










Synthesizing tree biodiversity data to understand global patterns and processes of vegetation

Gunnar Keppel^{1,2,3}  | Dylan Craven^{1,4,5,6}  | Patrick Weigelt^{1,7}  |
Stephen A. Smith⁸  | Masha T. van der Sande^{4,9,10,11,12}  | Brody Sandel¹³  |
Sam C. Levin^{4,5}  | Holger Kreft^{1,14}  | Tiffany M. Knight^{4,5,12} 

¹Biodiversity, Macroecology and Biogeography, University of Goettingen, Göttingen, Germany

²UniSA STEM, University of South Australia, Adelaide, SA, Australia

³Future Industries Institute, University of South Australia, Adelaide, SA, Australia

⁴German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany

⁵Department of Biologie, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany

⁶Centro de Modelación y Monitoreo de Ecosistemas, Universidad Mayor, Santiago, Chile

⁷Campus-Institut Data Science, Göttingen, Germany

⁸Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

⁹Institute for Global Ecology, Florida Institute of Technology, Melbourne, FL, USA

¹⁰Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

¹¹Forest Ecology and Forest Management Group, Wageningen University and Research, Wageningen, The Netherlands

¹²Department of Community Ecology, Helmholtz Centre for Environmental Research-UFZ, Halle (Saale), Germany

¹³Department of Biology, Santa Clara University, Santa Clara, CA, USA

¹⁴Centre of Biodiversity and Sustainable Land Use (CBL), University of Goettingen, Göttingen, Germany

Correspondence

Gunnar Keppel, UniSA STEM, University of South Australia, Room 2-50, F Building, Mawson Lakes Campus, 5095 South Australia, Australia.
Email: gunnar.keppel@unisa.edu.au

Tiffany M. Knight, German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Puschstraße 4, 04103 Leipzig, Germany.
Email: tiffany.knight@idiv.de

Funding information

TMK was supported by the Alexander von Humboldt Foundation in the framework of the Alexander von Humboldt Professorship and by the Helmholtz Recruitment Initiative of the Helmholtz Association. GK was partially supported by an Alexander von Humboldt Foundation fellowship. MvdS was

Abstract

Aims: Trees dominate the biomass in many ecosystems and are essential for ecosystem functioning and human well-being. They are also one of the best-studied functional groups of plants, with vast amounts of biodiversity data available in scattered sources. We here aim to illustrate that an efficient integration of these data could produce a more holistic understanding of vegetation.

Methods: To assess the extent of potential data integration, we use key databases of plant biodiversity to: (a) obtain a list of tree species and their distributions; (b) identify coverage of and gaps in different aspects of tree biodiversity data; and (c) discuss large-scale patterns of tree biodiversity in relation to vegetation.

Results: Our global list of trees included 58,044 species. Taxonomic coverage varies in three key databases, with data on the distribution, functional traits, and molecular sequences for about 84%, 45% and 44% of all tree species, which is >10% greater than

Holger Kreft and Tiffany M. Knight are joint senior authors. This article is a part of the Special Feature Macroecology of vegetation, edited by Meelis Pärtel, Francesco Maria Sabatini, Naia Morueta-Holme, Holger Kreft and Jürgen Dengler

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Vegetation Science* published by John Wiley & Sons Ltd on behalf of International Association for Vegetation Science



supported by the Rubicon research program with project number 019.171LW.023, the Veni research program with project number NOW-VI.Veni.192.027. and the project NWO-ALW.OP24, all financed by the Netherlands Organisation for Scientific Research (NWO).

Co-ordinating Editor: Meelis Pärtel

for plants overall. For 28% of all tree species, data are available in all three databases. However, less data are digitally accessible about the demography, ecological interactions, and socio-economic role of tree species. Integrating and imputing existing tree biodiversity data, mobilization of non-digitized resources and targeted data collection, especially in tropical countries, could help closing some of the remaining data gaps.

Conclusions: Due to their key ecosystem roles and having large amounts of accessible data, trees are a good model group for understanding vegetation patterns. Indeed, tree biodiversity data are already beginning to elucidate the community dynamics, functional diversity, evolutionary history and ecological interactions of vegetation, with great potential for future applications. An interoperable and openly accessible framework linking various databases would greatly benefit future macroecological studies and should be linked to a platform that makes information readily accessible to end users in biodiversity conservation and management.

KEYWORDS

biodiversity, biological databases, conservation, data integration and synthesis, forests, functional traits, IUCN Red List, macroecology, species distribution, tree diversity, vegetation, woodlands

1 | INTRODUCTION

Trees are the building blocks of many terrestrial ecosystems (Crowther et al., 2015; Bastin et al., 2017). In these ecosystems, trees are long-lived organisms (Chambers et al., 1998) that constitute the majority of living biomass, provide structure, habitats and resources for associated biota, and create unique microclimates (Keppel et al., 2017; Liu et al., 2018; Lutz et al., 2018). Trees play key roles in biogeochemical cycling and have a major effect on global climate through their role in carbon and water cycling (Spracklen et al., 2012; Lindenmayer & Laurance, 2017). Because of their extraordinary importance to ecosystems and to humans since prehistoric times (Fenning & Gershenson, 2002), trees are among the best-studied plant groups and, therefore, have the potential to provide important insights about large-scale patterns of vegetation biodiversity and composition.

Therefore, extensive data on tree biodiversity are available and advances in data science can now facilitate their collation and integration. Available data cover different aspects of biodiversity, ranging from taxonomic to socio-economic information (Figure 1) and are spread across multiple sources and locations, such as local and regional databases, literature and natural history collections (Beaman & Cellinese, 2012; Jetz et al., 2012; König et al., 2019). This information is being digitized to create comprehensive databases, an increasing proportion of which are open access (Appendix S1). Recent progress in big-data science, therefore, holds great promise for biodiversity studies and allows increasingly rapid integration of diverse data (Franklin et al., 2017; Forest et al., 2018; Allen et al., 2019; König et al., 2019).

Here we illustrate that advances in data science allow an integrated global synthesis of tree biodiversity data, and that such a

synthesis can lead to a more holistic understanding of vegetation. Specifically, we: (a) summarize key data available for different aspects of tree biodiversity and quantify their limitations with respect to coverage and accessibility; (b) evaluate the representation of trees in selected key databases; (c) identify the most important gaps in tree biodiversity data and outline steps required to fill them; and (d) explore potential applications of integrated tree biodiversity data to improve our understanding of vegetation patterns and processes.

2 | METHODS

Biodiversity is multifaceted, and we here consider nine key aspects (Figure 1). Of these aspects, a high-quality taxonomy is essential for holistic, integrated assessments of biodiversity. Tools such as the Taxonomic Name Resolution Service (TNRS; Boyle et al., 2013) and World Flora Online (<http://www.worldfloraonline.org/>) are available to help validating and standardizing the names of taxa. For example, we used the TNRS to link and integrate the species and distribution data from a global tree species list by Beech et al. (2017) with data from key biodiversity-related databases.

The global checklist (Beech et al., 2017) names 59,722 tree species with country-level distribution data (GlobalTreeSearch version 1.5; http://www.bgci.org/global_tree_search.php). Following the Global Tree Specialists Group of the International Union for the Conservation of Nature (IUCN), GlobalTreeSearch defines trees as woody plants growing at least 2 m high with a single stem or, if multi-stemmed, with at least one vertical stem being ≥ 5 cm in diameter at breast height. While this definition is widely used, it is more

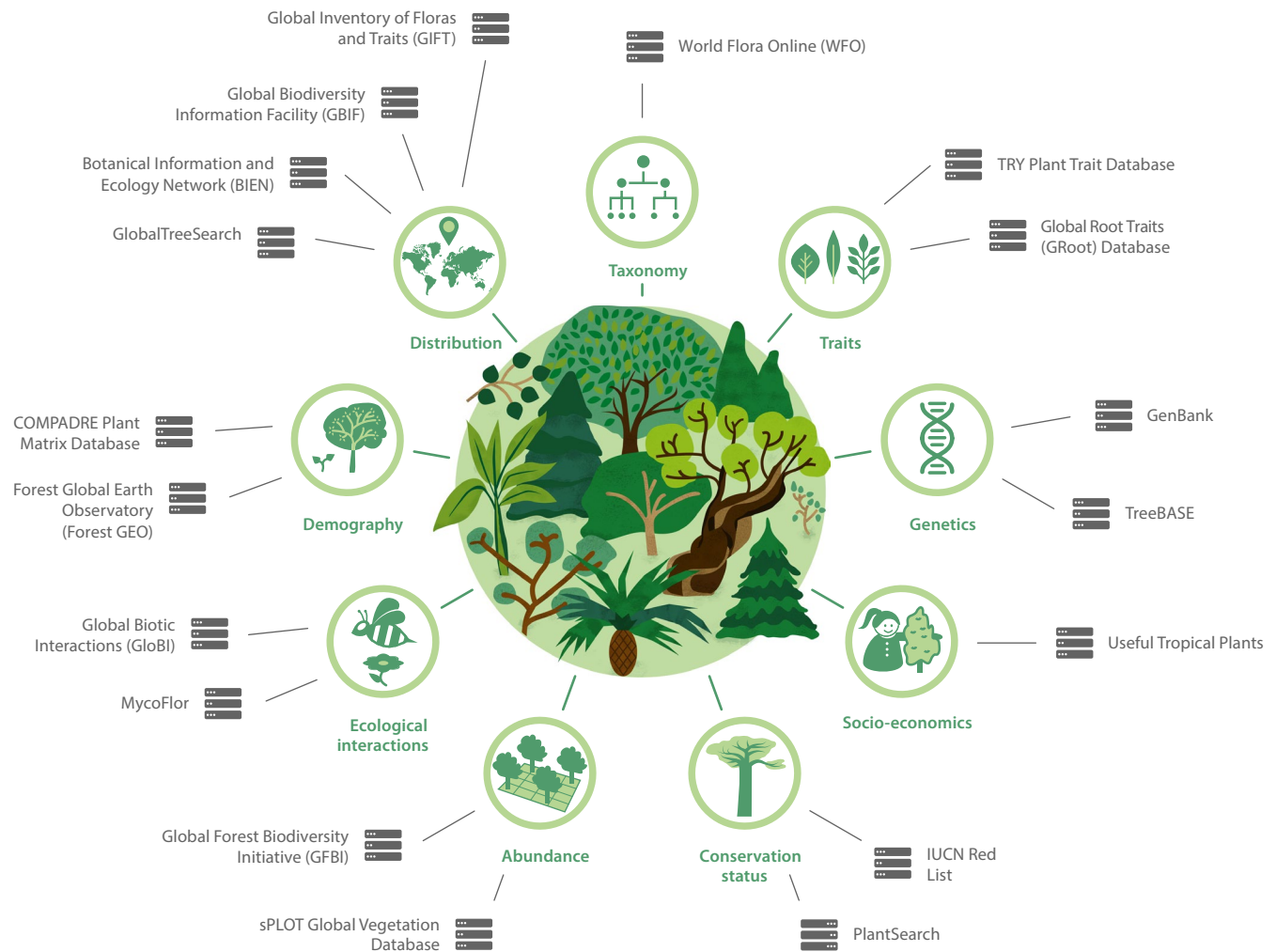


FIGURE 1 Overview of nine key aspects of biodiversity for macroecological studies and the major global databases (see Appendix S1 for more detail) with information for trees and other taxa. Note that many databases contain information on multiple aspects of biodiversity; here we only link each database to a single aspect

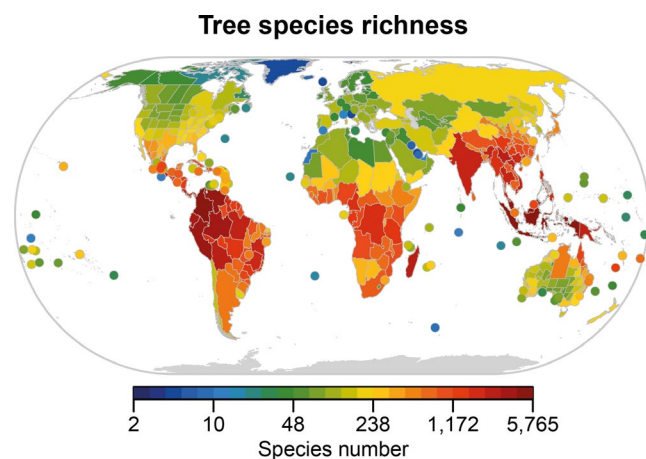


FIGURE 2 Global distribution of tree species diversity by region based on data from GlobalTreeSearch (http://www.bgci.org/global_tree_search.php), version 1.4, and the Global Inventory of Floras and Traits (GIFT; <http://gift.uni-goettingen.de>, Weigelt et al., 2020). See Appendix S2 for details on how the data were collated

inclusive than alternative definitions (cf. “freestanding plants” in Taseski et al., 2019). Here we combine data from GlobalTreeSearch with tree species and distribution data at finer spatial resolutions for large countries (i.e., USA, Canada, Mexico, Brazil, China and Australia) from the Global Inventory of Flora and Traits (GIFT) database (<http://gift.uni-goettingen.de>, Weigelt et al., 2020) using only records indicating the species to be native. The integration of tree data from these two databases and taxonomic standardization (see Appendix S2 for details) produced a final data set (the “tree list”) that included 58,044 tree species (Appendix S3) and distribution information across 463 geographic regions worldwide (Figure 2). We used this list to query various databases, including spatial occurrences from the Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>; GBIF occurrence download <https://doi.org/10.15468/dl.77gcvq>, accessed from R via *rgbif* [<https://github.com/ropensci/rgbif>] on 2021-03-16), publicly available traits from TRY version 5 (<https://www.try-db.org/TryWeb/Home.php>, Kattge et al., 2020), abundance data in sPlot 3.0 (<https://www.idiv.de/de/>

splot.html-, Bruelheide et al., 2019), molecular sequences from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), and conservation assessments from the IUCN Red List version 2020-3 (<https://www.iucnredlist.org/>; Appendix S2). We evaluated the completeness of publicly available tree biodiversity data by determining the number of species in the “tree list” for which data were available in the selected databases.

3 | UNDERSTANDING THE MACROECOLOGY OF VEGETATION USING EXISTING TREE BIODIVERSITY DATA

Below we discuss the completeness of data currently available in selected key databases for various aspects of tree biodiversity (Figure 1) and evaluate where gaps are present, focusing on global, geographically wide-ranging sources. Key databases are summarized in Appendix S1. Furthermore, we discuss and illustrate how the available tree biodiversity data could be applied to better understand macroecological patterns of vegetation. Where applicable, we also compare data availability for trees with that reported for all plants (Cornwell et al., 2019).

3.1 | Distribution and abundance

Providing a spatial dimension for biodiversity, distribution data are central to many applications in ecology, evolution and conservation (Franklin et al., 2013; Keppel et al., 2015; Daru et al., 2017). Several sources of occurrence data are available, and each has its challenges and limitations (Meyer et al., 2016; König et al., 2019). Published floras and checklists are often highly curated sources of information that provide the most complete occurrence records of taxa in a region or country, and an increasing number of these are digitally available in global databases (König et al., 2019). However, these have limited spatial precision, as many species do not occur throughout an entire range or region (Meyer et al., 2016). Herbarium and museum collections are the traditional source of the most precise geo-referenced tree occurrence data and the number of digitized specimens continues to increase (Lavoie, 2013; Soltis, 2017). Increasingly, citizen scientists produce high-quality data through various platforms that, when leveraged with user expertise and advanced artificial intelligence, provide information about the distribution and ecology of plants (Havens & Henderson, 2013; Van Horn et al., 2018).

Despite important gaps in occurrence data (Meyer et al., 2016; Serra-Diaz et al., 2017; Hortal et al., 2015), databases such as GBIF, Botanical Information and Ecology Network (BIEN; <http://bien.nceas.ucsb.edu/bien/>, Enquist et al., 2016) and GIFT (Weigelt et al., 2020) provide a good indication of the global distribution of tree species richness at the country scale. The largest of these databases, GBIF, currently holds >23 million georeferenced records for trees, with about 84.4% of all tree species having at least one record (Figure 3a). Hence, trees have about 11% better representation in

this database than plants in general (cf. Cornwell et al., 2019) and information for distribution at the country scale is available for all described tree species in GlobalTreeSearch (Beech et al., 2017). However, GBIF records frequently contain erroneous taxonomic and spatial information and may not capture the full extent of a species' range (Meyer et al., 2016; Zizka et al., 2020). Furthermore, GBIF has poor coverage for tree species on many islands, central Asia, the Middle East and central and northeast Africa (Figure 3a).

Databases with geo-referenced occurrence records, like GBIF, are particularly valuable for understanding large-scale vegetation patterns. Because a small number of tree species often dominates and defines vegetation, even in tropical rain forests (Keppel et al., 2011; Pitman et al., 2013; ter Steege et al., 2013), they can be good indicators of the distribution of vegetation. For example, the distribution records for the Australian desert oak, *Allocasuarina decaisneana*, suggest that the associated desert oak woodland vegetation may be more widespread than currently mapped (Figure 4). Spatial distribution data can also be used to predict the responses of species and vegetation to climatic change, provided relevant environmental data are available at grain sizes fine enough to capture habitat affinities of species (Franklin et al., 2013; Fourcade et al., 2018).

Forest plots and national forest inventories are frequently used primary sources of vegetation data and provide information on the abundance and distribution of tree species. Data on species abundance and distribution, and environmental data, can be integrated to determine how climatic conditions, dispersal barriers and/or biotic interactions may shape spatial patterns of abundance (Dallas et al., 2017; Copenhaver-Parry & Bell, 2018; Steidinger et al., 2019). Furthermore, abundance data can provide important insights into the dynamics of vegetation, particularly if related to disturbances, either in the form of time series from multiple censuses (Li et al., 2016), long-term historic data on abundance from fossil pollen records (van der Sande et al., 2019), or through comparing biodiversity change across sites with different disturbance intensities in a space-for-time substitution (Rozendaal et al., 2019; Ibanez et al., 2020).

While plot data and relevant databases are becoming increasingly available and comprehensive (e.g., Dengler et al., 2011; Bruelheide et al., 2019), they are spread thinly across a labyrinth of sources with different levels of access and are frequently reported in non-standard formats (Wiser, 2016). Furthermore, plot databases often have a strong spatial bias. For example, while sPlot (Bruelheide et al., 2019) includes at least one record for 25.6% of all tree species, its coverage is much better in Europe than elsewhere (Figure 3c). Like other databases, such as GFBI (Global Forest Biodiversity Initiative; <http://gfbinitiative.com/>), sPlot, therefore, has limited coverage of plot data in highly biodiverse regions, particularly Amazonia, southeast Asia, the Congo Basin and the southwest Pacific Islands, which is preventing a thorough understanding of these forests and their species (ter Steege et al., 2015; Tovo et al., 2017). At finer scales there may also be spatial biases, with more plot data believed to be available for locations that are easier to access or have experienced less anthropogenic disturbance (Jobe & White, 2009; Phillips et al., 2002).

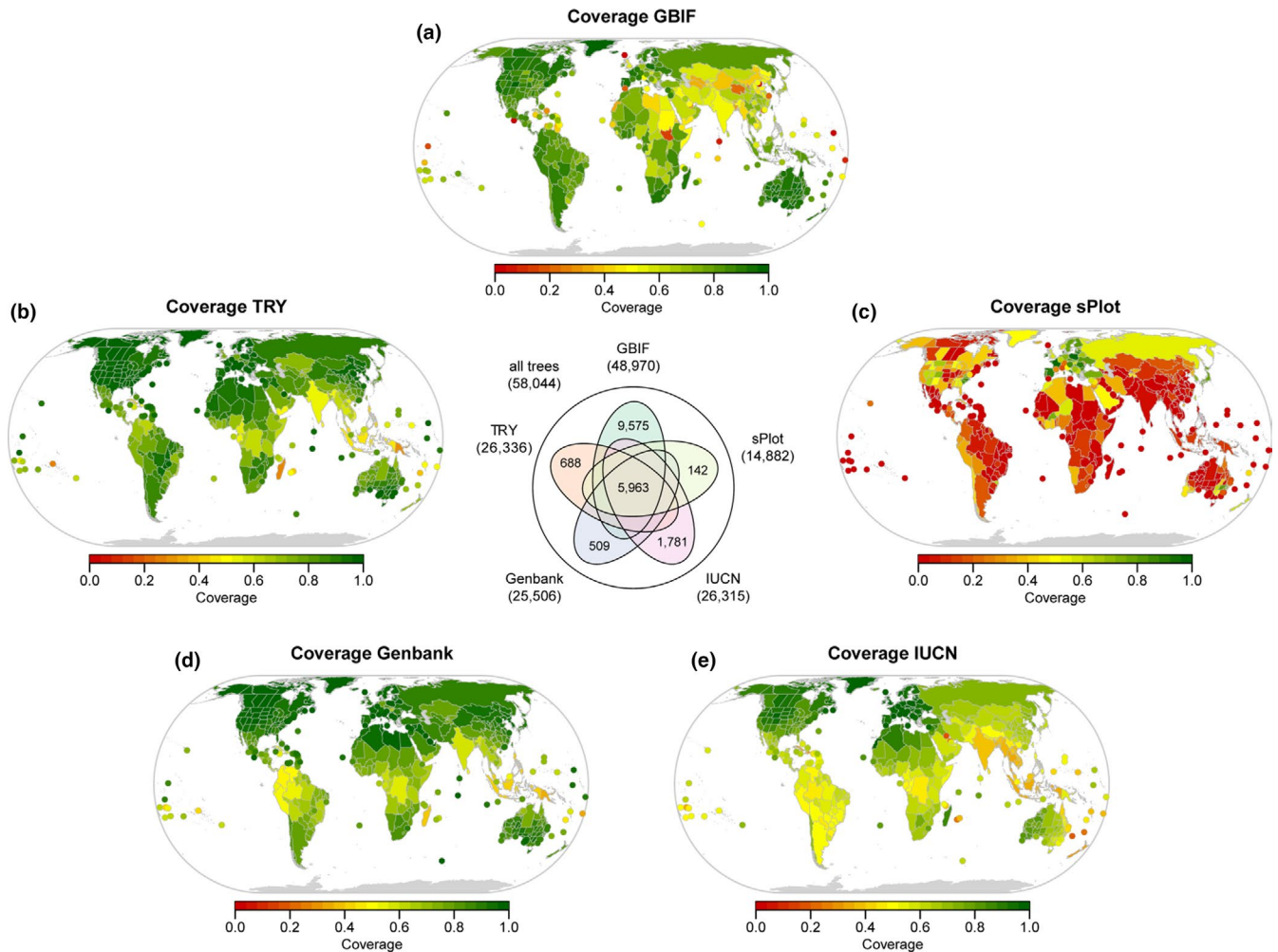


FIGURE 3 Completeness of global digitally accessible tree biodiversity data in key databases by region. Maps indicate the proportion of species covered per region with records for: (a) distribution in GBIF (<https://www.gbif.org/>); (b) at least three traits in TRY version 5 (<https://www.try-db.org/>); (c) at least one plot in sPlot 3.0 (<https://www.idiv.de/de/splot.html>); (d) at least one sequence in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>); and (e) assessment of conservation status in IUCN Red List version 2020-3 (<https://www.iucnredlist.org/>). In the central Venn diagram numbers of tree species unique to the various databases and the number of species represented in all databases are reported. Coverage = proportional coverage of tree species in a database with respect to total species richness in an administrative unit. See Appendix S1 for more details on databases and Appendix S2 for methods to produce values used in this figure

Other data inconsistencies and biases further limit our ability to understand spatial patterns of abundance. Plot sizes across forest community data sets vary considerably (Liang et al., 2016; Bruelheide et al., 2019) and this affects estimates of forest structure and dynamics (Wagner et al., 2010). In addition, variability in minimum tree size thresholds, sampled growth forms (e.g., lianas, palms and tree ferns), and taxonomy among forest plots, further complicate efforts to find generality in biodiversity patterns (Wiser, 2016; Muscarella et al., 2020).

3.2 | Demography and ecological interactions

Demographic information reveals how the environment influences vegetation and populations (Condit et al., 1999) through its relationship with the vital rates (survival, growth, reproduction) of

individuals (Caswell, 2001). Generally, models are used to connect vital rates of trees to population dynamics, as vital rates depend on the size of the individual (seedlings have low survivorship and fertility, while large trees have high survivorship and fertility) and individuals contribute unequally to population dynamics (Caswell, 2001; Ramula et al., 2009). However, limited demographic data of trees throughout their life cycle are currently available in the COMPADRE database (<https://compadre-db.org/>; Salguero-Gómez et al., 2015).

Forest plots with multiple censuses, such as those in the Forest Global Earth Observatory (ForestGEO; <https://forestgeo.si.edu/>; Anderson-Teixeira et al., 2015) can provide vital information about forest dynamics (Condit et al., 1999; Sullivan et al., 2020) and additional demographic data (e.g., Visser et al., 2016). Limitations of these data include that they are not typically available in an open access format and that individuals are often only measured in forest plots once they reach a minimum size. However, summary statistics

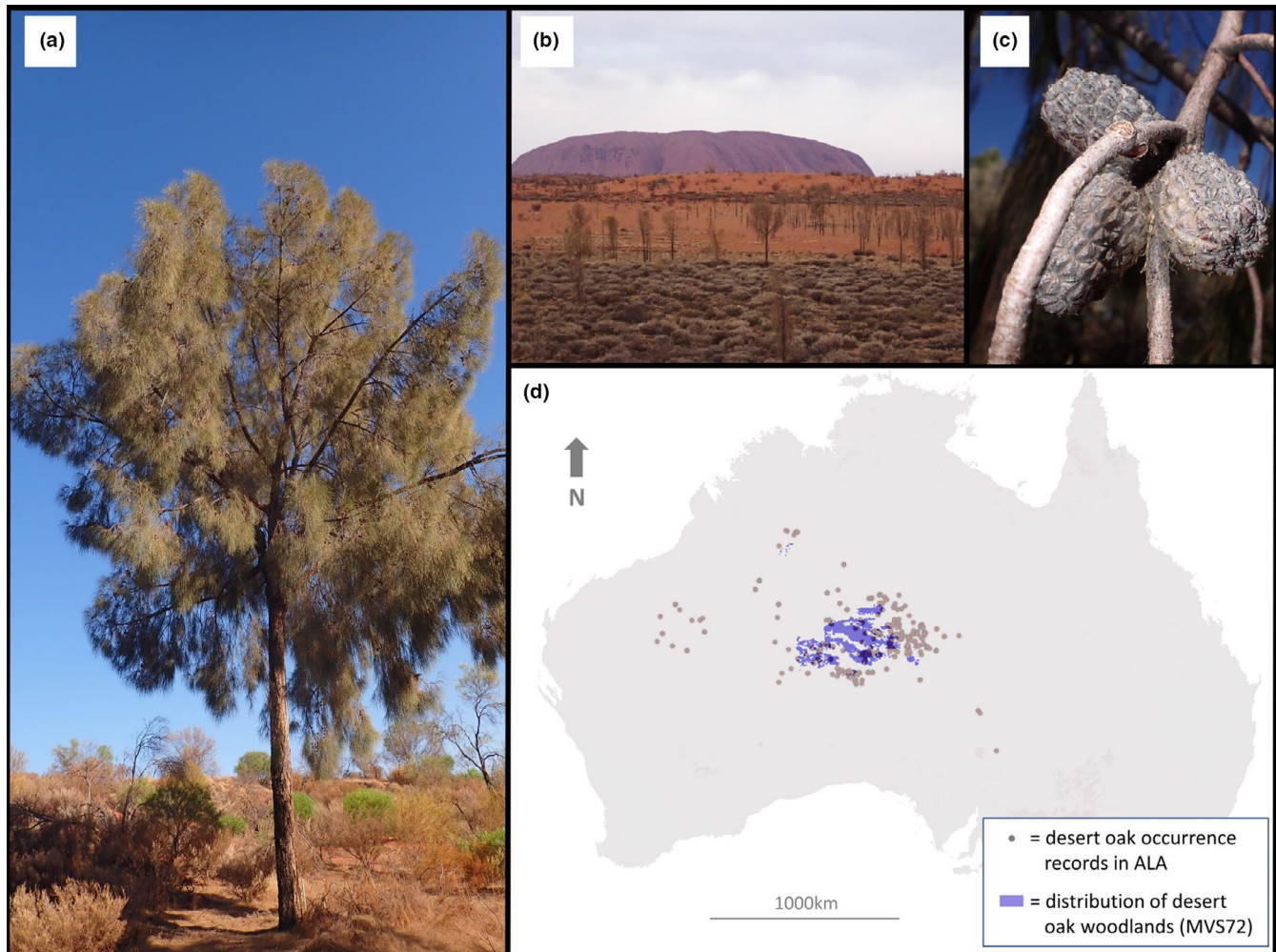


FIGURE 4 The Australian desert oak, *Allocasuarina decaisneana* (Casuarinaceae) is restricted to sandy desert environments in central Australia and exemplifies the close relationship between the distribution of tree species and related vegetation types. Adults (a) are often the only trees in sandy deserts, defining the desert oak woodlands major vegetation subgroup (MVS 72; Keith & Pellow, 2015). The species starts as feather-duster-like seedlings (b) that start branching after reaching below-ground water sources and has the largest fruits of the Casuarinaceae (c), suggesting unique functional traits. The occurrence records of the species (Atlas of Living Australia, ALA; <https://www.ala.org.au/>) suggest a wider distribution for MVS 72 than currently mapped (National Vegetation Information System V5.1, © Australian Government Department of Agriculture, Water and the Environment 2018; d)

can be synthesized from these data to gain an understanding of tree population dynamics. For example, two key trade-offs explained about a third of the variation in demographic traits in the tropical moist forests of Barro Colorado Island, Panama; fast-growing, light-demanding species with high mortality vs slow-growing, shade-tolerant species with lower mortality, and long-lived, tall species with low recruitment vs short-lived, shorter species with high recruitment. The resulting demographic model was able to accurately predict the changes in structure and composition during secondary succession in these forests (Rüger et al., 2020).

Trees interact with a variety of organisms, including other trees of the same or other species, pollinators, seed dispersers and predators, herbivores, parasites and symbionts. While some interactions of trees are relatively easy to quantify, others are laborious (e.g., below-ground functions) and rarely measured (Freschet et al., 2021). Many of these interactions are vital for maintaining biodiversity, the functioning of

ecosystems and for the performance and survival of trees (Neuschulz et al., 2016; Steidinger et al., 2019). However, few databases focus on ecological interactions (see Figure 1 and Appendix S1 for examples), and those that do primarily contain species of agricultural or agroforestry importance or are geographically restricted (e.g., MycoFlor; Hempel et al., 2013). Despite these limitations, the Global Biotic Interactions (GloBI) database (Poelen et al., 2014) is a useful tool for accessing available data sets, particularly as more are added in the future. While pollination and dispersal of trees have been studied extensively using direct and genetic approaches (Bacles et al., 2006; Bennett et al., 2018), only limited information for pollination is found in databases, preventing adequate assessment of phenomena like the global pollination crisis (Bartomeus et al., 2019). However, information on mycorrhizae and nitrogen fixation is more widely available (Appendix S1), likely due to the great importance of these interactions for plant survival and performance (Steidinger et al., 2019).



Forest plot data can reveal factors driving species co-occurrence, including ecological interactions such as competition and facilitation, which may co-vary with environmental conditions (Lankau et al., 2015; Steidinger et al., 2019). For example, in the Brazilian Atlantic forest, fragmentation was found to result in strong shifts toward tree species with less specialized pollinators and dispersers, which often require large stretches of high-quality forest (Girão et al., 2007; da Silva & Tabarelli, 2000). On a global scale, linking tree composition with associated symbionts revealed strong interactions between climate, microbial symbionts and trees. For example, ectomycorrhizal trees dominate at high latitudes and elevations, but arbuscular mycorrhizal trees in aseasonal, warm tropical forests (Steidinger et al., 2019).

3.3 | Functional trait and genetic data

Functional traits and genetics are important aspects of tree biodiversity and can provide insights into effects of biodiversity change on species, vegetation and ecosystem functioning (Dayrell et al., 2017; Echeverría-Londoño et al., 2018). Functional ecology has emerged as a dominant paradigm for understanding biophysical constraints on plant form and function, species- and community-level responses to environmental change, and ecosystem functioning in terrestrial ecosystems (Reich, 2014; Díaz et al., 2016; Dayrell et al., 2017; Gross et al., 2017). Molecular data have revolutionized our understanding of evolutionary relationships among species, populations and functional traits (Byrne et al., 2017; Dayrell et al., 2017; Sandel et al., 2019), identified cryptic species, and provided a deeper understanding of biodiversity (Turner et al., 2013; Eiserhardt et al., 2018; Forest et al., 2018; Sandel et al., 2020).

The compilation of open global plant trait databases (Appendix S1) and establishment of common measurement protocols (Pérez-Harguindeguy et al., 2013; Gallagher et al., 2020) have facilitated global syntheses of plant–environment relationships across time and space that provide insight to variation in key ecosystem functions and processes across ecological scales, e.g., leaves, individuals, communities, and ecosystems (Reich, 2014; Díaz et al., 2016; Echeverría-Londoño et al., 2018; van der Sande et al., 2020). Tree species are relatively well represented in trait databases, with 95.4% of all tree species having data for at least one trait in the TRY database (Kattge et al., 2020; Figure 2), compared to 35.5% of all plant species (Cornwell et al., 2019). However, for only about half of these species (45.4% of all tree species) are data available for at least three traits (i.e., at least one in addition to growth form and woodiness), with the tropics being particularly poorly sampled (Figure 3b). In addition, trait measurements in databases are typically reported as species' means, making it difficult to assess intra-specific variation when determining trait-based community processes (Violle et al., 2012).

Furthermore, trait databases are biased toward relatively easy to measure “effect” traits (i.e., specific leaf area, leaf dry matter content and leaf area) associated with a limited number of above-ground

ecosystem functions, principally carbon storage, growth/productivity and nutrient cycling. These traits are not direct measures of plant function, unlike photosynthesis or water-use efficiency (Figure 5), limiting current knowledge of tree function. They also do not capture information about below-ground and reproductive traits, which are associated with other key ecosystem functions (Girão et al., 2007; Ottaviani et al., 2017). However, there are efforts, such as GRooT (Global Root Traits database; <https://groot-database.github.io/GRooT/>), to address this gap (Klimešová et al., 2018; Guerrero-Ramírez et al., 2021).

Vegetation community data are increasingly being linked to plant functional traits, and this is providing novel insights into global patterns for traits and trait relationships (Bruehlheide et al., 2019; van der Sande et al., 2020). For example, climate (temperature variability and water availability) appears to be the key driver of functional diversity at the global scale (Wieczynski et al., 2019), while for communities in similar climatic and soil conditions, non-climatic factors, such as disturbance, fine-scale environmental heterogeneity and biological factors, appear to be more important (Bruehlheide et al., 2018). Another global-scale study found that invasive tree species are most abundant, if they are functionally similar to, but taller with higher seed production and wood density, than co-occurring native species (van der Sande et al., 2020).

Placement of tree species in a phylogenetic context should be based on DNA sequence data for each species and a well-supported and dated phylogeny that is readily updated when new data or methods become available (Eiserhardt et al., 2018). Many phylogenies, and the sequences they are derived from, are available from the TreeBASE database (<https://www.treebase.org/>). Although much of the molecular data gathered across the plant tree of life have been deposited into public databases, only 43.9% of all tree species are represented with any data in the GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>) and coverage is low in the tropics (Figure 3d). However, this representation is still >10% greater than that for plants in general (Cornwell et al., 2019).

Of those species that have sequences, not all have available data that are useful for phylogenetic analyses, as genes sequenced for some species may not be widely sampled. In fact, only 24% of taxa with any public sequence data can be confidently placed into a large phylogeny (Smith & Brown, 2018). In addition, population-level molecular data and information about intra-specific genetic diversity can be extremely valuable for conservation and understanding evolution (González-Martínez et al., 2006; Byrne et al., 2017). However, while genetic data are available for two or more populations of numerous species (Nason et al., 1997; González-Martínez et al., 2006), a comprehensive database for such information is lacking.

Phylogenetic data can provide important information about the genetic diversity and evolutionary history of vegetation. For example, rates of recent speciation have been found to be highest in less biodiverse communities (Schluter & Pennell, 2017; Igea & Tanetzap, 2020). Furthermore, phylogenetic endemism, an indicator of phylogenetic uniqueness of a community, tends to be higher in regions of relatively high climatic stability, greater geographic isolation and topographically

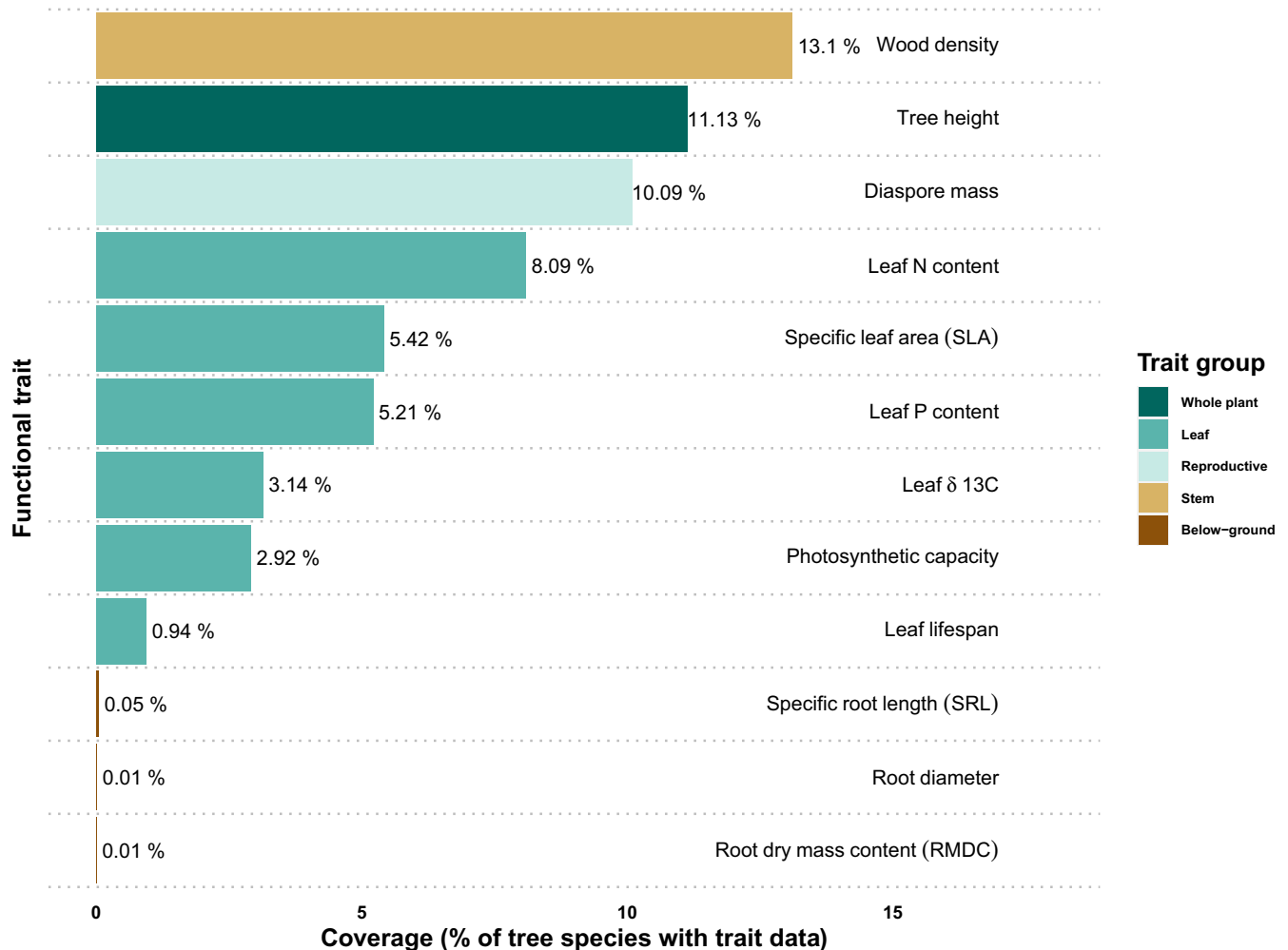


FIGURE 5 Available open-access trait data for tree species. Data were downloaded from the TRY database (<https://www.try-db.org/>), version 5, for traits associated with key axes of ecological variation and compared with our “tree list” (see Appendix S2)

more complex regions. Furthermore, the factors being most strongly related to phylogenetic endemism differ among phytogeographic regions in a manner that can be explained by their climatic histories (Sandel et al., 2020).

3.4 | Conservation status and socio-economic role

Trees are more threatened than ever by habitat destruction and degradation, overexploitation for timber and other products, displacement by invasive species, loss of pollinators and seed dispersers, and anthropogenic climate change (Allen et al., 2010; Crowther et al., 2015; Forest et al., 2018). It has been estimated that 15 million trees are cut down every year (Crowther et al., 2015) and that 2.3 million km², an area more than five times the size of France, of forest cover were lost from 2000 to 2012 (Hansen et al., 2013). A holistic understanding of landscapes, vegetation, and ecological communities without considering anthropogenic influences is, therefore, generally impossible (Clark, 1996; van der Sande et al., 2019).

The IUCN Red List (<https://www.iucnredlist.org/>) is the most comprehensive database for global conservation information and contains assessments for 45.3% of all tree species. Of the assessed species (Appendix S4), 9,854 (37.4%) are considered either threatened with extinction (9,792 species; “Critically Endangered”, “Endangered” and “Vulnerable” categories) or extinct (62 species; “Extinct” and “Extinct in the Wild” categories). The proportion of species assessed is lowest in the tropics and the Southern Hemisphere (Figure 3e), possibly due to more limited resources for conservation in some regions (e.g., Keppel et al., 2012). Additionally, national lists of threatened species are more relevant for achieving protection than the IUCN Red List for some countries, such as Australia (Schatz, 2009). An additional key resource for tree conservation is PlantSearch (https://tools.bgci.org/plant_search.php), a database of ex situ living plant and seed collections.

Trees have the potential to serve as indicators for the progress toward conservation goals and for the condition of vegetation in general. The number of species listed as threatened with extinction is already used as an indicator of the conservation value of vegetation



(e.g., Gillespie et al., 2014). As the Global Tree Assessment, an initiative to assess the IUCN Red List status for all known tree species (Newton et al., 2015), is moving closer to completion, tree data in the IUCN Red List database will become an increasingly representative tool for assessing the conservation status of, and threats to, vegetation.

An understanding of the socio-economic history of landscapes and trees is essential for understanding all aspects of the biodiversity of trees and vegetation (e.g., Clark, 1996; Anisi et al., 2021). Extensive socio-economic data for trees are available but are widely scattered across floristic works and monographs, most with a relatively narrow geographic focus (e.g., Thaman, 1992) or targeted at a specific type of usage (e.g., Van Wyk & Wink, 2017). The Useful Tropical Plants Database (<http://tropical.theferns.info/>; Appendix S1) is one of the few databases with a socio-economic focus that covers a broad geographic region, and there is urgent need to digitize and integrate available data.

4 | TOWARDS A GLOBAL SYNTHESIS OF TREE BIODIVERSITY

A wide range of data on tree biodiversity are, therefore, available and highly relevant for understanding patterns of vegetation. Although only 28.0% of all tree species (Figure 2) have at least some data in all three key databases (GBIF, GenBank and TRY) for distribution, phylogeny, and function, they are proportionally better represented (by about 10%) than plants in general (Cornwell et al., 2019). The ultimate goal should be to: (a) aggregate all available information about tree species; and (b) have comprehensive data for the distribution, patterns of abundance, phylogenetic relationships, population-level genetic data, key functional traits (response-and-effect traits), key ecological interactions, demographics and population dynamics, main human uses (timber, non-timber products), and key ecosystem functions for all tree species. Below we outline how such a global synthesis could be achieved, discuss important knowledge gaps that would need to be filled for this ambitious endeavor and illustrate potential applications of integrated, global tree biodiversity data using two examples.

4.1 | Aggregating and imputing existing data

All available data for each dimension of biodiversity (Figure 1) should be readily and freely available. Data integration is probably most urgent for the forest plot data sets that are currently spread across a wide variety of databases. Such an initiative should be inclusive and diverse, reflecting the opinions and needs of data owners and data users alike, and should fairly acknowledge the contributions of data owners (Tenopir et al., 2011; Gallagher et al., 2020).

It is important that the mobilization of biodiversity-related information continues (Lughadha & Miller, 2009; König et al., 2019). For example, the ongoing digitization of herbarium specimens is

providing critical information to investigate all aspects of tree biodiversity (Soltis, 2017). Libraries are another rich source of information and the increasing power of text mining might help to mobilize data on traits or economic uses of trees (Deans et al., 2012; König et al., 2019). Furthermore, there can be extensive local and traditional knowledge about trees with high relevance for tree biodiversity and conservation (e.g., Thaman, 1992).

Aggregating and integrating existing data would also increase our capacity to impute gaps in our knowledge. For example, functional traits can be strongly correlated (Reich, 2014; Díaz et al., 2016). Such correlations allow the imputation of missing trait values using gap-filling techniques (Schrodte et al., 2015; Echeverría-Londoño et al., 2018; Rüger et al., 2020), especially for traits that are phylogenetically conserved. For example, functional traits can predict demographic parameters such as generation time in vascular plants with reasonable accuracy (Salguero-Gómez, 2017). Therefore, there is the potential to use information on tree functional traits, such as wood density and specific leaf area, to estimate the generation times for tree species for which structured population models are unavailable.

The Open Tree of Life project illustrates the utility of data imputation. It uses publicly available phylogenetic data and taxonomic information to construct more comprehensive phylogenies, extrapolating for taxa that lack available sequence data by assuming monophyly of the recognized genera and/or families (Hinchliff et al., 2015). While there is no obvious taxonomic bias in those species that lack data, tropical species are less likely to be represented by sequences (Smith & Brown, 2018). Although many species have not been sampled well enough to confidently place them in a phylogeny without the use of taxonomic information, more than 90% of tree genera are represented in GenBank.

4.2 | Addressing remaining data needs

Glaring data gaps remain and there are limitations to what can be achieved through data integration and imputation (Moles, 2018). Many gaps clearly require additional data to be collected. For example, the placement of species in the Open Tree of Life phylogeny based on taxonomic information could be improved by specialists evaluating the morphology and genetics of understudied taxa and reporting results in a standardized format (Deans et al., 2012), particularly if focused on genera that currently lack any molecular data. Closing some gaps, such as the limited data in all aspects of tree biodiversity from the tropics, and our extremely limited knowledge of intraspecific differences, will require considerable investment in human and financial resources. Furthermore, investment in adequate maintenance of biodiversity collections would help prevent irreplaceable loss of preserved specimens and associated data (Escobar, 2018). However, declines in the number of botanists and reluctance to publish (by high-impact journals) and fund botanical work are starting to limit our ability to examine new or understudied taxa (Crisci et al., 2020), despite many tree species still awaiting description (Cheek et al., 2020; Slik et al., 2015).

Technological advances allow filling some of these gaps faster and cheaper than we could in the past. For example, DNA barcoding and other molecular techniques can be used for the identification of species and populations, with important applications for conservation, ecology, and evolution (Kress et al., 2015). Furthermore, high-throughput technologies such as near-infrared spectroscopy (NIRs) promise rapid, large-scale measurements of structural and chemical traits of leaves and wood (Ramirez et al., 2015). In addition, recent developments in remote-sensing techniques, such as LiDAR, allow mapping of trees and vegetation types (Schut et al., 2014; Vaughn et al., 2012) at large spatial extents, and faster quantification of forest functions, such as carbon storage (Saatchi et al., 2011) and leaf traits (Asner et al., 2017), especially when combined with statistical modeling (Butler et al., 2017). High-resolution, remotely-sensed imagery from satellites, planes, and unmanned aerial vehicles are making detailed mapping of individual trees increasingly feasible (Baena et al., 2017). Although these data have limits, e.g., the extent of imagery available and the number of species that can be unambiguously identified, they can play a key role mapping and detecting change in tree distributions (Vaughn et al., 2012; Baena et al., 2017) and generating geo-referenced occurrence records throughout the entire range of tree species.

4.3 | Potential applications

Global, integrated tree biodiversity data have wide-ranging applications and hold the potential for scientific breakthroughs in our understanding of tree and forest biodiversity. For example, it would allow a more thorough understanding of the global patterns of plant height. Moles et al. (2009), based on information from ca 10% of all tree species from published literature, determined latitudinal patterns and drivers of plant height. A considerably more refined and complete appraisal of global plant height patterns would now be possible by integrating published studies with available plot-based height measurements. Furthermore, using space-borne LiDAR imagery as part of the Global Ecosystem Dynamics Investigation (GEDI) or from interferometric synthetic aperture radar data from TanDEM-X would facilitate remotely-sensed estimates of forest structure and biomass of unprecedented accuracy (Qi et al., 2019), and could be integrated with actual plot-based measurements to provide ground-truthing and improved accuracy. Comparing such a global map of tree height with potential maximum tree height, which could be obtained by combining data on species distributions (e.g., from GBIF or GIFT) and maximum tree height (e.g., from TRY), could provide new insights into the factors limiting tree height when related to topographic, climatic, and geological conditions.

Refugia, i.e., places that provide buffering from landscape-scale trends in climate change, have facilitated the persistence of biodiversity in the past and are considered increasingly important for conservation (Keppel et al., 2015). However, we still know little about the ecological and evolutionary functioning of refugia and an integrated

functional traits and molecular approach has been proposed to address this (Keppel et al., 2018). By integrating data from GBIF, TRY and GenBank, this data gap could be addressed on a global scale by comparing the functional and phylogenetic characteristics inside and outside of refugia. Furthermore, data on the demography and conservation status of tree species could provide insights about the performance of tree populations within refugia.

5 | CONCLUSIONS

Considerable amounts of data for a global assessment of tree biodiversity are available and have important applications for understanding large-scale patterns and processes of vegetation. Although important data gaps remain, a comprehensive, integrated synthesis of tree biodiversity data is feasible. We show that such an approach, especially when combined with gap-filling approaches, can provide an increasingly complete understanding of biodiversity (species, functional and genetic) patterns, community dynamics, and ecological interactions of vegetation and biodiversity.

Future macroecological studies utilizing biodiversity data would greatly benefit from all accessible data being available in a framework based on database structures and standard terminology that facilitate interoperability. Such a framework could be enhanced by continuous integration of key resources from evolving taxonomies, fair and equitable data-sharing principles, and stored and new data from herbaria, museums, and libraries made available through ongoing data mobilization efforts. These efforts will require the development of interoperable standard terminology for all aspects of biodiversity (Figure 1) and could expand on existing standards. Due to the scale dependence of biodiversity and its environmental drivers (Chase et al., 2018; Ibanez et al., 2018), it is particularly important that data can be aggregated at multiple spatial grains and that the development of appropriate statistical techniques continues (Tovo et al., 2017; McGlinn et al., 2019).

An integrated, accessible framework would have broad applications, but outputs need to become more accessible to decision makers and end users. The framework, combined with imputation and modeling, would provide an increasingly comprehensive picture of biodiversity and macroecological patterns. Furthermore, data could be readily investigated to facilitate targeted and efficient data collection to fill gaps and move us closer to a complete picture. However, for these outputs to be efficiently applied in the management and conservation of biodiversity, a platform that makes this up-to-date information readily available to end users is needed. Therefore, the approach here used to comprehensively integrate existing tree data has broad applications for improved understanding and conservation of global biodiversity.

ACKNOWLEDGEMENTS

We thank the BGCI and its members for providing data on the global distributions of trees. We also appreciate the support of sPlot — the Global Vegetation-Plot database, a platform of iDiv — the German



Center for Integrative Biodiversity and support by the German Research Foundation (DFG FZT 118, 202548816). Dr. Francesco Maria Sabatini assisted with integrating data on the coverage of tree species in the sPlot database. We are also grateful to the anonymous reviewers who provided feedback on this manuscript.

AUTHOR CONTRIBUTIONS

The manuscript was conceived and outlined (including figures) by all authors during a workshop organized by TK on the biodiversity of trees in Leipzig, 28–31 August 2017. GK drafted and led the writing of the overall manuscript, with GK, DC, SS, MvdS, SCL and TK drafting sections of the manuscript. HK, GK and DC developed Figure 1, PW produced Figure 2 and Figure 3 based on data assembled by PW, GK, SS and DC. Figure 4 was produced by GK and Figure 5 by DC.

DATA AVAILABILITY STATEMENT

All data are provided in the supplementary materials.

ORCID

Gunnar Keppel  <https://orcid.org/0000-0001-7092-6149>

Dylan Craven  <https://orcid.org/0000-0003-3940-833X>

Patrick Weigelt  <https://orcid.org/0000-0002-2485-3708>

Stephen A. Smith  <https://orcid.org/0000-0003-2035-9531>

Masha T. van der Sande  <https://orcid.org/0000-0002-6845-2308>

Brody Sandel  <https://orcid.org/0000-0003-2162-6902>

Sam C. Levin  <https://orcid.org/0000-0002-3289-9925>

Holger Kreft  <https://orcid.org/0000-0003-4471-8236>

Tiffany M. Knight  <https://orcid.org/0000-0003-0318-1567>

REFERENCES

- Allen, C.D., Macalady, A.K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M. et al. (2010) A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, 259, 660–684.
- Allen, J.M., Folk, R.A., Soltis, P.S., Soltis, D.E. & Guralnick, R.P. (2019) Biodiversity synthesis across the green branches of the tree of life. *Nature Plants*, 5, 11–13.
- Anderson-Teixeira, K.J., Davies, S.J., Bennett, A.C., Gonzalez-Akre, E.B., Muller-Landau, H.C., Joseph Wright, S. et al. (2015) CTF- ForestGEO: A worldwide network monitoring forests in an era of global change. *Global Change Biology*, 21, 528–549.
- Anisi, R., de Souza, A., Brodie, G., Thaman, R., Peters, S., Jessop, L.W. et al. (2021) The impact of ebony wood harvesting on *Diospyros samoensis* (Ebenaceae) on Vangunu Island, Western Solomon Islands. *Pacific Conservation Biology*, <https://doi.org/10.1071/PC19052>. [Epub ahead of print].
- Asner, G.P., Martin, R.E., Knapp, D.E., Tupayachi, R., Anderson, C.B., Sinca, F. et al. (2017) Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. *Science*, 355, 385–389.
- Bacles, C.F.E., Lowe, A.J. & Ennos, R.A. (2006) Effective seed dispersal across a fragmented landscape. *Science*, 311, 628.
- Baena, S., Moat, J., Whaley, O. & Boyd, D.S. (2017) Identifying species from the air: UAVs and the very high resolution challenge for plant conservation. *PLoS One*, 12, e0188714.
- Bartomeus, I., Stavert, J.R., Ward, D. & Aguado, O. (2019) Historical collections as a tool for assessing the global pollination crisis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374, 20170389.
- Bastin, J.-F., Berrahmouni, N., Grainger, A., Maniatis, D., Mollicone, D., Moore, R. et al. (2017) The extent of forest in dryland biomes. *Science*, 356, 635–638.
- Beaman, R.S. & Cellinese, N. (2012) Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. *ZooKeys*, 209, 7–17.
- Beech, E., Rivers, M., Oldfield, S. & Smith, P.P. (2017) GlobalTreeSearch: The first complete global database of tree species and country distributions. *Journal of Sustainable Forestry*, 36, 454–489.
- Bennett, J.M., Steets, J.A., Burns, J.H., Durka, W., Vamosi, J.C., Arceo-Gómez, G. et al. (2018) GloPL, a global data base on pollen limitation of plant reproduction. *Scientific Data*, 5, 180249.
- Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J.A., Mozzherin, D., Rees, T. et al. (2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics*, 14, 16.
- Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B., Hennekens, S.M. et al. (2018) Global trait–environment relationships of plant communities. *Nature Ecology and Evolution*, 2, 1906–1917.
- Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S.M., Chytrý, M. et al. (2019) sPlot – A new tool for global vegetation analyses. *Journal of Vegetation Science*, 30, 161–186.
- Butler, E.E., Datta, A., Flores-Moreno, H., Chen, M., Wythers, K.R., Fazayeli, F. et al. (2017) Mapping local and global variability in plant trait distributions. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E10937–E10946.
- Byrne, M., Millar, M.A., Coates, D.J., Macdonald, B.M., McArthur, S.M., Zhou, M. et al. (2017) Refining expectations for environmental characteristics of refugia: Two ranges of differing elevation and topographical complexity are mesic refugia in an arid landscape. *Journal of Biogeography*, 44, 2539–2550.
- Caswell, H. (2001) *Matrix population models: Construction, analysis, and interpretation*, 2nd edition. Sunderland: Sinauer Associates.
- Chambers, J.Q., Higuchi, N. & Schimel, J.P. (1998) Ancient trees in Amazonia. *Science*, 391, 135–136.
- Chase, J.M., McGill, B.J., McGlinn, D.J., May, F., Blowes, S.A., Xiao, X. et al. (2018) Embracing scale-dependence to achieve a deeper understanding of biodiversity and its change across communities. *Ecology Letters*, 21, 1737–1751.
- Cheek, M., Nic Lughadha, E., Kirk, P., Lindon, H., Carretero, J., Looney, B. et al. (2020) New scientific discoveries: Plants and fungi. *Plant, People, Planet*, 2, 371–388.
- Clark, D.B. (1996) Abolishing virginity. *Journal of Tropical Ecology*, 12, 735–739.
- Condit, R., Ashton, P.S., Manokaran, N., LaFrankie, J.V., Hubbell, S.P. & Foster, R.B. (1999) Dynamics of the forest communities at Pasoh and Barro Colorado: Comparing two 50-ha plots. *Philosophical Transactions of the Royal Society, London, Series B: Biological Sciences*, 354, 1739–1748.
- Copenhaver-Parry, P.E. & Bell, D.M. (2018) Species interactions weakly modify climate-induced tree co-occurrence patterns. *Journal of Vegetation Science*, 29, 52–61.
- Cornwell, W.K., Pearse, W.D., Dalrymple, R.L. & Zanne, A.E. (2019) What we (don't) know about global plant diversity. *Ecography*, 42, 1819–1831.
- Crisci, J.V., Katinas, L., Apodaca, M.J. & Hoch, P.C. (2020) The end of botany. *Trends in Plant Science*, 25, 1173–1176.
- Crowther, T.W., Glick, H.B., Covey, K.R., Bettigole, C., Maynard, D.S., Thomas, S.M. et al. (2015) Mapping tree density at a global scale. *Nature*, 525, 201–205.
- Dallas, T., Decker, R.R., Hastings, A. & Anderson, M. (2017) Species are not most abundant in the centre of their geographic range or climatic niche. *Ecology Letters*, 20, 1526–1533.

- Daru, B.H., Elliott, T.L., Park, D.S. & Davies, T.J. (2017) Understanding the processes underpinning patterns of phylogenetic regionalization. *Trends in Ecology and Evolution*, 32, 845–860.
- Dayrell, R.L.C., Garcia, Q.S., Negreiros, D., Baskin, C.C., Baskin, J.M. & Silveira, F.A.O. (2017) Phylogeny strongly drives seed dormancy and quality in a climatically buffered hotspot for plant endemism. *Annals of Botany*, 119, 267–277.
- Deans, A.R., Yoder, M.J. & Balhoffer, J.P. (2012) Time to change how we describe biodiversity. *Trends in Ecology and Evolution*, 27, 78–84.
- Dengler, J., Jansen, F., Glöckler, F., Peet, R.K., De Cáceres, M., Chytrý, M. et al. (2011) The Global Index of Vegetation-Plot Databases (GIVD): A new resource for vegetation science. *Journal of Vegetation Science*, 22, 582–597.
- Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S. et al. (2016) The global spectrum of plant form and function. *Nature*, 529, 167–171.
- Echeverría-Londoño, S., Enquist, B.J., Neves, D.M., Violle, C., Boyle, B., Kraft, N.J.B. et al. (2018) Plant functional diversity and the biogeography of biomes in North and South America. *Frontiers in Ecology and Evolution*, 6, 219.
- Eiserhardt, W.L., Antonelli, A., Bennett, D.J., Botigué, L.R., Burleigh, J.G., Dodsworth, S. et al. (2018) A roadmap for global synthesis of the plant tree of life. *American Journal of Botany*, 105, 614–622.
- Enquist, B.J., Condit, R., Peet, R.K., Schildhauer, M. & Thiers, B.M. (2016) Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. *PeerJ Preprints*, 4, e2615v2612.
- Escobar, H. (2018) In a 'foretold tragedy', fire consumes Brazil museum. *Science*, 361, 960.
- Fenning, T.M. & Gershenzon, J. (2002) Where will the wood come from? Plantation forests and the role of biotechnology. *Trends in Biotechnology*, 20, 291–296.
- Forest, F., Moat, J., Baloch, E., Brummitt, N.A., Bachman, S.P., Ickert-Bond, S. et al. (2018) Gymnosperms on the EDGE. *Scientific Reports*, 8, 6053.
- Fourcade, Y., Besnard, A.G. & Secondi, J. (2018) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27, 245–256.
- Franklin, J., Davis, F.W., Ikegami, M., Syphard, A.D., Flint, L.E., Flint, A.L. et al. (2013) Modeling plant species distributions under future climates: How fine scale do climate projections need to be? *Global Change Biology*, 19, 473–483.
- Franklin, J., Serra-Díaz, J.M., Syphard, A.D. & Regan, H.M. (2017) Big data for forecasting the impacts of global change on plant communities. *Global Ecology and Biogeography*, 26, 6–17.
- Freschet, G.T., Roumet, C., Comas, L.H., Weemstra, M., Bengough, A.G., Rewald, B. et al. (2021) Root traits as drivers of plant and ecosystem functioning: current understanding, pitfalls and future research needs. *New Phytologist*, early view. <https://doi.org/10.1111/nph.17072>
- Gallagher, R.V., Falster, D.S., Maitner, B.S., Salguero-Gómez, R., Vandvik, V., Pearse, W.D. et al. (2020) Open Science principles for accelerating trait-based science across the Tree of Life. *Nature Ecology and Evolution*, 4, 294–303.
- Gillespie, T.W., O'Neill, K., Keppel, G., Pau, S., Meyer, J.-Y., Price, J.P. et al. (2014) Prioritizing conservation of tropical dry forests in the Pacific. *Oryx*, 48, 337–344.
- Girão, L.C., Lopes, A.V., Tabarelli, M. & Bruna, E.M. (2007) Changes in tree reproductive traits reduce functional diversity in a fragmented Atlantic forest landscape. *PLoS One*, 2, e908.
- González-Martínez, S.C., Krutovsky, K.V. & Neale, D.B. (2006) Forest-tree population genomics and adaptive evolution. *New Phytologist*, 170, 227–238.
- Gross, N., Bagousse-Pinguet, Y.L., Liancourt, P., Berdugo, M., Gotelli, N.J. & Maestre, F.T. (2017) Functional trait diversity maximizes ecosystem multifunctionality. *Nature Ecology and Evolution*, 1, 0132.
- Guerrero-Ramírez, N.R., Mommer, L., Freschet, G.T., Iversen, C.M., McCormack, M.L., Kattge, J. et al. (2021) Global root traits (GRooT) database. *Global Ecology and Biogeography*, 30, 25–37.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A. et al. (2013) High-resolution global maps of 21st-century forest cover change. *Science*, 342, 850–853.
- Havens, K. & Henderson, S. (2013) Citizen science takes root. *American Scientist*, 101, 378–385.
- Hempel, S., Götzenberger, L., Kühn, I., Michalski, S.G., Rillig, M.C., Zobel, M. et al. (2013) Mycorrhizas in the Central European flora: Relationships with plant life history traits and ecology. *Ecology*, 94, 1389–1399.
- Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M. et al. (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 12764–12769.
- Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46, 523–549.
- Ibanez, T., Keppel, G., Baider, C., Birkinshaw, C., Culmsee, H., Cordell, S. et al. (2018) Regional forcing explains local species diversity and turnover on tropical islands. *Global Ecology and Biogeography*, 27, 474–486.
- Ibanez, T., Keppel, G., Baider, C., Birkinshaw, C., Florens, F.B.V., Laidlaw, M. et al. (2020) Tropical cyclones and island area shape species abundance distributions of local tree communities. *Oikos*, 129, 1856–1866.
- Igea, J. & Tanentzap, A.J. (2020) Angiosperm speciation cools down in the tropics. *Ecology Letters*, 23, 692–700.
- Jetz, W., McPherson, J.M. & Guralnick, R.P. (2012) Integrating biodiversity distribution knowledge: Toward a global map of life. *Trends in Ecology and Evolution*, 27, 151–159.
- Jobe, R.T. & White, P.S. (2009) A new cost-distance model for human accessibility and an evaluation of accessibility bias in permanent vegetation plots in Great Smoky Mountains National Park, USA. *Journal of Vegetation Science*, 20, 1099–1109.
- Kattge, J., Bönsch, G., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P. et al. (2020) TRY plant trait database – enhanced coverage and open access. *Global Change Biology*, 26, 119–188.
- Keith, D.A. & Pellow, B.J. (2015) *Review of Australia's major vegetation classification and descriptions*. Sydney: Centre for Ecosystem Science, UNSW.
- Keppel, G., Anderson, S., Williams, C., Kleindorfer, S. & O'Connell, C. (2017) Microhabitats and canopy cover moderate high summer temperatures in a fragmented Mediterranean landscape. *PLoS One*, 12, e0183106.
- Keppel, G., Mokany, K., Wardell-Johnson, G.W., Phillips, B.L., Welbergen, J.A. & Reside, A.E. (2015) The capacity of refugia for conservation planning under climate change. *Frontiers in Ecology and the Environment*, 13, 106–112.
- Keppel, G., Morrison, C., Watling, D., Tuiwawa, M. & Rounds, I.A. (2012) Conservation in tropical Pacific Island countries: Why most current approaches are failing. *Conservation Letters*, 5, 256–265.
- Keppel, G., Ottaviani, G., Harrison, S., Wardell-Johnson, G.W., Marcantonio, M. & Mucina, L. (2018) Towards an eco-evolutionary understanding of endemism hotspots and refugia. *Annals of Botany*, 122, 927–934.
- Keppel, G., Tuiwawa, M.V., Naikatini, A. & Rounds, I.A. (2011) Microhabitat specialization of tropical rain forest canopy trees in the Sovi Basin, Viti Levu, Fiji Islands. *Journal of Tropical Ecology*, 27, 491–501.
- Klimešová, J., Martinková, J. & Ottaviani, G. (2018) Belowground plant functional ecology: Towards an integrated perspective. *Functional Ecology*, 32, 2115–2126.
- König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J. & Kreft, H. (2019) Biodiversity data integration – the significance of data resolution and domain. *PLoS Biology*, 17, e3000183.



- Kress, W.J., García-Robledo, C., Uriarte, M. & Erickson, D.L. (2015) DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution*, 30, 25–35.
- Lankau, R.A., Zhu, K. & Ordóñez, A. (2015) Mycorrhizal strategies of tree species correlate with trailing range edge responses to current and past climate change. *Ecology*, 96, 1451–1458.
- Lavoie, C. (2013) Biological collections in an ever changing world: Herbaria as tools for biogeographical and environmental studies. *Perspectives in Plant Ecology, Evolution and Systematics*, 15, 68–76.
- Li, S.-P., Cadotte, M.W., Meiners, S.J., Pu, Z., Fukami, T. & Jiang, L. (2016) Convergence and divergence in a long-term old-field succession: The importance of spatial scale and species abundance. *Ecology Letters*, 19, 1101–1109.
- Liang, J., Crowther, T.W., Picard, N., Wiser, S., Zhou, M., Alberti, G. et al. (2016) Positive biodiversity-productivity relationship predominant in global forests. *Science*, 354, aaf8957.
- Lindenmayer, D.B. & Laurance, W.F. (2017) The ecology, distribution, conservation and management of large old trees. *Biological Reviews*, 92, 1434–1458.
- Liu, X., Trogisch, S., He, J.-S., Niklaus, P.A., Bruehlheide, H., Tang, Z. et al. (2018) Tree species richness increases ecosystem carbon storage in subtropical forests. *Proceedings of the Royal Society B: Biological Sciences*, 285, 20181240.
- Lughadha, E.N. & Miller, C. (2009) Accelerating global access to plant diversity information. *Trends in Plant Science*, 14, 622–628.
- Lutz, J.A., Furniss, T.J., Johnson, D.J., Davies, S.J., Allen, D., Alonso, A. et al. (2018) Global importance of large-diameter trees. *Global Ecology and Biogeography*, 27, 849–864.
- McGlinn, D.J., Xiao, X., May, F., Gotelli, N.J., Engel, T., Blowes, S.A. et al. (2019) Measurement of Biodiversity (MoB): A method to separate the scale-dependent effects of species abundance distribution, density, and aggregation on diversity change. *Methods in Ecology and Evolution*, 10, 258–269.
- Meyer, C., Weigelt, P. & Kreft, H. (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, 19, 992–1006.
- Moles, A.T. (2018) Being John Harper: Using evolutionary ideas to improve understanding of global patterns in plant traits. *Journal of Ecology*, 106, 1–18.
- Moles, A.T., Warton, D.I., Warman, L., Swenson, N.G., Laffan, S.W., Zanne, A.E. et al. (2009) Global patterns in plant height. *Journal of Ecology*, 97, 923–932.
- Muscarella, R., Emilio, T., Phillips, O.L., Lewis, S.L., Slik, F., Baker, W.J. et al. (2020) The global abundance of tree palms. *Global Ecology and Biogeography*, 29, 1495–1514.
- Nason, J.D., Aldrich, P.R. & Hamrick, J.L. (1997) Dispersal and the dynamics of genetic structure in fragmented tropical tree populations. In: Laurance, W.F. & Bierregaard Jr., R.O. (Eds) *Tropical forest remnants. Ecology, Management, and conservation of fragmented communities*. Chicago: The University of Chicago Press, pp. 304–320.
- Neuschulz, E.L., Mueller, T., Schleuning, M. & Böhning-Gaese, K. (2016) Pollination and seed dispersal are the most threatened processes of plant regeneration. *Scientific Reports*, 6, 29839.
- Newton, A., Oldfield, S., Rivers, M., Mark, J., Schatz, G., Garavito, N.T. et al. (2015) Towards a global tree assessment. *Oryx*, 49, 410–415.
- Ottaviani, G., Martinková, J., Herben, T., Pausas, J.G. & Klimešová, J. (2017) On plant modularity traits: Functions and challenges. *Trends in Plant Science*, 22, 648–651.
- Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P. et al. (2013) New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, 61, 167–234.
- Phillips, O.L., Malhi, Y., Vinceti, B., Baker, T., Lewis, S.L., Higuchi, N. et al. (2002) Changes in growth of tropical forests: evaluating potential biases. *Ecological Applications*, 12, 576–587.
- Pitman, N.C.A., Silman, M.R. & Terborgh, J.W. (2013) Oligarchies in Amazonian tree communities: A ten-year review. *Ecography*, 36, 114–123.
- Poelen, J.H., Simons, J.D. & Mungall, C.J. (2014) Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics*, 24, 148–159.
- Qi, W., Lee, S.-K., Hancock, S., Luthcke, S., Tang, H., Armston, J. et al. (2019) Improved forest height estimation by fusion of simulated GEDI Lidar data and TanDEM-X InSAR data. *Remote Sensing of Environment*, 221, 621–634.
- Ramírez, J.A., Posada, J.M., Handa, I.T., Hoch, G., Vohland, M., Messier, C. et al. (2015) Near-infrared spectroscopy (NIRS) predicts non-structural carbohydrate concentrations in different tissue types of a broad range of tree species. *Methods in Ecology and Evolution*, 6, 1018–1025.
- Ramula, S., Rees, M. & Buckley, Y.M. (2009) Integral projection models perform better for small demographic data sets than matrix population models: a case study of two perennial herbs. *Journal of Applied Ecology*, 46, 1048–1053.
- Reich, P.B. (2014) The world-wide ‘fast-slow’ plant economics spectrum: a traits manifesto. *Journal of Ecology*, 102, 275–301.
- Rozendaal, D.M.A., Bongers, F., Aide, T.M., Alvarez-Dávila, E., Ascarrunz, N., Balvanera, P. et al. (2019) Biodiversity recovery of Neotropical secondary forests. *Science Advances*, 5, eaau3114.
- Rüger, N., Condit, R., Dent, D.H., DeWalt, S.J., Hubbell, S.P., Lichstein, J.W. et al. (2020) Demographic trade-offs predict tropical forest dynamics. *Science*, 368, 165–168.
- Saatchi, S.S., Harris, N.I., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W. et al. (2011) Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 9899–9904.
- Salguero-Gómez, R. (2017) Applications of the fast-slow continuum and reproductive strategy framework of plant life histories. *New Phytologist*, 213, 1618–1624.
- Salguero-Gómez, R., Jones, O.R., Archer, C.R., Buckley, Y.M., Che-Castaldo, J., Caswell, H. et al. (2015) The COMPADRE Plant Matrix Database: An open online repository for plant demography. *Journal of Ecology*, 103, 202–218.
- Sande, M.T., Bruehlheide, H., Dawson, W., Dengler, J., Essl, F., Field, R. et al. (2020) Similar factors underlie tree abundance in forests in native and alien ranges. *Global Ecology and Biogeography*, 29, 281–294.
- Sande, M.T., Gosling, W., Correa-Metrio, A., Prado-Junior, J., Poorter, L., Oliveira, R.S. et al. (2019) A 7000-year history of changing plant trait composition in an Amazonian landscape; the role of humans and climate. *Ecology Letters*, 22, 925–935.
- Sandel, B., Weigelt, P., Kreft, H., Keppel, G., van der Sande, M.T., Levin, S. et al. (2020) Current climate, isolation and history drive global patterns of tree phylogenetic endemism. *Global Ecology and Biogeography*, 29, 4–15.
- Schatz, G.E. (2009) Plants on the IUCN Red List: Setting priorities to inform conservation. *Trends in Plant Science*, 14, 638–642.
- Schluter, D. & Pennell, M.W. (2017) Speciation gradients and the distribution of biodiversity. *Nature*, 546, 48–55.
- Schrodt, F., Kattge, J., Shan, H., Fazayeli, F., Joswig, J., Banerjee, A. et al. (2015) BHPMF – a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecology and Biogeography*, 24, 1510–1521.
- Schut, A.G.T., Wardell-Johnson, G.W., Yates, C.J., Keppel, G., Baran, I., Franklin, S.E. et al. (2014) Rapid characterisation of vegetation structure to predict refugia and climate change impacts across a global biodiversity hotspot. *PLoS One*, 9, e82778.
- Serra-Díaz, J.M., Enquist, B.J., Maitner, B., Merow, C. & Svenning, J.-C. (2017) Big data of tree species distributions: How big and how good? *Forest Ecosystems*, 4, 30.

- da Silva, J.M.C. & Tabarelli, M. (2000) Tree species impoverishment and the future of the Atlantic forest of northeast Brazil. *Nature*, 404, 72–74.
- Slik, J.W.F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P. et al. (2015) An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 7472–7477.
- Smith, S.A. & Brown, J.W. (2018) Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, 105, 302–314.
- Soltis, P.S. (2017) Digitization of herbaria enables novel research. *American Journal of Botany*, 104, 1281–1284.
- Spracklen, D.V., Arnold, S.R. & Taylor, C.M. (2012) Observations of increased tropical rainfall preceded by air passage over forests. *Nature*, 489, 282–285.
- ter Steege, H., Pitman, N.C.A., Sabatier, D., Baraloto, C., Salomao, R.P., Guevara, J.E. et al. (2013) Hyperdominance in the Amazonian Tree Flora. *Science*, 342, 1243092.
- ter Steege, H., Pitman, N.C.A., Killeen, T.J., Laurance, W.F., Peres, C.A., Guevara, J.E. et al. (2015) Estimating the global conservation status of more than 15,000 Amazonian tree species. *Science Advances*, 1, e1500936.
- Steidinger, B.S., Crowther, T.W., Liang, J., Van Nuland, M.E., Werner, G.D.A., Reich, P.B. et al. (2019) Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature*, 569, 404–408.
- Sullivan, M.J.P., Lewis, S.L., Affum-Baffoe, K., Castilho, C., Costa, F., Sanchez, A.C. et al. (2020) Long-term thermal sensitivity of Earth's tropical forests. *Science*, 368, 869–874.
- Taseski, G.M., Beloe, C.J., Gallagher, R.V., Chan, J.Y., Dalrymple, R.L. & Cornwell, W.K. (2019) A global growth-form database for 143,616 vascular plant species. *Ecology*, 100, e02614.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E. et al. (2011) Data sharing by scientists: Practices and perceptions. *PLoS One*, 6, e21101.
- Thaman, R.R. (1992) Batiri kei Baravi: The ethnobotany of Pacific Island coastal plants. *Atoll Research Bulletin*, 361, 1–62.
- Tovo, A., Suweis, S., Formentin, M., Favretti, M., Volkov, I., Banavar, J.R. et al. (2017) Upscaling species richness and abundances in tropical forests. *Science Advances*, 3, e1701438.
- Turner, B., Paun, O., Munzinger, J., Duangjai, S., Chase, M.W. & Samuel, R. (2013) Analyses of amplified fragment length polymorphisms (AFLP) indicate rapid radiation of *Diospyros* species (Ebenaceae) endemic to New Caledonia. *BMC Evolutionary Biology*, 13, 269.
- Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A. et al. (2018) The iNaturalist species classification and detection dataset. In: *Proceedings. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR 2018*. Piscataway: Institute of Electrical and Electronics Engineers, pp. 8769–8778.
- Van Wyk, B.-E. & Wink, M. (2017) *Medicinal plants of the world*. Oxfordshire: CAB.
- Vaughn, N.R., Moskal, L.M. & Turnblom, E.C. (2012) Tree species detection accuracies using discrete point Lidar and airborne waveform Lidar. *Remote Sensing*, 4, 377–403.
- Violle, C., Enquist, B.J., McGill, B.J., Jiang, L., Albert, C.H., Hulshof, C. et al. (2012) The return of the variance: Intraspecific variability in community ecology. *Trends in Ecology and Evolution*, 27, 244–252.
- Visser, M.D., Bruijning, M., Wright, S.J., Muller-Landau, H.C., Jongejans, E., Comita, L.S. et al. (2016) Functional traits as predictors of vital rates across the life cycle of tropical trees. *Functional Ecology*, 30, 168–180.
- Wagner, F., Rutishauser, E., Blanc, L. & Hérault, B. (2010) Effects of plot size and census interval on descriptors of forest structure and dynamics. *Biotropica*, 42, 664–671.
- Weigelt, P., König, C. & Kreft, H. (2020) GIFT - A Global Inventory of Floras and Traits for macroecology and biogeography. *Journal of Biogeography*, 47, 16–43.
- Wieczynski, D.J., Boyle, B., Buzzard, V., Duran, S.M., Henderson, A.N., Hulshof, C.M. et al. (2019) Climate shapes and shifts functional biodiversity in forests worldwide. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 587–592.
- Wiser, S.K. (2016) Achievements and challenges in the integration, reuse and synthesis of vegetation plot data. *Journal of Vegetation Science*, 27, 868–879.
- Zizka, A., Carvalho, F.A., Calvente, A., Baez-Lizarazo, M.R., Cabral, A., Ramos Coelho, J.F. et al. (2020) No one-size-fits-all solution to clean GBIF. *PeerJ*, 8, e9916.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Appendix S1. Alphabetical list of key databases with global coverage that are relevant for trees

Appendix S2. Methodology for producing the tree list and matching records from key databases

Appendix S3. The 58,044 tree species considered in the analyses of this study and their presence in key databases

Appendix S4. Conservation status categories of tree species assessed for the IUCN Red List

How to cite this article: Keppel G, Craven D, Weigelt P, et al.

Synthesizing tree biodiversity data to understand global patterns and processes of vegetation. *J Veg Sci*.

2021;32:e13021. <https://doi.org/10.1111/jvs.13021>