenetically closely related genus *Oxytropis* is more diverse in the northern hemisphere, mainly in the countries Russia and China (Figure 4c), as well as the Kazakhstan botanical subdivision (Figure 4d). The *Acacia* species richness is higher in Australia, with a few occurring in other countries from Oceania and Southeast Asia (Figure 2a and 2b). The *Crotalaria* species richness is scattered, with many in Central and Eastern Africa and India (Figure 2c and 2d). The species richness of *Indigofera* is higher in the Old World, predominantly associated with the country South Africa but also high in South to East Africa, Southeast Asia, and Eastern Australia (Figure 3a and 3b). Notably, *Lupinus* and *Mimosa* species share the distribution essentially in the New World, but with different patterns. While *Lupinus* has many species in Peru and US (Figure 3c) and western botanical countries of the Americas (Figure 3d), Brazil stands out as the most species-rich country for *Mimosa*, mostly occurring in Brazil West-Central and Brazil Southeast (Figure 4a and 4b).

4 DISCUSSION

4.1 The current size of big genera

Four of the seven big legume genera were among the 57 big genera of angiosperms listed by Frodin (2004). The exception is to *Lupinus*, *Mimosa* and *Oxytropis*, which recently reached the threshold of 500 species. We found that some big legume genera were first described by Linnaeus (1753) in *Species Plantarum*, such as *Astragalus* (p. 755–762), *Lupinus* (p. 721–

722), *Mimosa* (p. 516–523), *Indigofera* (p. 751), *Cassia* (p. 376–380), *Crotalaria* (p. 714–716), while the data for *Acacia* were extracted from Miller (1754) in The Gardeners Dictionary, and *Oxytropis* from De Candolle (1802) in Astragalologia (p. 25–28). Additionally, it was noticeable that the genus *Cassia* L., present in the checklists since 1883, was not considered big in our query.

The size and order of some big legume genera changed noticeably since the last review (Frodin, 2004). The order changed because the genus *Chamaecrista* fell to the eight position, while *Lupinus*, *Oxytropis* and *Mimosa* overtook its place and surpassed the threshold of at least 500 species. The genus concept adopted for *Cassia* in the previous lists is the same as for the subtribe Cassiinae, which also comprises the much more species-rich genera *Senna* and *Chamaecrista*. This interpretation of genera for *Cassia* s.l. goes back to Linnaeus, who combined them in the beginning of 1700. Then, this combination continued in subsequent monographs to maintain stabilization, only changing with new knowledge associated with characters (Frodin, 2004). The generic concept for *Cassia* and related genera *Senna* and *Chamaecrista* used in POWO has been the same since Irwin & Barneby's (1982) last treatment of the South American Cassiinae, which has also been confirmed in all molecular phylogenetic analyses (Marazzi et al, 2006; Conceição et al., 2009).

It was proposed that the genus *Astragalus* itself is not more speciose than the others in the Astragalean clade (*Oxytropis* also included). Actually, the richness is a characteristic of this temperate herbaceous clade and is suggested to not seek explanations for the unusual diversity of the genus alone (Sanderson & Wojciechowski, 1996), but for the entire Astragaleae. It was already expected that *Astragalus* would be recognized as big. The species number of this genus is remarkable since its first description (De Candolle, 1802), with 142 species. Also, we found three new genera reaching the threshold. Now we know how much

the species richness of the genera *Mimosa* and *Lupinus* contributes to the Latin American diversity, a region that outlines ca. 749 new species being described annually and comprising at least a third of all vascular plant diversity (Raven *et al.*, 2020).

Not only the size of each genus changed since the last review. Historical phylogenetic conflicts affected the inference of species relationships and our knowledge about the biodiversity in the genus *Acacia*. After morphological studies and phylogenetic information indicating its lack of monophyly, *Acacia* underwent a taxonomic upheaval. Furthermore, a new type species was proposed to this genus (*A. penninervis* Sieber ex DC.) by Orchard & Maslin (2003) and the conservation of many Australian species was maintained (Orchard & Maslin, 2005). Thus, phylogenetic reconstruction showing *Acacia* s.l. as a polyphyletic group and its reduced number of species was provided by Miller & Siegler (2012). After the retypification, many names were synonymized to *Vachellia* Wight & Arn., *Senegalia* Raf., *Acaciella* Britton & Rose and *Mariosousa* Seigler & Ebinger (Seigler & Ebinger, 2010; Miller & Seigler, 2012), reducing the species number of the still big genus *Acacia* s.s and provoking debates about territorial chauvinism (Kull & Rangan, 2012). Changes in legume classification came up apace as a reflection of the non-monophyly of *Acacia* s.l. and to recircumscribe Caesalpinioideae.

Additionally, our results show that the changes in size of the big legume genera are mostly due to new species descriptions, excepting *Astragalus* and *Acacia* that have undergone more taxonomic changes. Many species were previously discovered and described with other names, and we only recognized the diversity of each genera much later (Figure 5). The most striking example is *Oxytropis*: while it appears that we only recognized ca. 90% of its diversity in 2000, actually many species already have been discovered and, therefore, synonymized in the future. The growing accumulation of new names that we have revealed

here show the results of the continuing investments in floristics inventories and taxonomic treatments and that those genera did not reach the nomenclatural stability threshold. It could be explained either because: (a) South America has many undescribed species diversity left to be discovered and many taxonomists studying these big genera; (b) freely available botanical databases (e.g. Tropicos, IPNI, speciesLink), allowing remote studies and new discoveries; (c) more recent collections available and in need for taxonomic description; (d) the faster electronic publication. These factors, together with the three main processes that lie behind the changes in genus size, may explain why the *Acacia* had a dramatic decrease in species diversity mostly due to splitting. Also, *Astragalus* is the biggest genus either considering all Leguminosae genera or in comparison to any other angiosperm genus (disregarding the composite genus *Hieracium* L. – due to many apomictic species) and is the one that is growing faster due either by new descriptions (e.g., Bacchetta & Brullo, 2006; Castillón *et al.*, 2016; Erkul *et al.*, 2022; Dönmez & Aydin, 2018) and taxonomic changes. However, several taxonomic changes also caused the reduction of the size of this genus (e.g., Choi *et al.*, 2015; Ekici *et al.*, 2015), since some species were synonymized.

4.2 Distribution patterns of big genera

The global distribution pattern of Leguminosae species richness reinforces the extraordinary ecological and evolutionary success of this family in occupying all major biomes (Schrire *et al.*, 2005). The diversity of life-forms, from herbs to giant long-lived trees, and ecological strategies helps to understand its cosmopolitan presence. Also, the legume distribution shown here is similar to what was found to vascular plant species (Brummitt *et al.*, 2020). Thereby, our results corroborate that Leguminosae have an incredible species richness in Amazon

Rainforest (BFG, 2015), indicated as the Brazil North region according to botanical subdivision.

Concerning the big legume genera, *Mimosa*, *Acacia*, *Crotalaria* and *Indigofera* are more species-rich in the tropical Grass and Succulent biomes, whereas *Astragalus*, *Oxytropis* and *Lupinus* are more diverse in the Temperate biome. Brown (2014) suggests that the great number of species occurrence in the tropics is because the temperature is the most important factor shaping the ecological and evolutionary rates, while temperate regions tend to contain younger species and higher levels of extinction. But our results show elevated species richness in the temperate zone too. The distribution of the big genera in temperate regions thus might be explained by phylogenetic niche conservatism (Wiens & Donoghue, 2004) or with dispersal capacity.

The species richness of the big genera *Mimosa* and *Acacia* s.s differs to the pattern of amphiatlantic disjunction narrowly associated to the Succulent biome described in Schrire et al. (2005). This difference is due to our maps showing their high richness in the regions where the tropical biomes Grass, Succulent and Rainforest (for *Mimosa*) and the Grass and Temperate biomes (for *Acacia*) are delimited. However, these findings subscribe to their prevalence in the Tropics. Regarding the evolutionary history of *Acacia*, the major diversification occurred during the emergence of the arid biome in Australia during the Miocene (Miller et al., 2013; Hua et al., 2022). Nonetheless, a recent research conducted with Australian species of *Acacia* and four concepts of biomes showed that this genus does not have biome conservatism across its clades (Dale et al., 2019). Thus, the pattern reconstructed here is a useful snapshot of the overall distribution for this genus. Regarding *Mimosa*, a genus of some widespread invasive weed species, we have shown that it is mainly restricted to tropical America, while others restricted its distribution to the Savanna domain in Brazil

entitled Cerrado (Barneby, 1991; Simon & Proença, 2000). Another tropical genus is *Crotalaria*, having a richness nucleus in the Central and Eastern Africa and occupying mostly both Grass and Rainforest biomes.

Finally, the genus *Indigofera* has diversified mainly in Africa (Figure 3b) but also achieving pantropical disjunct distribution throughout the succulent biome (Schrire *et al.*, 2005), where it has a great ecological predilection to open environments dominated by succulent-rich species. The oldest clades within this genus (Cape and Tethyan) originated in the semi-arid succulent biome and have limited dispersion than those species inhabiting Savanna and Tropical rainforests. Thus, time and dispersion limits might explain its spatial pattern of species richness according to what was proposed by Wiens & Donoghue (2004). Also, dispersal is the main biogeographic process explaining the distribution of *Indigofera*, and some species are endemic to high altitudes in Africa (e.g., *I. sokotrana* Vierh.), with many characters associated with this biome shift (Schrire *et al.*, 2009).

With respect to the temperate genera, we found that the highest center of diversity for *Astragalus* is in central Asia, in the northern temperate hemisphere of the Old World. The ecological predilection of *Astragalus* is mainly for Mediterranean habitats and *Oxytropis* occurs in cold and temperate grassland and in high altitudes. There are some species of *Astragalus* with occurrence in dry conditions (Podlech, 1996), but our results agree with Schrire *et al.* (2005), where it is shown that *Astragalus* and *Oxytropis* have a display in the temperate biome. Also, our results are similar to a recently published study focused on *Astragalus* in the Old World that shows more diversity and endemism in Iran, Turkey, and Afghanistan in Southwest Asia (Maassoumi & Ashouri, 2022).

The genus *Lupinus* is nested within the core Genistoids clade, one of the papilionoid clades that have high predominance in temperate regions (Schrire *et al.*, 2005). Here, exceptionally

high species richness was found for this genus in Peru. Additionally, its distribution in the western botanical countries of the Americas is similar to the Andes mountain chain disposal, denoting its presence in regions of elevated altitudes. Our results match with what was found by Hughes & Eastwood (2006), which considers *Lupinus* endemic and highly diverse in the tropical high Andes, and with Legumes of the World Online, which indicates species occurring mainly in California and in the Andes. Also, the shift from annual to perennial habit in alpine plants such as *Lupinus* suggests a causal relation with its high rates of diversification (Nevado et al., 2016) in those habitats. However, the existence of species in Central Brazil might be correlated to the adaptation to fire in the Cerrado lineages (Simon et al., 2009).

Several studies show the alpine regions of the world as the centers of endemism and auspicious to evolutive irradiations of our biodiversity. It is remarkable how much the endemism is correlated with mountain uplifts, specially of the Andes and Himalaya, and with climate changes (Hughes & Atchinson, 2015; Hua et al., 2022). Herein, we could notice species of *Oxytropis*, *Lupinus* and *Indigofera* with occurrence data in areas related to high altitude. Furthermore, other patterns arose regarding the distribution of the big legume genera. The Neotropical region is largely known as the most species rich in the world (Antonelli & Sanmartín, 2011), and we found here that some genera prevail in the Neotropics (*Lupinus* and *Mimosa*).

4.3. Big legume genera hold ecological, cultural and economic potential

In addition to playing an important ecological role across biomes and their biogeographic history, legume species greatly contribute to the global economy. The seven biggest legume genera harbor importance that run through cultures, with many species being used as

ornamental plants, food resource, medicine and as national symbol. However, the list of species approved to be used as medicine or under use as food resource is disproportionate in relation to the current size of these genera. It is exceptionally interesting when dealing with a group that has many species with the same use and other many disappearing due to climate change. In fact, this points out the great potential to be unveiled on big legume genera beyond the mainstream crops (Ulian *et al.*, 2020).

Due to years of degradation with pasturelands and agriculture, the tropical region dominated by Savanna includes many lands favourable to restoration, and the Brazilian Cerrado already was spotted with four areas of hotspots for restoration (Lewis *et al.*, 2022). A recent study suggested how diverse the seed-sources markets can be to restore the Brazilian Cerrado's biodiversity and that the main traded species for seed supply are legumes (Silva *et al.*, 2022). As described here, the extraordinary diversity of woody and herbaceous species of *Mimosa* and *Crotalaria*, predominantly associated with grasslands and savannas, can become key genera to the seed-based reforestation of this striking, yet highly threatened biome against global biome homogenization. Also, some studies show the potentialities of *Acacia* and *Mimosa* species in phytoremediation of soils contaminated with arsenic (Cipriani *et al.*, 2013), while others indicate a great potential in arsenic absorption by the highly nitrogen-fixing genus *Crotalaria* (Melo *et al.*, 2015).

Studies on nourishing and medicinal plants also pervade through legumes. The largest legume genus *Astragalus* also has a broad range of uses. It has great medicinal importance in treatment of neurological illnesses and has more than 200 compounds discovered (Elkader *et al.*, 2022). For example, the roots of *A. trimestris* L. are commonly used in traditional Chinese medicine, mostly documented to treat urinary and respiratory diseases and allowed to be used as functional food (Chau & Hu, 2006). Yet, pharmacological research has only

been performed for only a few species, while many other potentially useful species are still lacking information (Amiri *et al.*, 2020). Another member of Astragaleae is readily used in Mongolian traditional medicine. The wide application for *Oxytropis pseudoglandulosa* Gontsch. ex Grubov provoked suggestions for its inclusion in the food industry (Narangerel *et al.*, 2021), and its pharmacophylogenetics characteristics have already been assessed (Jia *et al.*, 2022). Phylogenetic tools can be addressed to provide explanatory power to ethnobotanical use (Saslis-Lagoudakis *et al.*, 2011a). As the cross-cultural phylogenetic patterns in ethnobotany is not random throughout the angiosperm phylogeny, Leguminosae was indicated for many medicinal uses and having more medicinal plants than other families (Saslis-Lagoudakis *et al.*, 2011b). More specifically, it makes sense to continue to look for shared chemistry and activity in species within this clade, since it already comprises many valuable species to traditional medicine. Although with good responses in humans, *Astragalus* and *Oxytropis* species are also called locoweeds due to its poison effect on livestock.

The third largest legume genus *Indigofera* is also highly economically important (Gerometta *et al.*, 2020). Some species are known to produce indigo dye, such as *I. tinctoria* Forssk (= *I. articulata* Gouan) and *I. suffruticosa* Mill., and as pasture and adornment (Burkill, 1995, p.857 apud Schrire, 2013, p. 282). Within this genus, the Pantropical clade (ca. 305 spp.) comprises more species with medicinal use, while the Paleotropical clade (185 spp.) have more species used as cattle feed (Schrire, 2013).

Legumes are tightly associated with culture and history, on which striking examples can be also found across the big legume genera. For example, the roots of the dry-forest inhabiting species *Mimosa tenuiflora* (Willd.) Poir. are used in the Pankararé indigenous sacred rituals of Northeast Brazil to produce the “wine of Jurema”, consumed to heal the body and soul

(Santos et al., 2012). Also, the species of other big genera, *Lupinus mutabilis* Sweet, is one crop of the hotspot of domesticated species with origin in the Incan civilization (Atchison et al., 2016). The Golden Wattle *Acacia pycnantha* Benth. is the Australian floral emblem due to its bloom display akin to the national colors, the wide distribution range across the country and its resilience in face of environmental changes. Then, this species is represented in the coat of arms of Australia and has a national day related to the beginning of its flowering period to promote plantation and heritage remembrance. Finally, there are many covered potentialities for those big genera that we can access in future research.

CONCLUDING REMARKS

Robust taxonomic information is fundamental in answering questions on different levels of biological sciences. Big legume genera are useful to ascertain many biological trends due to its broad geospatial range and species adapted to the major biomes. Further research could be conducted to investigate traits that enable species to occupy some specific niches and have great evolutive success, associated with a phylogenetic approach to address changes in the diversification rates through this family. It has been suggested that key apomorphies can led to species-rich clades (Sanderson & Donoghue, 1994), and this can be better understood by studying big legume genera. Also, we suggest future efforts in creating stable and natural infra-generic delimitations for these big genera.

Here, we explored the distribution of species richness in a global perspective with the focus on the biggest legume radiations. However, an interesting, still open topic to be approached in further studies is to explore and tackle species-poor or depauperate generic branches ("depauperons" sensu Donoghue & Sanderson, 2015) in the legume tree of life. We found here that more than a half of Leguminosae genera is monospecific or its species number is

less or equal than five species. This result shows how much species-poor genera are also pervasive in legumes. New directions can be settled with the data generated here, and their patterns of distribution combined with intrinsic dispersal capacity might hide some hints about its diversification. Finally, we suggest that their biogeographic history should be accessed with their phylogenies, in order to identify possible more diverse sister clades (Sanmartín & Meseguer, 2016). For example, in the highly diverse Papilionoideae, it is interesting that many depauperons are often remarkably marked by distinctive radial floral symmetry or non papilionate flowers (Cardoso et al., 2013).

Many studies were conducted to estimate biodiversity (Purvis & Hector, 2000), and others to identify priority species-rich environments and geopolitical regions to obtain more international investment (see Brummitt et al., 2020). But the estimation runs into difficulty when we consider incongruences in phylogenetic data and the lack of refinement in the distribution. What would be interesting now is to assess countries with floras partially or totally published online to refine the distribution highlighted here and assess the biome and phytogeographic domain predilection at a fine scale for all species. The Global Biodiversity Facility (GBIF, <https://www.gbif.org>) is another ongoing database that could contribute to addressing the distribution patterns of big genera.

ACKNOWLEDGEMENTS

We thank Dr. Haroldo Cavalcante de Lima for his precious suggestions to this paper. We also wish to express our gratitude to the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) for providing a scholarship for the first author and to the Programa de Apoio à Pós-Graduação (PROAP) - Financial Code 001, that financially supported part of this research. DC's research in plant biodiversity is supported by grants from CNPq (Research

Productivity Fellowship, grant no. 314187/2021-9), FAPESB (Edital Universal grant no. APP0037/2016) and Royal Society (Newton Advanced Fellowship grant no. NAF/R1/180331). This research is part of the first author's MSc Thesis developed at Programa de Pós-Graduação em Biodiversidade e Evolução and Universidade Federal da Bahia (UFBA).

CONFLICT OF INTEREST

The authors declare no conflict of interest. The conclusions in this publication are those of the authors and do not represent any official Kew's Plants of the World Online determination or policy.

DATA AVAILABILITY

The data supporting this research is retrieved from Kew's Plants of the World Online (<http://www.plantsoftheworldonline.org>).

REFERENCES

- Amiri, M. S., Joharchi, M. R., Nadaf, M. & Nasseh, Y. (2020). Ethnobotanical knowledge of *Astragalus* spp.: The world's largest genus of vascular plants. *Avicenna Journal of Phytomedicine*, **10**, 128–142. <https://doi.org/10.22038/AJP.2019.14240>
- Antonelli, A. & Sanmartín, I. (2011). Why are there so many plant species in the Neotropics? *Taxon*, **60**, 403–414. <https://doi.org/10.1002/tax.602010>
- Antonelli, A., Fry, C., Smith, R. J., Simmonds, M. S. J., Kersey, P. J., Pritchard, H. W., Abbo, M. S., Acedo, C., Adams, J., Ainsworth, A. M., Allkin, B., Annecke, W.,

- 492 Bachman, S. P., Bacon, K., Bárríos, S., Barstow, C., Battison, A., Bell, E., Bensusan, K.,
493 ... Zhang, B. G. (2020). *State of the World's Plants and Fungi 2020*. Royal Botanic
494 Gardens, Kew. <https://doi.org/10.34885/172>
- 495 Antonelli, A. (2022). The rise and fall of Neotropical biodiversity. *Botanical Journal of the*
496 *Linnean Society*, **199**, 8–24. <https://doi.org/10.1093/botlinnean/boab061>
- 497 Atchison, G. W., Nevado, B., Eastwood, R. J., Contreras-Ortiz, N., Reynel, C., Madriñán, S.,
498 Filatov, D. A. & Hughes, C. E. (2016). Lost crops of the Incas: Origins of domestication
499 of the Andean pulse crop tarwi, *Lupinus mutabilis*. *American Journal of Botany*, **103**,
500 1592–1606. <https://doi.org/10.3732/ajb.1600171>
- 501 Bacchetta, G. & Brullo, S. (2006). Taxonomic revision of the *Astragalus genargenteus*
502 complex (Fabaceae). *Willdenowia*, **36**, 157–167. <https://doi.org/10.3372/wi.36.36113>
- 503 Barneby, R. C. (1991). *Sensitivae censitae: A description of the genus Mimosa Linnaeus*
504 *(Mimosaceae) in the New World*. Memoirs of the New York Botanical Garden, ISSN
505 0071–5794; ISSN v. 65, 835p.
- 506 Bebbier, D. P., Carine, M. A., Wood, J. R., Wortley, A. H., Harris, D. J., Prance, G. T., ... &
507 Scotland, R. W. (2010). Herbaria are a major frontier for species discovery. *Proceedings*
508 *of the National Academy of Sciences*, **107**, 22169–22171.
509 <https://doi.org/10.1073/pnas.1011841108>
- 510 BFG - The Brazil Flora Group. (2015). Growing knowledge: An overview of seed plant
511 diversity in Brazil. *Rodriguésia*, **66**, 1085–1113. [https://doi.org/10.1590/2175-](https://doi.org/10.1590/2175-7860201566411)
512 [7860201566411](https://doi.org/10.1590/2175-7860201566411)
- 513 Brown, N. E. (1883). *The "Genera Plantarum"*. Gard. Chron., II, 19: 733–734.
- 514 Brown, J. H. (2014). Why are there so many species in the tropics? *Journal of biogeography*,
515 **41**, 8–22. <https://doi.org/10.1111/jbi.12228>

- 516 Brummitt, N., Araújo, A. C. & Harris, T. (2020). Areas of plant diversity—What do we
517 know? *Plants, People, Planet*, **3**, 33–44. <https://doi.org/10.1002/ppp3.10110>
- 518 Cardoso, D., Pennington, R. T., Queiroz, L. P., Boatwright, J. S., Van Wyk, B. -E.,
519 Wojciechowski, M. F. & Lavin, M. (2013). Reconstructing the deep-branching
520 relationships of the papilionoid legumes. *South African Journal of Botany*, **89**, 58–75.
521 <https://doi.org/10.1016/j.sajb.2013.05.001>
- 522 Castillón, E. E., González-Elizondo, M. S. & Quintanilla, J. Á. V. (2016). A new species of
523 *Astragalus* (Fabaceae, Faboideae) from Durango, Mexico. *Phytotaxa*, **288**, 91–95.
524 <https://doi.org/10.11646/phytotaxa.288.1.10>
- 525 Choi, I. S., Kim, S. Y. & Choi, B. H. (2015). A taxonomic revision of *Astragalus* L.
526 (Fabaceae) in Korea. *Korean Journal of Plant Taxonomy*, **45**, 227–238.
527 <https://doi.org/10.11110/kjpt.2015.45.3.227>
- 528 Conceição, A. S., Queiroz, L. P., Lewis, G. P., Andrade, M. J. G., Almeida, P. R. M.,
529 Schnadelbach, A. S. & van den Berg, C. (2009). Phylogeny of *Chamaecrista* Moench
530 (Leguminosae–Caesalpinioideae) based on nuclear and chloroplast DNA regions. *Taxon*,
531 **58**, 1168–1180. <https://doi.org/10.1002/tax.584010>
- 532 Chau, C. F. & Wu, S. H. (2006). The development of regulations of Chinese herbal medicines
533 for both medicinal and food uses. *Trends in Food Science & Technology*, **17**, 313–323.
534 <https://doi.org/10.1016/j.tifs.2005.12.005>
- 535 Cipriani, H. N., Dias, L. E., Costa, M. D., Campos, N. V., Azevedo, A. A., Gomes, R. J.,
536 Fialho, I.F. & Amezquita, S. P. M. (2013). Arsenic toxicity in *Acacia mangium* Willd.
537 and *Mimosa caesalpiniaefolia* Benth. seedlings. *Revista Brasileira de Ciência do Solo*,
538 **37**, 1423–1430. <https://doi.org/10.1590/S0100-06832013000500031>
- 539 Corner, E. J. H. (1961). *Evolution*. In McLeoud, A. M. & Cobley, L. S. (eds.). Contemporary
540 Botanical Thought: Oliver & Boyd, Edinburgh.

- 541 Dale, E. E., Larcombe, M. J., Lee, W. G. & Higgins, S. I. (2020). Diversification is
542 decoupled from biome fidelity: *Acacia*—a case study. *Journal of Biogeography*, **47**, 538–
543 552. <https://doi.org/10.1111/jbi.13768>
- 544 Dönmez, A. A. & Aydin, Z. U. (2018). *Astragalus ihsanalisii* (Fabaceae), a new species
545 from Erzurum province, E Turkey. *Willdenowia*, **48**, 399–404.
546 <https://doi.org/10.3372/wi.48.48309>
- 547 Donoghue, M. J. & Sanderson, M. J. (2015). Confluence, synnovation, and depauperons in
548 plant diversification. *New Phytologist*, **207**, 260–274. <https://doi.org/10.1111/nph.13367>
- 549 Ekici, M., Akan, H. & Aytac, Z. (2015). Taxonomic revision of *Astragalus* L. section
550 *Onobrychoidei* DC. (Fabaceae) in Turkey. *Turkish Journal of Botany*, **39**, 708–745.
551 <https://doi.org/10.3906/bot-1405-41>
- 552 Elkader, H. A. E. A., Essawy, A. E. & Al-Shami, A. S. (2022). *Astragalus* species:
553 Phytochemistry, biological actions and molecular mechanisms underlying their potential
554 neuroprotective effects on neurological diseases. *Phytochemistry*, **202**, 113293.
555 <https://doi.org/10.1016/j.phytochem.2022.113293>
- 556 Erkul, S. K., Duman, H. & Ateş, M. A. (2022). *Astragalus oksutdagensis* (Fabaceae), a new
557 species from Turkey. *Nordic Journal of Botany*, **2022**. <https://doi.org/10.1111/njb.03237>
- 558 Faria, S. M., Ringelberg, J. J., Gross, E., Koenen, E. J., Cardoso, D., Ametsitsi, G. K. D.,
559 Akomatey, J., Maluk, M., Tak, N., Gehlot, H. S. ... & James, E. K. (2022). The
560 innovation of the symbiosome has enhanced the evolutionary stability of nitrogen
561 fixation in legumes. *New Phytologist*, **235**, 2365–2377. <https://doi.org/10.1111/nph.18321>
- 562 Flora e Funga do Brasil. Jardim Botânico do Rio de Janeiro. Retrieved September 2022, from
563 <http://floradobrasil.jbrj.gov.br>

-
- 564 Forrest, L. L. & Hollingsworth, P. M. (2003). A recircumscription of *Begonia* based on
 565 nuclear ribosomal sequences. *Plant Systematics and Evolution*, **241**, 193–211.
 566 <https://doi.org/10.1007/s00606-002-0033-y>
- 567 Frodin, D. G. (2004). History and concepts of big plant genera. *Taxon*, **53**, 753–776.
 568 <https://doi.org/10.2307/4135449>
- 569 Gerometta, E., Grondin, I., Smadja, J., Frederich, M. & Gauvin-Bialecki, A. (2020). A review
 570 of traditional uses, phytochemistry and pharmacology of the genus *Indigofera*. *Journal of*
 571 *Ethnopharmacology*, **253**, 112608. <https://doi.org/10.1016/j.jep.2020.112608>
- 572 Giam, X. (2017). Global biodiversity loss from tropical deforestation. *Proceedings of the*
 573 *National Academy of Sciences*, **114**, 5775–5777.
 574 <https://doi.org/10.1073/pnas.1706264114>
- 575 Halse-Gramkow, M., Ernst, M., Rønsted, N., Dunn, R. R. & Saslis-Lagoudakis, C. H. (2016).
 576 Using evolutionary tools to search for novel psychoactive plants. *Plant Genetic*
 577 *Resources*, **14**, 246–255. <https://doi.org/10.1017/S1479262116000344>
- 578 Hua, X., Cardillo, M. & Bromham, L. (2022). Adapting to extremes: Reconstructing
 579 evolution in response to changing climate over time and space in the diverse Australian
 580 plant genus *Acacia*. *Journal of Biogeography*, **49**, 727–738.
 581 <https://doi.org/10.1111/jbi.14339>
- 582 Hughes, C. & Eastwood, R. (2006). Island radiation on a continental scale: Exceptional rates
 583 of plant diversification after uplift of the Andes. *Proceedings of the National Academy of*
 584 *Sciences*, **103**, 10334–10339. <https://doi.org/10.1073/pnas.0601928103>
- 585 Hughes, C. E. & Atchison, G. W. (2015). The ubiquity of alpine plant radiations: From the
 586 Andes to the Hengduan Mountains. *New Phytologist*, **207**, 275–282.
 587 <https://doi.org/10.1111/nph.13230>

- 588 Humphreys, A. M. & Linder, H. P. (2009). Concept versus data in delimitation of plant
589 genera. *Taxon*, **58**, 1054–1074. <https://doi.org/10.1002/tax.584002>
- 590 Irwin, H.S. & Barneby, R. C. (1982) The American Cassiinae: A synoptical revision of
591 Leguminosae tribe Cassieae subtribe Cassiinae in the New World. *Memoirs of the New*
592 *York Botanical Garden*, **35**, 1–918.
- 593 Jacobs, S.W.L., Gillespie, L.J. & Soreng, R.J. (2008). New combinations in *Hookerochloa*
594 and *Poa* (Gramineae). *Telopea*, **12**, 273–278. <https://doi.org/10.7751/telopea20085816>
- 595 Jia, X., Liu, Y., Wang, S., Ma, J., Yu, J., Yue, X., Zang, Y. & Wang, X. (2022). Screening of
596 metabolic markers present in *Oxytropis* by UHPLC-Q-TOF/MS and preliminary
597 pharmacophylogenetic investigation. *Frontiers in Plant Science*, **13**.
598 <https://doi.org/10.3389/fpls.2022.958460>
- 599 Kull, C. A., & Rangan, H. (2012). Science, sentiment and territorial chauvinism in the *Acacia*
600 name change debate. *Terra Australis*, **34**, 197–220.
601 <https://doi.org/10.22459/TA34.01.2012.09>
- 602 Linnaeus, C. (1753). *Species plantarum: exhibentes plantas rite cognitatas, ad genera relatas,*
603 *cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus,*
604 *secundum systema sexuale digestas.* Impensis Laurentii Salvii, Holmiae, vol. 2, 673 p.
605 Available at <https://www.biodiversitylibrary.org/item/13830>
- 606 Lewis, G. P., Schrire, B., Mackinder, B. & Lock, M. (2005). *Legumes of the World.* Royal
607 Botanic Gardens, Kew, 577p.
- 608 Lewis, K., Barros, F. D. V., Moonlight, P. W., Hill, T. C., Oliveira, R. S., Schmidt, I. B.,
609 Sampaio, A. B., Pennington, R. T. & Rowland, L. (2023). Identifying hotspots for
610 ecosystem restoration across heterogeneous tropical savannah-dominated regions.
611 *Philosophical Transactions of the Royal Society B*, **378**, 20210075.
612 <https://doi.org/10.1098/rstb.2021.0075>

- 613 LPWG - The Legume Phylogeny Working Group. (2013). Legume phylogeny and
614 classification in the 21st century: Progress, prospects and lessons for other species-rich
615 clades. *Taxon*, 62, 217–248. <https://doi.org/10.12705/622.8>
- 616 LPWG - The Legume Phylogeny Working Group. (2017). A new subfamily classification of
617 the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon*, 66, 44–
618 77. <https://doi.org/10.12705/661.3>
- 619 LPWG - Legume Phylogeny Working Group. (2022). The World Checklist of Vascular
620 Plants (WCVP): Fabaceae (R. Govaerts, Ed.). The Royal Botanic Gardens, Kew.
621 http://sftp.kew.org/pub/data_collaborations/Fabaceae/DwCA/
- 622 Maassoumi, A. A. & Ashouri, P. (2022). The hotspots and conservation gaps of the mega
623 genus *Astragalus* (Fabaceae) in the Old-World. *Biodiversity and Conservation*, 31, 1–21.
624 <https://doi.org/10.12705/661.3>
- 625 Mabberley, D. J. 1987. *The Plant-book*. Cambridge Univ. Press, Cambridge, Ed. 2, 680p.
- 626 Marazzi, B., Endress, P. K., Queiroz, L. P. & Conti, E. (2006). Phylogenetic relationships
627 within *Senna* (Leguminosae, Cassiinae) based on three chloroplast DNA regions:
628 Patterns in the evolution of floral symmetry and extrafloral nectaries. *American Journal*
629 *of Botany*, 93, 288–303. <https://doi.org/10.3732/ajb.93.2.288>
- 630 Melo, R. F., Dias, L. E. & Assis, I. R. (2015). Potencial de espécies de leguminosas para
631 recuperação de solo contaminado por arsênio. *Revista Brasileira de Geografia Física*, 08,
632 938–949.
- 633 Miller, P. 1754. *The Gardeners Dictionary*. Abridged Edition, London, Ed. 4, vol. 1.
- 634 Miller, J. T. & Seigler, D. S. (2012). Evolutionary and taxonomic relationships of *Acacia* s.l.
635 *Australian Systematic Botany*, 25, 217–224. <https://doi.org/10.1071/SB11042>
- 636 Miller, J. T., Murphy, D. J., Ho, S. Y., Cantrill, D. J., & Seigler, D. (2013). Comparative
637 dating of *Acacia*: combining fossils and multiple phylogenies to infer ages of clades with

- poor fossil records. *Australian Journal of Botany*, 61, 436–445.
<https://doi.org/10.1071/BT13149>
- Moonlight, P. W, Ardi, W., Arroyo Padilla, A., Chung, K. F., Fuller, D. et al. (2017).
 Dividing and conquering the fastest growing genus: Towards a natural sectional
 classification of the mega-diverse genus *Begonia* (Begoniaceae). *Taxon*, 67, 267–323.
<https://doi.org/10.12705/672.3>
- Narangerel, T., Bonikowski, R., Jastrąbek, K., Kunicka-Styczyńska, A., Plucińska, A.,
 Śmigielski, K., Majak, I., Bartos, A. & Leszczyńska, J. (2021). Chemical and biological
 characteristics of *Oxytropis pseudoglandulosa* plant of Mongolian origin. *Molecules*, 26,
 7573. <https://doi.org/10.3390/molecules26247573>
- Nevado, B., Atchison, G. W., Hughes, C. E., & Filatov, D. A. (2016). Widespread adaptive
 evolution during repeated evolutionary radiations in New World lupins. *Nature*
Communications, 7, 1–9. <https://doi.org/10.1038/ncomms12384>
- Orchard, A. E. & Maslin, B. R. (2003). Proposal to conserve the name *Acacia* (Leguminosae:
 Mimosoideae) with a conserved type. *Taxon*, 52, 362–363.
<https://doi.org/10.2307/3647418>
- Orchard, A. E. & Maslin, B. R. (2005). The case for conserving *Acacia* with a new type.
Taxon, 54, 509–512. <https://doi.org/10.2307/25065384>
- Podlech, D. (1986). Taxonomic and phytogeographical problems in *Astragalus* of the Old
 World and South-West Asia. *Proceedings of the Royal Society of Edinburgh, Section B:*
Biological Sciences, 89, 37–43. <https://doi.org/10.1017/S0269727000008885>
- POWO. (2022). Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew.
 Published on the Internet; Retrieved November 2022, from
<http://www.plantsoftheworldonline.org>

- 662 Purvis, A. & Hector, A. (2000). Getting the measure of biodiversity. *Nature*, 405, 212–219.
663 <https://doi.org/10.1038/35012221>
- 664 Ralimanana, H., Perrigo, A. L., Smith, R. J., Borrell, J. S., Faurby, S., Rajaonah, M. T.,
665 Randriamboavonjy, T., Vorontsova, M. S., Cooke, R. S., Phelps, L. N. ... & Antonelli, A.
666 (2022). Madagascar’s extraordinary biodiversity: Threats and opportunities. *Science*,
667 378, 1–13. <https://doi.org/10.1126/science.adf1466>
- 668 Raven, P. H., Gereau, R. E., Phillipson, P. B., Chatelain, C., Jenkins, C. N. & Ulloa Ulloa, C.
669 (2020). The distribution of biodiversity richness in the tropics. *Science Advances*, 6,
670 eabc6228. <https://doi.org/10.1126/sciadv.abc6228>
- 671 Rivers, M. C., Taylor, L., Brummitt, N. A., Meagher, T. R., Roberts, D. L. & Lughadha, E.
672 N. (2011). How many herbarium specimens are needed to detect threatened species?
673 *Biological Conservation*, 144, 2541–2547. <https://doi.org/10.1016/j.biocon.2011.07.014>
- 674 Sanmartín, I. & Meseguer, A. S. (2016). Extinction in phylogenetics and biogeography: from
675 timetrees to patterns of biotic assemblage. *Frontiers in Genetics*, 7, 1–17.
676 <https://doi.org/10.3389/fgene.2016.00035>
- 677 Sanderson, M. J. & Donoghue, M. J. (1994). Shifts in diversification rate with the origin of
678 angiosperms. *Science*, 264, 1590–1593. <https://doi.org/10.1126/science.264.5165.1590>
- 679 Sanderson, M. J. & Wojciechowski, M. F. (1996). Diversification rates in a temperate legume
680 clade: are there “so many species” of *Astragalus* (Fabaceae)? *American Journal of*
681 *Botany*, 83, 1488–1502. <https://doi.org/10.1002/j.1537-2197.1996.tb13942.x>
- 682 Santos, E. A., Santana, A. E. G., Oliveira, E. P. S., Santos, J. M. & Albuquerque, U. P.
683 (2012). Etnografia e etnobiologia sobre o uso da jurema-preta [*Mimosa tenuiflora*
684 (Willd.) Poir.] pelos índios Pankararé (Nordeste do Brasil). *Revista Ouricuri*, 2, 055–074.
- 685 Saslis-Lagoudakis, C. H., Klitgaard, B. B., Forest, F., Francis, L., Savolainen, V.,
686 Williamson, E. M. & Hawkins, J. A. (2011a). The use of phylogeny to interpret cross-

- cultural patterns in plant use and guide medicinal plant discovery: An example from
Pterocarpus (Leguminosae). PloS one, 6, e22275.
<https://doi.org/10.1371/journal.pone.0022275>
- Saslis-Lagoudakis, C. H., Williamson, E. M., Savolainen, V. & Hawkins, J. A. (2011b).
Cross-cultural comparison of three medicinal floras and implications for bioprospecting
strategies. Journal of Ethnopharmacology, 135, 476–487.
<https://doi.org/10.1016/j.jep.2011.03.044>
- Schrire, B. D., Lavin, M. & Lewis, G. P. (2005). Global distribution patterns of the
Leguminosae: Insights from recent phylogenies. Biologiske Skrifter, 55, 375–422.
- Schrire, B. D., Lavin, M., Barker, N. P. & Forest, F. (2009). Phylogeny of the tribe
Indigofereae (Leguminosae–Papilionoideae): Geographically structured more in
succulent-rich and temperate settings than in grass-rich environments. American Journal
of Botany, 96, 816–852. <https://doi.org/10.3732/ajb.0800185>
- Schrire, B. (2013). A review of tribe Indigofereae (Leguminosae–Papilionoideae) in Southern
Africa (including South Africa, Lesotho, Swaziland & Namibia; excluding Botswana).
South African Journal of Botany, 89, 281–283.
<https://doi.org/10.1016/j.sajb.2013.06.014>
- Seigler, D. S. & Ebinger, J. E. (2010). New combinations in Senegalia and Vachellia
(Fabaceae: Mimosoideae). Phytologia, 92, 92–95. <https://doi.org/10.1071/SB11042>
- Sharrock, S., Hoft, R. & Ferreira de Souza Dias, B. (2018). An overview of recent progress in
the implementation of the Global Strategy for Plant Conservation - a global perspective.
Rodriguesia, 69, 1589–1511. <https://doi.org/10.1590/2175-7860201869401>
- Silva, M. C., Moonlight, P., Oliveira, R. S., Pennington, R. T. & Rowland, L. (2022). Toward
diverse seed sourcing to upscale ecological restoration in the Brazilian Cerrado. Frontiers
in Ecology and Evolution, 10, 1–10. <https://doi.org/10.3389/fevo.2022.1045591>

- Simon, M. F., & Proença, C. (2000). Phytogeographic patterns of Mimosa (Mimosoideae, Leguminosae) in the Cerrado biome of Brazil: An indicator genus of high-altitude centers of endemism? *Biological Conservation*, 96, 279–296.
[https://doi.org/10.1016/S0006-3207\(00\)00085-9](https://doi.org/10.1016/S0006-3207(00)00085-9)
- Simon, M. F., Grether, R., Queiroz, L. P. de, Skema, C., Pennington, R. T. & Hughes, C. E. (2009). Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. *Proceedings of the National Academy of Sciences*, 106, 20359–20364. <https://doi.org/10.1073/pnas.0903410106>
- Stropp, J., Ladle, R. J., Emilio, T., Lessa, T., & Hortal, J. (2022). Taxonomic uncertainty and the challenge of estimating global species richness. *Journal of Biogeography*, 49, 1654–1656. <https://doi.org/10.1111/jbi.14463>
- Turland, N. J., Wiersema, J. H., Barrie, F. R., Greuter, W., Hawksworth, D. L., Herendeen, P. S., Knapp, S., Kusber, W. H., Li, D. Z., Marhold, K., May, T. W., McNeill, J., Monroe, A. M., Prado, J., Price, M. J. & Smith, G. F. (eds.). (2018). International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. *Regnum Vegetabile* 159. Glashütten: Koeltz Botanical Books. <https://doi.org/10.12705/Code.2018>
- Ulian, T., Diazgranados, M., Pironon, S., Padulosi, S., Liu, U., Davies, L., Howes, M-J. R., Borrell, J. S., Ondo, I., Pérez-Escobar, O. A. ... & Mattana, E. (2020). Unlocking plant resources to support food security and promote sustainable agriculture. *Plants, People, Planet*, 2, 421–445. <https://doi.org/10.1002/ppp3.10145>
- Werner, G. D., Cornwell, W. K., Cornelissen, J. H., & Kiers, E. T. (2015). Evolutionary signals of symbiotic persistence in the legume–rhizobia mutualism. *Proceedings of the National Academy of Sciences*, 112, 10262–10269.
<https://doi.org/10.1073/pnas.14240301>

-
- WFO. (2022). World Flora Online. Published on the Internet; Retrieved December 2022,
from <http://www.worldfloraonline.org>
- Wiens, J. J. & Donoghue, M. J. (2004). Historical biogeography, ecology and species
richness. *Trends in ecology & evolution*, 19, 639–644.
<https://doi.org/10.1016/j.tree.2004.09.011>
- Yahara, T., Javadi, F., Onoda, Y., Queiroz, L. P. de, Faith, D. P., Prado, D. E., Akasaka, M.,
Kadoya, T., Ishihama, F., Davies, S., Slik, J. W. F., Yi, T., Ma, K., Bin, C., Darnaedi, D.,
Pennington, R. T., Tuda, M., Shimada, M., Ito, M., Egan, A. N., Buerki, S., Raes, N.,
Kajita, T., Mimura, M., Tachida, H., Iwasa, Y., Smith, G. F., Victor, J. E., Nkonki, T.
(2013). Global legume diversity assessment: Concepts, key indicators, and strategies.
Taxon, 62, 249–266. <https://doi.org/10.12705/622.12>
- Zuanny, D., Vilela, B., Moonlight, P., Särkinen, T. & Cardoso, D. (in prep.). expowo: An R
package for mining plant diversity and distribution data.
<https://github.com/DBOSlab/expowo>

Tables

Table 1. | Comparison of the current species number of Leguminosae big genera extracted from Kew's POWO database and Brown (1883), Mabberley (1997) and Frodin (2004). Previous numbers extracted from Frodin's (2004) Appendices 1, 2 and 3. Genera in boldface are those current big genera present since the 1883 checklist.

Genus	Protologue	Brown (1883)	Polhill & Raven (1981)	Mabberley (1997)	Frodin (2004)	Lewis et al. (2005)	POWO (2023)
<i>Astragalus</i> L.	142	500	2000	1750	3270	2300	3066
<i>Acacia</i> Mill.*	24	432	1200	1200	1353	1450	1086
<i>Cassia</i> L.**	26	338	30	645	692	30	37
<i>Indigofera</i> L.	3	-	700	700	>700	700	731
<i>Crotalaria</i> L.	13	-	600	600	699	690	712
<i>Lupinus</i> L.	6	-	200	-	-	220- 230	604
<i>Oxytropis</i> DC.	33	-	300	-	-	300- 400	612
<i>Mimosa</i> L.	39	-	400- 450	-	-	490- 510	600

*The previous generic concept adopted for *Acacia* in the checklists of 1883, 1997 and 2004 also encompassed the genera *Senegalia* and *Vachelia*.

**The previous concept adopted for *Cassia* in the checklists of 1883, 1997 and 2004 also included *Senna* and *Chamaecrista*.

Figures

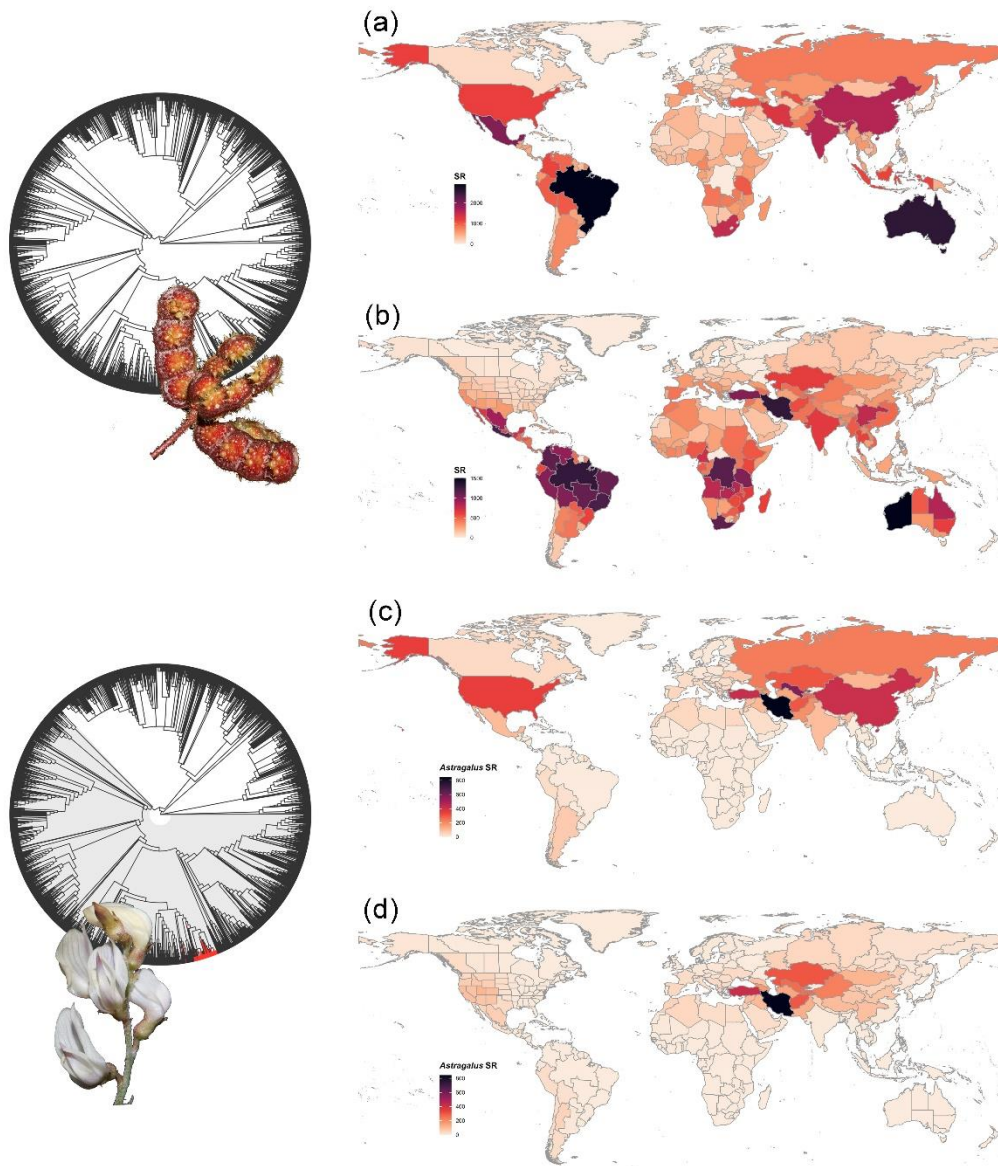


Figure 1 | Global distribution of all accepted species within the Leguminosae family according to native distribution across the political country division of the world (a) and across the botanical country subdivision (b); and the native distribution of *Astragalus* species according to political country division (c) and botanical countries (d) retrieved from Kew's Plants of the World Online. The global regions and associated scale bar are colored according to species richness: from beige (low species richness) to black (high species richness). On the left, it is displayed the bayesian consensus phylogenetic tree reconstruction with 3842 *matk* sequences of the family (top) extracted from LPWG (2017) and with the phylogenetic placement of the genus *Astragalus* in red (bottom). The legume subfamily is indicated by the light gray highlight in the phylogeny. Photos by Domingos Cardoso.

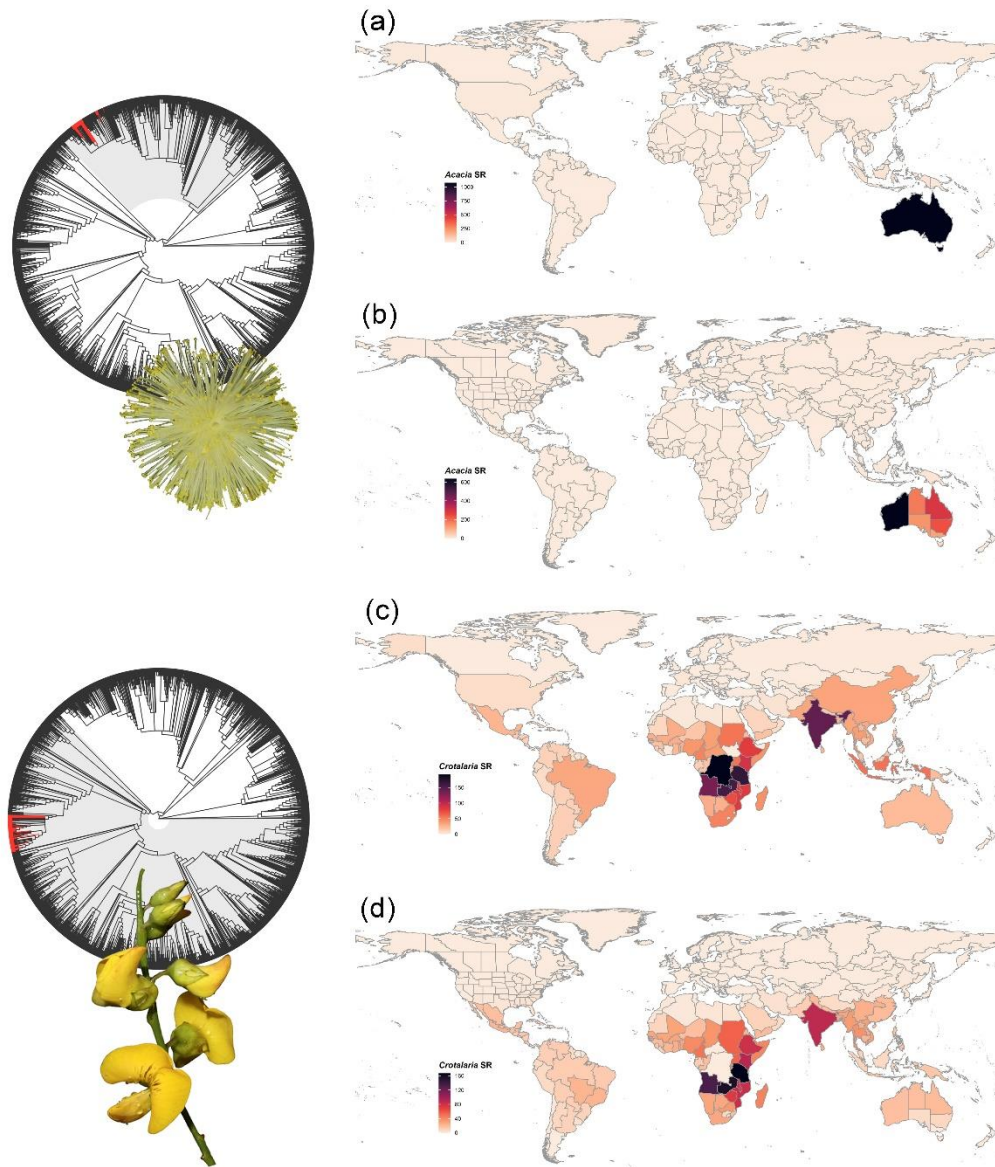
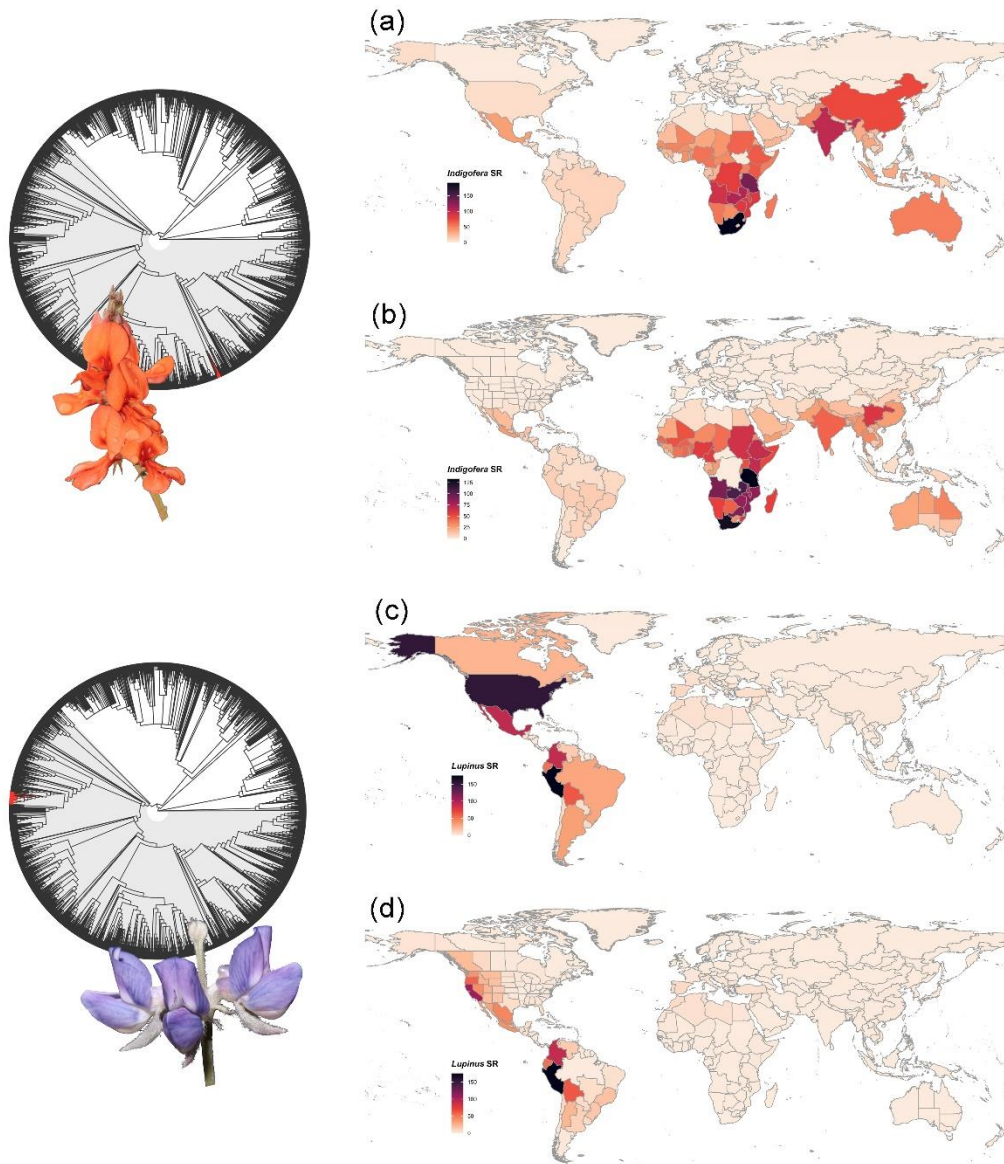


Figure 2 | Global distribution of all accepted species within the *Acacia* genus according to native distribution across the political country division of the world (a) and across the botanical subdivision (b); and the native distribution of *Crotalaria* species according to political country division of the world (c) and botanical subdivision (d) retrieved from Kew's Plants of the World Online. The global regions and associated scale bar are colored according to species richness: from beige (low species richness) to black (high species richness). On the left, it is displayed the bayesian consensus phylogenetic tree reconstruction with 3842 *matk* sequences of the family (LPWG, 2017) with the phylogenetic placement of the genus *Acacia* (top) and *Crotalaria* (bottom) in red. The respective legume subfamilies are indicated by the light gray highlight in the phylogeny. Photos by Domingos Cardoso.



790

791 **Figure 3** | Global distribution of all accepted species within the *Indigofera* genus according
 792 to native distribution across the political country division of the world (a) and across the
 793 botanical subdivision (b); and the native distribution of *Lupinus* species according to political
 794 country division of the world (c) and botanical subdivision (d) retrieved from Kew's Plants of
 795 the World Online. The global regions and associated scale bar are colored according to
 796 species richness: from beige (low species richness) to black (high species richness). On the
 797 left, it is displayed the bayesian consensus phylogenetic tree reconstruction with 3842 *matk*
 798 sequences of the family (LPWG, 2017) with the phylogenetic placement of the genus
 799 *Indigofera* (top) and *Lupinus* (bottom) in red. The respective legume subfamilies are
 800 indicated by the light gray highlight in the phylogeny. Photos by Domingos Cardoso.

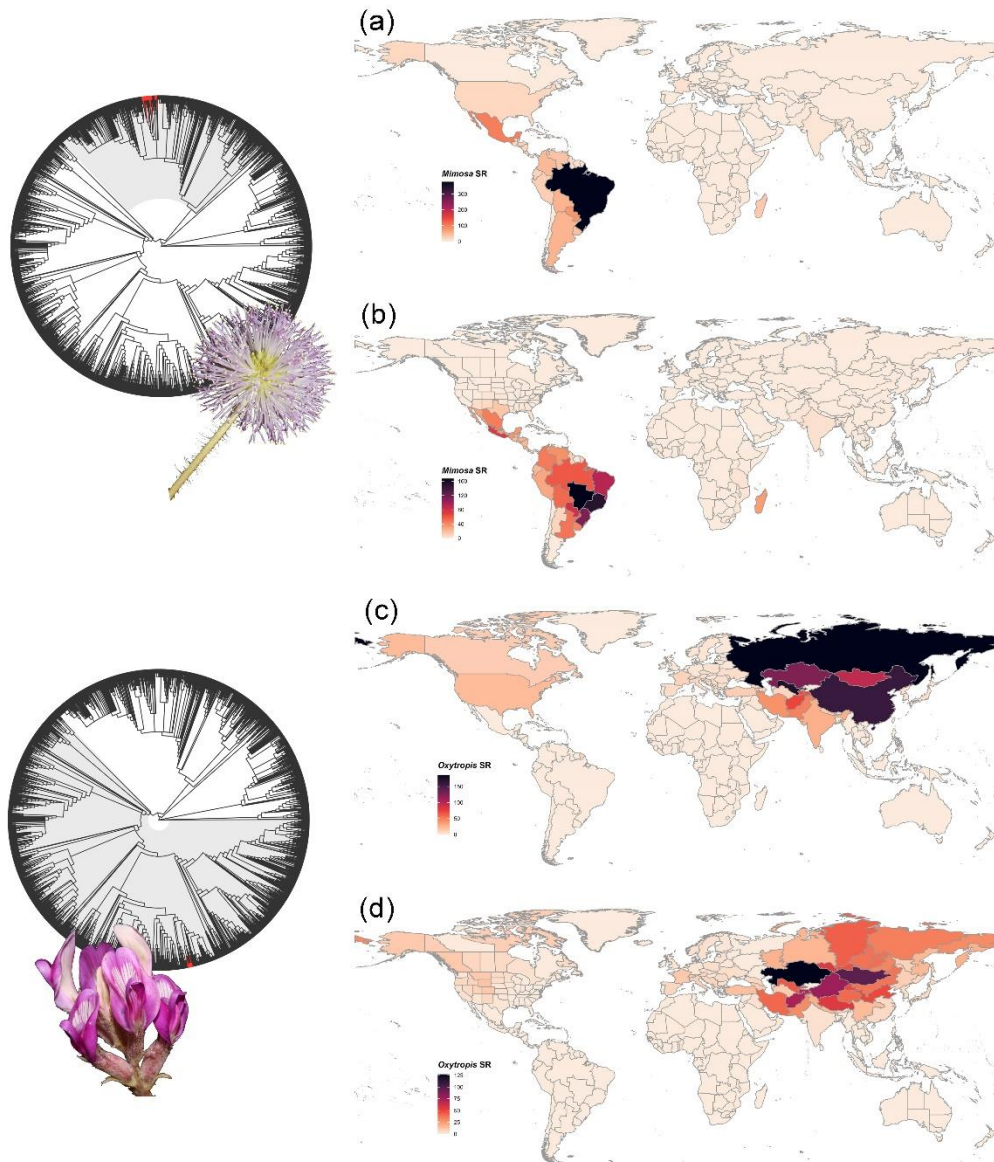


Figure 4 | Global distribution of all accepted species within the *Mimosa* genus according to native distribution across the political country division of the world (a) and across the botanical subdivision (b); and the native distribution of *Oxytropis* species according to political country division of the world (c) and botanical subdivision (d) retrieved from Kew's Plants of the World Online. The global regions and associated scale bar are colored according to species richness: from beige (low species richness) to black (high species richness). On the left, it is displayed the bayesian consensus phylogenetic tree reconstruction with 3842 *matk* sequences of the family (LPWG, 2017) with the phylogenetic placement of the genus *Mimosa* (top) and *Oxytropis* (bottom) in red. The respective legume subfamilies are indicated by the light gray highlight in the phylogeny. Photos by Domingos Cardoso.

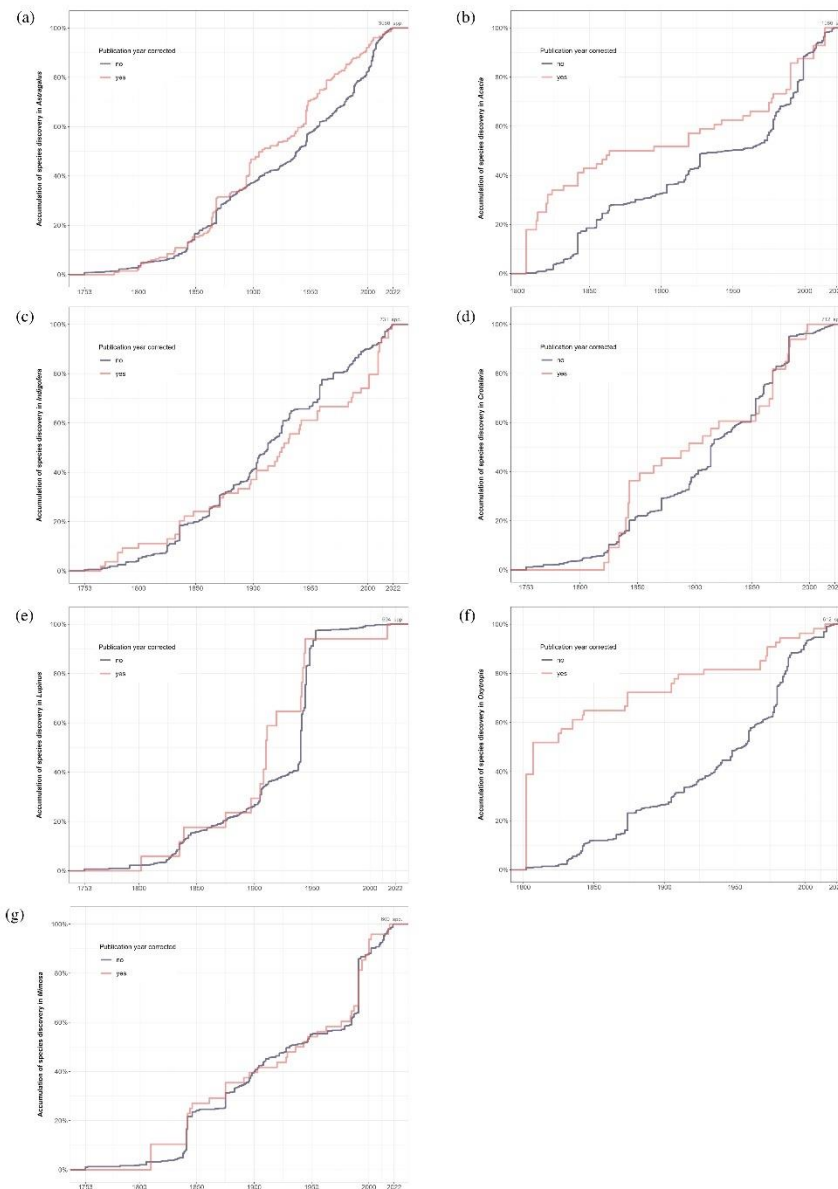


Figure 5 | Accumulation curves of species number due to new descriptions (dark gray) and taxonomic changes (light red) of all big genera in Leguminosae between 1753 and 2022 showing that changes have been made since the beginning of their taxonomic histories. The data without change in the year of publication is assigned in the legend as “no” and represents new descriptions/pure births, while the species data assigned as “yes” represents taxonomic changes. The respective graphs of each big legume genera are indicated as follows: (a) *Astragalus*, (b) *Acacia*, (c) *Indigofera*, (d) *Crotalaria*, (e) *Lupinus*, (f) *Oxytropis*, and (g) *Mimosa*.

822 **Supporting Information**

823

824 **Methods S1.** Custom R scripts to extract the complete list of Leguminosae species and to
825 search for the big legume genera from Kew’s Plants of the World Online (POWO) database,
826 to classify the nomenclatural history, and to generate the maps and the species accumulation
827 graphs. Each species from the query comes with taxonomic information (family, genus, taxon
828 name, authors, scientific name, and publication), native and introduced distribution
829 information (political country-level and botanical subdivision country-level), and
830 identification (Kew ID and POWO URI).

Supporting Information Methods S1

Article title: Big genera in Leguminosae: Overview of taxonomic changes and global distribution

Authors: Débora C. Zuanny and Domingos Cardoso

Contents

Reproducing the results

- Usage 1
- Extracting all Leguminosae/Fabaceae species 2
- Searching for the big legume genera and their nomenclatural data 3
- Plotting global distribution maps and graphs of the big legume genera 4

Usage

We have created **expowo** to automatize the process of extracting data from Kew's, Plants of The World Online database using R. Our goal is to pull species lists with all associated data of any plant family. The results present in the paper makes uses of this new R package. If you do not already have it, install **expowo** from CRAN and load it with:

```
install.packages("expowo")  
library(expowo)
```

Here we will show how to extract the data used in this paper, and then use this data to plot global distribution maps of species richness and the species accumulation curves.

Extracting all Leguminosae/Fabaceae species

To create the dataset with all legume species available at Plants of the World Online, use the function `powoSpecies`:

```
expowo::powoSpecies(family = "Fabaceae",
                    hybridspp = FALSE,
                    verbose = TRUE,
                    save = TRUE,
                    dir = "results_powoSpecies/",
                    filename = "Fabaceae_search")
```

Searching for the big legume genera and their nomenclatural data

To search for the big genera, use the function `megaGen` and set the threshold to 500 species and the argument `family` to "Fabaceae".

```
mega_Fab <- expowo::megaGen(family = "Fabaceae",
                           thld = 500,
                           verbose = TRUE,
                           save = TRUE,
                           dir = "results_megaGen/",
                           filename = "legume_big_genera_species")
```

With this dataset, it was possible to filter and gather nomenclatural information. The custom R script used to do so is present here:

```
# Load required packages
library(magrittr)
library(dplyr)
library(ggplot2)
library(cowplot)
library(viridis)

# Import the file
list.files()
df <- read.csv("legume_big_genera_species.csv")

df <- apply(df, 2, function(x) gsub("^$", NA, x))
df <- data.frame(df)

# Creating new columns for the year and corrected year of publication
df <- df %>% tibble::add_column(year = NA, .after = "publication")
df <- df %>% tibble::add_column(year_corrected = NA, .after = "year")
df <- df %>% tibble::add_column(number_synonyms = NA, .after = "accepted_name")

# Extracting the year from the full publication information
# Clean examples such as "non C. caerulea Jacq.", "non Engelh. (1898)"
df$publication <- gsub(",\\snon\\s[[:upper:]].*", "", df$publication)

# Get examples of not validly published names
tf <- grepl("[0-9][ ]", "\\s", df$publication)
```

```

p <- unique(gsub(".*[)]", "\\s", "", df$publication[tf]))

tf_inval <- !grepl(paste0(p, collapse = "|"), df$publication)

df$year[tf_inval] <- gsub("\\scontrary.*", "", df$publication[tf_inval])
df$year[tf_inval] <- gsub(".*[()|[]].*|.*\\.\\s|.*[.]", "", df$year[tf_inval])
df$year[tf_inval] <- gsub("unknown\\spublication|publication", "unknown",
  df$year[tf_inval])
df$year[!tf_inval] <- gsub(".*[()|,|.]", "", df$publication[!tf_inval])
df$year[!tf_inval] <- gsub(".*[.]\\.\\s|[]].*", "", df$year[!tf_inval])
df$year[!tf_inval] <- gsub("unknown\\spublication|publication", "unknown",
  df$year[!tf_inval])

# Extracting the correct year for species with synonyms
# Fill corrected year for species without synonyms
tf_accepted <- df$status %in% "Accepted"
# Get the accepted names for species without any synonym
tftf <- !df$scientific_name[tf_accepted] %in% df$accepted_name
df$year_corrected[tf_accepted][tftf] <- df$year[tf_accepted][tftf]

# Count number of synonyms
df$number_synonyms[tf_accepted][tftf] <- 0
df$number_synonyms[tf_accepted][!tftf] <- table(df$accepted_name)

# Extracting the correct year for species with synonyms
n_au <- unique(df$accepted_name[!is.na(df$accepted_name)])
for (i in seq_along(n_au)) {

  tf <- df$scientific_name[tf_accepted] %in% n_au[i]
  year_acc <- df$year[tf_accepted][tf]

  tftf <- df$accepted_name[tf_inval] %in% n_au[i]
  temp <- df[tf_inval, ][tftf, ]
  temp <- temp[!grepl("\\svar[.]|\\sf[.]\\s|\\ssubsp[.]\\s",
    temp$scientific_name), ]
  if(length(temp$year) == 0) {
    year_syn <- NA
  } else {
    year_syn <- min(temp$year)
  }

  if (year_acc < year_syn |
    is.na(year_syn)) {
    df$year_corrected[tf_accepted][tf] <- year_acc
  } else {
    df$year_corrected[tf_accepted][tf] <- year_syn
  }
}

# Identify each change in the corrected year of species discovery
df$year <- as.numeric(df$year)
tf <- is.na(df$year_corrected)

```

```

df$year_corrected[!tf] <- as.numeric(df$year_corrected[!tf])

df$year <- as.numeric(df$year)
df$year_corrected <- as.numeric(df$year_corrected)

tf <- df$year == df$year_corrected

df <- df %>% tibble::add_column(year_changed = NA, .after = "year_corrected")
df$year_changed[tf] <- "no"
df$year_changed[!tf] <- "yes"

# Final spreadsheet including the input df with only the accepted species and
# associated corrected year
df_accepted <- df %>%
  filter(status == "Accepted") %>%
  select(-c("status", "accepted_name"))

write.csv(df_accepted, "data_to_plot.csv")

```

Plotting global distribution maps and graphs of the big legume genera

To create publication-quality maps for each big legume genera with “rocket” color scale from Viridis, which composed the Figures 1-4, run the following R code:

```

# To create one map for each big genera according to political countries
expowo::powoMap(inputdf = mega_Fab,
  botctrs = FALSE,
  distcol = "native_to_country",
  taxclas = "genus",
  verbose = FALSE,
  save = TRUE,
  vir_color = "rocket",
  bre_color = NULL,
  leg_title = "SR",
  dpi = 600,
  dir = "results_powoMap/",
  filename = "Fab_global_richness_country_map",
  format = "jpg")

# To create one map for each big genera according to botanical subdivision
expowo::powoMap(inputdf = mega_Fab,
  botctrs = TRUE,
  distcol = "native_to_botanical_countries",
  taxclas = "genus",
  verbose = FALSE,
  save = TRUE,
  vir_color = "rocket",
  bre_color = NULL,

```

```

leg_title = "SR",
dpi = 600,
dir = "results_powoMap/",
filename = "Fab_global_richness_botcountry_map",
format = "jpg")

```

After running the previous steps of this section, use the dataframe “df_accepted” to plot the accumulation of species richness for each big legume genera.

Plot figures

```

scales::show_col(viridis(10, option = "magma"))
cols <- c("#180F3EFF", "#F1605DFF")

tax_group <- unique(df_accepted$genus)
for (i in seq_along(tax_group)) {

  temp_df_accepted <- df_accepted[df_accepted$genus %in% tax_group[i], ]
  l <- length(temp_df_accepted$species)

  plot <- ggplot(temp_df_accepted, aes(year, colour = year_changed))+
    #geom_vline(xintercept = 1850, colour = "red", linetype="dashed", size = 1,
    alpha=0.6)+
    stat_ecdf(linewidth = 1.5, alpha=0.6, geom = "step") +
    theme_bw() +
    ylab(eval(bquote(expression(
      bold("Accumulation of species discovery in")~bolditalic(. (tax_group[i])))))) +
    theme(axis.title.x=element_blank()) +
    scale_color_manual(values = cols) +
    scale_y_continuous(breaks=scales::pretty_breaks(n=5),
      labels=scales::percent_format(accuracy = 1)) +
    scale_x_continuous(breaks=c(1753, 1800, 1850, 1900, 1950, 2000,
      max(df$year[!is.na(df$year)]))) +
    theme(axis.title.y=element_text(size=14, margin=margin(0,12,0,0))) +
    theme(axis.text.x=element_text(size=14)) +
    theme(axis.text.y=element_text(size=14)) +
    theme(legend.title = element_text(size=14)) +
    theme(legend.position = c(0.2, 0.8),
      legend.key = element_rect(linewidth=10, linetype='blank'),
      legend.text = element_text(size=14),
      legend.key.size = unit(1, "cm"),
      legend.title = element_text(size=14)) +
    guides(alpha='none', colour=guide_legend("Publication year corrected"),
      size=14) +
    annotate("text", x = 2020, y = 1.02,
      label = paste(l, " spp."),
      colour = "black", alpha=0.6)

  save_plot(paste0("cumulative_discovery_", tax_group[i], ".jpg"), plot,
    ncol = 1, nrow = 1,
    base_height = 8.5,

```

```
    base_aspect_ratio = 1.3,  
    base_width = NULL)  
}
```

Considerações finais

- ❖ Foi criado um pacote de R para minerar o banco de dados taxonomicamente verificado Plants of the World Online, com um *site* associado;
- ❖ Geramos listas completas de espécies aceitas da flora do Brasil, Papua Nova Guiné, México, Malásia, Madagascar, Indonésia e Tailândia, e das famílias Leguminosae, Poaceae e Begoniaceae;
- ❖ A presente pesquisa definiu *Astragalus*, *Acacia*, *Indigofera*, *Crotalaria*, *Lupinus*, *Mimosa* e *Oxytropis* como os atuais grandes gêneros de Leguminosae;
- ❖ Fornecemos o número de espécies mais atual dos grandes gêneros de Leguminosae e representamos a sua distribuição global;
- ❖ Fornecemos o status dos grandes gêneros de Leguminosae, resumizamos o que variou desde a última revisão e sugerimos aplicações.

Apêndice

- ❖ Manual do pacote expowo disponível no CRAN.
- ❖ Impressão da página inicial do site associado ao pacote expowo (dboslab.github.io/expowo/), da página de referências, dos artigos de conjunto de dados associados, da função powoMap, da função powoGenera, da função powoSpecies, da powoFam, da megaGen, da powoSpDist, e da toptenGen.

Package ‘expowo’

February 22, 2023

Type Package

Title Data Mining of Plant Diversity and Distribution for R

Version 1.0

Maintainer Debora Zuanny <deborazuanny@gmail.com>

Description Produces diversity estimates and species lists with associated global distribution for any angiosperm family and genus from 'Plants of the World Online' database <<https://powo.science.kew.org/>>, by interacting with the source code of each plant taxon page, and creates global maps of species richness.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Depends R (>= 3.5.0)

Imports RColorBrewer, dplyr, magrittr, data.table, ggplot2,
rnatualearth, sp, sf, utils

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

URL <https://dboslab.github.io/expowo/>,
<https://github.com/dboslab/expowo>

BugReports <https://github.com/DBOSlab/expowo/issues>

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Debora Zuanny [aut, cre, cph] (<<https://orcid.org/0000-0001-6623-3788>>),
Domingos Cardoso [aut, ctb, cph]
(<<https://orcid.org/0000-0001-7072-2656>>)

Repository CRAN

Date/Publication 2023-02-22 15:10:08 UTC

R topics documented:

angioData	2
angioGenera	2
botregions	3
megaGen	3
POWOcodes	4
powoFam	5
powoGenera	6
powoMap	7
powoSpDist	10
powoSpecies	11
topGen	13
Index	15

angioData	<i>List of Angiosperm species</i>
-----------	-----------------------------------

Description

Complete list of all non-hybrid species and associated data of Lecythidaceae, Aristolochiaceae, Martyniaceae, Cabombaceae and Begoniaceae as retrieved from POWO database with the function powoSpecies of this package gathered at Nov 2022.

Usage

data(angioData)

Format

Dataframe

angioGenera	<i>List of Angiosperm genera</i>
-------------	----------------------------------

Description

Complete list of all genera and associated data of Lecythidaceae, Aristolochiaceae, Begoniaceae, Martyniaceae, Dipterocarpaceae, and Fagaceae as retrieved from POWO database with the function powoGenera of this package gathered at Jan 2023.

Usage

data(angioGenera)

Format

Dataframe

botregions	<i>Countries and associated classification of botanical divisions</i>
------------	---

Description

Countries and associated classification of botanical divisions according to the World Geographical Scheme for Recording Plant Distributions.

Usage

```
data(botregions)
```

Format

Table in .csv format

Source

<https://www.tdwg.org/standards/wgsrpd/>

megaGen	<i>Extract megadiverse genera from POWO</i>
---------	---

Description

Produces a CSV file listing all mega-diverse genera for any angiosperm family at [Plants of the World Online \(POWO\)](#) based on a provided numeric value as the threshold to be considered megadiverse. Frodin (2004) in Taxon suggests 500 species as the threshold.

Usage

```
megaGen(family, thld = 500, verbose = TRUE, save = FALSE, dir, filename)
```

Arguments

family	Either one family name or a vector of multiple families that is present in POWO.
thld	A defined threshold of species number for a genus to be considered megadiverse. The default value is 500 based on Frodin (2004) in Taxon.
verbose	Logical, if FALSE, the search results will not be printed in the console in full.
save	Logical, if FALSE, the searched results will not be saved on disk.
dir	Pathway to the computer's directory, where the file will be saved provided that the argument save is set up in TRUE. The default is to create a directory named results_megaGen and the searched results will be saved within a subfolder named by the current date.
filename	Name of the output file to be saved. The default is to create a file entitled output .

Value

Table in .csv format.

Author(s)

Debora Zuanny & Domingos Cardoso

See Also

[POW0codes](#)

Examples

```
library(expowo)

megaGen(family = "Cyperaceae",
        thld = 500,
        verbose = TRUE,
        save = FALSE,
        dir = "results_megaGen/",
        filename = "Cyperaceae_big_genera")
```

POW0codes

Complete list of angiosperm families and associated URI addresses

Description

Complete list of the APG IV-based families of flowering plants and associated URI addresses as retrieved with the function `get_pow` of the package **taxize**.

Usage

```
data(POW0codes)
```

Format

Dataframe

powoFam*Extract species number of any plant family from POWO*

Description

Produces a CSV file listing the number of species within the target botanical families of flowering plants available at [Plants of the World Online \(POWO\)](#).

Usage

```
powoFam(family, verbose = TRUE, save = FALSE, dir, filename)
```

Arguments

family	Either one family name or a vector of multiple families that is present in POWO.
verbose	Logical, if FALSE, the search results will not be printed in the console in full.
save	Logical, if FALSE, the search results will not be saved on disk.
dir	Pathway to the computer's directory, where the file will be saved provided that the argument save is set up in TRUE. The default is to create a directory named results_powoFam and the searched results will be saved within a subfolder named by the current date.
filename	Name of the output file to be saved. The default is to create a file entitled output .

Value

Table in .csv format.

Author(s)

Debora Zuanny & Domingos Cardoso

See Also

[megaGen](#)
[topGen](#)
[powoGenera](#)
[powoSpecies](#)
[powoMap](#)
[POWOCodes](#)

Examples

```
library(expowo)

powoFam(family = "Lecythidaceae",
        verbose = TRUE,
        save = FALSE,
        dir = "results_powoFam/",
        filename = "Lecythidaceae_spp_number")
```

powoGenera

Extract list of genera from POWO

Description

Produces a CSV file listing all genera with associated number of accepted species and geographical distribution for any angiosperm family at [Plants of the World Online \(POWO\)](#).

Usage

```
powoGenera(family, genus = NULL, country = NULL,
            verbose = TRUE, save = FALSE, dir, filename)
```

Arguments

family	Either one family name or a vector of multiple families that is present in POWO.
genus	Either one genus name or a vector of multiple genera that are present in POWO. If any genus name is not provided, then the function will search all accepted genera known for the target family.
country	Either one country name or a vector of multiple countries. If country names are provided, then the function will return only the genera that are native to such countries, according to POWO.
verbose	Logical, if FALSE, the searched results will not be printed in the console in full.
save	Logical, if FALSE, the searched results will not be saved on disk.
dir	Pathway to the computer's directory, where the file will be saved provided that the argument save is set up in TRUE. The default is to create a directory named results_powoGenera and the searched results will be saved within a subfolder named by the current date.
filename	Name of the output file to be saved. The default is to create a file entitled output .

Value

Table in .csv format.

Author(s)

Debora Zuanny & Domingos Cardoso

See Also

[megaGen](#)

[topGen](#)

[powoSpecies](#)

[powoFam](#)

[POWOCodes](#)

Examples

```
library(expowo)

powoGenera(family = "Lecythidaceae",
           verbose = TRUE,
           save = FALSE,
           dir = "results_powoGenera/",
           filename = "Lecythidaceae_genera")

powoGenera(family = "Lecythidaceae",
           genus = "Bertholletia",
           country = c("Argentina", "Brazil", "French Guiana"),
           verbose = TRUE,
           save = FALSE,
           dir = "results_powoGenera/",
           filename = "Lecythidaceae_search")
```

powoMap

Create global maps of species richness

Description

Produces global-scale maps of species richness at political country and botanical country levels. Despite being originally designed to create maps for all input data of any specified taxonomic group (genus or family) from the search results with `powoSpecies`, the function is also useful for any dataframe-formatted input data that has at least a column with species and one or two columns with associated distribution in the countries and/or botanical regions. Multiple richness maps for any different taxonomic groups within the input data can be produced automatically in a single run by just specifying a column name with the associated taxonomic classification.

Usage

```
powoMap(inputdf = NULL,
        botctrs = FALSE,
        distcol = NULL,
        taxclas = NULL,
        verbose = TRUE,
        save = FALSE,
        vir_color = "viridis",
        bre_color = NULL,
        leg_title = "SR",
        dpi = 600,
        dir = "results_powoMap/",
        filename = "global_richness_map",
        format = "jpg")
```

Arguments

inputdf	A dataframe with a species column and the associated global distribution at country or botanical country level. The species name must be as a binomial, i.e. must contain both the genus name and specific epithet, but the authorship is optional. Each species must be as a single row with its corresponding full distribution in all countries and/or botanical regions within a single cell of their respective columns, where the country names or botanical regions are separated by a comma. This is, for example, the standard dataframe from the search results with the function <code>powoSpecies</code> .
botctrs	Logical. If TRUE, the species richness maps will be created according to the botanical country subdivisions of the world. Also, a WGSRPD folder including the level 3 shapefile and the associated Brummitt's (2001) book fully describing the World Geographical Scheme for Recording Plant Distributions will be downloaded into the working directory. If you do not remove this folder or rename any of the contents, then the function will not download the same folder again. The default is FALSE.
distcol	Column name with the full global distribution data for each species at political country level or the level 3 of botanical subdivision of the World Geographical Scheme for Recording Plant Distributions. If the species distribution is given with botanical subdivisions, then you must also change the argument <code>botctrs</code> to TRUE. If the distribution is described only by political country names, then set <code>botctrs</code> to FALSE.
taxclas	A character vector with the column name for the corresponding taxonomic classification of each species in any higher taxonomic level. If provided, the function will produce, in a single run, all global richness maps for every distinct group within the input data. The default is NULL, then the function will generate only one global species richness map for the entire input data.
verbose	Logical. If FALSE, the map creation steps will not be printed in the console in full.
save	Logical. If FALSE, the global maps will not be saved on disk.

vir_color	A character vector with the name or code of any of the color palettes from Viridis package.
bre_color	A character vector with the name or code of any of the color palettes from RColorBrewer package.
leg_title	A character vector to be displayed in the output map as a legend. Default is to create a title called SR , an acronym for species richness.
dpi	One number in the range of 72-4000 referring to the image resolution in the format of dots per inch in the output file. Default is to create an output with 600 dpi.
dir	Pathway to the computer's directory, where the map file will be saved provided that the argument save is set up in TRUE. The default is to create a directory named results_powoMap/ and the search results will be saved within a sub-folder named by the current date.
filename	Name of the output file to be saved. The default is to create a file entitled global_richness_map .
format	A character vector related to the file format of the global map to be saved. The default is "jpg" to save the output in Joint Photographic Experts Group (.jpg), but you can also choose "pdf" to save in Portable Document Format (.pdf), "tiff" to save in Tag Image File Format (.tiff) or "png" to save in Portable Network Graphics (.png).

Value

One or a list of objects of class c("gg", "ggplot").

Author(s)

Debora Zuanny & Domingos Cardoso

See Also

[megaGen](#)
[topGen](#)
[powoSpecies](#)
[powoFam](#)
[powoGenera](#)

Examples

```
## Not run:
library(expowo)

mapspdist <- powoSpecies(family = "Martyniaceae",
                        hybridspp = FALSE,
                        country = NULL,
                        verbose = TRUE,
                        save = FALSE,
```

```

        dir = "results_powoSpecies/",
        filename = "Martyniaceae_spp")

# To create multiple maps for each genus within the input data according to
# political countries.
powoMap(inputdf = mapspdist,
        botctrs = FALSE,
        distcol = "native_to_country",
        taxclas = "genus",
        verbose = FALSE,
        save = FALSE,
        vir_color = "viridis",
        bre_color = NULL,
        leg_title = "SR",
        dpi = 600,
        dir = "results_powoMap/",
        filename = "global_richness_country_map",
        format = "jpg")

# To create multiple maps for each genus within the input data according to
# botanical country subdivisions.
powoMap(inputdf = mapspdist,
        botctrs = TRUE,
        distcol = "native_to_botanical_countries",
        taxclas = "genus",
        verbose = FALSE,
        save = FALSE,
        vir_color = "viridis",
        bre_color = NULL,
        leg_title = "SR",
        dpi = 600,
        dir = "results_powoMap/",
        filename = "global_richness_botcountry_map",
        format = "jpg")

## End(Not run)

```

powoSpDist

Extract list of species distribution from POWO

Description

Produces a CSV file listing the geographical distribution of all target species of flowering plants available at [Plants of the World Online \(POWO\)](#).

Usage

```

powoSpDist(family, species = NULL,
           verbose = TRUE, save = FALSE, dir, filename)

```

Arguments

family	Either one family name or a vector of multiple families that is present in POWO.
species	Either one non-hybrid species name or a vector of multiple species that are present in POWO. If any species name is not provided, then the function will search any species from all accepted genera known for the target family.
verbose	Logical, if FALSE, the search results will not be printed in the console in full.
save	Logical, if FALSE, the search results will not be saved on disk.
dir	Pathway to the computer's directory, where the file will be saved provided that the argument save is set up in TRUE. The default is to create a directory named results_powoSpDist and the search results will be saved within a subfolder named by the current date.
filename	Name of the output file to be saved. The default is to create a file entitled output .

Value

Table in .csv format.

Author(s)

Debora Zuanny & Domingos Cardoso

See Also

[POWOCodes](#)

Examples

```
library(expowo)

powoSpDist(family = "Lecythidaceae",
           species = "Lecythis pisonis",
           verbose = TRUE,
           save = FALSE,
           dir = "results_powoSpDist/",
           filename = "L_pisonis_distribution")
```

`powoSpecies`

Extract list of species from POWO

Description

Produces a CSV file listing all accepted species and associated geographical distribution from any target genus or family of flowering plants at [Plants of the World Online \(POWO\)](#).

Usage

```
powoSpecies(family, genus = NULL, hybridspp = FALSE, country = NULL,
            verbose = TRUE, save = FALSE, dir, filename)
```

Arguments

family	Either one family name or a vector of multiple families that is present in POWO.
genus	Either one genus name or a vector of multiple genera that are present in POWO. If any genus name is not provided, then the function will search any species from all accepted genera known for the target family.
hybridspp	Logical, if TRUE, the search results will include hybrid species.
country	Either one country name or a vector of multiple countries. If country names are provided, then the function will return only the species that are native to such countries, according to POWO.
verbose	Logical, if FALSE, the search results will not be printed in the console in full.
save	Logical, if FALSE, the search results will not be saved on disk.
dir	Pathway to the computer's directory, where the file will be saved provided that the argument save is set up in TRUE. The default is to create a directory named results_powoSpecies and the search results will be saved within a subfolder named by the current date.
filename	Name of the output file to be saved. The default is to create a file entitled output .

Value

Table in .csv format.

Author(s)

Debora Zuanny & Domingos Cardoso

See Also

[megaGen](#)
[topGen](#)
[powoFam](#)
[powoGenera](#)
[POW0codes](#)

Examples

```
library(expowo)

powoSpecies(family = "Martyniaceae",
            hybridspp = FALSE,
            country = c("Argentina", "Brazil", "French Guiana"),
```

```

verbose = TRUE,
save = FALSE,
dir = "results_powoSpecies/",
filename = "Martyniaceae_spp")

```

topGen

Extract the top most species-rich genera

Description

Produces a CSV file listing the top most diverse genera of any target botanical family of flowering plants at [Plants of the World Online \(POWO\)](#).

Usage

```
topGen(family, limit = 10, verbose = TRUE, save = FALSE, dir, filename)
```

Arguments

family	Either one family name or a vector of multiple families that is present in POWO.
limit	A defined numerical value to limit the most diverse genera to be selected within each plant family. The default is to select the top ten richest genera.
verbose	Logical, if FALSE, the search results will not be printed in the console in full.
save	Logical, if FALSE, the search results will not be saved on disk.
dir	Pathway to the computer's directory, where the file will be saved provided that the argument save is set up in TRUE. The default is to create a directory named results_topGen and the searched results will be saved within a subfolder named by the current date.
filename	Name of the output file to be saved. The default is to create a file entitled output .

Value

Table in .csv format.

Author(s)

Debora Zuanny & Domingos Cardoso

See Also

[POWOCodes](#)

Examples

```
library(expowo)

topGen(family = "Lecythidaceae",
      limit = 10,
      verbose = TRUE,
      save = FALSE,
      dir = "results_topGen/",
      filename = "Lecythidaceae_top_ten")
```

Index

angioData, [2](#)

angioGenera, [2](#)

botregions, [3](#)

megaGen, [3](#), [5](#), [7](#), [9](#), [12](#)

POW0codes, [4](#), [4](#), [5](#), [7](#), [11–13](#)

powoFam, [5](#), [7](#), [9](#), [12](#)

powoGenera, [5](#), [6](#), [9](#), [12](#)

powoMap, [5](#), [7](#)

powoSpDist, [10](#)

powoSpecies, [5](#), [7](#), [9](#), [11](#)

topGen, [5](#), [7](#), [9](#), [12](#), [13](#)

R-CMD-check **passing**

expowo



An R package for mining species list, diversity estimates, and distribution data for any genus or family of flowering plants from RBG Kew's [Plants of the World Online \(POWO\)](#).

Overview

The main goal of the **expowo** package is to retrieve information about the diversity and distribution of any plant family as publicly available at the taxonomically verified database [Plants of the World Online \(POWO\)](#). The package is intended to efficiently mine the content within the source HTML pages for any specific genus and family. It can return a comma-separated values (CSV) file with the number of accepted species and country-level distribution for any genus as well as the full list of accepted species in any genus or family, their authorship, original publication and global distribution. The latest major function implemented, `powoMap`, also can automatically create global maps for any taxon level.

Before using expowo

Update R and RStudio versions

Please make sure you have installed the latest versions of both R ([Mac OS](#), [Windows](#)) and RStudio ([Mac OS / Windows](#); choose the free version).

Installation

You can install the latest development version of **expowo** from [GitHub](#) using the [devtools](#) package with the following R code:

```
install.packages("devtools")
devtools::install_github("DBOSlab/expowo")
```

```
library(expowo)
```

OBS.: To download the development version, you will need to have the [Git](#) software installed. And if your operating system is Microsoft Windows, you will also need to download the Rtools.

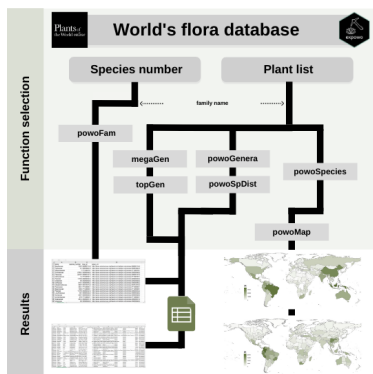
Otherwise, you will be able to install **expowo** more easily when it is available on CRAN, by just running the following R code:

```
install.packages("expowo")
```

```
library(expowo)
```

Usage

The package's major functions (`powoFam`, `powoGenera`, `powoSpecies`, `powoSpDist`, `megaGen`, and `topGen`) only require the name of the target family (or a vector with multiple family names). These major functions work with other minor functions (`getGenURI`, `getNum`, and `getDist`), with auxiliary and defensive functions to mine the plant data. Respectively, `getGenURI` mines the URI for each genus, `getNum` mines the total number of species within any genus, and `getDist` does a complete search for native and introduced country-level distribution for any genus and species. With this distribution, it is also possible to automatically generate global maps according to political country and botanical subdivision with the function `powoMap`. See the examples on how to use the **expowo**'s functions for mining basic information on the global plant diversity and distribution in the 'Articles' section in our [site](#).



Documentation

A detailed description of the **expowo**'s full functionality is available [here](#).

The **expowo** package is being continuously constructed. If you want to make suggestions, let us know! We hope it will be helpful for your botanical research!

Citation

Zuanny, D. & Cardoso, D. (2022). expowo: An R package for mining plant diversity and distribution data. <https://github.com/DBOSlab/expowo>



Links

[Browse source code](#)

[Report a bug](#)

License

[Full license](#)

[MIT](#) + file [LICENSE](#)

Citation

[Citing expowo](#)

Developers

Debora Zuanny

Author, maintainer, copyright holder

Domingos Cardoso

Author, contributor, copyright holder

Function reference

All functions

[angioData](#)

List of Angiosperm species

[angioGenera](#)

List of Angiosperm genera

[botregions](#)

Countries and associated classification of botanical divisions

[megaGen\(\)](#)

Extract megadiverse genera from POWO

[POW0codes](#)

Complete list of angiosperm families and associated URI addresses

[powoFam\(\)](#)

Extract species number of any plant family from POWO

[powoGenera\(\)](#)

Extract list of genera from POWO

[powoMap\(\)](#)

Create global maps of species richness

[powoSpDist\(\)](#)

Extract list of species distribution from POWO

[powoSpecies\(\)](#)

Extract list of species from POWO

[topGen\(\)](#)

Extract the top most species-rich genera

Associated data

Source: [vigettes/associated_data.Rmd](#)

Here in this article, we describe the data associated with `expowo` package that are helpful to enhance performance before and while mining the global diversity and distribution data.

There are three associated data package, all of which can be accessed and imported into the R environment by simply calling the `utils` function `data`. Below, we explain how to use the data package `POWOcodes`, `angioData`, and `botregions`.

Setup

Install the latest development version of **expowo** from [GitHub](#):

```
#install.packages("devtools")
devtools::install_github("DBO5lab/expowo")
```

```
library(expowo)
```

POWOcodes - A complete list of angiosperm families and associated URI addresses

The `POWOcodes` data is a dataframe containing a list of families of flowering plants available in the POWO database, according to [APG IV](#)'s classification. It was produced by using the functions `appFamilies` and `get_pow` of the package [taxize](#). `POWOcodes` data is very useful when the purpose is to query many plant families or all of them. In addition to the names of POWO's accepted plant families, which are required in any search with `expowo`'s functions, this data package has the corresponding family's URI, which is used internally in the function for opening a connection with each family page at Plants of The World Online.

To call this associated data to the R environment, use the code below:

```
utils::data("POWOcodes")
```

angioData - A dataframe of Angiosperm species

The `angioData` is a dataframe-formatted list of species of some families of flowering plants and associated data, as retrieved from POWO database using the `expowo`'s function `powoSpecies`. For publication purposes and package size limitation, we selected some families to include in this associated data and let the information easily available. Currently, there are species from five plant families as listed in Table 1. The last update of this data was at Nov 2022.

TABLE 1. The plant families and species number within `angioData`.

family	species_number	family	species_number
Lecythidaceae	381	Aristolochiaceae	727
Martyniaceae	14	Cabombaceae	7
Begoniaceae	1993		

In addition to all accepted species (excluding hybrid species) in the families above, the `angioData` also contains the publication, authorship, and global geographic distribution at country or botanical level for each species. The global classification of botanical divisions follows the [World Geographical Scheme](#) for Recording Plant Distributions, which is already associated with each taxon's distribution in POWO.

The example below shows how to call `angioData`:

```
utils::data("angioData")
```

angioGenera - A dataframe of Angiosperm genera

The `angioGenera` comprises the genera of some families of flowering plants and associated data as mined from POWO database using `powoGenera`. For publication purposes and package size limitation, we selected some families to include in this associated data and let the information easily available. Currently, there are six plant families (Lecythidaceae, Aristolochiaceae, Begoniaceae, Martyniaceae, Dipterocarpaceae, and Fagaceae) as listed in Table 2. The last update of this data was at Nov 2022.

TABLE 2. The plant families and genera within `angioGenera`.

family	scientific_name	publication	species_number
Aristolochiaceae	Aristolochia L.	Sp. Pl.: 960 (1753)	534
Aristolochiaceae	Asarum L.	Sp. Pl.: 442 (1753)	129
Aristolochiaceae	Euglypha Chodat & Hassl.	Bull. Herb. Boissier, ser. 2, 6: 138 (1906)	1
Aristolochiaceae	Hydnora Thunb.	Kongl. Vetensk. Acad. Handl. 36: 69 (1775)	7
Aristolochiaceae	Lactoris Phil.	Verh. K. K. Zool.-Bot. Ges. Wien 15: 521 (1865)	1
Aristolochiaceae	Prospanche de Bary	Abh. Naturf. Ges. Halle 10: 267 (1868)	7
Aristolochiaceae	Saruma Oliv.	Hooker's Icon. Pl. 19: t. 1895 (1889)	1
Aristolochiaceae	Thottea Rottb.	Nye Saml. Kongel. Danske Vidensk. Selsk. Skr. 2: 529 (1783)	45
Begoniaceae	Begonia L.	Sp. Pl.: 1056 (1753)	1992
Begoniaceae	Hillebrandia Oliv.	Trans. Linn. Soc. London 25: 361 (1866)	1
Martyniaceae	Craniolaria L.	Sp. Pl.: 618 (1753)	2
Martyniaceae	Holoregmia Nees	Flora 4: 300 (1821)	1
Martyniaceae	Ibicella (Stapf) Van Eselt.	Techn. Bull. New York Agric. Exp. Sta., Geneva 149: 31 (1929)	2
Martyniaceae	Martynia L.	Sp. Pl.: 618 (1753)	2
Martyniaceae	Probosidea Schmidel	Icon. Pl. Anal., ed. Keller: 49 (1763)	7
Dipterocarpaceae	Anisoptera Korth.	Verh. Nat. Gesch. Ned. Bezitt., Bot.: 65 (1841)	10
Dipterocarpaceae	Cotylelobium Pierre	Fl. Forest. Cochinch.: t. 235 (1889)	5
Dipterocarpaceae	Dipterocarpus C.F.Gaertn.	Suppl. Carp. 3: 50 (1805)	65

On this page

- Setup
- POWOcodes - A complete list of angiosperm families and associated URI addresses
- angioData - A dataframe of Angiosperm species
- angioGenera - A dataframe of Angiosperm genera
- botregions - Countries and associated classification of botanical divisions

Dipterocarpaceae	Erythrinocarpa Erythrinocarpeae	Suppl. Carp.: 2-42 (1807)	1
Dipterocarpaceae	Hopea Roxb.	Pl. Coromandel 3: 7 (1811), nom. cons.	114
Dipterocarpaceae	Marquesia Gilg	Bot. Jahrb. Syst. 40: 485 (1908)	3
Dipterocarpaceae	Monotes A.DC.	A.P.de Candolle, Prodr. 16(2): 623 (1868)	23
Dipterocarpaceae	Neobalanocarpus P.S.Ashton	Fl. Males. 9: 388 (1982)	1
Dipterocarpaceae	Parashorea Kurz	J. Asiat. Soc. Bengal, Pt. 2, Nat. Hist. 39: 65 (1870)	13
Dipterocarpaceae	Pseudomonotes A.C.Londoño, E.Alvarez & Forero	Brittonia 47: 230 (1995)	1
Dipterocarpaceae	Shorea Roxb. ex C.F.Gaertn.	Suppl. Carp.: 47 (1805)	189
Dipterocarpaceae	Stemonoporus Thwaites	Hooker's J. Bot. Kew Gard. Misc. 6: 67 (1854)	26
Dipterocarpaceae	Upuna Symington	Bull. Jard. Bot. Buitenzorg, ser. 3, 17: 88 (1941)	1
Dipterocarpaceae	Vateria L.	Sp. Pl.: 515 (1753)	3
Dipterocarpaceae	Vateriopsis F.Heim	Rech. Dipterocarp.: 94 (1892)	1
Dipterocarpaceae	Vatica L.	Mant. Pl. 2: 152 (1771)	77
Fagaceae	Castanea Mill.	Gard. Dict. Abr. ed. 4: s.p. (1754)	8
Fagaceae	Castanopsis (D.Don) Spach	Hist. Nat. Veg. 11: 185 (1842), nom. cons.	143
Fagaceae	Chrysopsis Hjelmq.	Bot. Not. Suppl. 2, 1: 117 (1948)	2
Fagaceae	Fagus L.	Sp. Pl.: 997 (1753)	12
Fagaceae	Lithocarpus Blume	Bijdr. Fl. Ned. Ind.: 526 (1826)	343
Fagaceae	Notholithocarpus Manos, Cannon & S.H.Ch	Madrono 55: 188 (2008)	1
Fagaceae	Quercus L.	Sp. Pl.: 994 (1753)	464
Fagaceae	Trigonobalanus Forman	Taxon 11: 140 (1962)	3
Lecythidaceae	Allantoma Miers	Trans. Linn. Soc. London 30: 291 (1874)	8
Lecythidaceae	Asteranthos Desf.	Mem. Mus. Hist. Nat. 6: 9 (1820)	1
Lecythidaceae	Barringtonia J.R.Forst. & G.Forst.	Char. Gen. Pl.: 75 (1776), nom. cons.	72
Lecythidaceae	Bertholletia Bonpl.	F.W.H.Avon Humboldt & A.J.A.Bonpland, Pl. Aequinoct. 1: 122 (1807)	1
Lecythidaceae	Brazzeia Baill.	Bull. Mens. Soc. Linn. Paris 1: 609 (1886)	3
Lecythidaceae	Careya Roxb.	Fl. Ind. ed. 1832. 2: 636 (1832)	3
Lecythidaceae	Cariniana Casar.	Nov. Stirp. Bras. 4: 35 (1842)	9
Lecythidaceae	Chydenanthus Miers	Trans. Linn. Soc. London, Bot. 1: 111 (1875)	1
Lecythidaceae	Corythophora R.Knuth	H.G.A.Engler (ed.), Pflanzenr., IV, 219a: 50 (1939)	4
Lecythidaceae	Couratari Aubl.	Hist. Pl. Guiane 2: 723 (1775)	19
Lecythidaceae	Couroupita Aubl.	Hist. Pl. Guiane 2: 708 (1775)	3
Lecythidaceae	Crateranthus Baker f.	A.B.Rendle & al., Cat. Pl. Obern: 35 (1913)	4
Lecythidaceae	Eschweilera Mart. ex DC.	Prodr. 3: 293 (1828)	99
Lecythidaceae	Foetidia Comm. ex Lam.	Encycl. 2: 457 (1788)	18
Lecythidaceae	Grias L.	Syst. Nat. ed. 10. 2: 1075 (1759)	12
Lecythidaceae	Gustavia L.	Pl. Surin. 12: 17 (1775), nom. cons.	46
Lecythidaceae	Lecythis Loeff.	Iter Hispan.: 189 (1758)	30
Lecythidaceae	Napoleonaea P.Beauv.	Napoleone: 1 (1804)	17
Lecythidaceae	Oubanguia Baill.	Bull. Mens. Soc. Linn. Paris 2: 869 (1890)	3
Lecythidaceae	Petersianthus Merr.	Philipp. J. Sci., C 11: 200 (1916)	2
Lecythidaceae	Pierina Engl.	Bot. Jahrb. Syst. 43: 374 (1909)	1
Lecythidaceae	Planchonia Blume	Fl. Seres Jard. Eur. 7: 24 (1851)	8
Lecythidaceae	Rhaptopetalum Oliv.	J. Linn. Soc., Bot. 8: 159 (1865)	13
Lecythidaceae	Scytopetalum Pierre ex Engl.	H.G.A.Engler & K.A.E.Prantl, Nat. Pflanzenfam., Nachtr. 1: 244 (1897)	4

In addition to list all accepted genera in the above families, the dataframe-formatted `angioGenera` also contains publication, authorship, species number and global geographic distribution at country or botanical level for each genus. The global classification of botanical divisions follows the [World Geographical Scheme](#) for Recording Plant Distributions, which is already associated with each taxon's distribution in POWO.

The example below shows how to call `angioGenera`:

```
utils::data("angioGenera")
```

botregions - Countries and associated classification of botanical divisions

The `botregions` data package is a dataframe containing all countries of the World and associated classification of botanical divisions according to the [World Geographical Scheme](#) for Recording Plant Distributions. It was built using a custom script with specific functions from the packages `rgdal`, `ggplot2`, `broom`, `sf` and `raster` (original script can be made available upon request). This dataset is useful when the purpose is to query the distribution data by both divisions of the world and for the `expowo`'s functions convert the country names.

To call this associated data to your R environment, use the code below:

```
utils::data("botregions")
```



Mapping global distribution for any taxonomic level

Source: [vignettes/mapping_distribution.Rmd](#)

Here in this article, we show how to use the package's function `powoMap` for mapping the global distribution for any genus or family of flowering plants either at political country or botanical country level, following the [World Geographical Scheme for Recording Plant Distributions](#). The `powoMap` can use as input data the dataframe-formatted resulting query from `powoSpecies` function or a external spreadsheet. Then, `powoMap` can automatically create one global distribution map for the entire group (e.g., order, family) or multiple maps for each desired taxonomic level within the input data (e.g., genus). The map is colored according to the species richness using either [Viridis](#) or [RColorBrewer](#) color palettes.

Setup

Install the latest development version of **expowo** from [GitHub](#):

```
#install.packages("devtools")
devtools::install_github("DBOSlab/expowo")
```

```
library(expowo)
```

1. Mining species distribution from POWO to use as input data in powoMap

To create global distribution maps using `powoMap` function, you can either import a spreadsheet or use the package's `powoSpecies` function to get distribution information for the species of the target family or genus. It is possible to adjust one of the arguments in `powoSpecies` function to extract data only from a genus or list of genera. [See further details on how to do so in another article here](#).

The example below shows how to mine the distribution for an specific genus of Lecythidaceae. By defining a vector within the argument `genus`, here the function will search for all species and associated distribution data for just the genus *Cariniana*. Note that we create an object called `mapspdist`, which is the dataframe-formatted input data to create the map. The output shown here (TABLE 1) is a simplified version, where we have removed some columns so as to focus on the display of just the distribution data.

```
mapspdist <- powoSpecies(family = "Lecythidaceae",
                        genus = "Cariniana",
                        hybridsp = FALSE,
                        country = NULL,
                        verbose = TRUE,
                        save = FALSE,
                        dir = "results_powoSpecies/",
                        filename = "Lecythidaceae_Cariniana")
```

family	taxon_name	native_to_country	native_to_botanical_countries	kew_id
Lecythidaceae	Cariniana domestica	Bolivia, Brazil	Bolivia, Brazil North, Brazil West-Central	592121-1
Lecythidaceae	Cariniana estrellensis	Bolivia, Brazil, Paraguay	Bolivia, Brazil North, Brazil Northeast, Brazil South, Brazil Southeast, Brazil West-Central, Paraguay	47516-2
Lecythidaceae	Cariniana ianeirensis	Bolivia, Brazil	Bolivia, Brazil Southeast	47518-2
Lecythidaceae	Cariniana legalis	Brazil	Brazil Northeast, Brazil Southeast	592126-1
Lecythidaceae	Cariniana micrantha	Bolivia, Brazil	Bolivia, Brazil North, Brazil West-Central	47522-2
Lecythidaceae	Cariniana parvifolia	Brazil	Brazil Southeast	302134-2
Lecythidaceae	Cariniana penduliflora	Brazil	Brazil North	47527-2
Lecythidaceae	Cariniana pyriformis	Colombia, Venezuela	Colombia, Venezuela	592127-1
Lecythidaceae	Cariniana rubra	Brazil	Brazil North, Brazil Southeast, Brazil West-Central	592128-1

TABLE 1. A general `powoSpecies` search for mining distribution of the Lecythidaceae genus *Cariniana*.

2. Creating global-scale map for one genus to show species richness at political country level

To generate a global distribution map for one genus that show species richness at political country level, you can use the previously generated `mapspdist` object as input data of the `powoMap` function and set the `distcol` argument as "native_to_country". This is the very column of the dataframe-formatted `mapspdist` object that is associated to species distribution at political country.

The example below shows how to create the global distribution map of *Cariniana*'s species richness as colored by the `viridis` palette of the **Viridis** package. If you set up both the arguments `vir_color` and `bre_color` with any of the **Viridis** and **RColorBrewer** color palettes, then two maps colored respectively by these color vectors will be automatically generated. Different file formats and resolutions (e.g. JPG, PDF, PNG, TIFF) are also supported by setting the arguments `format` and `dpi`.

```
powoMap(inputdf = mapspdist,
        botctrs = FALSE,
        distcol = "native_to_country",
        taxclas = "genus",
        verbose = FALSE,
        save = FALSE,
        vir_color = "viridis",
        bre_color = NULL,
        leg_title = "SR",
```

On this page

Setup

1. Mining species distribution from POWO to use as input data in powoMap
2. Creating global-scale map for one genus to show species richness at political country level
3. Creating global-scale map for one genus to show species richness at botanical country level
4. Creating global-scale map of species richness for an entire family
5. Using an external spreadsheet as input data in powoMap

```
filename = "global_richness_country_map",
format = "jpg")
```



FIGURE 1. Global species richness of the genus *Cariniana* at country level and colored with viridis scale.

3. Creating global-scale map for one genus to show species richness at botanical country level

To generate a global distribution map for one genus that show species richness at botanical country level, adjust the argument `botctrs` within the `powoMap` function to `TRUE`. Then, use the previously generated `mapspdist` object as input data in `inputdf` and set the `distcol` argument to the name of the column with distribution according to botanical countries between quotation marks. This is the column of the dataframe-formatted `mapspdist` object or from an external spreadsheet.

The example below shows how to create the global distribution map of *Cariniana*'s species richness according to botanical subdivision and colored by the `viridis` palette of the **Viridis** package. If you set up both the arguments `vir_color` and `bre_color` with any of the **Viridis** and **RColorBrewer** color palettes, then two maps colored respectively by these color vectors will be automatically generated. Different file formats and resolutions (e.g. JPG, PDF, PNG, TIFF) are also supported by setting the arguments `format` and `dpi`.

```
powoMap(inputdf = mapspdist,
        botctrs = TRUE,
        distcol = "native_to_botanical_countries",
        taxclas = "genus",
        verbose = FALSE,
        save = FALSE,
        vir_color = "viridis",
        bre_color = NULL,
        leg_title = "SR",
        dpi = 600,
        dir = "results_powoMap/",
        filename = "global_richness_botcountry_map",
        format = "jpg")
```

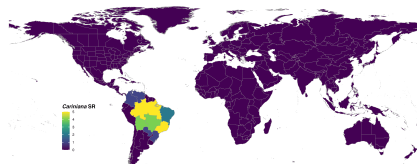


FIGURE 2. Global species richness of the genus *Cariniana* at botanical country level and colored with viridis scale.

4. Creating global-scale map of species richness for an entire family

To generate a family-level global distribution map of species richness, use `powoSpecies` function without any genus constraint so as to query a full species list of the desired family and the associated distribution data. Then, the resulting query object should be used as input data of the `powoMap` function.

```
mapspdist <- powoSpecies(family = "Lecythidaceae",
                        genus = NULL,
                        hybridsp = FALSE,
                        country = NULL,
                        verbose = FALSE,
                        save = FALSE,
                        dir = "results_powoSpecies/",
                        filename = "Lecythidaceae")
```

The example below uses the previously generated dataframe-formatted `mapspdist` object as input data to `powoMap` function, with the `distcol` argument set as `"native_to_country"`, the `taxclas` argument set as `"family"`, the `vir_color` argument set as `"viridis"`, and the `bre_color` argument set as `"Spectral"`, so as to automatically produce two **global distribution maps** of all *Lecythidaceae*'s species richness at **country level**, which are distinctly colored by `viridis` and `RColorBrewer` color palettes.

Note that if you had set here the `taxclas` as `"genus"`, then the `powoMap` function would have produced individual distribution maps of species richness for every single genus within *Lecythidaceae*, all of them being stored in the desired directory at `dir` argument.

```
powoMap(inputdf = mapspdist,
        botctrs = FALSE,
        distcol = "native_to_country",
        taxclas = "family",
        verbose = FALSE,
        save = FALSE,
        vir_color = "viridis",
        bre_color = "Spectral",
        leg_title = "SR",
        dpi = 600,
        dir = "results_powoMap/",
        filename = "global_richness_botcountry_map",
        format = "jpg")
```

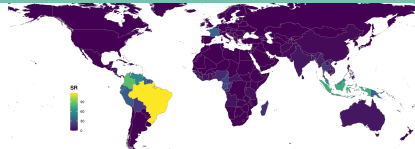


FIGURE 3. Global species richness of Lecythidaceae at country level and colored with viridis scale.

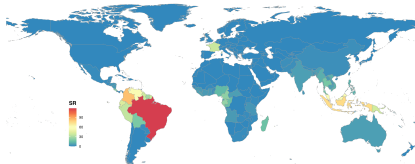


FIGURE 4. Global species richness of Lecythidaceae at country level and colored with Spectral scale.

To create global maps of Lecythidaceae according to botanical subdivision, change the argument `distcol` to "native_to_botanical_countries", the `vir_color` argument to "viridis", and the `bre_color` argument to "Spectral". This automatically produce two global distribution maps of Lecythidaceae's species richness at botanical country-level, which are distinctly colored by Viridis and RColorBrewer color palettes.

Note that if you had set here the `taxclas` as "genus", then the `powoMap` function would have produced individual distribution maps of species richness for every single genus within Lecythidaceae, all of them being stored in the desired directory at `dir` argument.

```
powoMap(inputdf = mapspdist,
        botctrs = TRUE,
        distcol = "native_to_botanical_countries",
        taxclas = "family",
        verbose = FALSE,
        save = FALSE,
        vir_color = "viridis",
        bre_color = "Spectral",
        leg_title = "SR",
        dpi = 600,
        dir = "results_powoMap/",
        filename = "global_richness_botcountry_map",
        format = "jpg")
```

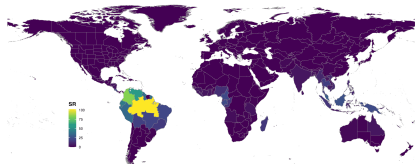


FIGURE 5. Global species richness of Lecythidaceae at botanical country level and colored with viridis scale.

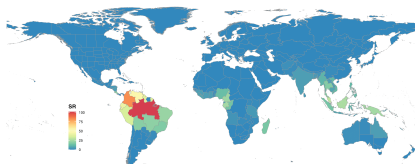


FIGURE 6. Global species richness of Lecythidaceae at botanical country level and colored with Spectral scale.

5. Using an external spreadsheet as input data in powoMap

It is possible to use data from other databases in order to create global maps using our `powoMap` function. To perform this task, the spreadsheet must have at least a column with species names and one or two columns with associated distribution in the countries and/or botanical regions. The species names must be binomial, and the country names must be written in full and separated by a comma and a space right after to the `powoMap` function works properly. The dataframe generated by our function `powoSpecies` can be used as a model to create this standard spreadsheet (see the examples above).

To use your external dataset as an imported spreadsheet, you must provide the name of the object in the argument `inputdf`. The `powoMap` function also will need information about the taxon level to generate the maps. To do so, the user must provide the name of the correspondent column in the argument `taxclas` (e.g., "family", "genus"). To indicate the column with the distribution, change the `distcol` argument. If the distribution is according to political country names (e.g., Brazil, Peru), you just have to write the name of the column with this data. Otherwise, if you need global maps according to the TDWG botanical subdivision level 3, you must indicate in which column this data (e.g., Brazil North, Peru) is stored and also change the argument `botctrs` from FALSE to TRUE.

Mining genus list for any plant family

Source: [vignettes/mining_accepted_genera.Rmd](#)

Here in this article, we show how to use the package's function `powoGenera` for mining all accepted genera for any family of flowering plants.

Setup

Install the latest development version of **expowo** from [GitHub](#):

```
#install.packages("devtools")
devtools::install_github("DBOSlab/expowo")

library(expowo)
```

Mining all accepted genera for any angiosperm family

The function `powoGenera` returns a dataframe or saves a CSV file listing all genera with associated number of accepted species and their global geographic distribution at country or botanical level. The global classification of botanical divisions follows the [World Geographical Scheme for Recording Plant Distributions](#), which is already associated with each taxon's distribution in POWO.

```
ABM <- powoGenera(family = c("Aristolochiaceae", "Begoniaceae", "Martyniaceae"),
  verbose = FALSE,
  save = FALSE,
  dir = "results_powoGenera/",
  filename = "Ar-ist_Begon_Martyniaceae_search")
```

family	genus	authors	scientific_name	publication	species_number	kew_id	powo_uri
Aristolochiaceae	Aristolochia	L.	Aristolochia L.	Sp. Pl.: 960 (1753)	534	330834-2	http://www.
Aristolochiaceae	Asarum	L.	Asarum L.	Sp. Pl.: 442 (1753)	129	3127-1	http://www.
Aristolochiaceae	Euglypha	Chodat & Hassl.	Euglypha Chodat & Hassl.	Bull. Herb. Boissier, sér. 2, 6: 138 (1906)	1	295147-2	http://www.
Aristolochiaceae	Hydnora	Thunb.	Hydnora Thunb.	Kongl. Vetensk. Acad. Handl. 36: 69 (1775)	7	30159275-2	http://www.
Aristolochiaceae	Lactoris	Phil.	Lactoris Phil.	Verh. K. K. Zool.-Bot. Ges. Wien 15: 521 (1865)	1	32025-1	http://www.
Aristolochiaceae	Prosopanche	de Bary	Prosopanche de Bary	Abh. Naturf. Ges. Halle 10: 267 (1868)	7	14132-1	http://www.
Aristolochiaceae	Saruma	Oliv.	Saruma Oliv.	Hooker's Icon. Pl. 19: t. 1895 (1889)	1	3165-1	http://www.
Aristolochiaceae	Thottea	Rottb.	Thottea Rottb.	Nye Saml. Kongel. Danske Vidensk. Selsk. Skr. 2: 529 (1783)	45	3169-1	http://www.
Begoniaceae	Begonia	L.	Begonia L.	Sp. Pl.: 1056 (1753)	1992	327437-2	http://www.
Begoniaceae	Hillebrandia	Oliv.	Hillebrandia Oliv.	Trans. Linn. Soc. London 25: 361 (1866)	1	3888-1	http://www.
Martyniaceae	Craniolaria	L.	Craniolaria L.	Sp. Pl.: 618 (1753)	2	31831-1	http://www.
Martyniaceae	Holoregmia	Nees	Holoregmia Nees	Flora 4: 300 (1821)	1	327223-2	http://www.
Martyniaceae	Ibicella	(Stapf) Van Eselt.	Ibicella (Stapf) Van Eselt.	Techn. Bull. New York Agric. Exp. Sta., Geneva 149: 31 (1929)	2	26127-1	http://www.
Martyniaceae	Martynia	L.	Martynia L.	Sp. Pl.: 618 (1753)	2	30002379-2	http://www.
Martyniaceae	Proboscoidea	Schmidel	Proboscoidea Schmidel	Icon. Pl. Anal., ed. Keller: 49 (1763)	7	30064181-2	http://www.

TABLE 1. A general `powoGenera` search for mining all accepted genera for some specific angiosperm families.

Mining all accepted genera for all angiosperm families

To mine a global genus checklist with associated species number and distribution for all families of flowering plants, we recommend to load the dataframe-formatted data object called `POWOCodes` that comes associated with the **expowo** package. Because the `POWOCodes` data object already contains the URI addresses for all angiosperms families recognized in the [POWO](#) database, you just need to run the code below.

- On this page
- Setup
 - Mining all accepted genera for any angiosperm family
 - Mining all accepted genera for all angiosperm families
 - Narrowing down the `powoGenera` search based on a specified country vector
 - Narrowing down the `powoGenera` search based on a specified genus vector

POWOCodes object.

```
data(POWOCodes)

ALL_gen <- powoGenera(POWOCodes$family,
  verbose = TRUE,
  save = FALSE,
  dir = "results_powoGenera/",
  filename = "all_angiosperm_genera")
```

Narrowing down the powoGenera search based on a specified country vector

You can also narrow down the search for all accepted genera of any family so as to focus on just a particular country or a list of countries. You just need to define a vector of country names in the argument `country`. In the example below, see that we have originally searched for the genera within the families Aristolochiaceae, Lecythidaceae, Fagaceae, and Dipterocarpaceae, but the function only returned a list of genera of Aristolochiaceae and Lecythidaceae, because the remaining families do not have any genera recorded in the list of countries of interest, i.e. Argentina, Brazil or French Guiana.

```
ADFL <- powoGenera(family =
  c("Aristolochiaceae", "Dipterocarpaceae", "Fagaceae", "Lecythidaceae"),
  country = c("Argentina", "Brazil", "French Guiana"),
  verbose = FALSE,
  save = FALSE,
  dir = "results_powoGenera/",
  filename = "country_constrained_search")
```

family	genus	authors	scientific_name	publication	species_number
Aristolochiaceae	Aristolochia	L.	Aristolochia L.	Sp. Pl.: 960 (1753)	534
Aristolochiaceae	Euglypha	Chodat & Hassl.	Euglypha Chodat & Hassl.	Bull. Herb. Boissier, sér. 2, 6: 138 (1906)	1
Aristolochiaceae	Prosopanche	de Bary	Prosopanche de Bary	Abh. Naturf. Ges. Halle 10: 267 (1868)	7
Lecythidaceae	Allantoma	Miers	Allantoma Miers	Trans. Linn. Soc. London 30: 291 (1874)	8
Lecythidaceae	Asteranthos	Desf.	Asteranthos Desf.	Mém. Mus. Hist. Nat. 6: 9 (1820)	1
Lecythidaceae	Bertholletia	Bonpl.	Bertholletia Bonpl.	F.W.H.A.von Humboldt & A.J.A.Bonpland, Pl. Aequinoct. 1: 122 (1807)	1
Lecythidaceae	Cariniana	Casar.	Cariniana Casar.	Nov. Stirp. Bras. 4: 35 (1842)	9
Lecythidaceae	Corythophora	R.Knuth	Corythophora R.Knuth	H.G.A.Engler (ed.), Pflanzenz., IV, 219a: 50 (1939)	4
Lecythidaceae	Couratari	Aubl.	Couratari Aubl.	Hist. Pl. Guiane 2: 723 (1775)	19
Lecythidaceae	Couroupita	Aubl.	Couroupita Aubl.	Hist. Pl. Guiane 2: 708 (1775)	3
Lecythidaceae	Eschweilera	Mart. ex DC.	Eschweilera Mart. ex DC.	Prodr. 3: 293 (1828)	99
Lecythidaceae	Gustavia	L.	Gustavia L.	Pl. Surin. 12: 17 (1775), nom. cons.	46
Lecythidaceae	Lecythis	Loefl.	Lecythis Loeffl.	Iter Hispan.: 189 (1758)	30

TABLE 2. A `powoGenera` search based on a specified country vector.

Narrowing down the powoGenera search based on a specified genus vector

You may want to retrieve information for just one or a list of accepted genera from a given country (or from a list of countries). Just like before, you only need to define a vector of genus names in the argument `genus` and a vector of country names in the argument `country`. In the example below, see that we have again searched for just the genera *Asarum* and *Bertholletia* of the families Aristolochiaceae and Lecythidaceae, but the function only returned the Lecythidaceae genus *Bertholletia*, because *Asarum* does not occur in any of the provided list of countries, i.e. Brazil or French Guiana.

```
AL <- powoGenera(family = c("Aristolochiaceae", "Lecythidaceae"),
  genus = c("Asarum", "Bertholletia"),
  country = c("Brazil", "French Guiana"),
  verbose = TRUE,
  save = FALSE,
  dir = "results_powoGenera/",
  filename = "genus_country_constrained_search")
```

family	genus	authors	scientific_name	publication	species_number	native_to_country
Lecythidaceae	Bertholletia	Bonpl.	Bertholletia Bonpl.	F.W.H.A.von Humboldt & A.J.A.Bonpland, Pl. Aequinoct. 1: 122 (1807)	1	Bolivia, Brazil, Colombia, French Guiana, Guyana, Suriname, Venezuela

TABLE 3. A `powoGenera` search based on specified genus and country vectors.



Mining species list for any plant family

Source: [vignettes/mining_accepted_species.Rmd](#)

Here in this article, we show how to use the package's function `powoSpecies` for mining all accepted species for any genus or family of flowering plants.

Setup

Install the latest development version of **expowo** from [GitHub](#):

```
#install.packages("devtools")
devtools::install_github("DBOSlab/expowo")
```

```
library(expowo)
```

Mining all accepted species for any angiosperm family

The function `powoSpecies` returns a dataframe or saves a CSV file listing all accepted species (including hybrid species or not), their publication, authorship, and global geographic distribution at country or botanical level for any family of flowering plants. The global classification of botanical divisions follows the [World Geographical Scheme](#) for Recording Plant Distributions, which is already associated with each taxon's distribution in the POWO.

The example below shows how to mine all accepted species for a specified vector of families. The output shown here (TABLE 1) is a simplified version, where we have removed some columns (genus, epithet and authors) so as to focus on the display of the taxonomic and distribution data.

```
res <- powoSpecies(family = c("Cabombaceae", "Martyniaceae"),
  hybridspp = FALSE,
  verbose = TRUE,
  save = FALSE,
  dir = "results_powoSpecies/",
  filename = "Cabon_Martyniaceae_search")
```

family	scientific_name	publication	native_to_country	native_to_botanical_countries	kew_id
Martyniaceae	Craniolaria annua L.	Sp. Pl. 618 (1753)	Colombia, Dominican Republic, Guyana, Venezuela, Venezuelan Antilles	Colombia, Dominican Republic, Guyana, Venezuela, Venezuelan Antilles	675800-1
Martyniaceae	Craniolaria integrifolia Cham.	Linnaea 7: 725 (1832)	Argentina, Bolivia, Brazil, Paraguay, Uruguay	Argentina Northeast, Argentina Northwest, Bolivia, Brazil Northeast, Brazil South, Brazil Southeast, Brazil West-Central, Paraguay, Uruguay	675805-1
Martyniaceae	Holoregmia viscida Nees	Flora 4: 300 (1821)	Brazil	Brazil Northeast	327210-2
Martyniaceae	Ibicella lutea (Lindl.) Van Eselt.	Techn. Bull. New York Agric. Exp. Sta., Geneva 149: 34 (1929)	Argentina, Brazil, Paraguay, Uruguay	Argentina Northeast, Argentina Northwest, Brazil Northeast, Brazil South, Brazil Southeast, Paraguay, Uruguay	565368-1
Martyniaceae	Ibicella parodii Abbiatti	Notas Mus. La Plata, Bot. 4: 458 (1939)	Argentina, Bolivia	Argentina Northeast, Argentina Northwest, Bolivia	127285-2
Martyniaceae	Martynia annua L.	Sp. Pl. 618 (1753)	Bahamas, Belize, Costa Rica, Cuba, Dominican Republic, El Salvador, Guatemala, Haiti, Honduras, Leeward Is., Mexico, Nicaragua, Puerto Rico, Trinidad-Tobago, Windward Is.	Bahamas, Belize, Costa Rica, Cuba, Dominican Republic, El Salvador, Guatemala, Haiti, Honduras, Leeward Is., Mexico Central, Mexico Gulf, Mexico Northeast, Mexico Northwest, Mexico Southeast, Mexico Southwest, Nicaragua, Puerto Rico, Trinidad-Tobago, Windward Is.	675840-1
Martyniaceae	Martynia palmeri S.Watson	Proc. Amer. Acad. Arts 24: 66 (1889)	Mexico	Mexico Northwest	153752-2
Martyniaceae	Proboscidea althaeifolia (Benth.) Decne.	Ann. Sci. Nat., Bot., ser. 5, 3: 324 (1865)	Mexico, Peru, United States	Arizona, California, Mexico Northeast, Mexico Northwest, New Mexico, Peru, Texas	675893-1
Martyniaceae	Proboscidea diversifolia Hevly	Brittonia 21: 311 (1970)	Mexico	Mexico Southwest	283779-2
Martyniaceae	Proboscidea louisianica (Mill.) Thell.	Mem. Soc. Sci. Nat. Math. Cherbourg 38: 480 (1912)	Mexico, United States	California, Colorado, Florida, Georgia, Kansas, Kentucky, Louisiana, Mexico Central, Mexico Gulf, Mexico Northeast, Mexico Northwest, Mexico Southwest, Mississippi, Missouri, Nebraska, New Jersey, New Mexico, North Carolina, Oregon, Pennsylvania, South Carolina, South Dakota, Texas, Vermont, Virginia, Washington, Texas, Wyoming	210335-2
Martyniaceae	Proboscidea parviflora (Wootton) Wootton & Standl.	Contr. U.S. Natl. Herb. 19: 602 (1915)	Mexico, United States	Arizona, California, Mexico Central, Mexico Northeast, Mexico Northwest, Nevada, New Mexico, Texas, Utah	210338-2
Martyniaceae	Proboscidea sabulosa Correll	Rhodora 68: 426 (1966)	United States	New Mexico, Texas	210340-2
Martyniaceae	Proboscidea spicata Correll	Wrightia 4: 77 (1968)	Mexico	Mexico Northeast	210342-2
Martyniaceae	Proboscidea triloba (Schlttdl. & Cham.) Decne.	Ann. Sci. Nat., Bot., ser. 5, 3: 326 (1865)	Guatemala, Mexico	Guatemala, Mexico Central, Mexico Gulf, Mexico Northeast, Mexico Southeast, Mexico Southwest	675906-1

On this page

Setup

Mining all accepted species for any angiosperm family

Narrowing down the powoSpecies search based on a specified country vector

Narrowing down the powoSpecies search based on a specified genus and country vector

Mining all accepted species for all angiosperm families

Cabombaceae	<i>Ardisia schreberi</i> J.F.Gmel.	Syst. Nat. ed. 13(bis): 853 (1791)	Angola, Australia, Bangladesh, Belize, Botswana, Burundi, Cameroon, Canada, China, Cuba, District of Columbia, Dominican Republic, El Salvador, Ethiopia, Guatemala, Guyana, India, Jamaica, Japan, Kenya, Mexico, Nigeria, North Korea, South Korea, Quebec, Rhode I., Russia, South Africa, Taiwan, Tanzania, Uganda, United States, Venezuela, Zaire, Zambia, Zimbabwe	Albania, Arkansas, Assam, Bangladesh, Belize, Botswana, British Columbia, Burundi, California, Cameroon, China South-Central, China Southeast, Connecticut, Cuba, Delaware, District of Columbia, Dominican Republic, El Salvador, Florida, Georgia, Guatemala, Guyana, Idaho, Illinois, India, Indiana, Iowa, Jamaica, Japan, Kansas, Kentucky, Kenya, Korea, Louisiana, Maine, Manitoba, Maryland, Massachusetts, Mexico Gulf, Mexico Northeast, Mexico Southeast, Mexico Southwest, Michigan, Minnesota, Mississippi, Missouri, Montana, New Brunswick, New Hampshire, New Jersey, New South Wales, New York, Newfoundland, Nigeria, North Carolina, Northern Provinces, Nova Scotia, Ohio, Oklahoma, Ontario, Oregon, Pennsylvania, Primorye, Queensland, Quebec, Rhode I., South Carolina, Taiwan, Tanzania, Tennessee, Texas, Uganda, Venezuela, Vermont, Victoria, Virginia, Washington, West Virginia, Wisconsin, Zambia, Zaire, Zimbabwe	1
Cabombaceae	<i>Cabomba aquatica</i> Aubl.	Hist. Pl. Guiane 1: 321 (1775)	Bolivia, Brazil, Colombia, French Guiana, Guyana, Peru, Suriname, Venezuela	Bolivia, Brazil North, Brazil Northeast, Brazil Southeast, Brazil West-Central, Colombia, French Guiana, Guyana, Peru, Suriname, Venezuela	605272-1
Cabombaceae	<i>Cabomba caroliniana</i> A.Gray	Ann. Lyceum Nat. Hist. New York 4: 47 (1837)	Argentina, Bolivia, Brazil, District of Columbia, Mexico, Paraguay, Rhode I., United States, Uruguay	Alabama, Argentina Northeast, Argentina South, Arkansas, Bolivia, Brazil South, Brazil Southeast, Connecticut, Delaware, District of Columbia, Florida, Georgia, Illinois, Indiana, Kentucky, Louisiana, Maryland, Massachusetts, Mexico Northeast, Michigan, Mississippi, Missouri, New Hampshire, New Jersey, New York, North Carolina, Ohio, Oklahoma, Paraguay, Pennsylvania, Rhode I., South Carolina, Tennessee, Texas, Uruguay, Virginia	273581-2
Cabombaceae	<i>Cabomba furcata</i> Schult. & Schult.f.	Syst. Veg. 7: 1379 (1830)	Bolivia, Brazil, Colombia, Costa Rica, Cuba, Ecuador, Guyana, Honduras, Panama, Peru, Suriname, Trinidad-Tobago, Venezuela	Bolivia, Brazil North, Brazil Northeast, Brazil Southeast, Brazil West-Central, Colombia, Costa Rica, Cuba, Ecuador, Guyana, Honduras, Panama, Peru, Suriname, Trinidad-Tobago, Venezuela	273584-2
Cabombaceae	<i>Cabomba haynesii</i> Wiersma	Ann. Missouri Bot. Gard. 76: 1167 (1989)	Brazil, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Honduras, Jamaica, Nicaragua, Panama, Puerto Rico, Trinidad-Tobago, Venezuela	Brazil West-Central, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Honduras, Jamaica, Nicaragua, Panama, Puerto Rico, Trinidad-Tobago, Venezuela	273585-2
Cabombaceae	<i>Cabomba palaeiformis</i> Fassett	Castanea 18: 127 (1953)	Belize, Costa Rica, Guatemala, Mexico, Nicaragua	Belize, Costa Rica, Guatemala, Mexico Central, Mexico Gulf, Mexico Northeast, Mexico Southeast, Mexico Southwest, Nicaragua	39039-2
Cabombaceae	<i>Cabomba schwartzii</i> Rataj	Acta Amazonica 7: 143 (1977)	Brazil	Brazil North	128749-1

TABLE 1. A general `powoSpecies` search for mining all accepted species for some specific angiosperm families.

Narrowing down the `powoSpecies` search based on a specified country vector

You can also narrow down the species checklist search of any family by focusing on just a particular country or a list of countries. You just need to define a vector of country names in the argument `. In the example below, note that we have originally searched for the species within the families Aristolochiaceae, Cabombaceae, and Martyniaceae, but the function returned a smaller species list, because many species in the searched families are not recorded in any of the specified vector of country, i.e. Brazil or Colombia.`

```
ACM <- powoSpecies(family = c("Aristolochiaceae", "Cabombaceae", "Martyniaceae"),
  hybridspp = FALSE,
  country = c("Brazil", "Colombia"),
  verbose = FALSE,
  save = FALSE,
  dir = "results_powoSpecies/",
  filename = "country_constrained_search")
```

family	scientific_name	publication	native_to_country	native_to_botanical_countries
Aristolochiaceae	<i>Aristolochia arcuata</i> Mast.	C.F.P.von Martius & auct. suc. (eds.), Fl. Bras. 4(2): 101 (1875)	Brazil	Brazil South, Brazil Southeast, Brazil West-Central
Aristolochiaceae	<i>Aristolochia assisii</i> J.Freitas, Lirio & F.Gonzalez	Phytotaxa 163: 263 (2014)	Brazil	Brazil Northeast, Brazil Southeast
Aristolochiaceae	<i>Aristolochia bahiensis</i> F.Gonzalez	Brittonia 50: 8 (1998)	Brazil	Brazil Northeast, Brazil Southeast
Aristolochiaceae	<i>Aristolochia birostris</i> Duch.	Ann. Sci. Nat. Bot., ser. 4, 2: 60 (1854)	Brazil	Brazil Northeast
Aristolochiaceae	<i>Aristolochia brunneomaculata</i> L.S.Abreu & Giul.	Sitientibus Ser. Cl. Biol. 16: 1 (2016)	Brazil	Brazil Northeast
Aristolochiaceae	<i>Aristolochia burchellii</i> Mast.	C.F.P.von Martius & auct. suc. (eds.), Fl. Bras. 4(2): 90 (1875)	Brazil	Brazil North
Aristolochiaceae	<i>Aristolochia cabreriae</i> Ahumada	Folia Bot. Geobot. Corrent. 2: 1 (1989)	Brazil	Brazil West-Central
Aristolochiaceae	<i>Aristolochia chrysoclhora</i> Barb.Rodr.	Vellozia 2: 23A (1885-1888 publ.1891)	Brazil	Brazil North
Aristolochiaceae	<i>Aristolochia colombiana</i>	Revista Acad. Colomb. Ci. Exact. 90:	Colombia	Colombia

Aristolochiaceae	<i>Aristolochia colossifolia</i> Hoehne	Arq. Bot. Estado Sao Paulo 2: 171 (1952)	Colombia	Colombia
Aristolochiaceae	<i>Aristolochia cordigera</i> (Klotzsch) Duch.	A.P.de Candolle, Prodr. 15(1): 455 (1864)	Brazil	Brazil North, Brazil Northeast, Brazil Southeast, Brazil West-Central
Aristolochiaceae	<i>Aristolochia cymbifera</i> Mart.	Nov. Gen. Sp. Pl. 1: 76 (1824)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia cynanchifolia</i> Mart.	Nov. Gen. Sp. Pl. 1: 77 (1824)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia echinata</i> Barb.Rodr.	Pl. Jard. Rio de Janeiro 5: 9 (1896)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia eriantha</i> Mart.	Nov. Gen. Sp. Pl. 1: 78 (1824)	Brazil	Brazil Northeast
Aristolochiaceae	<i>Aristolochia filipendulina</i> Duch.	Ann. Sci. Nat., Bot., ser. 4, 2: 70 (1854)	Brazil	Brazil West-Central
Aristolochiaceae	<i>Aristolochia gardneri</i> Duch.	Ann. Sci. Nat., Bot., ser. 4, 2: 69 (1854)	Brazil	Brazil North, Brazil Northeast, Brazil West-Central
Aristolochiaceae	<i>Aristolochia gehrtii</i> Hoehne	Arq. Bot. Estado Sao Paulo, n.s., f.m., 1: 45 (1939)	Brazil	Brazil Southeast, Brazil West-Central
Aristolochiaceae	<i>Aristolochia gigantea</i> Mart. & Zucc.	Flora 7(1, Beil.): 139 (1824)	Brazil	Brazil Northeast, Brazil South, Brazil Southeast
Aristolochiaceae	<i>Aristolochia ginzbergeri</i> Ahumada	Darwiniana 21: 78 (1977 publ. 1978)	Brazil	Brazil North
Aristolochiaceae	<i>Aristolochia gracilipedunculata</i> F.Gonzalez	Brittonia 63: 431 (2011)	Brazil	Brazil Northeast, Brazil Southeast
Aristolochiaceae	<i>Aristolochia gracilis</i> Duch.	A.P.de Candolle, Prodr. 15(1): 460 (1864)	Brazil	Brazil Southeast, Brazil West-Central
Aristolochiaceae	<i>Aristolochia hilariana</i> Duch.	A.P.de Candolle, Prodr. 15(1): 451 (1864)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia hispida</i> Pohl ex Duch.	A.P.de Candolle, Prodr. 15(1): 498 (1864)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia huberiana</i> S.Moore	J. Bot. 53: 6 (1915)	Brazil	Brazil North
Aristolochiaceae	<i>Aristolochia huebneriana</i> O.C.Schmidt	Repert. Spec. Nov. Regni Veg. 40: 133 (1936)	Brazil	Brazil North
Aristolochiaceae	<i>Aristolochia hypoglauca</i> Kuhlhm.	Arq. Inst. Biol. Veg. 3: 45 (1936)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia insolita</i> J.Freitas & M.Peixoto	Syst. Bot. 42: 169 (2017)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia leratuma</i> F.Gonzalez & Palbon-Mora	Caldasia 39(1): 52 (2017)	Colombia	Colombia
Aristolochiaceae	<i>Aristolochia kruckhoffii</i> O.C.Schmidt	Repert. Spec. Nov. Regni Veg. 45: 53 (1938)	Brazil	Brazil North
Aristolochiaceae	<i>Aristolochia labiata</i> Willd.	Mem. Soc. Imp. Naturalistes Moscou 2: 101 (1809)	Brazil	Brazil North, Brazil Northeast, Brazil South, Brazil Southeast, Brazil West-Central
Aristolochiaceae	<i>Aristolochia lanceolatorata</i> S.Moore	J. Bot. 53: 4 (1915)	Brazil	Brazil North
Aristolochiaceae	<i>Aristolochia limai</i> Hoehne	Arq. Bot. Estado Sao Paulo 1: 91 (1942)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia longispaululata</i> F.Gonzalez	Novon 10: 371 (2000)	Brazil	Brazil Northeast
Aristolochiaceae	<i>Aristolochia lorenae</i> J.Freitas & F.Gonzalez	Syst. Bot. 44: 101 (2019)	Brazil	Brazil North
Aristolochiaceae	<i>Aristolochia lutescens</i> Duch.	A.P.de Candolle, Prodr. 15(1): 461 (1864)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia manaosensis</i> Ahumada	Darwiniana 21: 77 (1977 publ. 1978)	Brazil	Brazil North
Aristolochiaceae	<i>Aristolochia marianensis</i> Ahumada	Hickenia 1: 171 (1979)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia medicinalis</i> R.E.Schult.	Bot. Mus. Leaf. 21: 266 (1967)	Colombia	Colombia
Aristolochiaceae	<i>Aristolochia melastoma</i> Silva Manso ex Duch.	A.P.de Candolle, Prodr. 15(1): 461 (1864)	Brazil	Brazil North, Brazil South, Brazil Southeast, Brazil West-Central
Aristolochiaceae	<i>Aristolochia morae</i> F.Gonzalez	Revista Acad. Colomb. Ci. Exact. 71: 500 (1993)	Colombia	Colombia
Aristolochiaceae	<i>Aristolochia nevesamondiana</i> Hoehne	Arch. Bot. Sao Paulo 1: 7 (1925)	Brazil	Brazil Southeast
Aristolochiaceae	<i>Aristolochia odora</i> Steud.	Nomencl. Bot., ed. 2, 1: 133 (1840)	Brazil	Brazil Northeast, Brazil Southeast
Aristolochiaceae	<i>Aristolochia papillaris</i> Mast.	C.F.P.von Martius & auct. suc. (eds.), Fl. Bras. 4(2): 100 (1875)	Brazil	Brazil North, Brazil Northeast, Brazil South, Brazil Southeast
Aristolochiaceae	<i>Aristolochia paulistana</i> Hoehne	Arch. Bot. Sao Paulo 1: 13 (1925)	Brazil	Brazil Northeast, Brazil South, Brazil Southeast
Aristolochiaceae	<i>Aristolochia pohliana</i> Duch.	A.P.de Candolle, Prodr. 15(1): 496 (1864)	Brazil	Brazil Northeast, Brazil Southeast, Brazil West-Central
Aristolochiaceae	<i>Aristolochia pubescens</i> Willd. ex Duch.	A.P.de Candolle, Prodr. 15(1): 450 (1864)	Brazil	Brazil Northeast

Aristolochiaceae	Aristolochia rugosa Mart.	Nov. Gen. Sp. Pl. 1: 78 (1824)	Brazil	Brazil Southeast
Aristolochiaceae	Aristolochia robertii Ahumada	Fl. Illustr. Catar. 1(Arist.): 35 (1975)	Brazil	Brazil South
Aristolochiaceae	Aristolochia rumicifolia Mart.	Nov. Gen. Sp. Pl. 1: 79 (1824)	Brazil	Brazil Southeast
Aristolochiaceae	Aristolochia sepicola Mast.	C.F.P.von Martius & auct. suc. (eds.), Fl. Bras. 4(2): 100 (1875)	Brazil	Brazil Northeast
Aristolochiaceae	Aristolochia setosa Duch.	A.P.de Candolle, Prodr. 15(1): 496 (1864)	Brazil	Brazil Northeast, Brazil West-Central
Aristolochiaceae	Aristolochia smilacina Duch.	A.P.de Candolle, Prodr. 15(1): 459 (1864)	Brazil	Brazil Northeast, Brazil Southeast, Brazil West-Central
Aristolochiaceae	Aristolochia subglobosa J.Freitas, Lirio & F.Gonzalez	Phytotaxa 124: 56 (2013)	Brazil	Brazil Southeast
Aristolochiaceae	Aristolochia tamnifolia (Klotzsch) Duch.	A.P.de Candolle, Prodr. 15(1): 448 (1864)	Brazil	Brazil Northeast, Brazil Southeast
Aristolochiaceae	Aristolochia trulliformis Mast.	C.F.P.von Martius & auct. suc. (eds.), Fl. Bras. 4(2): 101 (1875)	Brazil	Brazil North, Brazil Northeast
Aristolochiaceae	Aristolochia urbaniana Taub.	Beibl. Bot. Jahrb. Syst. 27: 14 (1890)	Brazil	Brazil Southeast
Aristolochiaceae	Aristolochia warmingii Mast.	C.F.P.von Martius & auct. suc. (eds.), Fl. Bras. 4(2): 109 (1875)	Brazil	Brazil North, Brazil Northeast, Brazil Southeast, Brazil West-Central
Aristolochiaceae	Aristolochia wendeliana Hoehne	Arq. Bot. Estado Sao Paulo, n.s., Em., 1: 45 (1939)	Brazil	Brazil South, Brazil Southeast
Aristolochiaceae	Aristolochia zebrina J.Freitas & F.Gonzalez	Nordic J. Bot. 34: 54 (2016)	Brazil	Brazil Southeast
Aristolochiaceae	Prosopanche caatingicola R.F.Machado & L.P.Queiroz	Phytotaxa 75: 59 (2012)	Brazil	Brazil Northeast
Aristolochiaceae	Prosopanche cocucci Tav.de Carvalho, Zachia & Mariath	Phytotaxa 521: 178 (2021)	Brazil	Brazil South
Aristolochiaceae	Prosopanche demogorgoni Funez	Phytotaxa 422: 94 (2019)	Brazil	Brazil South
Martyniaceae	Holoregmia viscida Nees	Flora 4: 300 (1821)	Brazil	Brazil Northeast
Cabombaceae	Cabomba schwartzii Rataj	Acta Amazonica 7: 143 (1977)	Brazil	Brazil North

TABLE 2. A `powoSpecies` search based on a specified country vector.

Narrowing down the `powoSpecies` search based on a specified genus and country vector

You may want to retrieve information for just one or a list of accepted genera from a given country (or from a list of countries). Just like before, you only need to define a vector of genus names in the argument and a vector of country names in the argument. In the example below, see that we have searched for just the genera *Proboscidea* and *Cabomba* of the families **Martyniaceae** and **Cabombaceae**, but the function only returned the **Cabombaceae** genus *Cabomba*, because *Proboscidea* does not occur in any of the provided list of countries, i.e. **Argentina**, **French Guiana**, or **Brazil**.

```
MC <- powoSpecies(family = c("Martyniaceae", "Cabombaceae"),
  genus = c("Proboscidea", "Cabomba"),
  hybridspp = FALSE,
  country = c("Argentina", "French Guiana", "Brazil"),
  verbose = TRUE,
  save = FALSE,
  dir = "results_powoSpecies/",
  filename = "country_and_genus_constrained_search")
```

family	taxon_name	authors	scientific_name	publication	native_to_country	native_to_botanical_coun
Cabombaceae	Cabomba schwartzii	Rataj	Cabomba schwartzii Rataj	Acta Amazonica 7: 143 (1977)	Brazil	Brazil North

TABLE 3. A `powoSpecies` search based on specified genus and country vectors.

Mining all accepted species for all angiosperm families

To mine a species checklist for all families of flowering plants, you recommend to load the dataframe-formatted data object called `POWOCodes` that comes associated with the **expowo** package. The `POWOCodes` data object already contains the URI addresses for all angiosperms families recognized in the [POWO](#) database, so you just need to call it to your R environment.

The example below shows how to mine a global checklist of all accepted species of flowering plants by using the vector of all angiosperm families and associated URI addresses stored in the `POWOCodes` object.

```
utils::data("POWOCodes")

ALL_sp <- powoSpecies(POWOCodes$family,
  hybridspp = TRUE,
  verbose = TRUE,
  save = FALSE,
  dir = "results_powoSpecies/",
  filename = "all_angiosperm_species")
```



Mining species number of any plant family

Source: [vignettes/mining_family_species_number.Rmd](#)

Here in this article, we show how to use the package's function `powoFam` for mining the species number for any family of flowering plants. By providing a family name or a vector of family names, the `powoFam` function visits each POWO page of the genera within the queried family(ies) to extract the accepted species number. Then, it sums these values to deliver the total species number within each plant family.

Setup

Install the latest development version of **expowo** from [GitHub](#):

```
#install.packages("devtools")
devtools::install_github("D805lab/expowo")
```

```
library(expowo)
```

Mining the total species number for any angiosperm family

The function `powoFam` returns a dataframe or saves a CSV file listing the total species number (excluding hybrid species) for the queried plant families.

The example below shows how to mine the total species number by defining a vector of three families: **Cabombaceae**, **Lecythidaceae**, and **Martyniaceae**. The resulting output (TABLE 1) is also saved directly in the specified directory **results_powoFam**.

```
CLM <- powoFam(family = c("Cabombaceae", "Lecythidaceae", "Martyniaceae"),
               verbose = TRUE,
               save = FALSE,
               dir = "results_powoFam/",
               filename = "Camb_Lecy_Martyniaceae_diversity")
```

family	species_number	kew_id	powo_uri
Cabombaceae	7	77126578-1	https://powo.science.kew.org/taxon/um:lsid:jnpi.org:names:77126578-1
Lecythidaceae	381	30000348-2	https://powo.science.kew.org/taxon/um:lsid:jnpi.org:names:30000348-2
Martyniaceae	14	77126797-1	https://powo.science.kew.org/taxon/um:lsid:jnpi.org:names:77126797-1

TABLE 1. A general `powoFam` search for mining the total species number of three angiosperm families.

Mining species number accross all angiosperm families

To mine the species number of all families of flowering plants, we recommend to load the data frame-formatted object called `POWOCodes` that comes associated with the **expowo** package. The `POWOCodes` data object already contains the URI addresses for all angiosperms families recognized in the [POWO](#) database, so you just need to call it to your R environment.

The example below shows how to mine the species number of all plant families by using the vector of all angiosperm families and associated URI addresses stored in the `POWOCodes` object.

```
utils::data(POWOCodes)

ALL_fam <- powoFam(POWOCodes$family,
                  verbose = TRUE,
                  save = FALSE,
                  dir = "results_powoFam/",
                  filename = "all_angiosperms_species_number")
```

On this page

Setup

Mining the total species number for any angiosperm family

Mining species number accross all angiosperm families



Mining megadiverse genera of flowering plants

Source: [vignettes/mining_megadiverse_genera.Rmd](#)

Here in this article, we show how to use the package's function `megaGen` for mining the megadiverse genera for any family of flowering plants. Frodin (2004) in Taxon suggests 500 species as the threshold for a genus to be considered megadiverse. The function `megaGen` uses any integer numeric value in the argument `thld` to define the threshold for a megadiverse genus.

Setup

Install the latest development version of **expowo** from [GitHub](#):

```
#install.packages("devtools")
devtools::install_github("D805lab/expowo")
```

```
library(expowo)
```

Mining megadiverse genera for any angiosperm family

The function `megaGen` uses a specified threshold number to return a dataframe or save a CSV file listing all megadiverse genera for any family of flowering plants and their associated number of accepted species. In this example, we performed a search for three plant families using a threshold of 500 species, but the output only contains results for **Aristolochia** and **Begonia** of **Aristolochiaceae** and **Begoniaceae** families.

```
ABM <- megaGen(family = c("Aristolochiaceae", "Begoniaceae", "Martyniaceae"),
  thld = 500,
  verbose = TRUE,
  save = FALSE,
  dir = "results_megaGen/",
  filename = "Arist_Begon_Martyniaceae_search")
```

family	genus	authors	scientific_name	species_number	kew_id	powo_uri
Aristolochiaceae	Aristolochia	L.	Aristolochia L.	534	330834-2	http://www.plantsoftheworld.org/
Begoniaceae	Begonia	L.	Begonia L.	1992	327437-2	http://www.plantsoftheworld.org/

TABLE 1. A general `megaGen` search to mining the megadiverse genera for some specific angiosperm families, based on a spec



Mining megadiverse genera accross all angiosperm families

To mine a global checklist of megadiverse genera for all families of flowering plants, including their associated species number, we recommend to load the dataframe-formatted data object called `POWOcodes` that comes associated with the **expowo** package. The `POWOcodes` data object already contains the URI addresses for all angiosperms families recognized in the [POWO](#) database, you just need to call it to your R environment.

The example below shows how to mine a global checklist of all most speciose genera of flowering plants by using the vector of all angiosperm families and associated URI addresses stored in the `POWOcodes` object.

```
data(POWOcodes)

ALL_megaGen <- megaGen(POWOcodes$family,
  thld = 500,
  verbose = TRUE,
  save = FALSE,
  dir = "results_megaGen/",
  filename = "all_megadiverse_angiosperm_genera")
```

On this page

Setup

Mining megadiverse genera for any angiosperm family

Mining megadiverse genera accross all angiosperm families



Mining distribution for any plant species

Source: [vignettes/mining_species_distribution.Rmd](#)

Here in this article, we show how to use the package's function `powoSpDist` for mining the entire distribution for any species of flowering plants. The `powoSpDist` uses the argument `species` to set the name or list of species names, and then the function access the auxiliary functions `getGenURI` and `getDist` to catch their respective distribution.

Setup

Install the latest development version of **expowo** from [GitHub](#):

```
#install.packages("devtools")
devtools::install_github("D805lab/expowo")
```

```
library(expowo)
```

Mining the distribution for any angiosperm species

The function `powoSpDist` returns a dataframe or saves a CSV file listing the entire global distribution at country or botanical level for any of the given species scientific name(s) (excluding hybrid species), their genus, publication and authorship. The global classification of botanical divisions follows the [World Geographical Scheme](#) for Recording Plant Distributions, which is already associated with each taxon's distribution in the POWO.

The example below shows how to mine the distribution for a specified vector of two species. The output shown here (a table) is simplified by removing the family and distribution according to botanical subdivision columns.

```
BL_dist <- powoSpDist(family = c("Begoniaceae", "Lecythidaceae"),
  species = c("Hillebrandia sandwicensis", "Lecythis pisonis"),
  verbose = TRUE,
  save = FALSE,
  dir = "results_powoSpDist/",
  filename = "Begoniaceae_Lecythidaceae")
```

family	taxon_name	authors	publication	native_to_country	introduced_to_country	kew_id
Lecythidaceae	Lecythis pisonis	Cambess.	A.F.C.P.de Saint-Hilaire & al., Fl. Bras. Merid. 2: 377 (1833)	Brazil, Colombia, Peru	Trinidad-Tobago	597315-1
Begoniaceae	Hillebrandia sandwicensis	Oliv.	Trans. Linn. Soc. London 25: 361 (1866)	United States	unknown	106195-1

TABLE 1. A general `powoSpDist` search for mining distribution of two angiosperm species.

Mining distribution for all angiosperm species

To mine the distribution checklist for all accepted species of flowering plants, you can load the dataframe-formatted data object called `POWOcodes` that comes associated with the **expowo** package. The `POWOcodes` data object already contains the URI addresses for all angiosperms families recognized in the [POWO](#) database, you just need to call it to your R environment.

The example below shows how to mine the global distribution of all accepted non hybrid species of flowering plants by using the vector of all angiosperm families and associated URI addresses stored in the `POWOcodes` object. Note that the argument `species` should be as `NULL`, so that the search is not constrained to any particular species, i.e. the function will return global distribution for all species in all queried families.

```
data(POWOcodes)

ALL_dist <- powoSpDist(POWOcodes$family,
  species = NULL,
  verbose = TRUE,
  save = FALSE,
  dir = "results_powoSpDist/",
  filename = "all_angiosperm_distribution")
```

On this page

Setup

Mining the distribution for any angiosperm species

Mining distribution for all angiosperm species



Mining the top richest genera of flowering plants

Source: [vignettes/mining_toprichest_genera.Rmd](#)

Here in this article, we show how to use the package's function `topGen` for mining the top most species rich genera for any family of flowering plants.

Setup

Install the latest development version of **expowo** from [GitHub](#):

```
#install.packages("devtools")
devtools::install_github("DBOSlab/expowo")
```

```
library(expowo)
```

Mining the top richest genera for any angiosperm family

The function `topGen` is relatively similar to the `megaGen`, but instead of using a specific threshold of species number to be considered a big genera, it saves a CSV file listing the top most species-rich genera of any target angiosperm family, and their associated number of accepted species based on an integer number set in the argument `limit`. In the example below, we used the default to search for the top ten richest genera (i.e., when the argument `limit` is set as `NULL`, the function searches for the 10 most diverse genera within a plant family). Some columns were omitted to display the results, and the rows show the top ten of the three chosen families: Bignoniaceae, Solanaceae and Lecythidaceae. Note that the table does not have 30 rows, what would be expected if we multiply each family ($n = 3$) by ten. It is not an error, since some plant families do not have many genera, which is the case of Begoniaceae with 2 genera.

```
ABL_top <- topGen(family = c("Aristolochiaceae", "Begoniaceae", "Lecythidaceae"),
  limit = NULL,
  verbose = TRUE,
  save = FALSE,
  dir = "results_topGen/",
  filename = "Aristo_Bego_Lecythidaceae_search")
```

family	genus	species_number
Aristolochiaceae	Aristolochia	534
Aristolochiaceae	Asarum	129
Aristolochiaceae	Thottea	45
Aristolochiaceae	Hydnora	7
Aristolochiaceae	Prosopanche	7
Aristolochiaceae	Euglypha	1
Aristolochiaceae	Lactoris	1
Aristolochiaceae	Saruma	1
Begoniaceae	Begonia	1992
Begoniaceae	Hillebrandia	1
Lecythidaceae	Eschweillera	99
Lecythidaceae	Barringtonia	72
Lecythidaceae	Gustavia	46
Lecythidaceae	Lecythis	30
Lecythidaceae	Couratari	19
Lecythidaceae	Foetidia	18
Lecythidaceae	Napoleonaea	17
Lecythidaceae	Rhaptopetalum	13
Lecythidaceae	Grias	12
Lecythidaceae	Cariniana	9

TABLE 1. A general `topGen` search to mining the top most species rich genera for some specific angiosperm families.

Mining the top richest genera accross all angiosperm families

To mine a global checklist of the top species-richest genera for all families of flowering plants, including their associated species number, we recommend to load the dataframe-formatted data object called `POWOcodes` that comes associated with the **expowo** package. The `POWOcodes` data object already contains the URI addresses for all angiosperms families recognized in the [POWO](#) database, so you just need to call it to your R environment.

The example below shows how to mine all top most species-rich genera of flowering plants by using the vector of all angiosperm families, the associated URI addresses stored in the `POWOcodes` object and the limit of 10 genera for each family.

```
data(POWOcodes)

ALL_top <- topGen(POWOcodes$family,
  limit = 10,
  verbose = TRUE,
  save = FALSE,
  dir = "results_topGen/",
  filename = "all_toprichest_angiosperm_genera")
```

On this page

Setup

Mining the top richest genera for any angiosperm family

Mining the top richest genera accross all angiosperm families

