CENTRE OF BIODIVERSITY AND SUSTAINABLE LAND USE

SECTION: BIODIVERSITY, ECOLOGY AND NATURE CONSERVATION

THE ASSEMBLY OF ISLAND FLORAS FROM A MACROECOLOGICAL PERSPECTIVE

DISSERTATION

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

im Promotionsprogramm „Biodiversity, Ecology and Evolution“
der Georg-August University School of Science (GAUSS)

vorgelegt von

Christian König

aus

Chemnitz

Göttingen, 2018

Betreuungsausschuss:

Prof. Dr. Holger Kreft – Biodiversity, Macroecology & Biogeography, Georg-August-Universität Göttingen

Prof. Dr. Erwin Bergmeier, Abteilung Vegetationsanalyse & Phytodiversität, Georg-August-Universität Göttingen

Dr. Patrick Weigelt, Biodiversity, Macroecology & Biogeography, Georg-August-Universität Göttingen

Mitglieder der Prüfungskommission

Referent: Prof. Dr. Holger Kreft – Biodiversity, Macroecology & Biogeography, Georg-August-Universität Göttingen

Korreferent: Dr. Erwin Bergmeier, Abteilung Vegetationsanalyse & Phytodiversität, Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Dirk Hölscher

Prof. Dr. Mark Maraun

Prof. Dr. Teja Tscharntke

Prof. Dr. Stefan Scheu

Tag der mündlichen Prüfung: 25.10.2018

Biodiversity starts in the distant past and it points toward the future.

Frans Lanting

Thesis abstract

Islands have always played a central role in ecology and biogeography. On the one hand, island biotas are ecologically unique by featuring exceptionally high rates of endemism and remarkable evolutionary adaptations while being generally poor in species. On the other hand, the geographical, climatic, and geological diversity of islands across the globe facilitates the detailed study of abiotic and biotic factors that have shaped these extraordinary assemblages. Many findings from island biogeography have led to general ecological insights in the past, e.g. the dynamic regulation of species diversity via immigration, extinction and speciation. Today, the increasing availability of ecological data allows going beyond species numbers and resolving the identities, functional traits and phylogenetic relationships of individual species at the global scale. This opens new and promising ways of inquiry in the field of island biogeography and bears great potential for understanding the ecological processes shaping island biodiversity at a deeper level.

The objective of the present thesis is twofold. First, I aim to identify and address challenges in the utilization of global plant diversity data that currently impede the effectiveness of macroecological approaches in (island) biogeographical research. Second, I endeavour to utilize these insights to conduct large-scale, data-driven analyses of plant diversity that examine the ecological and biogeographical mechanisms underlying the assembly of island floras. Consequently, the chapters of this thesis are arranged into a conceptual part (Chapters 1 and 2) and an empirical part (Chapters 3 and 4).

In Chapter 1, I develop a novel conceptualization of ecological data types according to their domain and resolution. Focusing on data from two domains, species distributions and functional traits, I show that existing digital infrastructures are generally more advanced for disaggregated data types, e.g. point occurrence records, vegetation plots and individual-level trait measurement, than for aggregated data types, e.g. regional checklists or species-level functional traits). I discuss the need for the integration of aggregated data types into the macroecological data landscape and demonstrate the potential of this approach with three case studies. In Chapter 2, I present the GIFT database, a novel resource for macroecological analyses of global plant diversity. GIFT implements many of the concepts outlined in chapter 1 and achieves nearly global coverage in terms of plant distributions and several key functional traits. The chapter provides extensive information on the design and internal processing workflows of the database, and describes the geographical, taxonomic and functional coverage of GIFT.

In Chapter 3, I use data from GIFT to assess global patterns in the beta diversity of island and mainland floras. To this end, I model species turnover, i.e. the richness-insensitive component of beta diversity, as a function of pairwise geographical distance and climatic differences between floristic regions. I show that, on average, island floras are more similar to each other than mainland floras and that species turnover among island assemblages is mostly determined by climatic conditions rather than by geographic distance. These findings suggest that island floras sample largely from a limited set of widespread, dispersive species, while less dispersive taxonomic groups tend to be rare on islands and hence contribute little to species turnover. This interpretation is substantiated by the turnover patterns observed for varyingly dispersive taxonomic and functional groups, and represents a strong basis for the quantitative evaluation of dispersal and environmental filters during island colonization. In Chapter 4, I further examine ecological filters during island colonization by providing the first global, quantitative analysis of island disharmony – a phenomenon that describes the biased representation of higher taxa on islands compared to their mainland source regions. I develop a novel method for the statistical estimation of island-specific source regions as well as two measures that quantify the overall compositional disharmony of an island flora and the global over- or under-representation of individual families on islands. Analyzing these two measures as a function of island- and family-specific characteristics, respectively, reveals that the overall disharmony of island floras is closely linked to island area, isolation, and climatic conditions, whereas the global over- or under-representation of individual families shows little systematic variation with family-level functional traits or family size. These findings provide a comprehensive basis for understanding the island- and taxon-specific factors that determine assembly processes on islands, but at the same time highlight the need for a stronger utilization of functional and phylogenetic approaches in island biogeography.

In conclusion, the present thesis makes several important contributions to the fields of macroecology and island biogeography. In a broader context, I identify aggregated data types as a rich, but under-utilized source of plant diversity information with great potential for improving global data coverage and representativeness. The effectiveness of a targeted integration of aggregated data is demonstrated by the GIFT database, which describes global plant diversity at an unprecedented level of completeness and constitutes an invaluable resource for future macroecological research. In a more specific context, my research on the beta diversity and disharmony of island floras provides comprehensive new insights into fundamental ecological processes that govern the assembly of island floras. In particular, the proposed method for a statistical estimation of island source regions as well as my findings on the relative roles of dispersal, environmental and biotic filtering address key problems in island biogeography.

Zusammenfassung

In der ökologischen und biogeographischen Forschung nehmen Inseln seit jeher eine besondere Rolle ein. Zum einen besitzen Inseln ökologisch faszinierende Artgemeinschaften, welche sich oft durch einen hohen Grad an Endemismus und außergewöhnliche Adaptionen bei gleichzeitig relativ geringer Gesamtartenzahl auszeichnen. Zum anderen ermöglicht die geographische, klimatische und geologische Vielfalt der weltweit mehr als 100.000 Inseln vergleichende Studien zu den ökologischen und evolutionären Faktoren, die zum Entstehen dieser Artgemeinschaften beigetragen haben. Die Erkenntnisse der Inselbiogeographie sind dabei nicht nur auf Inseln beschränkt, sondern führten in der Vergangenheit immer wieder zu grundlegenden Einsichten in ökologische Zusammenhänge, wie etwa der dynamischen Regulierung von Artenvielfalt durch Immigrations-, Extinktions- und Artbildungsprozesse. Heute erlaubt die steigende Verfügbarkeit ökologischer Daten außerdem über die Betrachtung von Artenvielfalt hinaus zu gehen, und die Identitäten, funktionellen Eigenschaften und phylogenetischen Beziehungen einzelner Arten aufzuschlüsseln. Dadurch gewinnen makroökologische Methoden in der inselbiogeographischen Forschung zunehmend an Relevanz und versprechen ein tieferes Verständnis ökologischer Prozesse auf Inseln. Auf diesem Weg sind allerdings noch mehrere Hürden zu überwinden.

Mit der vorliegenden Arbeit verfolge ich zwei Ziele. Zum einen möchte aktuelle Defizite in der Verfüg- und Verwendbarkeit von Biodiversitätsdaten identifizieren und somit zu einer allgemein effektiveren Nutzung makroökologischer Ansätze in der Inselbiogeographie beitragen. Zum anderen möchte ich die daraus gewonnen Erkenntnisse in der Analyse globaler Pflanzendiversitätsmuster umsetzen, um neue Einsichten in die Entstehung und Zusammensetzung von Inselfloren zu gewinnen. Die vier Forschungskapitel gliedern sich dementsprechend in einen konzeptionellen (Kapitel 1 und 2) und einen empirischen Teil (Kapitel 3 und 4).

In Kapitel 1 erarbeite ich eine Einteilung ökologischer Datentypen auf Grundlage von Datenauflösung und -domäne. Ich kann zeigen, dass die existierende digitale Infrastruktur in den Schlüsseldomänen der Verbreitung und funktionellen Eigenschaften von Arten deutlich ausgereifter für disaggregierte Datentypen (z.B. Punktvorkommen, Vegetationsplots, Einzelmessungen von funktionellen Eigenschaften) als für aggregierte Datentypen ist (z.B. regionale Checklisten oder Mittelwerte funktioneller Eigenschaften). Im Weiteren diskutiere ich die Notwendigkeit einer stärkeren Integration von aggregierten Datentypen in die makroökologische Datenlandschaft, und belege das Potenzial eines solchen Vorgehens anhand von drei makroökologischen Fallstudien. In Kapitel 2 stelle ich GIFT vor, eine neue Datenbank zur makroökologischen Analyse von Pflanzendiversität. GIFT setzt viele der in Kapitel 1 erarbeiteten Konzepte zur Integration globaler Biodiversitätsdaten um, und erreicht nahezu globale Abdeckung hinsichtlich floristischer Verbreitungsdaten sowie

bestimmter funktioneller Eigenschaften von Planzen. Das Kapitel stellt umfassende Informationen zum Aufbau der Datenbank zusammen, erläutert automatisierte Abläufe zur Verarbeitung ökologischer Daten und präsentiert detaillierte Statistiken zur geographischen, taxonomischen und funktionellen Abdeckung von GIFT.

In Kapitel 3 nutze ich GIFT um globale Muster in der kompositionellen Ähnlichkeit von Insel- und Festlandsfloren zu untersuchen. Hierzu modelliere ich den Arten-turnover, d.h. den von der Gesamtartenzahl unbeeinflussten Teil der Beta-Diversität zweier Floren, in Abhängigkeit von geographischer Distanz und klimatischen Variablen. Ich zeige, dass sich Inselfloren im Mittel ähnlicher sind als Festlandsfloren und der Arten-turnover auf Inseln weniger von geographischer Distanz als vielmehr von klimatischen Bedingungen bestimmt wird. Die Ergebnisse legen nahe, dass Inseln mehrheitlich durch eine begrenzte Gruppe von Arten kolonisiert werden, die sich verlässlich über weite Entfernungen ausbreiten können, während Arten mit schlechteren Ausbreitungsfähigkeiten nur selten auf Inseln vertreten sind und daher wenig zum Arten-turnover beitragen. Diese Interpretation wird vom entsprechenden Verhalten sich unterschiedlich gut ausbreitender taxonomischer und funktioneller Gruppen gestützt und liefert eine wichtige Grundlage zur quantitativen Bewertung von Ausbreitungs- und Umwelt-Filtern bei der Kolonisierung von Inseln. Auch in Kapitel 4 untersuche ich ökologische Filtereffekte während der Kolonisierung von Inseln und präsentiere die erste globale, quantitative Analyse von Insel-„Disharmonie“ – einem Konzept, das die proportionale Über- oder Unterrepräsentation bestimmter Taxa auf Inseln im Vergleich zu deren Ursprungsregionen auf dem Festland beschreibt. Dazu entwickle ich einen neuen Ansatz zur statistischen Abschätzung der geographischen Ursprungsregionen von Inselfloren sowie zwei Maße zur Quantifizierung der Disharmonie einer Flora als Ganzes und der globalen relativen Häufigkeit einzelner Pflanzenfamilien auf Inseln. Die Analyse dieser Maße in Abhängigkeit insel- beziehungsweise familienspezifischer Eigenschaften zeigt, dass die Disharmonie von Inselfloren insgesamt stark durch Inselgröße, -isolation und -klima bestimmt wird, während die Repräsentation einzelner Familien kaum anhand funktioneller Eigenschaften oder der Familiengröße vorhersagbar ist. Dieser Ergebnisse liefern wichtige Beiträge zum Verständnis insel- und taxon-spezifischer Faktoren bei der Zusammensetzung von Inselfloren. Gleichzeitig hebt die Studie das hohe Potential einer verstärkten Einbindung funktioneller und phylogenetischer Ansätze in die makroökologische (Insel-)forschung hervor.

Zusammenfassend leistet die vorliegende Dissertation mehrere wichtige Beiträge zur makroökologischen und inselbiogeographischen Forschung. Im erweiterten Kontext identifiziere ich aggregierte Datentypen als reichhaltige, jedoch vernachlässigte Quelle von Informationen zur globalen Pflanzendiversität, welche massiv zu einer verbesserten Datenabdeckung und -repräsentativität beitragen kann. Die von mir vorgestellte GIFT Datenbank demonstriert das Potenzial einer stärkeren Integration aggregierter Datentypen in die makroökologische Forschung und bildet die globale Pflanzenvielfalt in teilweise unerreichtem Umfang ab. GIFT wird daher auch zukünftig als Grundlage wichtiger

makroökologischer Analysen dienen. Im konkreten inselbiogeographischen Kontext verhilft meine Forschung bezüglich Betadiversität und Disharmonie zu neuen Einsichten in grundlegende ökologische Prozesse bei der Entstehung und Entwicklung von Inselfloren. Insbesondere die von mir entwickelte Methode zur Abschätzung der geographischen Ursprünge von Insel-Artgemeinschaften, sowie meine Erkenntnisse zu den relativen Beiträgen von Ausbreitungs-, Umwelt-, und Interaktionsfiltern bei der Kolonisierung von Inseln stellen wichtige Fortschritte in Kernbereichen der Inselbiogeographie dar.

Table of contents

List of figures

List of tables

Author Contributions

Chapter 1: Global integration of plant diversity data – a functional perspective

Christian König[1], Patrick Weigelt[1], Julian Schrader[1], Amanda Taylor[1], Jens Kattge[2] and Holger Kreft[1].

HK and CK conceived the project. PW and CK collected data and developed the database. CK designed the case studies and performed the analyses. CK led the writing with major contributions from all authors.

Manuscript under review at *PLOS Biology*.

Chapter 2: GIFT - A Global Inventory of Floras and Traits for macroecology and biogeography

Patrick Weigelt[1], Christian König[1] and Holger Kreft[1].

PW and HK conceived the idea of the GIFT database. All authors led the collection of checklist and trait data. PW and CK developed the workflows for importing and processing data in GIFT and for calculating derived variables. PW and CK performed the analyses presented in this manuscript and all authors contributed to writing the manuscript.

Manuscript under review at *Ecology and Evolution*.

Chapter 3: Dissecting global turnover in vascular plants.

Christian König[1], Patrick Weigelt[1] and Holger Kreft[1].

All authors conceived the project. CK performed the analyses and led the writing with significant contributions from PW and HK.

Published by John Wiley & Sons Ltd.: König, C., Weigelt, P. & Kreft, H. (2017) Dissecting global turnover in vascular plants. *Global Ecology and Biogeography*, **26**, 228–242. Definitive version available under: https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.12536

Chapter 4: Source pools and disharmony of the world's island floras

König, Christian[1], Patrick Weigelt[1], Amanda Taylor[1], Anke Stein[3], Wayne Dawson[4], Franz Essl[5,6], Jan Pergl[7], Petr Pyšek[7,8], Mark van Kleunen[3,9], Marten Winter[10], Cyrille Chatelain[11], Jan Wieringa[12,13], Pavel Krestov[14] and Holger Kreft[1]

HK, PW and CK conceived the project. CK designed the methodology, performed the analyses and led the writing with significant contributions from all authors.

Unpublished manuscript.

Author affiliations

[1] Biodiversity, Macroecology & Biogeography Group, University of Goettingen, 37077 Goettingen, Germany

[2] Research Group Functional Biogeography, Max Planck Institute for Biogeochemistry, Jena, Germany

[3] Ecology, Department of Biology, University of Konstanz, D-78457 Konstanz, Germany

[4] Department of Biosciences, Durham University, South Road, Durham DH1 3LE, United Kingdom

[5] Division of Conservation Biology, Vegetation and Landscape Ecology, University of Vienna, 1030 Wien, Austria

[6] Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Matieland 7602, South Africa

[7] Czech Academy of Sciences, Institute of Botany, Department of Invasion Ecology, CZ-252 43 Průhonice, Czech Republic

[8] Department of Ecology, Faculty of Science, Charles University, Viničná 7, CZ-128 44 Prague, Czech Republic

[9] Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou 318000, China

[10] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany

[11] Conservatoire et Jardin botaniques de la Ville de Genève, 1202 Genève, Switzerland

[12] Naturalis Biodiversity Centre, Darwinweg 2, 2333 CR Leiden, Netherlands

[13] Wageningen University, Biosystematics Group, Droevendaalsesteeg 1, 6708 PB Wageningen, Netherlands

[14] Botanical Garden-Institute of the Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, 690024, Russia

General Introduction

Historical biogeography and the significance of islands

Terrestrial plant life has endured more than 400 million years of geological, environmental, and geographical change (Morris *et al.*, 2018). The effects of this eventful past are preserved in the complex distribution and striking variation of today's plant diversity. Among the three to four hundred thousand species of extant vascular plants (Christenhusz & Byng, 2016; Willis, 2017), examples range from miniscule aquatic herbs (*Wolffia arrhiza,* Díaz *et al.*, 2016) to giant forest trees (*Sequoia sempervirens,* Díaz *et al.*, 2016), from narrow-ranged endemics (*Erica capensis,* Helme & Trinder-Smith, 2006) to global cosmopolitans (*Phragmites australis,* Eller *et al.*, 2017), and from ancient evolutionary relics (*Amborella trichipoda,* Poncet *et al.*, 2013) to members of recent radiations (*Lupinus semperflorens,* Hughes & Eastwood, 2006). Understanding how such diversity patterns vary in space and time is among the most fundamental questions in ecology (Pennisi, 2005; Sutherland *et al.*, 2013). The respective scientific discipline, focusing on the systematic investigation of spatiotemporal variations in biodiversity, is termed biogeography (Lomolino *et al.*, 2016).

In the 18th century, early naturalists started to realize that the spatial distribution of species is highly structured. Carl Linnaeus (1707-1778) noted that species are adapted to certain environments and do not occur outside their preferred range of conditions. Georges-Louis Leclerc, Comte de Buffon (1707-1788) added to this observation that distant locations generally harbor distinct sets of species, irrespective of their climatic and environmental similarity (Lomolino *et al.*, 2016). Subsequently, eminent researchers such as Johann Reinhold Forster (1729-1798), Sir Joseph Banks (1743-1820), Augustin-Pyrame de Candolle (1778-1841) and, especially, Alexander von Humboldt (1769-1859) further consolidated the emerging field of biogeography, documenting latitudinal and elevational gradients in species diversity, defining biogeographic regions, and expressing first ideas of mutual interactions influencing the distribution of species (Lomolino *et al.*, 2016, see also Plate 1). These contributions greatly helped to understand the relationship between species distributions and contemporary environmental conditions, but could not sufficiently explain biogeographical patterns such as the abrupt faunal change within the Malay Archipelago or the unique biotas of oceanic islands. These and other observations were finally put into perspective by Charles Darwin (1809-1882) and Alfred Russel Wallace (1823-1913). Their independent discovery of evolution by means of natural selection (Darwin & Wallace, 1858; Darwin, 1859) provided the key to understanding species distributions – in fact, species themselves – as the current endpoints in a series of past geological, climatic and ecological dynamics. It is not a coincidence that the ideas of both Darwin and Wallace were substantially inspired by observations they had made on islands.

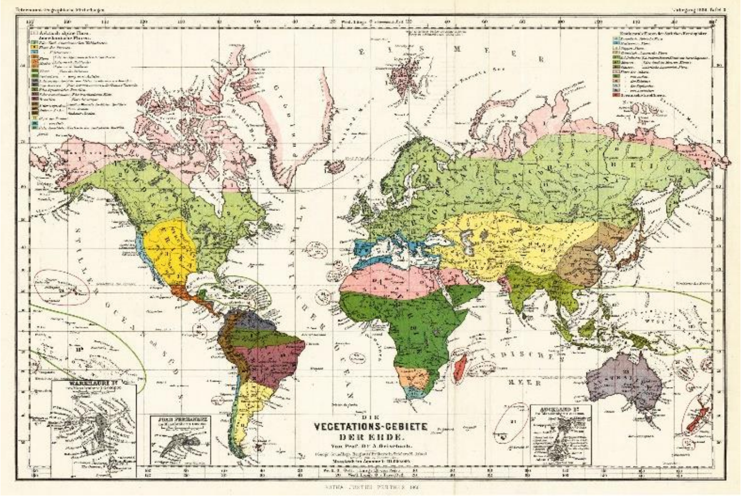


Plate 1: Seminal works by early biogeographers on the distribution of plant diversity. Top: Elevational zonation of the Ecuadorian Andes including Mt. Chimborazo (Humboldt, 1805-1834). Left: Global floristic regionalization (Grisebach, 1866). Grisebach acknowledged the uniqueness of island floras by placing them in a separate category (zone 24: "Oceanische Inselfloren").

Islands are exceptionally informative subjects of biogeographical research. Islands are characterized by isolated, comparatively simple biotas, well-defined geographical boundaries (Gillespie, 2007), and feature a large range of climatic (e.g. temperature, precipitation, seasonality), geographical (e.g. area, elevation, isolation) and historical (e.g. island age, geological origin, Pleistocene impacts) conditions (Weigelt *et al.*, 2013). This makes islands ideal model systems for studying evolutionary, ecological and biogeographical processes at large spatial scales, where experimental manipulations are infeasible (Vitousek, 2002; Whittaker & Fernández-Palacios, 2007; Whittaker *et al.*, 2017). Moreover, islands disproportionately contribute to global biodiversity (Myers *et al.*, 2000; Barthlott *et al.*, 2005) and feature some of the highest endemism rates worldwide (Kier *et al.*, 2009) while being known hotspots of biological invasions and species extinctions (Sax & Gaines, 2008; van Kleunen *et al.*, 2015). For these reasons, islands are highly relevant study systems from both a methodological and ecological point of view.

The unique properties of islands inspired another seminal work that holds relevance to this date: the equilibrium theory of island biogeography (ETIB, MacArthur & Wilson, 1963, 1967). Similar to the theory of evolution by natural selection, the ETIB laid out a radically new perspective that describes complex biotic patterns as the outcome of only a few fundamental processes. According to the ETIB, the species number of an island arises dynamically from opposing rates of immigration and extinction that vary with island isolation and area, respectively. The simple yet elegant mathematical formulation of the model prompted a shift towards a more quantitative approach to ecology and biogeography (Simberloff, 1969; Levin, 1974b; Connor & McCoy, 1979; Hubbell, 2001). In fact, the simplicity of the ETIB was the key to its immense success, as it provided a generalizable framework for predicting species richness across different taxa and geographical settings (Simberloff, 1974; Santos *et al.*, 2016), including insular habitats on the mainland such as mountain tops (Brown, 1971), lakes (Browne, 1981), or forest fragments (Harris, 1984).

Many aspects of island biodiversity, however, remained beyond the scope of the ETIB and its extensions. In particular, compositional and morphological features of island biota proved notoriously hard to predict from analytical models, as they result from a complex interplay of island- and taxon-specific characteristics, evolutionary dynamics, and stochastic events (Whittaker & Fernández-Palacios, 2007). Scientific progress on these more intricate aspects of island biodiversity therefore was based on natural-historical observations and conceptual models. Especially the work of Carlquist (1965, 1966a, 1966b, 1966c, 1966d, 1974) greatly advanced the understanding of assembly processes on islands. Accordingly, species immigration and extinction are characterized by selective ecological filters (Carlquist, 1965): on the one hand, dispersal filtering prevents species with poor dispersal abilities from crossing the open sea; on the other hand, environmental filtering prevents the establishment of species that cannot persist under the predominant biotic and abiotic conditions of the island. Successful colonizers find themselves in a new ecological and evolutionary arena and – given a sufficient amount of time and reproductive isolation from the mainland – potentially diversify and/or adapt to the local conditions. This sequential view of assembly processes has helped to understand many peculiar features of island biota such as the over- or under-representation of certain taxa (Carlquist, 1965; Hoekstra & Fagan, 1998) or common evolutionary trends (e.g. insular woodiness or loss of dispersal capacity; Carlquist, 1966b, 1970; Whittaker & Fernández-Palacios, 2007). Furthermore, it provided a framework for deriving testable hypotheses regarding the taxonomic, functional, and phylogenetic composition of island biota (Midway & Hodge, 2012).

Biogeographical research has impacted our understanding of the natural world in many ways, and islands have played a central role in this process. Due to their geographical isolation and ecological simplicity, islands represent excellent study systems, which helped uncover fundamental mechanisms of evolution (natural selection), community assembly (immigration, extinction and speciation). However, the two classical research paradigms in (island) biogeography, natural history (Humboldt, 1805-1834; Wallace, 1881; Carlquist, 1965) and mathematical modelling (Arrhenius, 1921; MacArthur & Wilson, 1967; Hubbell, 2001), have been unable to fully bridge the gap between detailed descriptions and robust generalizations. Consequently, a novel approach – rigorously quantitative yet capable of resolving the complexities of ecological systems – was required.

The macroecological approach

Macroecology seeks to understand ecological phenomena at large spatiotemporal scales by analyzing emergent statistical patterns in the distribution, abundance and diversity of organisms (Brown & Maurer, 1989; Brown, 1995; Kent, 2005). This data-driven approach offers a powerful toolkit for island biogeographical research (Kueffer *et al.*, 2014). Kreft *et al.* (2008), for example, analyzed the effects of area, isolation, climate, topography and geology on the number of native vascular plant species in 1458 island and mainland floras, providing the first quantitative synthesis on the drivers of insular species richness. Their finding that on islands, but not on the mainland, area is the most important predictor of species richness showed that area-mediated effects on species richness – e.g. speciation rate, extinction rate, or carrying capacity – differ in strength across geographical settings. Macroecological approaches have also been critical for testing theoretical frameworks such as the general dynamic model of island biogeography (Whittaker *et al.*, 2008), which postulates that rates of key ecological processes on islands vary over geological timescales. The major prediction of this model, that species richness follows a hump-shaped relationship with island age, has been empirically confirmed for multiple archipelagos and taxa (Whittaker *et al.*, 2008; Cameron *et al.*, 2013; Lenzner *et al.*, 2017).

The power of the macroecological approach is manifest most clearly when looking not just at species numbers, but also at species composition. Knowing *which* species occur in a given geographical area, and not just how many, opens up entirely new research avenues. Species identities establish a link to the wealth of species-specific information on functional traits, taxonomic and phylogenetic relationships, biotic interactions, and abiotic preferences that constitute the basis for a statistical (i.e. macroecological) characterization of species assemblages. This makes aspects of island biodiversity that used to be too complex for analytical models tangible. In recent years, the potential of species-level macroecological approaches has been demonstrated by numerous studies, for example on the beta diversity (Stuart *et al.*, 2012; Cabral *et al.*, 2014), functional characteristics (Santos *et al.*, 2015; Whittaker *et al.*, 2014), or phylogenetic structure (Cardillo *et al.*, 2008; Weigelt *et al.*, 2015) of island biotas. However, many fundamental questions in island biology and biogeography remain to be addressed (Patiño *et al.*, 2017).

The focus on statistical patterns makes macroecology a particularly data-intensive discipline, whose capacity to produce novel ecological insights is highly dependent on the availability and quality of ecological data (Kueffer *et al.*, 2014). The rise of macroecology within the last two decades (Beck *et al.*, 2012) has been enabled and accompanied by the rapid growth of ecological databases. Today, unprecedented amounts of data on the spatial distribution (Global Biodiversity Information Facility, GBIF, 2018; Map of Life, Jetz *et al.*, 2012), functional traits (TRY, Kattge *et al.*, 2011a), taxonomic affiliations (TPL, The Plant List, 2013; TNRS, Boyle *et al.*, 2013) and (phylo-)genetic relationships (Genbank, Benson *et al.*, 2005; TreeBASE, Piel *et al.*, 2009) of plant species are available. Moreover, modern geospatial data products allow for a global characterization of abiotic, biotic and socioeconomic variables with high accuracy and at high spatiotemporal resolutions (e.g. Karger *et al.*, 2017; Hengl *et al.*, 2017; Copernicus Global Land Service, 2018).

Despite these developments, our knowledge of biodiversity continues to be limited by the lack of ecological data (Taugourdeau *et al.*, 2014; Hortal *et al.*, 2015). Some data limitations are inevitable and arise from fundamental constraints (e.g. in terms of money, time, labor, etc.) on the spatiotemporal resolution at which biodiversity can be measured (Hortal, 2008), but others can be overcome by a coordinated utilization and integration of existing data resources. One potential area of improvement is the common practice of using local (i.e. highly resolved) diversity data such as point occurrences or vegetation plots to address questions at continental or global scales (see e.g. Moles *et al.*, 2007; Moles *et al.*, 2009; Morueta-Holme *et al.*, 2013; Vellend *et al.*, 2013). This mismatch in scales entails two pitfalls that may compromise the reliability of ecological inferences. First, highly resolved diversity data are particularly affected by the above-mentioned constraints on the ability to measure biodiversity, and therefore exhibit severe deficits in terms of large-scale geographical, temporal and taxonomic coverage (Gonzalez *et al.*, 2016; Meyer *et al.*, 2016). Second, highly resolved diversity data reflect local ecological processes and do not scale up to large geographical extents, where other factors such as climate and biogeographical history regulate biodiversity (Huston, 1999; Hortal, 2008, but see e.g. Azaele *et al.*, 2015). A viable way to overcome these drawbacks is to align the scale of the analyzed data with that of the research question, which emphasizes a stronger utilization of relatively coarse-grained, but sufficiently complete and representative diversity data to address macroecological problems.

Study outline

With the present thesis, I aim at elucidating the assembly of island floras from a macroecological perspective, with a particular focus on the taxonomic and functional composition of island plant assemblages. The four research chapters recapitulate major steps towards this objective.

In Chapter 1, I provide a general perspective on the opportunities and challenges of data integration for macroecological research. I examine the availability, applicability and utilization of different types of plant diversity data and show that (1) the macroecological data landscape is dominated by disaggregated data (e.g. point occurrence records, trait measurements) as opposed to aggregated data (e.g. species checklists, taxonomic monographs), and that (2) major data providers mostly focus on a single domain of data (e.g. distributions, functional traits, genetic sequences). I argue that a stronger integration of data across domains and different levels of aggregation has considerable potential for improving data coverage and representativeness at global scales. I describe generalizable strategies for the effective collection, mobilization, imputation and integration ecological data with a particular focus on plant distributions and functional traits. Finally, I present three case studies that highlight the potential of macroecological data integration for answering fundamental ecological and (island) biogeographical questions.

In Chapter 2, I present the Global Inventory of Floras and Traits (GIFT) database. GIFT represents the basis of all empirical studies in this thesis (Chapters 3 and 4, case studies in Chapter 1) and implements many concepts and ideas outlined in Chapter 1, in particular the utilization of aggregated data (e.g. species checklists and Floras) and the integration of data from multiple domains (e.g. species distributions, functional traits, taxonomic and phylogenetic information, geographical characteristics). The chapter provides detailed information on the technical design, processing workflows and data coverage of GIFT.

In Chapter 3, I assess the drivers of species turnover among vascular plant assemblages on islands and the mainland. I use generalized linear models to compare the distance decay of similarity, i.e. species turnover as a function of geographic distance among sites, for different taxonomic and functional plant groups (angiosperms, gymnosperms, pteridophytes, trees, shrubs, herbs) on islands and the mainland. I then apply generalized dissimilarity models to quantify the unique effects of geographic distance and climatic variables in creating species turnover among island and mainland assemblages, respectively. Finally, I present a global prediction of species turnover across a high-resolution equal-area grid.

In Chapter 4, I address the phenomenon of island disharmony, the biased representation of higher taxa on islands compared to their mainland source regions. I present a novel method for identifying island-specific species source regions and develop a measure that quantifies the compositional disharmony of a given island flora. I analyze this measure for 320 islands as a function of important island biogeographical variables (distance from the mainland, area, geological origin, climatic conditions), providing the first global, quantitative assessment of island disharmony to date. Furthermore, I analyze the global over- or under-representation of 450 vascular plant families on islands as a function of family-specific characteristics that presumably affect colonization success (range size, species number and age; functional traits related to dispersal ability, reproduction and life-history).

1 Global integration of plant diversity data – the significance of data resolution and domain

Christian König, Patrick Weigelt, Julian Schrader, Amanda Taylor, Jens Kattge and Holger Kreft

1.1 Abstract

Recent years have seen an explosion in the availability of biodiversity data describing the distribution, function, and evolutionary history of life on earth. Integrating these heterogeneous data remains a challenge due to large variations in observational scales, collection purposes and terminologies. While seminal projects for the integration of disaggregated biodiversity data (e.g. point occurrence records, trait measurements) have been established, aggregated data types (e.g. Floras, taxonomic monographs) have received less such attention, leaving a major source of information on global biodiversity largely untapped. Focusing on plant distributions and functional traits, we here demonstrate the synergies arising from a more tight integration of biodiversity data across domains and resolutions. To this end, we lay out effective strategies for data collection, mobilization, imputation, and sharing, and summarize existing frameworks for scalable and integrative biodiversity research. In three case studies related to the global distribution of plant growth forms, the latitudinal gradient of seed mass, and the global prevalence of insular woodiness, we highlight the potential of aggregated data for biodiversity research and improving the representativeness and completeness of biodiversity data in general. Our results show the need for a more extensive use of available data resources for achieving a both precise and general picture of global biodiversity.

1.2 Introduction

Minimizing the negative ecological impacts of habitat loss (Watson *et al.*, 2016), climate change (Pachauri *et al.*, 2014), and species invasion (Seebens *et al.*, 2017) is one of the major challenges of this century and requires a detailed understanding of global biodiversity (Kerr *et al.*, 2007; Barnard & Thuiller, 2008). In this context, vascular plants constitute a critical group, as they are key providers of biochemical energy and habitat structure. At the same time, the sheer magnitude of plant diversity renders an exhaustive assessment of even its most basic dimensions, e.g. the number of extant species, difficult (Brown & Lomolino, 1998). This effect is further amplified when looking at more complex, often interdependent aspects such as species distributions, functional traits, or phylogenetic relationships, and becomes increasingly pervasive at small informational grain sizes (Hortal *et al.*, 2015). Despite these existing shortfalls in on our knowledge of global plant diversity, recent years have seen an explosion in both the availability (Kattge *et al.*, 2011a; GBIF, 2018; Maitner *et al.*, 2018) and large-scale utilization (Zanne *et al.*, 2014; Díaz *et al.*, 2016; König *et al.*, 2017; Butler *et al.*, 2017; Smith & Brown, 2018) of plant diversity data. This data-driven paradigm has been recognized as key for reducing the shortfalls in biodiversity knowledge and building a sufficiently robust understanding of global biodiversity to address the pressing challenges imposed by global change (Kelling *et al.*, 2009; Hampton *et al.*, 2013).

Biogeography is a key discipline for the integration of heterogeneous biodiversity data, as it brings together the two principal dimensions of ecology – the organism and the environment – at large spatiotemporal scales. Biogeographical data can therefore be integrated with a variety of organismic (e.g. taxonomic, functional, phylogenetic) and environmental (e.g. climate, soil, topography) information. A particularly promising branch of biogeography is functional biogeography. Functional biogeography focuses on documenting and understanding the geographical variation in traits, utilizing ideas, concepts, and methods from a variety of disciplines including ecosystem ecology, evolutionary biology, earth sciences, and ecoinformatics (Violle *et al.*, 2014). In particular, functional biogeography adds a spatial dimension to functional ecology and is thus relevant for a variety of research areas, in which adopting a functional perspective has stimulated substantial scientific progress, e.g. community ecology (McGill *et al.*, 2006; Stegen & Swenson, 2009), biodiversity research (Petchey & Gaston, 2002; Lamanna *et al.*, 2014), ecosystem ecology (Díaz *et al.*, 2007; Bello *et al.*, 2010), or conservation biology (Cadotte *et al.*, 2011; Ostertag *et al.*, 2015). Moreover, the integration of species distributions and functional traits opens up new and interesting research questions: How are different aspects of functional diversity distributed in space? Is there a consistent relationship between functional diversity and ecosystem functioning across habitats, ecosystems, or biomes? Which functional properties are particularly sensitive to climate and land-use changes, and where do they occur most frequently?

Data-driven functional biogeography – and biodiversity research in general – has to bridge the gap between fine-scale precision and global representativeness. This gap is reflected by the variety of existing data types, ranging from highly resolved point occurrence records and trait measurements to relatively coarse, but also more representative data types such as Floras and taxonomic monographs. Consequently, the integration of biodiversity data across multiple resolutions is crucial for overcoming the deficits of individual data types and constitutes a key requirement for developing a deeper understanding of global biodiversity (Jetz *et al.*, 2012). This poses new scientific challenges, e.g. with respect to data sharing and collaborative research (Hampton *et al.*, 2015; Michener, 2015a), the representativeness of large-scale datasets (Engemann *et al.*, 2015; Meyer *et al.*, 2016), or the effective integration of multiple data types (Jetz *et al.*, 2012; La Salle *et al.*, 2016).

Focusing on plant distributions and functional traits, our aim here is to help address these challenges in order to realize the full potential of plant diversity data. First, we characterize common data types with respect to their informational resolution and domain, and highlight general trade-offs across biodiversity data. Based on that, we outline strategies for the effective utilization and integration of plant diversity data across domains and resolutions. We provide suggestions for improving data collection, identify potentials for data mobilization, and describe methods for filling data gaps through imputation. Furthermore, we discuss methodological, sociocultural, and information technological barriers that currently impede the large-scale integration of biodiversity data. We present three case studies based on the Global Inventory of Floras and Traits database (see Box A1.1, Chapter 2), a novel resource for functional biogeography, to demonstrate how already the integration of selected aggregated data types allows tackling fundamental questions in ecology and biogeography related to (1) the global distribution of plant growth forms, (2) the latitudinal gradient in seed mass and (3) the prevalence of insular woodiness on oceanic islands.

1.3 Data as key to global plant ecology

1.3.1 Data domains, types and resolution

Biodiversity science can be organized into different domains that cover distinct spheres of knowledge, e.g. of the taxonomic classification, geographical distribution, functional traits or abiotic tolerances of organisms (Hortal *et al.*, 2015). A domain is typically associated with a set of domain-specific data types (Figure 1.1). Species distributions, for example, can be represented by point occurrences, plot networks, checklists, or expert range maps. Functional trait data may come in the form of field measurements for individual plants, or as aggregated values for populations, species, or higher taxonomic groups (e.g. genera or families). In addition, some biodiversity data types combine information from multiple domains, e.g. regional Floras representing a source of both distributional and functional information.

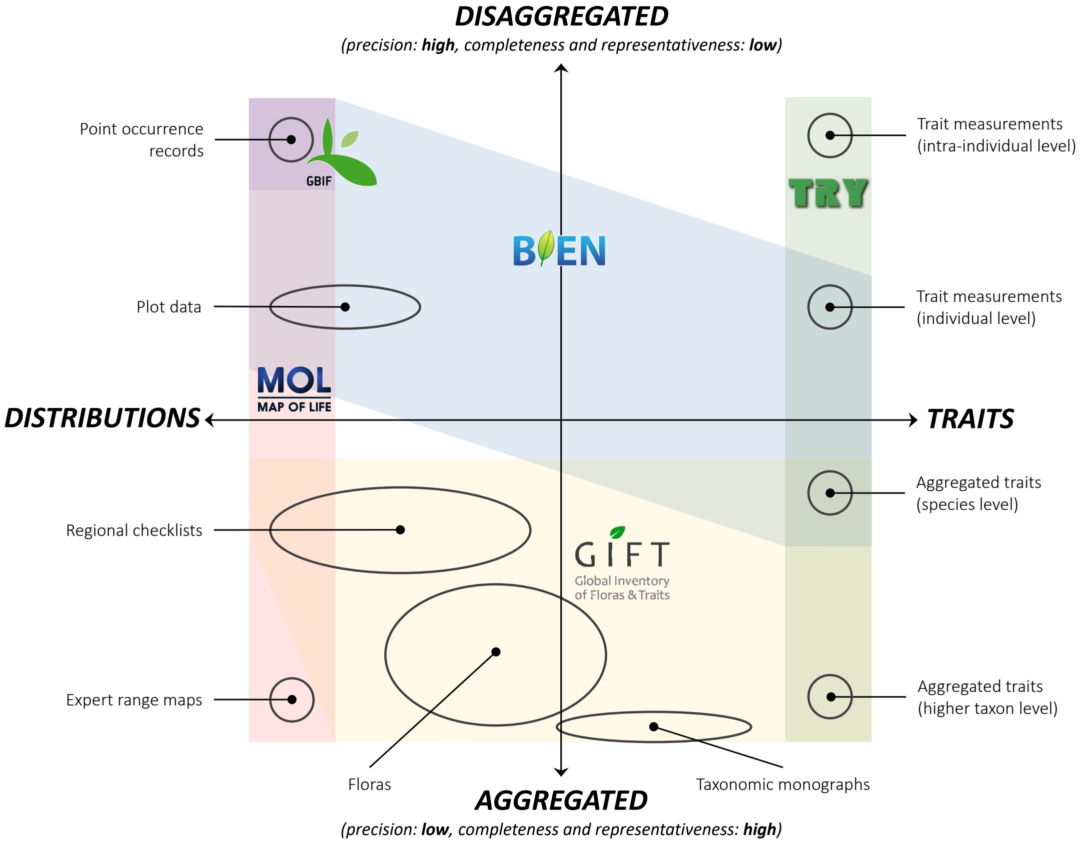


Figure 1.1: Selected biodiversity data types, arranged according to their primary domain (species distributions vs. functional traits) and informational resolution (disaggregated vs. aggregated). Existing projects that integrate global plant diversity data are often domain-specific (e.g. Map of Life: Jetz et al., 2012; TRY: Kattge et al., 2011a, GBIF, 2018) or focus on the disaggregated end of the data spectrum (e.g. BIEN: Enquist et al., 2016). Complementing the ecological data landscape with aggregated data (e.g. GIFT, see Chapter 2) creates strong synergies and facilitates biodiversity data integration across domains and resolutions.

Across different data types, there is a trade-off between high informational resolution on the one hand, and completeness and representativeness on the other (Rondinini *et al.*, 2006). This trade-off is important, because data resolution affects the *precision* (i.e. certainty) of ecological inferences, whereas data representativeness affects their *accuracy* (i.e. correctness) (Walther & Moore, 2005; Hortal *et al.*, 2015). Disaggregated data, e.g. point occurrences or trait measurements, generally have a high informational resolution, which is necessary to address questions at the level of populations or communities (Bolnick *et al.*, 2011; Meyer *et al.*, 2018). However, at macroecological scales, disaggregated data often exhibit deficits in terms of completeness and representativeness (Schrodt *et al.*, 2015; Engemann *et al.*, 2015; Meyer *et al.*, 2016). In contrast, aggregated data, e.g. regional floras and checklists, or taxonomic monographs, provide a mostly complete and representative account of their subject region or taxon (Frodin, 2001; Farjon, 2010) but are limited in their capacity to resolve fine-grained ecological information (Figure 1.1).

Major projects for biodiversity data integration focus primarily, though not exclusively, on the disaggregated end of the data spectrum, e.g. the Global Biodiversity Information Facility

(GBIF) for species occurrence records, TRY for primary trait data or the Botanical Information and Ecology Network (BIEN) for primary data on New World plant distributions and functional traits (see also Figure 1.1). A systematic compilation of existing aggregated plant diversity data to complement these initiatives is still missing. GIFT, the *Global Inventory of Floras and Traits database* (Chapter 2), is a contribution towards filling this gap and building a robust baseline for global plant diversity research.

1.3.2 Data collection and processing

The integration of biodiversity data starts in the field – with the primary biodiversity data collected in surveys, experiments, citizen science projects and other research efforts. Such data is usually specifically tailored to answer a particular research question. Thus, robust ecological generalizations require large quantities of (disaggregated) primary or (aggregated) derived data that is organized and integrated in comprehensive biodiversity databases. The quality and coverage of such databases can be greatly improved when primary research projects put strong emphasis on the utility and re-usability of collected data for secondary scientific purposes (Michener & Jones, 2012).

The utility of primary data for data integration efforts can be increased in several ways. First, focusing on regions, ecosystems, plant groups, or functional traits that are currently underrepresented in global biodiversity databases increases the general interest in the collected data as well as the study itself. Coverage analyses based on integrated biodiversity resources can provide guidance by identifying knowledge gaps and setting research priorities (Meyer *et al.*, 2016). Second, cross-institutional coordination of research projects creates synergies through standardized methods and complementary research foci. Research networks such as the International Long Term Ecological Research Network (ILTER, see Vanderbilt & Gaiser, 2017) provide an ideal framework to utilize these synergetic effects (Peters *et al.*, 2014b). Third, an efficient study design helps to maximize the data output given the available resources. This can be aided, for instance, by statistical power analyses (Johnson *et al.*, 2015), optimizing study logistics and surveying effort (Moore & McCarthy, 2016), and cooperating closely with local field guides and botanists (Elbroch *et al.*, 2011). Throughout the process of data collection, digital solutions such as Open Data Kit (Brunette *et al.*, 2013) can help to conveniently enter, cross-check, annotate and aggregate field data. This increases data integrity and provides crucial meta-information for later quality assessments and integration efforts.

The re-usability of primary data can be ensured by adopting existing data standards and protocols. The Plant List (2013) provides a widely-accepted basis for resolving and standardizing plant species names. Software packages such as *taxonstand* (Cayuela *et al.*, 2012), *taxize* (Chamberlain & Szöcs, 2013) or the *taxonomic name resolution service (*Boyle *et al.*, 2013) help to utilize The Plant List and other authoritative taxonomic resources to resolve thousands of species names at a time. With respect to functional traits, defined measurement protocols (Pérez-Harguindeguy *et al.*, 2013) and terminologies (Garnier *et al.*, 2017) facilitate interoperability across research projects. The exchange of diversity data is supported by data standards like the Darwin Core Archive (Wieczorek *et al.*, 2012) or the Humboldt Core Archive (Guralnick *et al.*, 2017). Finally, innovative publishing frameworks such as the Biodiversity Data Journal (Pensoft, 2017) or the GBIF Integrated Publishing Toolkit (GBIF, 2018) allow for a quick publication of standardized and easily accessible datasets.

1.3.3 Data mobilization

The increasing digitization of scientific collections and literature has set ecology up for the age of "Big Data" (Hampton *et al.*, 2013). The Global Biodiversity Information Facility (GBIF, 2018), for example, currently provides access to more than 208 million occurrence records of vascular plants, 62 million of which are derived from preserved herbarium specimens. While this is a substantial achievement, specimen records encode more than just distributional information (Beaman & Cellinese, 2012). In particular, the (semi-)automated extraction of traits from herbarium specimens represents an area of largely unused potential. Standardized measurements on collected plant material may be incorporated into digitization workflows, potentially yielding thousands of geographically defined records of e.g. specific leaf area (Queenborough & Porras, 2014) or phenological plant information (Gallinat *et al.*, 2018). Also, images of already digitized specimens can be used to retrieve certain functional traits, e.g. leaf size (Corney *et al.*, 2012). Nonetheless, the set of traits that can be (non-destructively) obtained from herbarium specimens excludes many important characteristics, e.g. plant growth form, vegetative height, or stem specific density.

Another way to mobilize substantial amounts of ecological data – mainly from the aggregated end of the data spectrum – lies in the botanical literature. Generations of botanists have produced thousands of Floras, species checklists, and taxonomic monographs. Vascular plants are among the most intensively studied groups, and with some exceptions, almost any region on earth has been subject to some form of floristic inventory (Frodin, 2001). Such resources provide expert-validated distributional information, often including the biogeographical status of the listed species (e.g. endemic, native, introduced). Moreover, descriptions of general morphology, life history, flowers, fruits, seeds, phenology and other features of the covered taxa are often available. Massive efforts to make biodiversity literature digitally available and searchable are underway (e.g. www.biodiversitylibrary.org, www.plantsoftheworldonline.org) and machine learning algorithms are becoming increasingly successful at extracting information from loosely structured text data (Collobert *et al.*, 2011; LeCun *et al.*, 2015). Considering the wealth of information contained in published floristic literature, the development of general, scalable methods for data extraction seems to be central for improving the coverage of biodiversity databases. Machine learning techniques such as recurrent neural networks might be particularly suited for this task. First studies using machine learning to extract trait information from floristic descriptions show promising results (Hoehndorf *et al.*, 2016).

1.3.4 Data imputation

Data imputation is a technique where missing or inconsistent data items are replaced with estimated values (OECD, 2013) and represents an inexpensive yet powerful way to improve data coverage in ecological datasets. A conceptual distinction can be made between logical and statistical imputation methods (Figure 1.2).

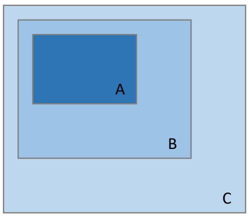
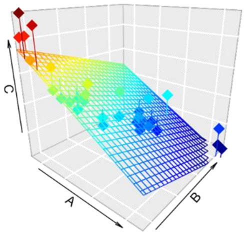
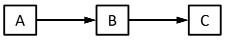
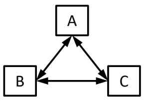
	Logical imputation	**Statistical imputation**
Data relationship	Hierarchical (one-to-many) or bijective (one-to-one) 	Correlative (many-to-many) 
Imputation method	Logical deduction 	Statistical prediction 
Gap-filling potential	Limited	Very high
Certainty of results	Very high (depending on correctness of input data and specified relationships)	Variable (depending on correlative structure of input data and model performance)
Applications (examples)	Hierarchical deduction of categorical traits („tree“ → „woody“) or occurrence information („occurs in Yasuni National Park“ → „occurs in Ecuador“)	Bayesian Hierarchical Probabilistic Matrix Factorization (Schrodt, 2015), Multiple Imputation by Chained Equations (Azur, 2011)

Figure 1.2: Comparison of logical and statistical data imputation. Logical imputation infers a limited quantity of highly certain data (e.g. deducing woodiness status from growth form), whereas statistical imputation yields large quantities of less certain data (e.g. predicting a suite of functional traits from sparse records).

Logical imputation uses unequivocal relationships among data to infer new values. This is possible either when data is categorically nested, e.g. trees always being woody (Beentje, 2016), or linked by mathematical relationships, e.g. leaf mass per unit area (LMA) being the inverse of specific leaf area (SLA). While the considerations underlying logical imputation seem rather trivial, the approach has yet to be widely used for complementing plant diversity data. Applications of logical imputation include, for example, (1) the propagation of information from complex functional traits to more simple ones (see Figure A1.2), (2) the imputation of species-level traits when a higher taxon is known to be uniform with respect to that trait, or (3) the improvement of regional species checklists based on geographically nested occurrence records or plot data. The main advantage of logical imputation is that the results can be treated with the same certainty as the underlying data. This makes it a particularly suitable approach for building and extending repositories of primary data. At the same time, logical imputation helps to harmonize data that uses differing terminologies by embedding it in a logical hierarchy (e.g. bee-pollination, insect-pollination, and animal-pollination form nested subsets of pollination syndromes). However, considering that such clear hierarchical relationships are scarce among biodiversity data, the gap-filling potential of logical imputation is limited.

Statistical imputation, on the other hand, utilizes correlative relationships among data to predict new values. As statistical imputation is based on statistical models, it can incorporate a variety of additional data to refine prediction accuracy. Gap filling techniques for functional traits, for example, take into account trait-trait, trait-environment, and trait-phylogeny relationships to predict full trait matrices from sparse data (Penone *et al.*, 2014; Schrodt *et al.*, 2015). Analogous to that, species distribution models make use of environmental information, species-specific characteristics, or biotic interactions to predict continuous species distributions from point occurrence records (Elith & Leathwick, 2009; Peterson, 2011). Statistical imputation methods allow for the prediction of any number of missing values, but the accuracy of these predictions is always dependent on the quality (i.e. correctness, representativeness, and completeness) of predictor variables as well as the performance of the underlying statistical model. Hence, statistical imputation is a valuable tool for improving data coverage in specific use cases (Paine *et al.*, 2011; Syfert *et al.*, 2014; Díaz *et al.*, 2016), but cannot be considered an expansion of primary data.

Strong synergies arise from combining logical imputation, which maximizes the amount of *quasi*-primary data, with statistical imputation, which may utilize this additional data to improve prediction accuracy. The potential of logical imputation for deducing simple functional traits such as woodiness or growth form is substantial (see case studies in Chapter 1.4). While improved knowledge on these traits is of broad ecological interest in itself (McGlone *et al.*, 2015; Beech *et al.*, 2017), it might be particularly useful to enhance the performance of statistical imputation techniques (van Buuren & Groothuis-Oudshoorn, 2011; Schrodt *et al.*, 2015). Similarly, logically imputed distributional information can help to improve species distribution models, e.g. by flagging and removing inconsistent occurrence records (Jetz *et al.*, 2012) or deriving often-required pseudo-absences for species distribution models from regional checklists (VanDerWal *et al.*, 2009; Barbet-Massin *et al.*, 2012).

1.3.5 Data sharing

Data sharing is a basic condition for the global integration of plant diversity data. Ecology, as opposed to, e.g., taxonomy or molecular biology, lacks a long-standing culture of data sharing (Reichman *et al.*, 2011; Hampton *et al.*, 2013) and although open science initiatives start to gain traction in ecology, considerable institutional and sociocultural challenges remain (Michener, 2015a; Gewin, 2016). Publishers, universities, and funding agencies have a central responsibility for creating an environment where data sharing is a scientific asset, not a disadvantage. Corresponding measures comprise a range of obligations and incentives for data sharing (Whitlock, 2011; Kattge *et al.*, 2014). One example for an effective obligation is that many journals now require all data that were used to conduct a study to be stored in open repositories (Mills *et al.*, 2015). Likewise, funding agencies strive to improve data quality and long-term accessibility by requiring data management plans (Michener, 2015b). The most important measure, however, is the establishment of adequate incentives for data sharing, which is primarily a matter of increasing the academic credit gained from doing so. Dataset citations are an effective way of incentivizing and acknowledging data contributions, but also alternative impact measures and a stronger appreciation of data as scientific output can help to open up ecological research culture (Kattge *et al.*, 2014; Gewin, 2016).

1.3.6 Data integration

Biodiversity data are typically collated and integrated in domain-specific databases that allow fast extraction, exploration, and visualization of highly normalized d ata. This approach has transformed the ecological research landscape in the past decades and acted as a catalyst of ecological knowledge synthesis (Kelling *et al.*, 2009). However, the scope of any single project is bound to a limited amount of technical, financial and human resources. The challenge of building a scalable, dynamic e-infrastructure that integrates the wealth of existing environmental and ecological data is therefore best realized by bundling existing efforts within a unifying framework (Peters *et al.*, 2014b; La Salle *et al.*, 2016).

The idea of a distributed network aims to organize data, resources and expertise from diverse data holders in a single, collaborative infrastructure that allows for the seamless discovery, acquisition, citation and (re-)use of data (Michener *et al.*, 2011; Peters *et al.*, 2014b). A shared data portal acts as a central access point, while more specialized databases remain generally in charge of data aggregation and warehousing (Michener & Jones, 2012). This organizational model, in principle, has the potential to integrate the heterogeneous ecological data landscape, but is also strongly dependent on the broad adoption of data standards. These include, but are not necessarily restricted to: (1) universal identifiers ranging from standardized species names to digital identifiers for documents, data and persons (e.g. DOIs, LSIDs, ORCIDs) (Page, 2008), (2) compatible database structures and the implementation of standardized APIs and exchange formats (Kattge *et al.*, 2011b), (3) rich and well-structured meta-data (Reichman *et al.*, 2011; Fegraus *et al.*, 2005), and (4) the formalization of existing ecological concepts in ontologies and thesauri (Mouquet *et al.*, 2015; Garnier *et al.*, 2017).

The Data Observation Network for Earth (https://www.dataone.org, Michener *et al.*, 2011) already provides the basic infrastructure for building an open and distributed network of biodiversity data holders. However, currently the majority of member nodes consists of generic data repositories (e.g. DRYAD) and regional projects (e.g. USGS), whereas the participation of major aggregators of global plant diversity data (e.g. GBIF) has yet to be realized. Consequently, DataONE currently does not leverage the full potential of its powerful organizational model (Reichman *et al.*, 2011; Michener *et al.*, 2011). Some of the future challenges for distributed infrastructures such as DataONE are, for example, the continuing promotion and development of data standards, the improvement of web-based visualization and analysis capabilities, the incorporation of machine learning for improved data discovery and utilization (Peters *et al.*, 2014a), and the robust implementation of dynamic cross-checking and data imputation workflows for parallel data streams.

1.4 Case studies

We conducted three case studies that demonstrate the potential of large-scale data integration in plant diversity research. Each case study addresses a central topic in functional biogeography using data from GIFT (Chapter 2). Considering that GIFT's main focus lies on aggregated data on plant distributions and functional traits, these examples merely provide an outlook of what the full integration of biodiversity data across domains and resolutions can achieve.

1.4.1 Global patterns in plant growth form

The grouping of plants into plant functional types such as growth forms captures fundamental axes of ecological variation in a uniquely simple way (Leishman & Westoby, 1992; Díaz *et al.*, 2016). Consequently, knowledge of plant growth form constitutes an important aspect in many ecological applications, ranging from local studies of plant diversity (Knapp *et al.*, 2008; Madrigal-González *et al.*, 2017) to dynamic global vegetation models (Prentice *et al.*, 2007; Wullschleger *et al.*, 2014). However, despite being a relatively simple and easily determinable trait, data on growth form is still surprisingly scarce and scattered both taxonomically and geographically. Here, we demonstrate the opportunities arising from a systematic collection of growth form data.

We combined angiosperm checklists and growth form data (herb/shrub/tree) available in GIFT. Oceanic islands and units with more than 33 % of species lacking growth form information were excluded. From the remaining 818 regional checklists, we included only those species with known growth form status, yielding 1,472,024 species-by-sites combinations and 162,300 unique species. We used this dataset for predicting growth form spectra for 6495 equal-area grid cells (~ 23,300 km^2 each) using multinomial logistic regression (nnet R-package, Venables & Ripley, 2002) based on contemporary climatic conditions. Since our objective was predictive accuracy, not statistical inference, we did not account for collinearity among predictors and used all 19 bioclimatic variables from the CHELSA climate layers (Karger *et al.*, 2017). We weighted each observation by the inverse log-area of the corresponding geographical region to account for the decreasing representativeness of averaged climatic conditions for larger, climatically more variable regions.

Globally, herbs represented the most frequent growth form (Figure 1.3A&C), accounting for 68 % of species-by-sites combinations and 56 % of species. Shrubs and trees were less abundant with 17 % and 18 % of species-by-sites combinations, and 23 % and 21 % of species, respectively. Regionally, however, shrubs and trees reached relatively high proportions, particularly in Australian scrublands (Figure 1.3E) and the Amazon rainforest (Figure 1.3G). Except for local deviations, e.g. in the shrub-dominated ecosystems of Western Australia, our predictions of global patterns in growth form composition were in strong agreement with the observed data (McFadden's Pseudo-R^2 = 0.91). Additionally, our results are supported by an independent analysis of Engemann *et al.* (2016), which revealed similar geographical trends in growth form composition for North- and South America.

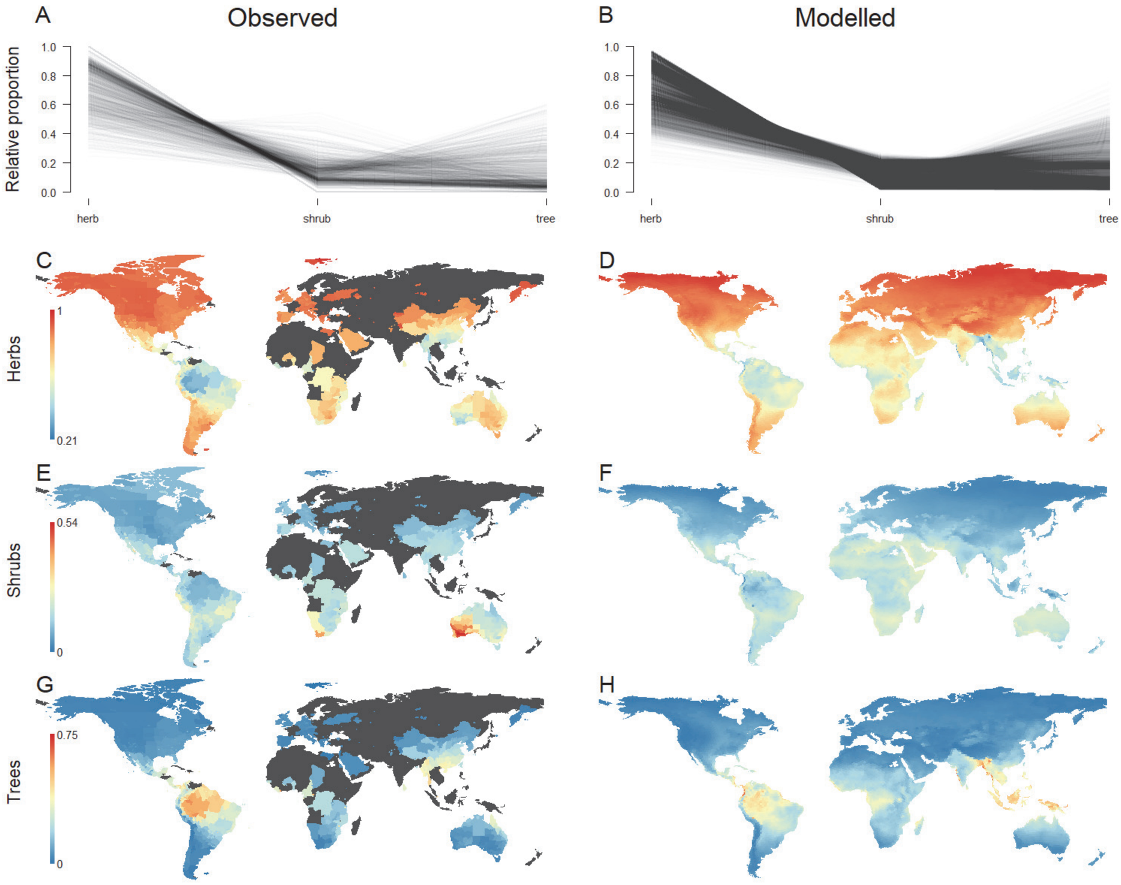


Figure 1.3: The global composition in plant growth form as observed for 818 angiosperm floras (left) and modelled for 6495 equal-area grid cells of approximately 23,300 km^2 each (right). Upper plots summarize the overall growth form composition across all observed (A) and modelled (B) geographical units, with each line representing a single flora. Lower plots (C-H) show the observed and modelled geographic variation in the proportion of herbs, shrubs, and trees individually. Note that the range of values varies across growth forms.

This case study has two implications. First, a characterization of all plant species with respect to fundamental categorical plant traits such as growth form is within reach when exploiting the full potential of data mobilization and imputation. Second, even spatially coarse-grained data may contain enough information to derive reasonably accurate predictions at finer grain sizes. Consequently, improving knowledge on coarse, yet ecologically informative traits will allow for an increasingly accurate functional description of plant assemblages worldwide and improve our understanding of their responses under altered environmental conditions.

1.4.2 The latitudinal gradient in seed mass revisited

Latitude is strongly correlated with several ecologically relevant environmental characteristics, e.g. temperature, precipitation, seasonality, and long-term climatic stability. Hence, many aspects of biodiversity including geographic range size (Stevens, 1989), net primary productivity (Cramer *et al.*, 1999), and species diversity (Hillebrand, 2004) show systematic variation along latitude. Also some plant traits vary strongly with latitude. Moles *et al.* (2007) analysed the latitudinal variation in seed mass based on a dataset of 11,481 species-by-sites combinations. They found a 320-fold decrease in seed mass between the equator and 60 degrees latitude as well as a sudden, 7-fold drop at 23 degrees latitude. These results were linked to changes in vegetation type and growth form composition, leading the authors to posit an abrupt change in plant strategy at the edge of the tropics. Here, our aim is to replicate these findings.

We extracted species lists for all mainland units in GIFT where a complete survey of seed plants was available. In cases where geographical units overlapped by more than 5 %, we removed the larger unit if floristic data was available at a higher spatial resolution (e.g. preferring federal state- over country-level data); otherwise we removed the smaller unit (e.g. preferring country-level data over a single national park inventory). Furthermore, we only kept species with information on both seed mass or growth form, yielding a final data set of 519,812 species-by-region combinations and 563 distinct geographical units. In re-assessing the relationship between seed mass and latitude, we followed the methodology of Moles *et al.* (2007) and used linear regression and piecewise linear regression.

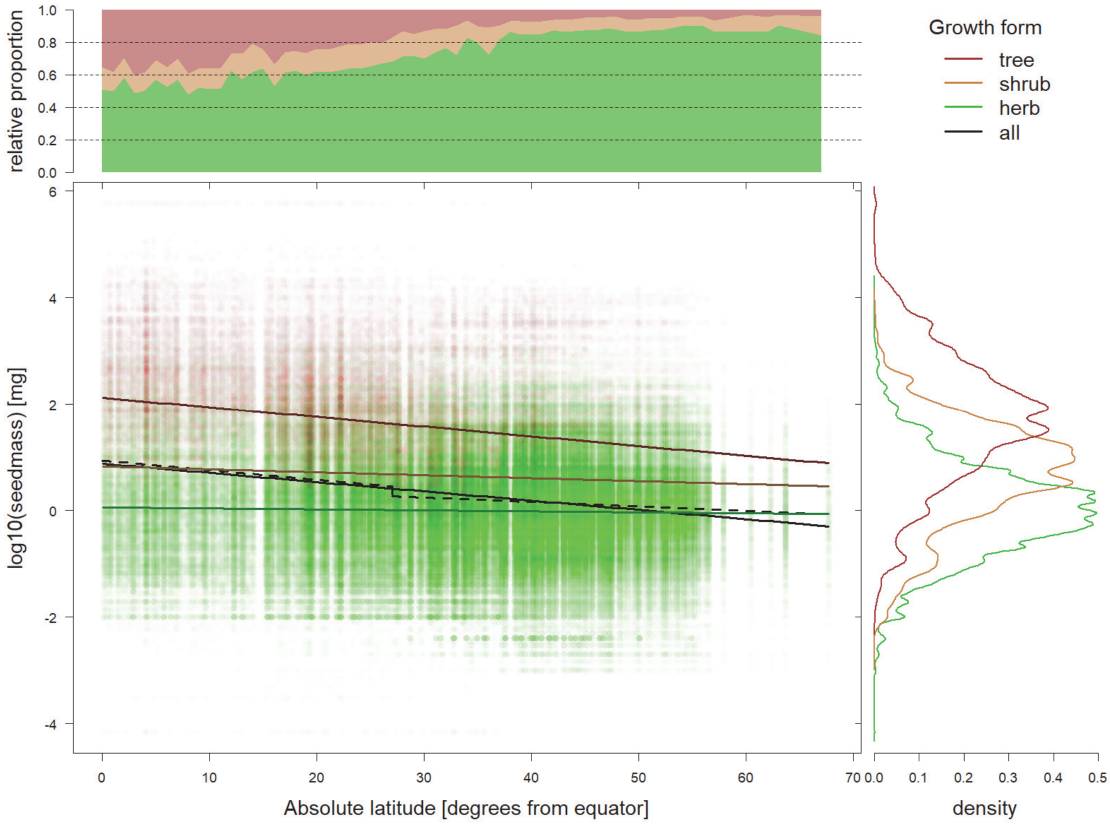


Figure 1.4: Latitudinal gradient in seed mass for 519,812 species-sites combinations. Piecewise regression (dashed black line) was implemented following Moles *et al.* (2007) and compared against linear models for the entire data set (solid black line) and individual growth forms (coloured lines). Upper plot shows the relative proportion of growth forms in each 1-degree latitudinal band. Right-hand plot depicts the frequency distribution of seed mass for individual growth forms.

We found that the estimated decrease in mean seed mass between the equator and 60 degrees latitude was only 11-fold according to simple linear regression (Figure 1.4, solid black line) and 8.8-fold according piecewise linear regression, the latter indicating a 1.5-fold drop at 27 degrees latitude (Figure 1.4, dashed black line). In both cases, the explanatory power was low ($R^2_{linear} = 0.045$, $R^2_{piecewise} = 0.048$), reflecting the presence of substantial variation in seed mass at any given latitude. The latitudinal response of individual growth forms was even weaker than the overall effect (see coloured lines vs. black line in Figure 1.4), while the logarithmic mean seed mass per growth form (herbs: 0.99 mg, shrubs: 4.59 mg, trees: 48.95 mg, Figure 1.4, right-hand plot) differed significantly (Kolmogorov–Smirnov test, $p < 0.001$). Consequently, the overall poleward decrease in seed mass seems to be mostly driven by the gradual replacement of large-seeded trees by small-seeded herbs (Figure 1.4, upper plot). In conclusion, our results suggest that the latitudinal gradient in seed mass is considerably less steep than previously reported (Moles *et al.*, 2007) and lacks a pronounced drop at the edge of the tropics.

This case study illustrates that that the quantification of large-scale diversity patterns is highly dependent on the representativeness of the underlying data. In this respect, functional representativeness has been a largely neglected dimension of sample quality. Indeed, the data that generated the original results show a much higher proportion of tree-dominated biomes and, additionally, of tree species at tropical latitudes compared to ours (Moles *et al.*, 2007). Integrated biodiversity resources with broad data coverage can help to detect and resolve such latent biases in macroecological datasets.

1.4.3 A global assessment of insular woodiness

In our last case study, we examine the prominent island syndrome of insular woodiness, the tendency of primarily herbaceous plant lineages to adopt a woody habit on islands. Explanations for this condition include the competitive advantage arising from a higher stature (Darwin, 1859), the increased pollination probability due to an extended lifespan (Wallace, 1878), and the reduced physiological stress due to moderated climate on islands (Carlquist, 1974). The generality of island syndromes such as insular woodiness is regarded as one of the most fundamental questions in island biology (Patiño *et al.*, 2017). Here, we tackle this question and explore patterns in woodiness of island floras.

We selected a set of twelve globally representative oceanic islands with a substantial number (> 40) of endemic plant species from GIFT. We focused on seed plants because extant spore-bearing plants do not exhibit secondary growth, which is a precondition for developing woodiness (Ragni & Greb, 2018). Based on the biogeographical status, we grouped species on each island into native non-endemics (species whose natural range includes, but is not restricted to the respective island or island group) and endemics (species whose range is restricted to the respective island or island group). We then contrasted endemics and non-endemics on each island with respect to the proportion of woody vs. non-woody species and the proportions of different life forms *sensu* Raunkiær (1934), assuming that trait syndromes of endemic species are the outcome of adaptive processes to local biotic and abiotic conditions.

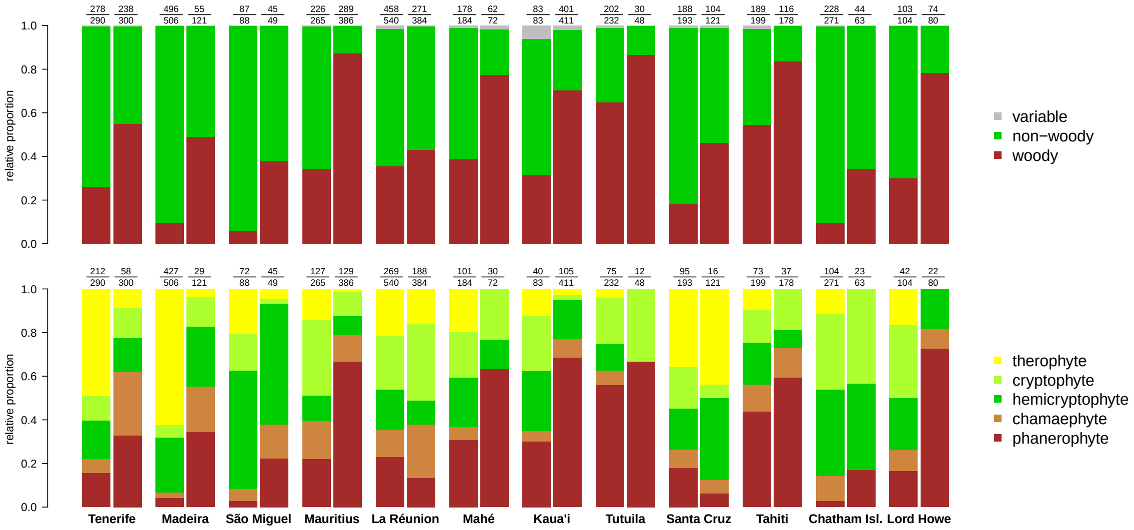


Figure 1.5: Proportions of woody vs non-woody species and Raunkiær life forms among seed plants on twelve oceanic islands. For each island, species were classified into native non-endemics (left-hand bars) and endemics (right-hand bars). Numbers above bars denote the number of species with known trait status and the total number of species for each group per island.

On all investigated islands except La Réunion, endemics showed a significantly higher proportion of woody species compared to native non-endemics (χ^2 test of homogeneity at $\alpha = 0.05$, see Figure 1.5). Likewise, woody life forms (phanerophytes and chamaephytes) were strongly overrepresented among island endemics. Moreover, we found the differential representation of life forms to be highly collinear with their approximate position the rK-spectrum: therophytes (strongly r-selected) showed the largest overall decrease, while phanerophytes (strongly K-selected) showed the largest overall increase between native non-endemics and endemics (Figure 1.5). We did not perform statistical tests on the proportion of life forms due to the relatively low data coverage for endemic species.

This study illustrates that data integration bears great potential for examining long-standing ecological and biogeographical questions from a data-driven perspective. Our findings suggest that insular woodiness is indeed a widespread phenomenon, occurring under a wide range of climatic conditions and spatial settings. Although an altered functional composition of island endemics may have other causes than adaptation, e.g. higher diversification rates of woody colonizers or relictual populations of woody clades, our results are in line with molecular studies that focus on trait shifts that occurred after island colonization (Lens *et al.*, 2013; García-Verdugo *et al.*, 2014).

1.5 Conclusion and future directions

The availability, quality and interoperability of data is paramount to the progress of biogeography and ecology as increasingly data-intensive disciplines (Michener & Jones, 2012; Hampton *et al.*, 2013; Franklin *et al.*, 2017). Here, we demonstrated how the explicit consideration of data resolution offers new perspectives on the compilation and integration of plant diversity data. Our results show that a near-complete collection of coarse-grained plant distributions and basic functional traits is within reach, when exploiting the full potential of data mobilization and imputation. This offers new opportunities for plant diversity research in general.

Currently, studies and projects for the integration of global plant diversity are mostly based on disaggregated data. While this approach has been a highly successful line of research (Swenson *et al.*, 2012; Moles *et al.*, 2014; Díaz *et al.*, 2016), the pervasiveness of biases and gaps in disaggregated biodiversity data is of increasing concern to ecologists (Boakes *et al.*, 2010; Engemann *et al.*, 2015; Sandel *et al.*, 2015; Meyer *et al.*, 2016). We have shown that systematic utilization of aggregated data can help address this problem. First, aggregated data provide a coarse but more complete and less biased picture of geographical variation in taxonomic, functional and phylogenetic diversity. This offers much-needed baselines against which the completeness of disaggregated data can be evaluated in order to quantify and map gaps in global biodiversity knowledge (Hortal *et al.*, 2015; Franklin *et al.*, 2017). Second, aggregated data provide prior information about the geographical and statistical distribution of more highly resolved, but potentially incomplete or biased ecological variables. This knowledge can be used, for instance, to inform functional biogeographical analyses (see case study 2), to improve species distribution and niche models (Merow *et al.*, 2016), or to parametrize ancestral state reconstructions (Pagel *et al.*, 2004) and dynamic global vegetation models (Scheiter *et al.*, 2013). Third, aggregated data capitalize on expert knowledge to compensate for the varying availability and quality of primary (disaggregated) data. Consequently, aggregated data types are not mere compilations of disaggregated data, but provide valuable additional information, e.g. reliable species absences or uniform functional traits for higher taxa. The potential of utilizing aggregated biodiversity data extends to other clades for which a wealth of literature exists, e.g. mammals, birds, or many groups of arthropods.

Data integration is potent not only across resolutions, but also across domains. Satellite-borne, multispectral imagery is a crucial component of biodiversity research, providing global high-resolution data of e.g. net primary productivity, vegetation cover or canopy height (Kuenzer *et al.*, 2014). Advanced instruments will soon enable the derivation of similar data products for selected functional traits, which helps tracking changes in the biosphere at increasing spatial and temporal resolutions. Nevertheless, the identification of individual plants from space remains impossible for most practical purposes, which highlights the need for integrating *in-situ* and satellite-borne data to address ecological questions at global scales

(Jetz *et al.*, 2016). Vegetation plot databases are another key source of plant diversity data, holding crucial information on species abundances and co-occurrences. BIEN demonstrates how the integration of specimen- and plot data with taxonomic, functional and phylogenetic information helps bridging the gap between local-, regional- and continental-scale ecological processes (Blonder *et al.*, 2015; Engemann *et al.*, 2016). Furthermore, biogeographical analyses could benefit from integrating contemporary species distributions with fossil records and phylogenies, and conservation planning could be aided by bringing together ecological, environmental and socioeconomic data within a consistent framework – the potential of cross-domain data integration remains to be fully explored.

The unparalleled pressure on our global biosphere renders a full utilization of all available biodiversity data imperative. Rapid advancements in information technology have brought down the technological barriers to this objective. It is now up to ecologists to keep pace with this development, and to work collaboratively on creating infrastructures for the integration of biodiversity data that bridge the gap between fine-scale precision and global representativeness.

2 GIFT – A Global Inventory of Floras and Traits for macroecology and biogeography

Patrick Weigelt, Christian König and Holger Kreft

2.1 Abstract

To understand the evolutionary history and geographic distribution of plant life on Earth, we need to integrate high-quality and global-scale distribution data with functional and phylogenetic information. Large-scale distribution data for plants are, however, often restricted to either certain taxonomic groups or geographic regions. For example, range maps only exist for a small subset of all plant species and digitally available point-occurrence information is strongly biased geographically and taxonomically. An alternative, currently rarely used source of information is represented by regional Floras and checklists, which contain highly curated information about the species found in clearly defined areas, and which together cover virtually the entire global land surface. Here we report on our recent efforts to mobilize this information for macroecological and biogeographical analyses in the GIFT database, the Global Inventory of Floras and Traits. GIFT integrates species distributions of land plants (focusing on vascular plants) with trait and phylogenetic information as well as region-level geographic, environmental and socio-economic data. GIFT currently holds species lists for 2,893 regions across the whole globe including ~315,000 taxonomically standardized species names (i.e. c. 80% of all known land plant species) and ~3 million species-by-region occurrences. In addition, GIFT contains information about the floristic status (native, endemic, alien and naturalized) and takes advantage of the wealth of trait information in the regional Floras, complemented by data from global trait databases. Utilizing hierarchical and taxonomic trait imputation, GIFT holds information for 83 functional traits and more than 2.3 million trait-by-species combinations and achieves unprecedented coverage in categorical traits such as woodiness (~233,000 spp.) or growth form (~213,000 spp.). Here we present the structure, content and automated workflows of GIFT and a corresponding web-interface (http://gift.uni-goettingen.de) as proof of concept for the feasibility and potential of mobilizing aggregated biodiversity data for global macroecological and biogeographical research.

2.2 Introduction

Worldwide, about 382,000 vascular plant species form the basis of our terrestrial biosphere and provide key ecosystem services to humanity (Willis, 2017). Despite the long history of botanical exploration of our planet, the global distribution is only known for a subset of all plant species at comparatively coarse spatial grains (e.g. WCSP, 2014). In contrast to smaller and better known taxa like birds and mammals (BirdLife International, 2018; IUCN, 2018), high-quality species-level range maps or atlas data of plants are only available for certain well-studied groups (e.g. conifers in Farjon & Filer, 2013; cacti in Barthlott *et al.*, 2015) or confined regions (e.g. Europe in Tutin *et al.*, 1964–1980). Many research questions at the forefront of biogeography and macroecology, however, require a detailed knowledge of global plant distributions and, additionally, of species-level functional traits and phylogenetic relationships (e.g. Morueta-Holme *et al.*, 2013; Weigelt *et al.*, 2015; König *et al.*, 2017).

Several national and international initiatives focus on mobilizing and aggregating plant distribution information. For instance, the Global Biodiversity Information Facility (GBIF, 2018) provides access to ~214 million point occurrences of vascular plant species from herbarium records and observations. These records are invaluable for plant ecology and conservation-related research, as they provide information about key aspects of species identity, time and place (Powney & Isaac, 2015). However, taxonomic, geographical and temporal biases (Hortal *et al.*, 2015; Meyer *et al.*, 2016) as well as the lack of important meta-information, like, for example, the floristic status at a given location (native, non-native, naturalized, etc.), limit their usefulness for macroecological research. An alternative source of information are Floras and checklists which, in contrast, present highly curated accounts of the plant species known to occur in a certain region. Floras and checklists are often based on decades to centuries of exploration and regional botanical work, and have profited from the expertise of generations of botanists. They aim at providing (near-)complete floristic inventories for a given region and thus provide information on species presences and their floristic status, and additionally allow for the inference of local species absences (Lobo *et al.*, 2010; Jetz *et al.*, 2012). So far, extensive compilations of plant checklists exist only for certain geographic regions (e.g. Ulloa Ulloa *et al.*, 2017), taxonomic groups (e.g. Flann, 2009; WCSP, 2014), functional types (e.g. BGCI, 2017), or, for example, naturalized alien plants (van Kleunen *et al.*, 2015; Pyšek *et al.*, 2017).

In light of the increasing availability of biodiversity data, it is a major challenge to integrate various data types and to link data from different ecological domains representing species distributions, functional traits, phylogenetic relationships or environmental characteristics for analyses and cross-validation (see Chapter 1). Initiatives that integrate different types of distribution data with additional biotic or abiotic information are currently most comprehensive for particular geographic regions (e.g. BIEN for the Americas; Enquist *et al.*, 2016) or other taxa (e.g. Map of Life for vertebrates; Jetz *et al.*, 2012). However, the wealth of aggregated information in regional Floras and checklists (Frodin, 2001) allows for a near-global characterization of plant distributions. In combination with functional traits from the botanical literature or large trait databases (e.g. Kattge *et al.*, 2011a; Royal Botanic Gardens Kew, 2016) and ever-growing species-level phylogenies (e.g. Smith & Brown, 2018), this represents a promising basis for macroecological and biogeographic research.

Here, we present GIFT, the Global Inventory of Floras and Traits database, a new resource designed to integrate species distribution data and functional traits of vascular plants from regional Floras and checklists with phylogenetic information and geographic, environmental, and socio-economic characteristics (Figure 2.1). As such, the database architecture, workflows, and data of GIFT facilitate a wide array of macroecological and biogeographical analyses and may help to extent and validate other plant distribution and trait data resources. The general concepts outlined here may serve as a role-model for aggregated species checklist and trait databases for other major taxonomic groups.

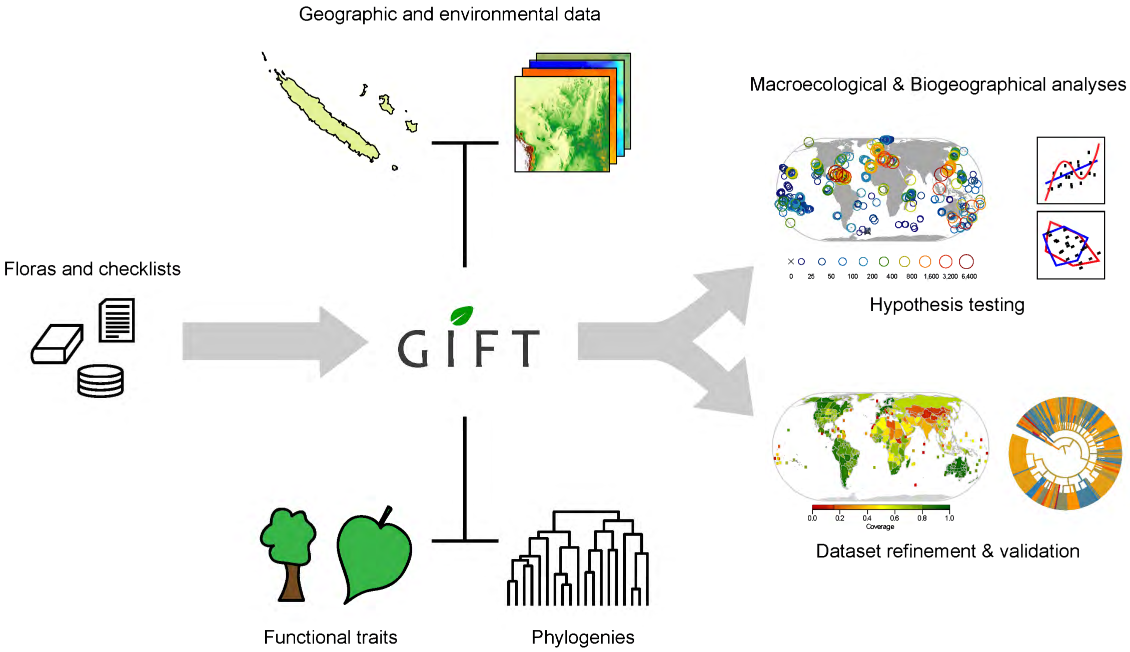


Figure 2.1: Conceptual framework of the Global Inventory of Floras and Traits database (GIFT). The core information in GIFT are species occurrences in geographic regions (islands, political units, protected areas, biogeographical regions) based on Floras and checklist. At the level of the geographical regions, this information is linked to physical geographic, bioclimatic, and socioeconomic properties. At the level of the species, functional traits, taxonomic placement, and phylogenetic relationships are linked. This integration of species distribution data in the form of full regional inventories and regional and species characteristics allows for a wide variety of macroecological and biogeographical analyses of taxonomic, phylogenetic, and functional diversity as well as for the refinement and validation of other plant distribution and trait datasets.

2.3 Content and structure of GIFT

2.3.1 Overview

Regional Floras and checklists are a rich source of information on species distributions that often also contain detailed descriptions of species traits (Palmer & Richardson, 2012). In the past, a lot of the knowledge hidden in printed books (Frodin, 2001) resided in libraries and was difficult to access for ecological research. These resources, however, are increasingly being made digitally available (e.g. Zuloaga *et al.*, 2004; Acevedo-Rodríguez & Strong, 2007) and modern regional inventory projects are set up as digital databases right from the start (e.g. Brach & Song, 2006; Jardim Botânico do Rio de Janeiro, 2016). In GIFT, we make use of this wealth of available information, and collate and mobilize plant species lists and trait information from published and unpublished Floras, catalogs, checklists and online databases into one global database.

The original checklist data in GIFT consist of species names from the literature, their occurrences in the regional species lists and original trait information (yellow boxes in Figure 2.2). All this information is linked to meta-data on the included literature references, species lists, traits, and geographic entities (white boxes in Figure 2.2). Transparent automated workflows allow a fast and reliable integration of new datasets and provide extensive derived information (blue boxes in Figure 2.2): (1) taxonomic match-up with taxonomic resources and name standardization, (2) taxon placement according to a taxonomic backbone and phylogeny, (3) trait standardization and hierarchical and taxonomic trait derivation, (4) calculation of regional summary statistics like species richness or trait coverage and (5) extraction and computation of geographic, environmental and socio-economic regional characteristics. Based on this generic database framework, GIFT can be queried for complete species checklists of a certain taxonomic or functional group and floristic status (e.g. 'native angiosperms' or 'naturalized trees'). Alternatively, one can extract e.g. the distribution or trait information for a given set of species, or regional characteristics such as environmental data, species numbers, and trait coverage per taxon, floristic status, region. GIFT is stored on a MySQL 5.5.43 database server. Workflows for preparing, importing, processing, extracting, and visualizing data are written in the R statistical programming language (R Core Team, 2018).

2.3.2 Checklists

GIFT currently contains 3,826 species lists referring to 2,893 different geographic regions which are based on 429 original checklist data sources (Figure 2.2). A full and up-to-date list of all data citations and their bibliographic references is available at the GIFT website (http://gift.uni-goettingen.de) and publications based on GIFT are requested to cite the checklist resources the analyses are based upon (e.g in Weigelt *et al.*, 2016). Checklists and inventories stem from publically available sources as well as from unpublished sources (4.2%), for which usage conditions are stored in the database. Meta-data on references and species lists further specify the type of the reference as provided (Flora, checklist, catalogue, identification key, survey, etc.), the taxonomic and floristic scope of the reference (e.g. all native and naturalized angiosperms), whether the species' floristic status is indicated and which functional traits are reported.

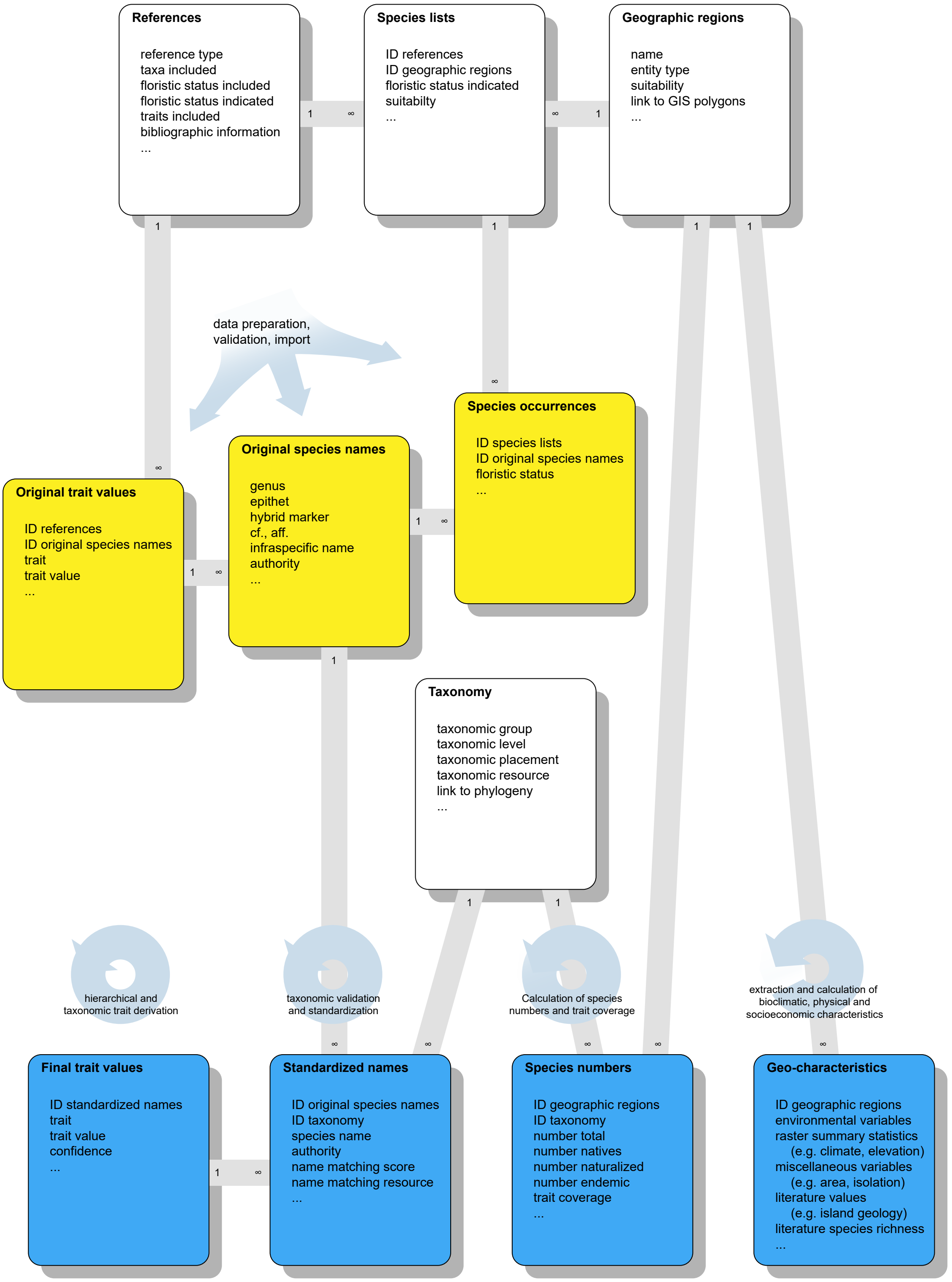


Figure 2.2: Simplified structure of the Global Inventory of Floras and Traits database (GIFT). Metadata on literature references, species lists and geographic regions builds the backbone of the database (top row). A reference can include several species lists (e.g. for different sub-regions) and a geographic region can be covered by several lists and references. Species lists vary in taxonomic and floristic scope (e.g. all native and naturalized angiosperms) and in the information content (floristic status, functional traits).

The actual distribution information is kept in a separate table that links the taxonomic names to the species lists they occur in and via those to the geographic regions (Figure 2.2, 'species occurrences'). For each species occurrence, we indicate, if possible, whether the taxon is native or not. For native species, we indicate if they are endemic to the geographic entity of the species list or to the geographic entity of the entire reference. For non-native species, we indicate whether they are naturalized or not. In all cases, we also indicate whether this information is questionable according to the literature source. Via the species names, occurrences are linked to functional traits as well as to the taxonomic and phylogenetic backbone. Via the geographic regions, species and traits are linked to regional geographical characteristics (Figure 2.2). Routines for exporting checklists from GIFT and their meta-data as Darwin Core (Wieczorek *et al.*, 2012) and Humboldt Core (Guralnick *et al.*, 2017) archives, respectively, are currently being developed.

2.3.3 Species names and taxonomic standardization

All species names enter the database in their original form including infraspecific information and author names where available. Species names derived from heterogeneous resources, referring to various geographic regions and published over a timespan of about one hundred years, inevitably vary considerably in the taxonomic concepts (Jansen & Dengler, 2010). To compare species identities across different resources, we therefore submit all non-hybrid species names to a semi-automated taxonomic standardization and validation procedure based on taxonomic information provided by The Plant List 1.1 (TPL; The Plant List, 2013) and additional resources available via iPlant's Taxonomic Name Resolution Service (TNRS; Boyle *et al.*, 2013). This procedure was exclusively developed to meet the needs of the GIFT database and has already been applied and described in Meyer *et al.* (2016).

First, all genus names not occurring in TPL are corrected manually based on literature and online resources (e.g. Mabberley, 2008; IPNI, 2012). Entries that cannot be assigned to an established genus name at all (valid or not) are excluded from further steps. Second, all species names are compared automatically to all taxonomic names available for a particular genus in TPL based on pairwise orthographic distances (generalized Levenshtein distance; Levenshtein, 1966) between species epithets, infraspecific names, author names and the entire species names. We use both the absolute orthographic distance, which is the number of changes needed to transform one character string into the other (Levenshtein, 1966), and the relative orthographic distance, which relates the absolute orthographic distance to the length of the longer input string. Based on the orthographic distances of an original species name to all congeneric species listed in TPL, we determine the final working name hierarchically: First, we choose the best-matching species epithet. If multiple epithets match equally well, we choose those with best-matching infraspecific names (if infraspecific name available and if absolute orthographic distance < 4 and relative orthographic distance < 0.3), and then those with best-matching author names (if author names available and relative orthographic distance < 0.5). The specific matching thresholds at each step were set in consideration of the balance between the number of names that cannot be matched and the number of names that are matched to the wrong species. Synonyms are linked to their accepted species names as suggested by TPL. If several names match equally well and lead to different accepted binomial species names, we first remove illegitimate and invalid names, then synonyms and then accepted names with poorer overall orthographic distance. In addition, all names are resolved using the TNRS *application programming interface (API)*, which returns similar statistics on the name matching and the status of the matched names like the above-described approach using TPL. For choosing standardized binomial working names we give priority to TPL over TNRS, because of the possibility of adjusting our TPL name matching approach. If a name does not match any name via TPL or TNRS with a relative orthographic distance < 0.25 for either the epithet or the full name, we keep the original name as working name. If not stated otherwise summary statistics below are based on these standardized binomial working names.

All original names, orthographic distances, matched names and meta-information about the matching are stored in the database (Figure 2.2). Thus, the taxonomic standardization in GIFT is fully transparent and repeatable whenever taxonomic resources are updated or extended. Moreover, the stored information can be used to filter out names that did not match, matched only to a certain degree, or that do not lead to an accepted species name, allowing for rigorous sensitivity analyses of the effects of taxonomic uncertainties on the outcome of macroecological and biogeographical analyses.

2.3.4 Taxonomic backbone and phylogeny

All species working names are linked to a taxonomic backbone via their genus names. The taxonomy is based on the Angiosperm Phylogeny Group IV system for angiosperms (The Angiosperm Phylogeny Group, 2016), and on the Angiosperm Phylogeny Website version 13 (Stevens, 2001 onwards) and The Plant List 1.1 for gymnosperms, pteridophytes and bryophytes (The Plant List, 2013). Based on the taxonomic backbone, the database can be queried in two directions. First, species lists can be extracted including only species that belong to a certain taxon (e.g. only angiosperms). Second, geographic units can be chosen for which species lists cover a complete taxon of interest (e.g. all regions with Bixaceae checklists). In combination, species lists of a certain taxon can be produced for all regions where the required data is available. In addition, species-level functional traits can be aggregated at any desired taxonomic level and trait information for broad taxonomic groups can be used to derive species-level information for traits that are consistent across a larger taxonomic groups.

All seed plant species are linked to a global phylogeny with 353,185 terminal taxa (Smith & Brown, 2018) for phylo-geographical analyses. Two versions of this phylogeny are included in the database in tabular form to extract checklist and trait information for particular clades and to visualize trait and taxonomic coverage across the phylogeny. In one version, species in GIFT not included in the phylogeny were added replacing all members of the genera they belong to with polytomies (Pearse *et al.*, 2015) and in the other version missing species were excluded to keep detailed phylogenetic relationships among the species included. In addition, all vascular plant species in GIFT are linked to a phylogeny with fewer terminal taxa but broader phylogenetic extent (i.e. including pteridophytes; Qian & Jin, 2016), which was used here to assess taxonomic coverage of distribution and trait data at the family level.

2.3.5 Functional traits

Species in GIFT are linked to functional trait information from currently 155 original resources. Most trait resources are Floras and checklists for which annotated information on traits have been extracted, but also large trait compilations with or without spatial context are incorporated (e.g. Zotz, 2013; BGCI, 2017). The range of functional traits currently covered by GIFT reflects different aspects of plant morphology, life history, reproduction, physiology, genetics, and ecology (Table A2.1). The main focus lies on trait information aggregated at the species level, making GIFT a valuable complementary resource to initiatives that collate large amounts of trait measurements at the individual level (e.g. Kattge *et al.*, 2011a).

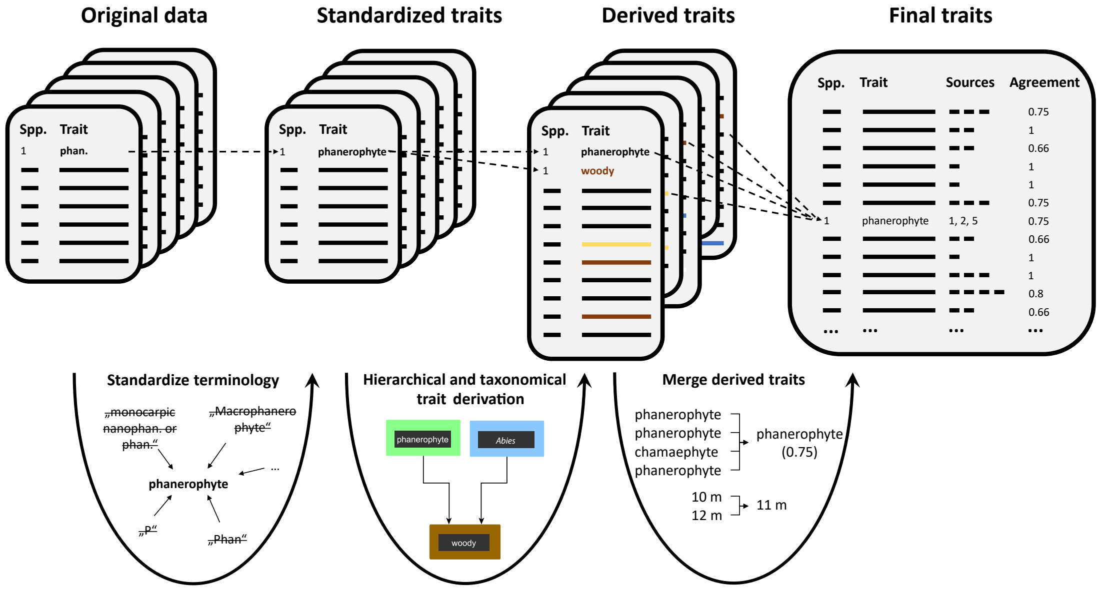


Figure 2.3: Trait processing in GIFT. Original trait records entering GIFT are subjected to three processing steps: (1) Trait values are standardized with respect to language, terminology and measurement unit. (2) Additional trait values are derived hierarchically for traits that are logically nested (Figure A1.2) and taxonomically for species that belong to taxonomic groups that are uniform with respect to a particular trait. (3) Derived trait values are aggregated at the species level based on the consensus among resources (categorical traits) or summary statistics are computed based on the original values (numerical traits).

Many trait resources provide equivalent information in various languages, using different terminologies or measurement units. The first step of trait processing in GIFT (Figure 2.3) is therefore the standardization of primary trait data according to pertinent trait literature

(Pérez-Harguindeguy *et al.*, 2013; Garnier *et al.*, 2017) using defined categorical levels (categorical traits) and units of measurement (numerical traits) (Table A2.1). To retain the maximum information provided by the original resources, many categorical traits are defined at multiple levels of detail (e.g. life form 1 and life form 2, Figure A1.2). In the second step of trait processing, the standardized trait values are subjected to a hierarchical derivation procedure (Figure 2.3). This procedure makes use of logical links and nestedness among many functional traits based on their definitions. For example, the value "tree" in the trait growth form implies the value "woody" in the trait woodiness. As such, hierarchical trait derivation increases the amount of trait information in GIFT. In addition, hierarchical derivation ensures compatibility among different levels of detail of the same trait by automatically deriving values in coarser variations from available information in more detailed ones. We organize such hierarchical relationships between traits in a directed graph that can be easily traversed to fill data gaps (Figure A1.2) A tabular version of the graph is stored in the database (Table A2.2). A similar derivation approach is implemented for taxonomic groups that are uniform with respect to a particular trait (taxonomic trait derivation). In this case, the basis of derivation is not the logical hierarchy of trait values, but that of taxonomic groups. For example, the genus *Abies* consists exclusively of monoecious trees (Farjon, 2010). Thus, all species of *Abies* can be characterized with respect to growth form "tree" and reproductive mode "monoecious" based only on their taxonomic position. Subsequently, the taxonomically derived species-level traits are subjected to the hierarchical derivation as outlined above.

To obtain a single, standardized value per trait and species, original and derived trait values are aggregated based on the consensus among resources (categorical variables, 66 % consensus threshold) or summary statistics are computed based on the original values (numerical variables, currently: mean, minimum and maximum) (Figure 2.3). Trait derivation and aggregation for a given species are repeated each time a new trait record enters the database, such that the final values and proportions of supporting resources are continually re-evaluated in the light of new information. Throughout the entire procedure of trait processing, information can be traced back to their original reference and unstandardized value. This provides a basis for implementing advanced gap-filling (Schrodt *et al.*, 2015) and aggregation techniques (Kattge *et al.*, 2011b) in the future.

2.3.6 Geographic regions

Each regional species checklist in GIFT unambiguously refers to a geographic region, e.g. an island, archipelago, political or biogeographical unit, or protected area. Spatial polygons come from the Biodiversity Information Standards Working Group (TDWG, 2007), the GADM database of Global Administrative Areas (Hijmans *et al.*, 2009), single island polygons extracted from GADM (see Weigelt *et al.*, 2013), the Global Island Database (UNEP-WCMC, 2013), the World Database on Protected Areas (UNEP-WCMC, 2014), or are digitized manually according to the checklist literature resources. The regions vary in area by

13 orders of magnitude, ranging from tiny islands to large countries and botanical continents (Figure 2.4). Regions may be nested or overlap with each other. The degree of overlap is calculated automatically and recorded in the database to allow spatial aggregation of information or the exclusion of overlapping entities analyses.

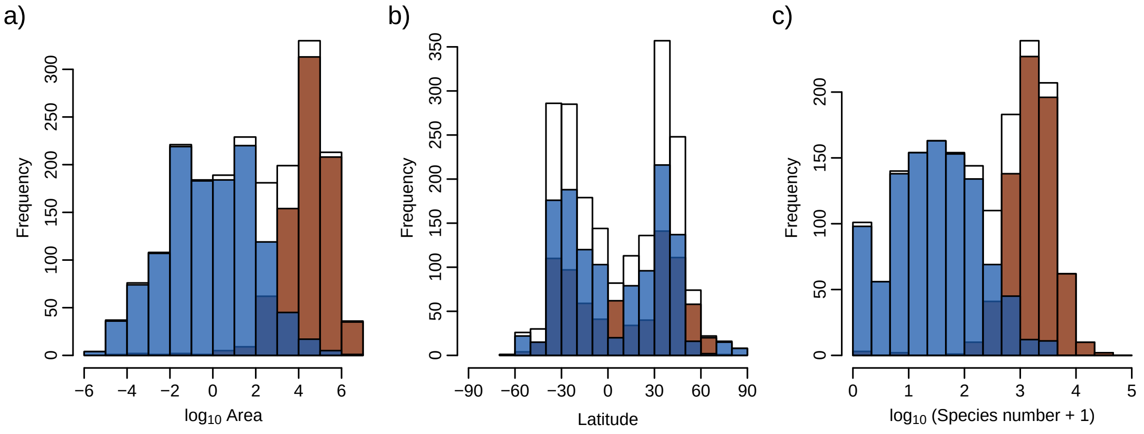


Figure 2.4: Frequency distributions of 2007 geographic regions in GIFT with information on native vascular plant species composition and spatial properties for a) region area (km^2), b) latitude and c) species richness of native vascular plants. Blue bars = islands, brown = mainland regions, white = total, dark blue = overlap of mainland and island bars.

For each geographic region, a suite of currently 123 physical geographic, bioclimatic and socio-economic characteristics is computed for macroecological analyses based on the regions' spatial information and additional spatial datasets (Table A2.3). Specifically, this includes (1) characteristics based on the spatial polygon itself like its area, centroid coordinates and geographic extent, (2) summary statistics (15 quantiles including minimum, median and maximum, mean, standard deviation, mode, number of unique values, Shannon diversity and number of cells) derived from raster layers like digital elevation models (Danielson & Gesch, 2011), global climate layers (e.g. CHELSA; Karger *et al.*, 2017) or human population density (Doxsey-Whitfield *et al.*, 2015), and (3) miscellaneous metrics calculated from additional spatial resources like biogeographic region affinity (Takhtajan, 1986) or island isolation (Weigelt & Kreft, 2013) (Figure A2.1c).

For families and higher taxonomic groups, we automatically calculate the number of all species, native species, naturalized species and endemic species for all regions that are covered by checklists for the given combination of taxonomic group and floristics status. Additionally, we calculate for the same taxonomic groups and floristic subsets the percent trait coverage for all functional traits covered by the database. It is hence possible to extract and visualize for which regions and taxa what information in terms of species checklists and functional traits is available. Based on the various checklist resources available for each geographic region, we decide whether the checklist information should theoretically completely cover the whole native vascular or angiosperm flora. This, however, is a rough and subjective estimate and given the huge amount of unexplored plant diversity especially in the Tropics and only partial completeness of the according Floras, it needs further evaluation. To this end, we are currently incorporating species numbers and richness estimates from the literature (Frodin, 2001; Kreft & Jetz, 2007) into the database and develop workflows to compare them to the species numbers derived from the checklists. Regions deviating considerably from these literature values can be excluded from analyses if needed. The same way regional trait coverage can be used to exclude regions with little trait information from analyses on trait patterns.

2.3.7 Versioning

This paper describes GIFT version 1.0. New data is incorporated in chunks and each time new data is added or workflows are modified, a new version is released. Changes will be documented at http://gift.uni-goettingen.de/about. Old versions are backed up and can be restored to reproduce analyses carried out on old versions of the database.

2.4 Current state

2.4.1 Geographic coverage

Initially, GIFT started with the collection of Floras and checklists for oceanic islands and the basic workflows have been developed for various projects focusing on island plant diversity (Weigelt *et al.*, 2015; Weigelt *et al.*, 2016). Island floras usually host a comparatively limited set of species and have clearly defined geographic boundaries. As such, they have attracted a lot of scientific interest in the past, leading to a high availability of island Floras and checklists. GIFT therefore offers a very comprehensive overview over the floristic composition of 1,845 of the world's islands (Figure 2.4), which has already led to a variety of studies on island biodiversity patterns and their determinants (e.g. Weigelt *et al.*, 2013; Cabral *et al.*, 2014; Weigelt *et al.*, 2015; Weigelt *et al.*, 2016; Lenzner *et al.*, 2017). More recently, GIFT has been expanded to cover 1,048 mainland regions, allowing for comparative analyses of continental and insular floras (König *et al.*, 2017) and exhaustive studies of global plant diversity.

Table 2.1: Current coverage of GIFT for selected major plant groups in terms of number of regions with supposedly full inventories for native species, unstandardized taxonomic names, standardized species names, species with resolved taxonomy, and trait records.

Taxonomic group	Regions	Names	Species	Species resolved	Trait records
Embryophyta	53	717117	324136	277580	2307100
Tracheophyta	2062	714781	322002	275610	2306973
Pteridophyta	2079	24241	11888	8408	54772
Gymnospermae	2211	4031	1151	1051	12352
Angiospermae	2218	686509	308963	266151	2239849
Orchidaceae	2478	64508	28155	27029	192332
Asteraceae	2218	58492	27755	24450	167300
Fabaceae	2218	46999	21000	18416	145895
Poaceae	2431	38464	12368	11215	130492
Rubiaceae	2431	29485	14260	13545	96684
Lamiaceae	2431	18120	7882	7560	61708

In total, GIFT currently includes 2,893 geographic regions covering the whole globe. Of those regions, 2,062 have at least one checklist for all native vascular plants, together covering all floristic kingdoms and biomes and 79.1 % of the earth's land surface excluding Antarctica (Figure 2.5a). After removing overlapping entities to avoid pseudo-replication, up to 1,841 regions and 58.2 % land surface coverage remain when prioritizing small entities (> 100 km²) over large entities, and 1,555 regions and 73.1 % land surface coverage remain when prioritizing large entities over small entities (single islands always prioritized over island groups). Geographic coverage varies with focal taxonomic group (Figure 2.5, Table 2.1) and floristic status (Figure A2.2), and is highest for native species. Largest gaps for native vascular plant floras are currently located in Tropical Africa, the Middle East, Central Eurasia, and South East Asia (Figure 2.5). Data gaps in GIFT do not necessarily represent true knowledge gaps. Floras of the countries of the former USSR, West Africa, Madagascar, Java and India, for example, are available and are currently in the process of being incorporated.

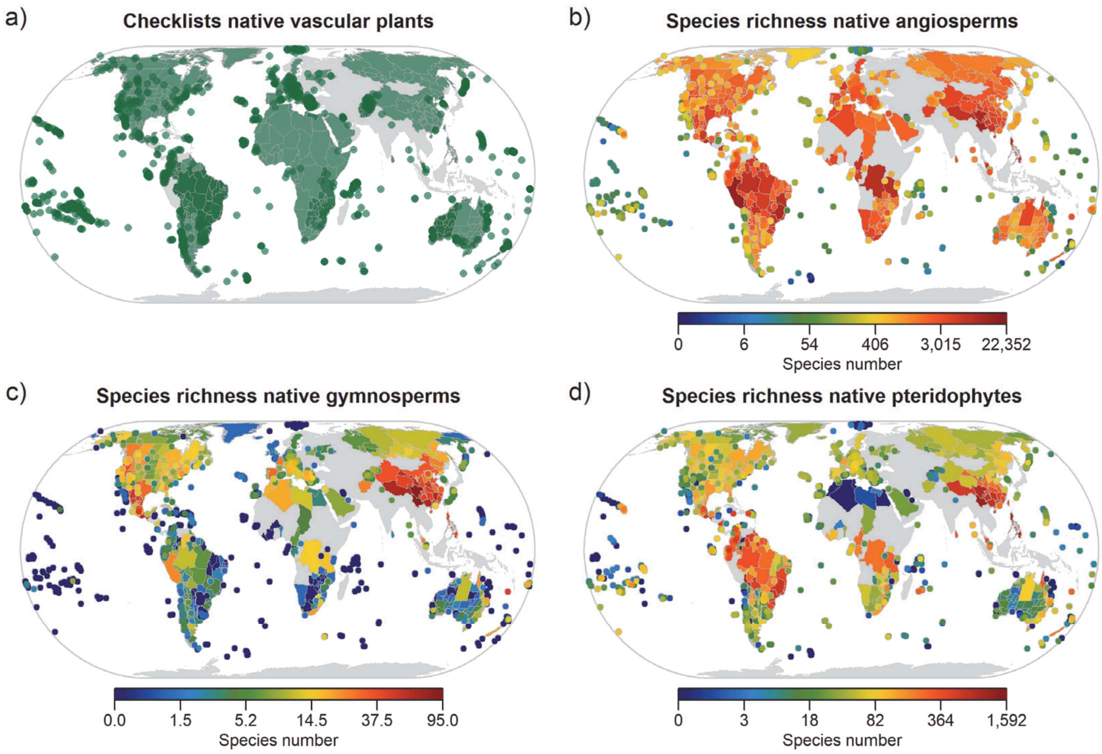


Figure 2.5: Spatial coverage of checklist data currently stored in GIFT. a) Regions with checklist data for native vascular plants. Darker green shade indicates overlapping regions. b-d) Checklist coverage and species richness of major taxonomic groups for regions with theoretically complete inventories. Regions <25,000 km² are plotted as points.

Since many Floras refer to entire countries of various sizes, and some of the resources in GIFT use broad distributional classifications (e.g. WCSP, 2014; BGCI, 2017), most mainland regions in GIFT are relatively large, especially in comparison to an average island (Figure 2.4). However, we aim to also include mainland regions of small sizes like protected areas and small political units, since smaller units span smaller environmental gradients, and should thus provide a tighter link between taxonomic, functional and phylogenetic species composition and aggregated abiotic conditions (Pearson & Dawson, 2003).

2.4.2 Taxonomic coverage

GIFT currently includes 324,136 taxonomically standardized species from all major groups of land plants (Embryophyta), 277,580 of which are resolved to accepted species names. The focus for the collection of species lists and traits lies on vascular plants (Tracheophyta, 322,002 species) and in particular on angiosperms (Angiospermae; 308,963 species; Table 2.1). On average, 79.3% of all accepted species per plant family according to TPL are covered by distribution data. Taxonomic coverage of distribution data does not show a significant phylogenetic signal (Abouheif's Cmean = 0.03, p = 0.142; Abouheif, 1999), i.e. it exhibits no detectable taxonomic bias (Figure 2.6). The 324,136 species names in total derive from 717,117 unstandardized original names (after genus name correction and exclusion of hybrid names) that differ in spelling or in the availability of author names or infraspecific information. 98.2 % of all original names could be matched and standardized to an existing species name using our approach to match TPL or using the TNRS API. For 90.5 % of all names, the synonymy could be resolved. Only 3.6 % of all working names are names that were adopted unchanged from the original names because they could not be adequately matched to taxonomic resources.

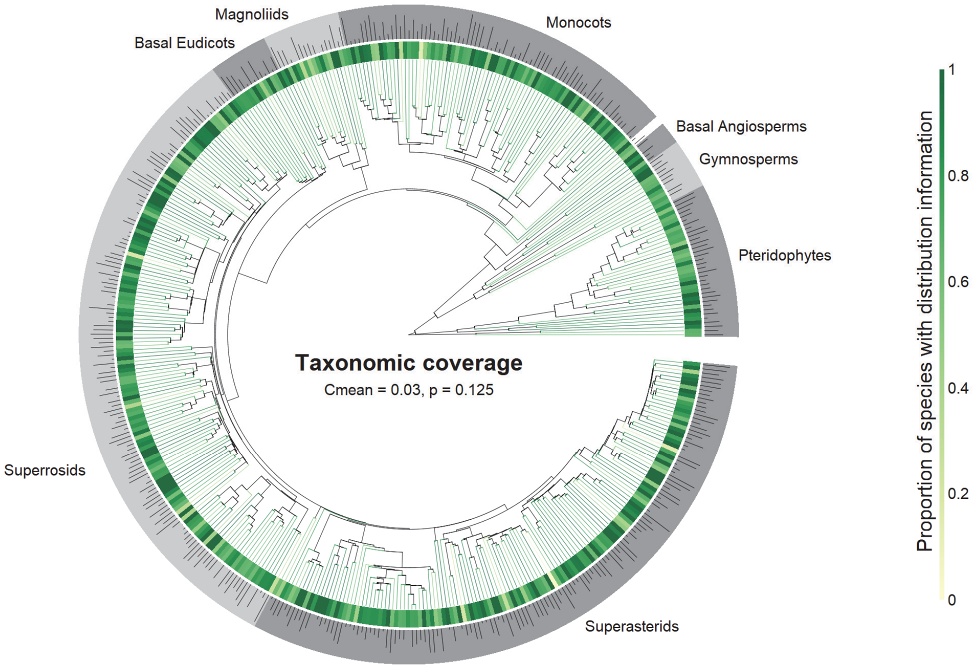


Figure 2.6: Taxonomic coverage of distribution data in GIFT at the family level. Tip color and inner ring color indicate the proportion of species with distribution information relative to all species of a given family, the grey outer ring delimits major clades of vascular plants. The height of bars in the outer ring is proportional to log10 total family species richness. Phylogenetic signal in taxonomic coverage was assessed as Abouheif's Cmean, a measure of phylogenetic autocorrelation based on the sum of the successive squared differences between values of neighbouring tips in the phylogeny (Abouheif, 1999).

2.4.3 Trait coverage

In total, there are 3,475,337 original trait records referring to 550,892 original taxon names. Hierarchical trait derivation yields an additional 1,261,718 trait records. After aggregating original and derived trait records, i.e. resolving species names and combining trait records for identical species, 2,307,100 species-trait combinations for 267,978 standardized species remain for ecological analyses (Table 2.1).

The majority of trait information in GIFT refers to morphological characteristics such as woodiness (234,214 species) climbing habit (223,280 species), or growth form (213,372 species)(Table A2.1). Life history traits such as life form (100,607 species) or life cycle (84,206 species) are the second most common trait category. Other categorical traits are considerably rarer, e.g. photosynthetic pathway (31,534 species), dispersal syndrome (8,204 species), or pollination syndrome (4,511 species). Also quantitative traits such as maximum plant height (53,449 species), mean seed mass (23,874 species) or mean specific leaf area (2,304 species) are comparatively poorly covered.

To illustrate patterns in the geographic and taxonomic trait coverage of GIFT, we use the overall coverage across all traits as well as four exemplary traits (growth form, plant height, life form and seed mass). Geographically, most trait information per species is available in Europe and some comparatively species-poor temperate islands (Figure 2.7a). Also, non-tropical parts of the Americas, Africa and Australia are well covered, whereas tropical regions in Africa and South-East Asia are least well covered with respect to their plant functional characteristics. However, geographic coverage varies strongly among individual functional traits. Frequent traits such as growth form are available for most species in almost every floristic region, whereas the coverage of comparatively less well-covered traits is strongly dependent on the geographic scope of the main contributing resources (Figure 2.7). Life form *sensu* Raunkiær (1934), for example, is widely available throughout Europe but rarely reported for species in other regions of the world (Figure 2.7d). Likewise, plant height and seed mass exhibit uneven geographical coverage distributions, with highest coverage in Australia, South Africa, and Europe (Figure 2.7c&e).

The taxonomic coverage of trait information in GIFT bears little, though significant, phylogenetic signal overall ($C_{mean} = 0.21$, $p < 0.001$, Figure A2.3a), but reveals interesting patterns when examined at the level of individual traits (Figure A2.3b-e). For example, plant height is very well covered for the graminid clade (leftmost group within the monocots, Figure A2.3c), and Raunkiær life form is particularly well covered in gymnosperms and monocots (Figure A2.3d).

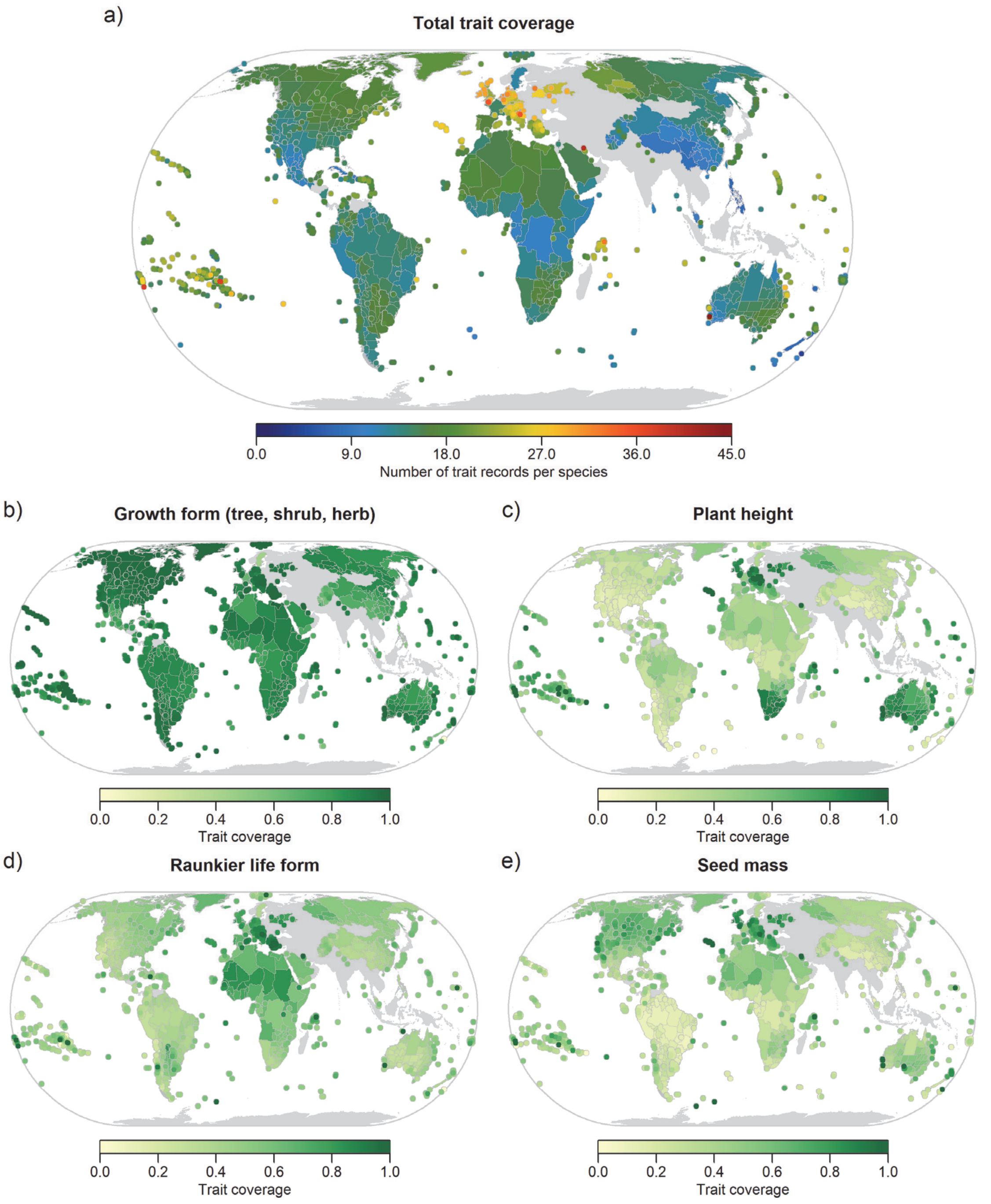


Figure 2.7: Geographical trait coverage of GIFT. (a) The total number of trait records per native angiosperm species per region and (b-e) trait coverage per region (number of native angiosperm species with trait information/number of all native angiosperm species) for exemplary traits with characteristic geographic patterns in coverage. Regions <25,000 km² are plotted as points.

2.4.4 Web interface

An overview of the current content of GIFT is available through a web interface at http://gift.uni-goettingen.de. It provides summary statistics and allows producing customized richness and trait coverage maps for every combination of taxonomic group and floristic subset based on the species numbers and trait coverage values in the database. It is possible to see for which regions and taxa what information in terms of species checklists and functional traits is available and to browse the bibliographic references.

2.5 Applications and outlook

GIFT takes advantage of the wealth and quality of curated Floras and plant checklists, integrated with diverse information on species characteristics and their environment. The integration of these data and the workflows described above allow exploring and testing macroecological and biogeographical hypotheses and improving existing distribution and trait datasets (Figure 2.1). The combination of floristic composition, environmental properties, functional traits and phylogenetic information allows moving beyond species richness as a proxy for biodiversity (Barthlott *et al.*, 2005; Kreft & Jetz, 2007; Kreft *et al.*, 2008) and considering other facets of diversity. Examining the drivers of taxonomic, functional and phylogenetic diversity and turnover (Qian *et al.*, 2013; Lamanna *et al.*, 2014; Weigelt *et al.*, 2015; König *et al.*, 2017), for example, may help to disentangle the mechanisms underlying global plant diversity more directly (Graham *et al.*, 2014).

The integration of plant distribution data and functional traits in GIFT opens up new avenues in the emerging field of functional biogeography. Functional biogeography combines the mechanistic focus of functional ecology with the large eco-evolutionary scales of biogeography (Violle *et al.*, 2014) and thus provides a direct link between measures of organismal performance and a wide range of abiotic and biotic conditions. Although functional biogeographical approaches already provided significant insights into patterns and drivers of functional diversity (Moles *et al.*, 2014; Reichstein *et al.*, 2014; Engemann *et al.*, 2016; Butler *et al.*, 2017), the availability and representativeness of data on plant traits and distributions remains a limiting factor. Together with distribution and floristic status information available in GIFT, functional traits may help to better understand the biogeographic history of plant life on Earth and its anthropogenic stressors. Analyses of endemic species and their traits, for example, can shed light on the evolution of new species and their contribution to current biogeographic patterns (Weigelt *et al.*, 2016). Naturalized alien species and their traits help to understand the role of humans in changing plant assemblages and may teach us how new habitats and regions are colonized. Knowledge on the composition of native vs. alien floras (see www.glonaf.org, van Kleunen *et al.*, 2015; Pyšek *et al.*, 2017) allows to tackle pressing questions in invasion ecology, for example what native floras are susceptible to plant invasions and how regional plant composition changes due to the naturalization of alien species (Winter *et al.*, 2009).

Apart from direct use as data source for macroecological or biogeographical research, GIFT is also a valuable resource to validate or expand other distribution or trait datasets (Figure 2.1). Having near-global and full taxonomic coverage of distribution data (Figure 2.5) and several functional traits (Table A2.1), GIFT can help to assess the representativeness of macroecological datasets and to overcome data limitations to find answers to fundamental questions in functional biogeography and macroecology (Scheffer *et al.*, 2014; FitzJohn *et al.*, 2014, see also case studies in Chapter 1). It may for example help to estimate whether data from resources like GBIF or TRY are sufficiently complete or representative for analyses of a given taxon, region or functional group. Alternatively, GIFT can also be used to infer the floristic status of plant point occurrences (e.g. to tell apart native and non-native species), to identify unlikely or dubious occurrences or to infer local species absences. The latter may be particularly useful for species distribution modelling where random pseudo-absences are commonly used when true absences are not known (Lobo *et al.*, 2010; Barbet-Massin *et al.*, 2012). Furthermore, GIFT can be used to define regional species pools of local plant communities (Karger *et al.*, 2016), for example, for identifying likely source regions of species that colonize oceanic islands (see Chapter 4). Defining the regional species pool or inferring the floristic status may not only be important for macroecological studies, but also for field projects at the local to regional scale.

The mid-term goal of GIFT is to reach full global coverage of vascular plant checklists. Already now, 79.1 % of the global land surface is covered and further Floras and checklists covering missing parts are currently processed. Realistically, GIFT will reach about 90% spatial coverage by the end of 2018 and will serve as a representative resource for analyses of global plant diversity. In the meantime, regions already covered by coarse geographical units will be complemented by finer-scale data, and new literature resources will be included to update outdated checklists. Once the availability of checklists per region has further increased, workflows to spatially aggregate them will be developed. This will include the identification of conflicting information and choice of the best and most up-to-date information as well as derivation of the floristic status from small to large regions and vice-versa.

A major challenge regards the evaluation of checklist quality and completeness in GIFT. The species richness data sets currently being included allow for a comparison of expected and observed species numbers, but also the integration of other data like, for example, point occurrence information as provided by GBIF or vegetation plot data may help to estimate completeness of the regional checklists in GIFT and eventually update them. Furthermore, the lack of cosmopolitan or regionally common species in checklists, an uneven representation of expected higher taxa, or deviances from expected ecological relationships like, for example, the species area relationship or the latitudinal diversity gradient may be used to flag potentially incomplete checklists (Santos *et al.*, 2010). Regions with incomplete checklists can then be excluded from analyses or survey effort can be included in statistical models and data acquisition can be prioritized for those regions.

In conclusion, GIFT offers a novel integrated database framework to study the geographic distribution of plant life across the globe. The integration of regional plant checklists with functional traits, phylogenetic relationships and regional environmental characteristics allows for a swift extraction of macroecological datasets for hypothesis testing and the validation and extension of alternative resources. In addition, the outlined database framework can serve as an example for other taxa with insufficiently complete information at the level of individual species and for an integration of comparable data types such as vegetation plots or surveys. The spatially nested structure of regions in GIFT allows for an ongoing inclusion of resources to improve inventory quality and spatial resolution in future database releases.

3 Dissecting global turnover in vascular plants

Christian König, Patrick Weigelt and Holger Kreft

3.1 Abstract

Aim: To provide a global assessment of compositional turnover in vascular plants across geographic settings, taxa, and functional groups. We tested whether turnover and its spatial and environmental drivers are affected by the geographical setting and whether taxonomic and functional groups exhibit specific turnover patterns that are associated with their ecological characteristics.

Location: Global.

Methods: We collated a global dataset of vascular plant checklists comprising 258 island and 346 mainland units. We created subsets based on the geographical setting of study units (mainland, islands, different island types) as well as taxonomic and functional properties of species (angiosperms, gymnosperms, pteridophytes, trees, shrubs, herbs). For the entire dataset and each subset, the distance decay of similarity was assessed using generalized linear models. To disentangle the relative importance of spatial and environmental drivers of turnover, we employed generalized dissimilarity models. Finally, the model results were used to predict compositional similarity of vascular plants across a global grid.

Results: The distance decay of similarity was stronger for mainland units than for islands. Among taxonomic and functional groups, the rate of decay was lowest for pteridophytes and highest for shrubs. Partitioning of turnover into distance- and environment-related effects revealed fundamental differences between mainland and island systems, with geographic distance being more important on the mainland than on islands. This trend was consistent across taxonomic and functional groups.

Main conclusions: Our results reveal the important role of geographical context in shaping beta diversity patterns. We argue that geographical settings are characterized by specific configurations of ecological filters that strongly impact the magnitude and structure of turnover. Moreover, taxonomic and functional groups are differentially successful in passing these filters, resulting in group- and setting-specific turnover patterns. Exploring these interdependencies for different taxa and geographical settings at different scales will help to improve our understanding of beta diversity.

3.2 Introduction

Beta diversity, the variation in community composition among sites (Whittaker, 1960), is a central concept in ecology and biogeography as it provides a key link for understanding the relationships between species and their environment. Three main explanations for the origin of beta diversity have been proposed: 1) the uniformity hypothesis (Pitman *et al.*, 1999), stating that compositional variation arises as sampling artifact due to different local abundances of species; 2) the neutrality hypothesis (Hubbell, 2001), stating that compositional variation is created through stochastic, spatially limited dispersal, and speciation; and 3) the niche-assembly hypothesis (Whittaker, 1956; Hutchinson, 1957), stating that environmental factors determine the presence or absence of species and hence the compositional variation among sites. While the uniformity hypothesis predicts overall low beta diversity and has received little empirical support, numerous studies have contrasted the role of neutral vs. niche-based processes (Condit *et al.*, 2002; Fargione *et al.*, 2003; Leibold & McPeek, 2006; Baldeck *et al.*, 2013). It is generally asserted that a strong spatial signal in community similarity indicates a strong role of dispersal limitation, thus supporting the neutrality hypothesis, whereas a strong environmental signal indicates strong habitat filtering, thus supporting the niche-assembly hypothesis (Legendre *et al.*, 2005). Importantly, both hypotheses are not mutually exclusive and neutral and niche-based process may act jointly in natural systems (Chase, 2014). The challenge in understanding beta diversity is therefore not to find a single universal explanation for all observable phenomena, but to evaluate the relative role of individual drivers and processes at different scales and in varying contexts.

Perhaps the most widely used approach to examine determinants of beta diversity is to track the change in compositional similarity along gradients of geographical or environmental separation. The rate at which compositional similarity decreases along such a gradient indicates the importance of the considered factor for species composition (Tuomisto *et al.*, 2003; Soininen *et al.*, 2007). Geographic distance is most frequently chosen as predictor of compositional similarity (distance decay of similarity, Nekola & White, 1999) because it can be measured easily and ecological theory provides mechanistic explanations, e.g. spatially structured speciation or dispersal patterns (Hubbell, 2001), upon which testable hypotheses can be formulated. However, as Nekola & White (1999) pointed out, not only compositional but also environmental similarity is negatively correlated with geographic distance, making it difficult to disentangle the unique contributions of spatial and environmental factors to beta diversity. Moreover and unlike geographic distance, there is no canonical measure of environmental distance because relevant target variables are both scale- and taxon-dependent and differ in their relative importance. Several methods have been proposed to address these issues including multiple regression on distance matrices (Manly, 1986), variation partitioning (Borcard *et al.*, 1992), and generalized dissimilarity modelling (Ferrier *et al.*, 2007).

The diverse approaches to analyzing beta diversity are often applied to different subsets such as taxonomic groups (Tuomisto *et al.*, 2003; Ruokolainen *et al.*, 2007), functional groups (Bin *et al.*, 2010; López-Martínez *et al.*, 2013), or geographical settings (Fattorini, 2010; Stuart *et al.*, 2012), where each subset serves as an independent test case under different conditions. Comparing species groups with differential dispersal abilities, for example, allows inference about the strength of dispersal-related (distance-dependent) determinants of beta diversity. Contrasting geographical settings with distinct properties, on the other hand, may yield insights on the relative importance of assembly processes under different environmental conditions. However, studies on beta diversity that take advantage of species-level traits as well as environmental factors are profoundly rare.

Here, we present a global analysis of vascular plant beta diversity using a large floristic data set of 604 checklists comprising 149,475 species. We focused on the turnover (i.e. richness-insensitive) component of beta diversity and examined different geographic, taxonomic, and functional subsets in order to identify the factors determining turnover at global scales. To assess the general strength of turnover, we analyzed the distance decay of similarity in each subset. We then applied generalized dissimilarity models with an additional set of environmental predictors to disentangle the relative importance of spatial and environmental drivers, and to predict floristic similarity across a global equal-area grid. We hypothesize that turnover and the importance of spatial and environmental variables vary (1) across geographic settings, indicating that different geographical settings impose distinctive combinations of dispersal and environmental filters, and (2) across taxonomic and functional groups, indicating that groups are differentially successful in passing these filters. In particular, we expect geographic distance (i.e. dispersal filtering) to be the predominant driver in island systems and environmental conditions (i.e. environmental filtering) to be predominant drivers in mainland systems. On the other hand, we expect groups with, on average, good dispersal abilities to be less sensitive to geographic distance. Likewise, groups composed of species with large environmental tolerances should be less sensitive to environmental distances.

3.3 Methods

3.3.1 Species data

Building on previous work of Weigelt *et al.* (2015), we assembled vascular plant checklists from published floras, checklists and online databases (see Appendix 1 – data sources). We standardized species names and higher taxonomic ranks according to The Plant List (2013) and, for species not listed in The Plant List, using iPlant's taxonomic name resolution service (Boyle *et al.*, 2013). Our inclusion criteria for checklists were that (1) the checklist was sufficiently complete, i.e. the reported species number fell into a reasonable range for the given biome and area (Kreft & Jetz, 2007), (2) the floristic status (native vs. alien) of all species was documented, (3) the checklist contained more than 30 native vascular plant species to allow robust statistical analyses, and (4) the corresponding geographical unit could be clearly defined and had a size between 1 and 500,000 km^2. To increase comparability among and environmental consistency within operational geographical units (hereafter OGUs), we focused on collecting mainland checklists for relatively small regions such as protected areas or federal states. The final dataset (see Table A3.2) consisted of 258 island and 346 mainland OGUs containing a total of 149,475 native plant species (all aliens were excluded) and 771,403 species records (Figure A3.1).

We collected morphology-related trait information from the original checklists and additionally large trait databases (see Table A3.2). Since terminology of trait information differed greatly across resources, we translated the original descriptions into a hierarchical system of growth- and lifeform traits with defined values for each level (Figure A3.2). We then used values from highly resolved levels to deduce missing values for coarser levels. Where resources provided conflicting trait values for a given species (less than 5 % of cases), we used a threshold of at least 66 % agreement to assign a final value to the species. Although more detailed traits such as life form *sensu* Raunkiær (1934) are ecologically more informative, we chose growth form (values: herb, shrub and tree) as target variable in all trait-related analyses, because it represented the best compromise between data coverage and ecological relevance. This way, we could assign growth forms to a total 102,809 out of 149,475 species and achieved an average coverage of 81.9 % and 87.1 % for island and mainland OGUs, respectively.

3.3.2 Abiotic data

Spatial polygons for OGUs were assembled from the GADM database of Global Administrative Areas (Hijmans *et al.*, 2009) for political units and islands, and from the World Database of Protected Areas (UNEP-WCMC, 2014) for protected areas. Some OGUs were digitized manually on the basis of information given in the respective publication (maps, coordinates). The geographic distance between OGUs was calculated as the distance between the polygon mass centroids on a sphere with the earth's radius. Environmental data were extracted from 30 arc-seconds WorldClim raster layers (Hijmans *et al.*, 2005) for annual mean temperature, annual mean precipitation, temperature seasonality, precipitation seasonality and elevation. These variables have been shown to capture major ecologically relevant axes of environmental space (Moser *et al.*, 2005; Kreft & Jetz, 2007; Weigelt & Kreft, 2013). Because of the skewed distribution of raster cell values, we did not use the mean but the median of all cells overlapping with an OGU as environmental predictors. Pairwise correlation plots and variance inflation factors did not indicate problematic levels of multicollinearity among the predictor variables (Figure A3.3). Island geology, distinguishing between continental islands (i.e. shelf islands and continental fragments), oceanic islands (i.e. islands of volcanic origin or uplifted by tectonic processes), and atolls (i.e. flat, ring-shaped coral islands) was ascertained through extensive literature research.

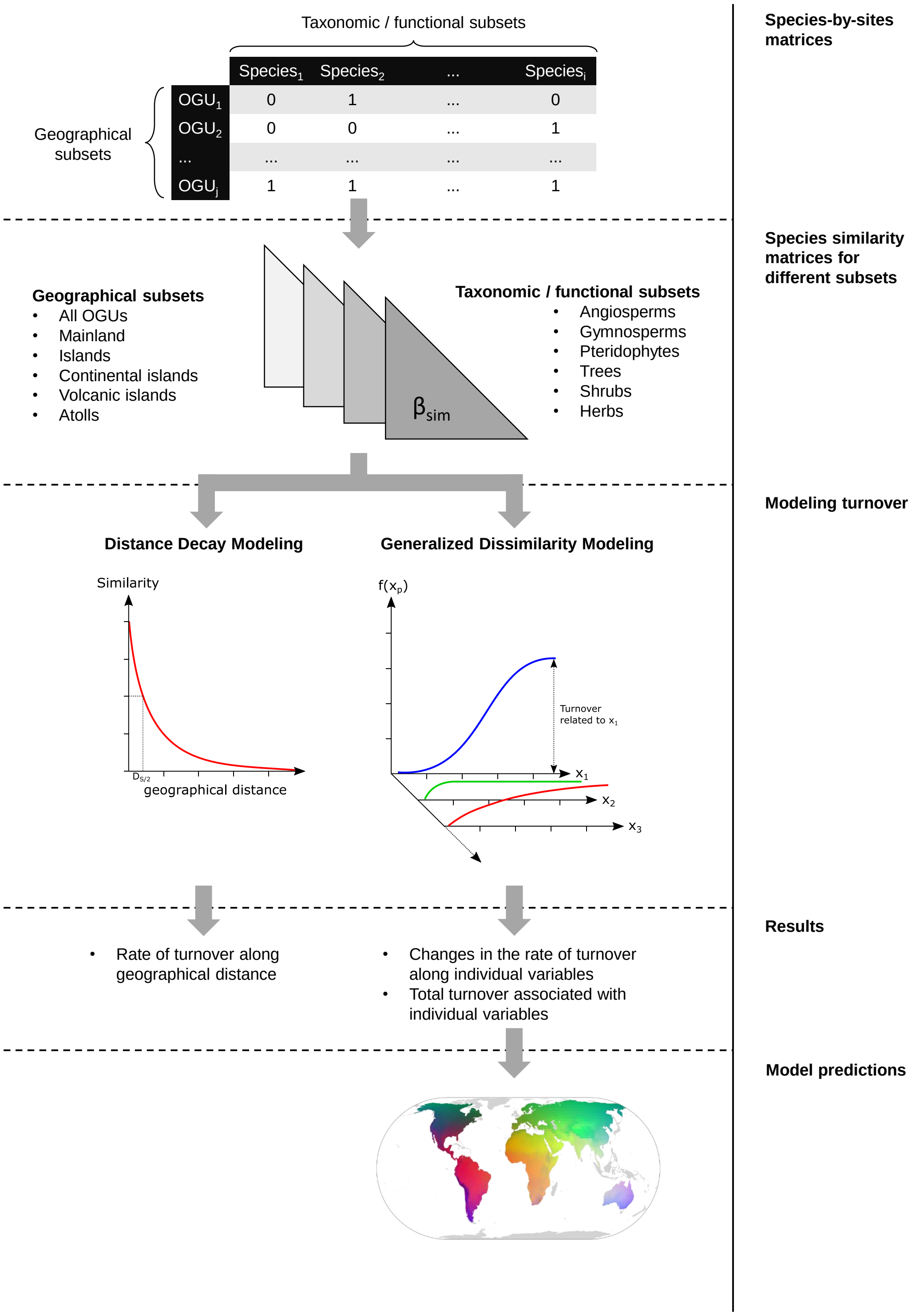
3.3.3 Compositional similarity

Beta diversity describes two independent sources of compositional variation: The replacement or *turnover* component, reflecting the amount of change in species identities among study sizes, and the *nestedness* component, reflecting the compositional variation owing to differences in species numbers (Lennon *et al.*, 2001; Baselga, 2010; Legendre, 2014). We quantified compositional variation using the β_{sim}-index which is insensitive to richness differences and thus only quantifies the turnover component of beta diversity (Baselga, 2010). This is a crucial property when comparing units of unequal area, as we do here, because of the strong dependence of species number on area (Arrhenius, 1921; Connor & McCoy, 1979). The index is defined as $\beta_{sim} = \frac{\min(b,c)}{\min(b,c)+a}$ where a is the number of species common to both comparison units and b and c are the number of species unique to the first and second unit, respectively.

We prepared a binary species-by-sites matrix containing all species and all OGUs. Based on geographical, taxonomic and functional criteria, we produced different subsets of the full matrix (Figure 3.1). We derived row-wise subsets for mainland units, islands, and particular island types as well as column-wise subsets for angiosperms, gymnosperms, pteridophytes, trees, shrubs, and herbs. We also produced separate island- and mainland subsets for all taxonomic and functional groups. For each subset, we computed separate β_{sim}-similarity matrices. If an OGU had less than 10 species of a particular taxonomic or functional subset, it was excluded from the respective similarity matrix to avoid bias emerging from low sample size. Moreover, we excluded OGUs from the functional subsets when the growth form was known for less than 80 % of the species.

3.3.4 Analysis

Turnover was assessed in a two-step approach (Figure 3.1). First, we were interested in the general pattern of turnover for the different geographical, taxonomic and functional subsets and applied single-predictor log-binomial generalized linear models (GLM), producing a best-fit decay curve of compositional similarity against geographical distance. The intercept was fixed at 1, assuming complete similarity at 0 km distance. Following Soininen *et al.* (2007), we used the *halving distance* ($D_{S/2}$), i.e. the distance after which a given similarity value is predicted to decrease by 50 percent, as a measure of the turnover rate. Due to the constant rate of decay assumed in the model, the halving distance holds as descriptor of turnover per unit distance over the whole gradient. Acknowledging that the non-independence of observations in similarity matrices leads to correct parameter estimates but invalid test statistics (Lichstein, 2007), we estimated confidence intervals for the GLMs using a bootstrapping approach. We randomly excluded 50 % of OGUs from each similarity matrix and used the 2.5 and 97.5 percentiles of the sample distribution of estimated coefficients from 250 individual runs.



	$Species_1$	$Species_2$	...	$Species_i$
OGU_1	0	1	...	0
OGU_2	0	0	...	1
...	...	...	...	...
OGU_j	1	1	...	1

Figure 3.1: Framework of this study for analyzing global turnover of vascular plants. Subsets of the species-by-sites matrix were produced on the basis of geographical, taxonomic and functional criteria. For each subset, turnover along geographical and environmental gradients was assessed using single predictor log-binomial generalized linear models and multi-predictor generalized dissimilarity models (GDM). Abbreviations: $D_{S/2}$, distance after which similarity decreases by 50% (halving distance); $x_{1...p}$, GDM predictor variables; $f(x_{1...p})$, GDM transformation functions for $x_{1...p}$ that maximize the deviance explained by full model; OGU, operational geographical unit.

In the second part of the analysis, we applied generalized dissimilarity modelling (GDM) (Ferrier *et al.*, 2007) to investigate spatial and environmental drivers of turnover. GDM is specifically designed to analyze and predict large-scale patterns in beta diversity with respect to a set of explanatory variables. The model accounts (1) for variations in the rate of turnover along individual variables and (2) for the curvilinear relationship between compositional similarity and gradients of geographic or environmental separation (Ferrier *et al.*, 2007). This is achieved by producing a set of monotonically increasing I-spline basis functions for each predictor and reformulating pairwise differences among sites as the differences along the I-spline basis functions. Coefficients for each basis function are calculated using maximum likelihood estimation in an inverse log-binomial GLM with compositional similarity as response and the set of reformulated distances as predictors. The sum of fitted I-spline basis functions per predictor yields the final transformation function, representing the best supported relationship between observed compositional similarity and pairwise distances along the considered predictor (Fitzpatrick *et al.*, 2013). The transformation functions provide two important pieces of information (Figure 3.1): First, the total height of the curve serves as a proxy for the turnover associated with the respective environmental variable when all other variables are held constant. Second, the shape of the curve reflects the variation in the rate of turnover along the considered variable (Ferrier *et al.*, 2007).

Generalized dissimilarity models for each subset were fitted with the R-package *gdm* (Manion *et al.*, 2015), using a geographical distance matrix and the untransformed vectors of mean annual temperature, annual precipitation, temperature seasonality, precipitation seasonality, area and elevation as predictor variables of compositional similarity. We used the default setting of three I-spline basis functions per predictor. Significance testing of variables was realized by a combination of Monte Carlo sampling and stepwise backward elimination as implemented in the gdm.varImp-function. We performed 250 permutations per step until only significant ($\alpha = 0.05$) variables remained in the model.

We used two different approaches to assess the importance of predictor variables which reflect the *level importance* and *dispersion importance* (Achen, 1982; Grömping, 2006) with respect to compositional similarity. Dispersion importance, i.e. the amount of variance explained, was quantified by fitting separate GDMs using the full set of significant predictors (full model), only geographic distance (distance-only model), and only environmental variables (environment-only model). Based on the respective *deviance explained*, the metric of model fit in GDM, we partitioned the variation in compositional similarity into purely spatial, purely environmental, shared and unexplained effects (Borcard *et al.*, 1992; Legendre, 2008). For assessing level importance, i.e. a predictor's importance for the response's mean, we followed Fitzpatrick *et al.* (2013) and used the height of significant transformation functions in the full model. To improve comparability between methods, heights were linearly rescaled so that their sum equals the proportion of deviance explained by the model.

Finally, we used GDM to derive predictions of compositional similarity across a global grid of 2986 equal-area grid cells (c. 50,000 km² per cell) on the basis of the distance-only, environment-only and full model for the mainland subset. Observed and predicted pairwise similarities were subjected to non-metric multidimensional scaling (NMDS) using the metaMDS function of the R-package *vegan* (Oksanen *et al.*, 2013). This yielded a quasi-optimal representation of pairwise similarities in a three-dimensional space. The orientation of the NMDS objects was aligned perpendicular to the vectors of mean annual temperature and mean annual precipitation. The resulting vectors of x, y and z coordinates were individually rescaled to a range between 0 and 1 and projected into the red-green-blue color-space. This allowed us to visually assess the model performance and to create a first ever global map of floristic similarity.

All analyses were carried out with the R statistical programming language v 3.2.3 (R Core Team, 2017).

3.4 Results

Across all subsets, compositional similarity showed a pronounced decay with geographic distance, but the rate of decay (quantified as halving distance, $D_{S/2}$) differed strongly. The overall similarity of plant species assemblages declined by 50 percent every 1,576 km (Figure 3.2a). Island OGUs ($D_{S/2}$ = 1,840 km) showed a lower turnover rate than mainland OGUs ($D_{S/2}$ = 1216 km). This pattern was consistent across taxonomic and functional subsets and most pronounced for trees (Figure A3.4 and Table A3.1). Also when inspecting the overall turnover rates for different island types (Figure 3.2b), neither continental ($D_{S/2}$ = 1,384 km) nor oceanic islands ($D_{S/2}$ = 1,628 km) exhibited a higher turnover rate than mainland OGUs. Moreover, we found very low turnover for atolls ($D_{S/2}$ = 5,967 km). Taxonomic and functional groups exhibited different distance decay patterns as well (Figure 3.2c). The halving distance of angiosperms ($D_{S/2}$ = 1,544 km), with 95 % of all species records the largest taxonomic subset, was very close to that of the full dataset (see Figure 3.2c vs. 3.2a). Gymnosperms ($D_{S/2}$ = 1,476 km) showed similarly high turnover as angiosperms, while pteridophytes ($D_{S/2}$ = 2,156 km) had the lowest turnover rate. Among functional groups (Figure 3.2d), shrubs ($D_{S/2}$ = 826 km) had a higher turnover rate than trees and herbs ($D_{S/2}$ = 1,163 km and 1,523 km, respectively).

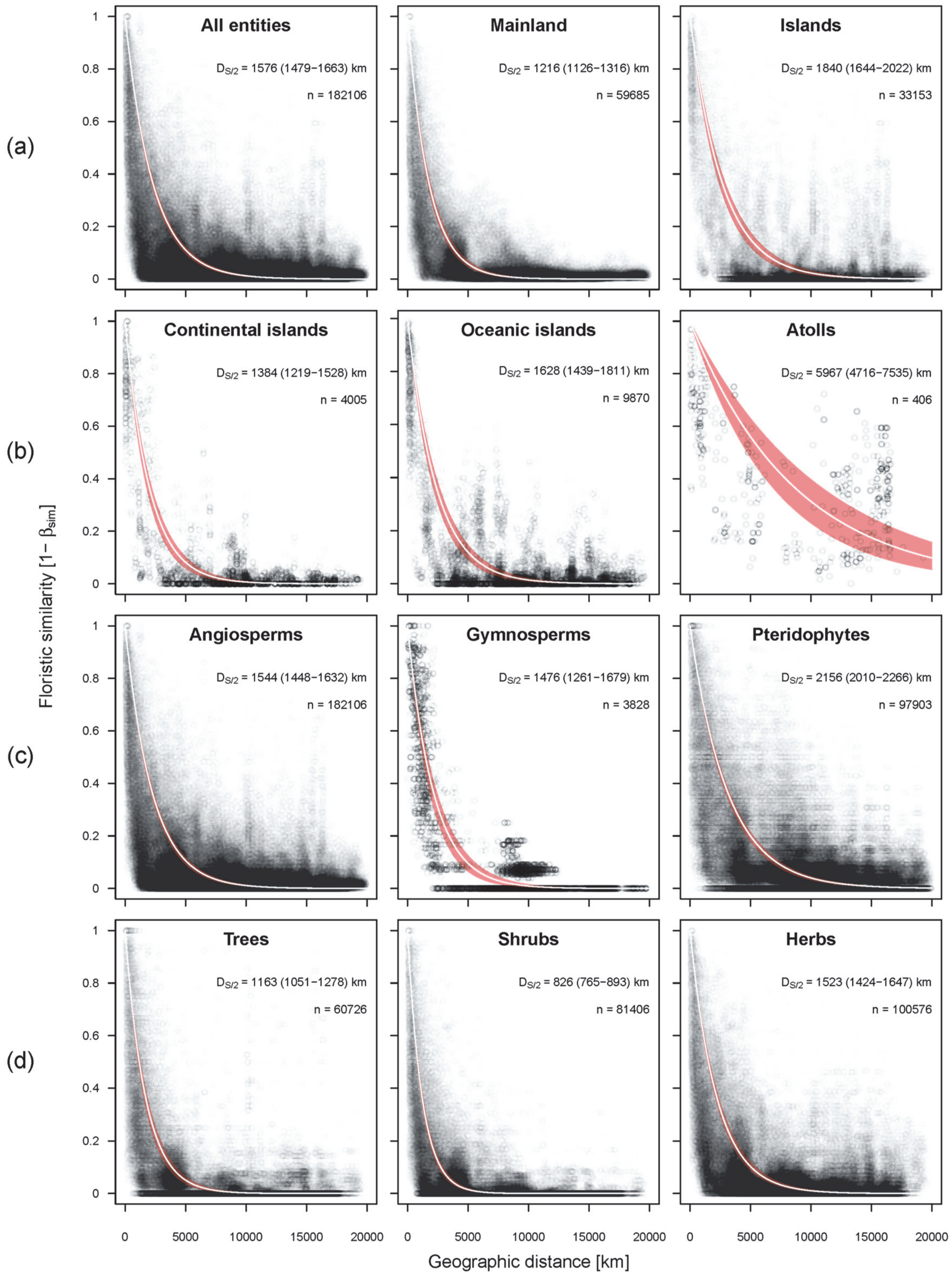


Figure 3.2: Distance decay of similarity for different subsets based on (a) geographical setting, (b) island type, (c) taxonomic group and (d) functional group. For each subset, $D_{S/2}$ denotes the distance after which similarity decreases by 50 % (halving distance) and n denotes the number of unique pairwise combinations within the subset. Regression lines (white) were fitted using a log-binomial generalized linear model (GLM) with an intercept of 1. Confidence intervals (red) were computed by subsampling the data 250 times, refitting the model and taking the 2.5 and 97.5 percentiles of the sampling distribution of coefficient estimates.

According to generalized dissimilarity modelling, geographic distance and environmental variables accounted for a combined 56-85 % of compositional variation in the investigated subsets. The shapes of transformation functions in the fitted GDMs were broadly similar, indicating no fundamentally different responses along geographical and environmental gradients (Figure A3.6). However, the relative importance of predictor variables differed strongly across subsets (Figure 3.3). The variance-based approach generally tended to assign a higher importance to geographic distance than the transformation-based approach, such that the shared effects drew more strongly from the contribution of environmental variables. Contrary to our expectations, both methods concordantly showed a higher importance of geographic distance on the mainland, whereas environmental variables were considerably more important on islands. Looking at particular island types, we found the importance of environmental variables vs. geographic distance to increase from continental islands to oceanic islands to atolls (Figure 3.3). Moreover, atolls stood out due to strong effects of precipitation seasonality and the non-significance of all other environmental variables. However, it must be acknowledged that atolls, owing to their specific ontogeny, span only small environmental gradients. For instance, being located in the tropics, they cover a confined temperature range.

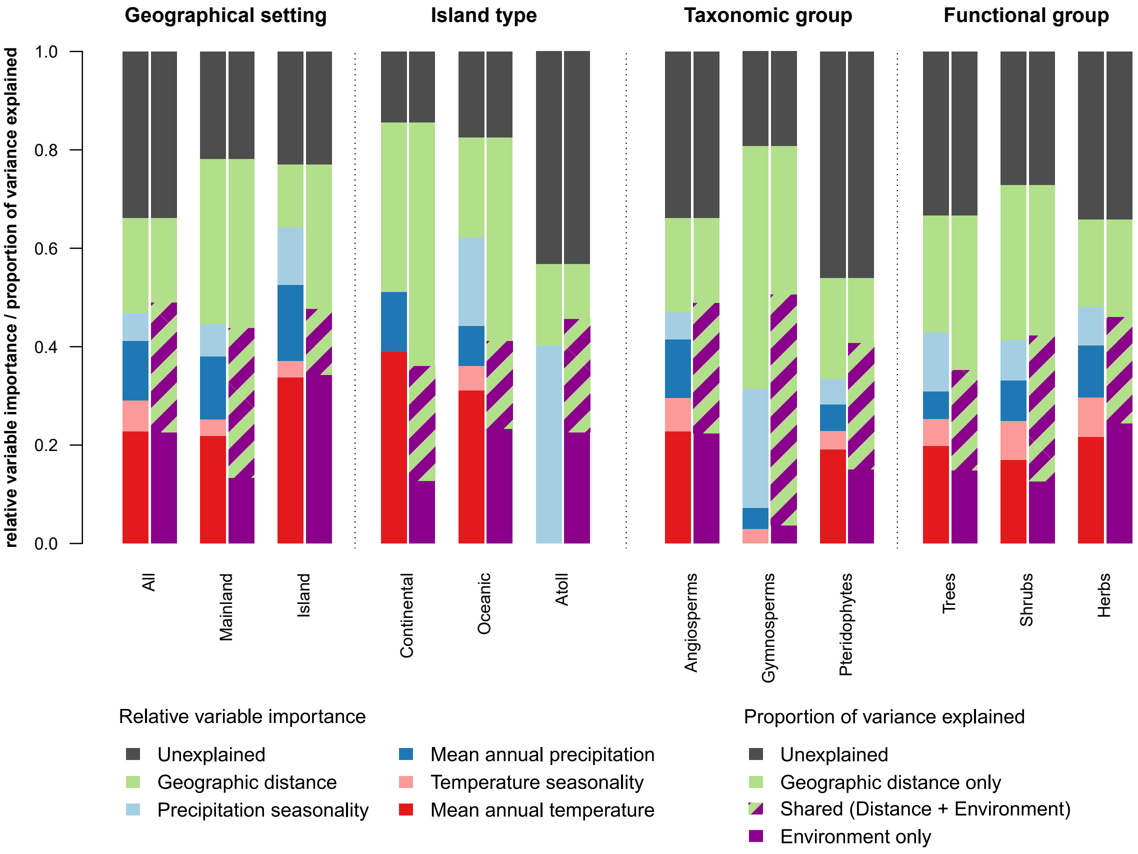


Figure 3.3: Turnover partitioning for taxonomic and functional groups using generalized dissimilarity modelling (GDM). Results are shown for different subsets of the entire dataset based on geographical setting, island type, taxonomic, and functional group. For each subset, two alternative measures of variable importance are presented where the left-hand side is based on the height of GDM transformation functions and the right-hand side is based on variation partitioning.

Taxonomic and functional subsets were quite similar regarding the relative importance of predictor variables (Figure 3.3). Only turnover in gymnosperms exhibited a substantially larger sensitivity to geographical distance compared to other subsets. However, we obtained more heterogeneous results when analyzing turnover of individual groups for island and mainland OGUs separately (Figure A3.5). Here, turnover again showed a consistent shift towards stronger environmental control and lower importance of geographic distance on islands compared to the mainland. Tree turnover exhibited the strongest contrast, while the differences were least pronounced for pteridophytes. The shift in variable importance was generally stronger in the transformation-based as compared to the variance-based approach. For shrubs, the variance-based approach even indicated a higher importance of geographic distance on islands. No GDM could be fitted for gymnosperms on islands because only seven OGUs met our inclusion criteria.

Our predictions of global compositional similarity (Figure 3.4) illustrate how turnover results from a combination of spatial and environmental determinants. Noticeably, the predictions of the environment-only model closely resemble biomes and reflect the expected compositional similarity in the absence of dispersal limitation and biogeographical history. The distance-only model, in contrast, depicts the expected similarity in absence of environmental limitations. A visual comparison of observed vs. predicted similarities (Figure 3.4) confirms that GDM is able to accurately model large-scale turnover patterns in continuous environments, albeit predictions became less accurate in regions with a unique biogeographical history, e.g. Australia and South Africa.

3.5 Discussion

In general, global turnover in vascular plants was remarkably well explained by a combination of geographic distance and a small set of climatic predictors. However, our analyses revealed considerable differences among geographical settings as well as functional and taxonomic groups in the total amount of compositional turnover and the relative importance of its spatial and environmental drivers.

3.5.1 Turnover as result of filtering processes

Unexpectedly, turnover among islands was lower than among mainland OGUs and generally more associated with environmental variables than with geographic distance (Figures 3.2 and 3.3). This finding seems to contradict the general notion of islands as isolated microcosms, featuring unique biota (Gillespie, 2007). However, these results become comprehensible when shifting the perspective from iconic, isolated archipelagos such as Hawaii towards a more general view on islands. Islands are surrounded by inhospitable open sea which presents a strong dispersal filter to the immigrating species (Carlquist, 1965; Whittaker & Fernández-Palacios, 2007). On the other hand, local conditions on islands impose an environmental filter upon arriving species' climatic tolerances, habitat preferences, or reproductive strategies (Carlquist, 1965; Weigelt *et al.*, 2015). A lower turnover rate among islands can thus be explained by these filtering processes reducing the set of potential colonizers and thereby homogenizing the composition of realized species assemblages. Additionally, the particularly strong dispersal filter associated with islands may be passed only by dispersive – and therefore rather distance-insensitive – species, which explains the reduced importance of geographic distance for compositional turnover (Figures 3.3 and A3.5).

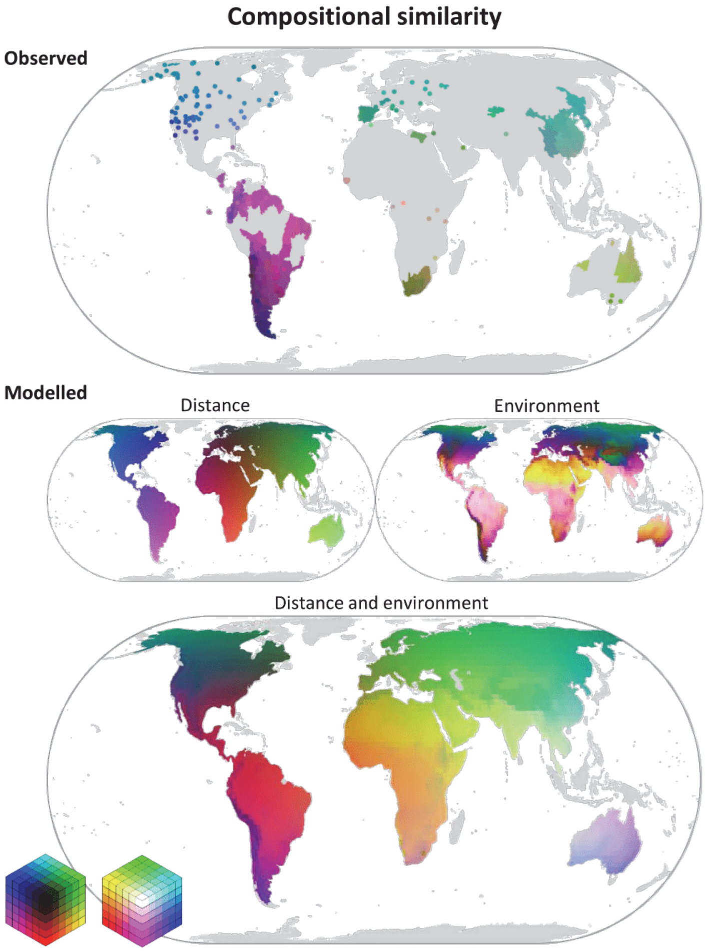


Figure 3.4: Predicted compositional similarity of vascular plants for 2,986 equal-area grid cells of c. 50,000 km² size. Predictions are derived from a generalized dissimilarity model fitted with geographical, environmental, and floristic information for 346 mainland operational geographical units. Predicted similarities were subjected to a three-dimensional non-metric multidimensional scaling (NMDS), standardized, and projected into RGB-color space. Similar colors indicate grid cells with similar predicted species composition. Upper panel shows observed similarities. Center left- and right-hand panels show model predictions for geographic distance and environmental conditions only. Lower panel depicts the predicted similarities based on the full model. Projection: Eckert IV.

Atolls make a strong case for the interpretation of turnover being the result of filtering processes. The conditions on atolls impose both very strong dispersal and environmental filters that may be passed only by few species that are both highly dispersive and adapted to tropical temperatures, coastal conditions (e.g. high salinity, high levels of disturbance), and small population sizes. Consequently, we found very low turnover among atolls and a low sensitivity to geographical distance (Figures 3.2 and 3.3), confirming previous, more descriptive analyses (Stoddard, 1992). Systems that exhibit relaxed filtering such as oceanic islands (reduced environmental filtering) or continental islands (reduced dispersal and environmental filtering), however, sample from a larger pool of potential colonizers and are thus characterized by higher turnover rates. This rationale may also help to understand other prominent biogeographical patterns including latitudinal or elevational gradients in beta diversity (Qian & Ricklefs, 2007; Tello *et al.*, 2015). For example, it suggests that the size of the species pool (gamma diversity) is an inherent property of the geographical or environmental setting and that correcting for variations in pool size (Kraft *et al.*, 2011) might lead to incorrect conclusions regarding the role of assembly processes.

The above considerations are not addressing the main reason why island biota are regarded as unique: endemism. Islands are hotspots of *in-situ* diversification (Kier *et al.*, 2009) and endemic species are potentially a strong driver of turnover among islands. The impact of endemism on turnover is proportional to the fraction of endemic species in a flora. While this fraction may be quite high for large, isolated islands, the majority of islands represents smaller and less isolated geographical entities (Weigelt *et al.*, 2013). From the 115 island OGUs in our dataset with available information on endemism status, 90 had less than 10 % endemics. Thus, the native flora of most islands consists predominantly of rather widespread non-endemic species and is a product of strong dispersal and environmental filtering, resulting in reduced turnover. Supporting this, the most frequently shared species among islands in our dataset were invariably pantropical coastal and littoral plants such as *Scaevola taccada, Cassytha filiformis, Fimbristylis cymosa,* or *Ipomoea pes-caprae*. While studies on other organismic groups report different results (Stuart *et al.*, 2012), the overall effect of endemism on the compositional structure of plant assemblages on islands may be smaller than expected.

3.5.2 The role of species attributes for turnover

As illustrated by the small set of widespread littoral plants on islands, the ability to pass ecological filters depends on species attributes like dispersal ability or environmental tolerance. This is reflected by group-specific turnover patterns. According to our analyses, turnover rates were lowest for pteridophytes, intermediate for angiosperms, gymnosperms and herbs, high for trees and very high for shrubs (Figures 3.2c and 3.2d). Similar results at smaller scales were reported by Tuomisto *et al.* (2003) and Qian (2009) for angiosperms vs. pteridophytes, and by López-Martínez *et al.* (2013) for shrubs vs. trees. Our findings align well with the average propagule size, and thus approximate dispersal ability, of the investigated groups (Levin, 1974a; Westoby *et al.*, 1996; Moles, 2005) and further confirm a negative relationship between dispersal ability and turnover (Nekola & White, 1999; Tuomisto *et al.*, 2003; Qian, 2009).

The relative importance of geographic distance and environmental variables for turnover was rather similar across functional and taxonomic groups (Figure 3.3). When contrasting group responses in mainland and island systems, however, we found a consistently elevated importance of environmental variables and reduced importance of geographic distance on islands (Figure A3.5). Notably, the shift in variable importance was strongest for groups with high sensitivity to geographic distance on the mainland, suggesting that these are more strongly affected by the increased level of dispersal filtering on islands. Consequentially, only a small, highly dispersive fraction of these groups is represented on islands, leading to the inversion of variable importance and the low relative importance of geographic distance.

Gymnosperms on islands mark an extreme case of this effect: here, the strength of dispersal filtering apparently surpasses the group's dispersal abilities to such an extent, that it is almost entirely absent from this geographical setting (Figures A3.4 and A3.5).

While our results demonstrate a clear relationship between turnover and group-specific dispersal abilities, the linkage between turnover and environmental group characteristics is less evident. However, assumptions regarding group differences in dispersal ability can be reasonably justified on the basis of average propagule sizes and characteristic reproductive strategies. In contrast, a similar evaluation of the considered groups with respect to environmental tolerances is more difficult. A rigorous test of the relationships between species characteristics and environmental filtering would require a detailed *a priori* definition of groups based on features that are causally linked to niche breadth and independent of dispersal ability. Given the recent progress in estimating niche dimensions (Kearney *et al.*, 2010; Blonder *et al.*, 2014; Díaz *et al.*, 2016) and the increasing availability of functional trait data (Kattge *et al.*, 2014), we are confident that such analyses are soon within reach.

3.5.3 The origin of beta diversity

Although estimates of variable importance were plausible in the context of specific geographical settings and species groups, it is difficult to draw a general conclusion regarding the balance of spatial and environmental drivers of vascular plant turnover. The two measures of variable importance produced slightly different results, with the variance-based approach putting more weight on geographic distance than the transformation-based approach. Considering the way these methods work, this difference exposes highlights some important properties of large-scale beta-diversity patterns. Most of the systematic change in compositional similarity occurred within the first 3,000-5,000 km, whereas pairwise similarity was generally low thereafter (Figure 3.2). Hence, for the 75-85 % of site-pairs that lie beyond that threshold, almost complete turnover is predicted very accurately by geographic distance alone. This is reflected by the high importance of geographic distance in the variance-based approach. In contrast, the transformation-based approach is more robust against the high number of low values of compositional similarity at large geographic distances: If a further change in a predictor does not affect compositional similarity, the transformation function flattens out and the variable importance remains constant. This, however, does not allow addressing multicollinearity among predictors by estimating shared effects. In summary, the variance-based approach is best understood as reflecting the *overall* prediction accuracy of a variable, whereas the transformation-based approach indicates of a variables' importance for *systematic* changes in compositional similarity.

Relating this to our results, both methods clearly show that environmental and spatial factors act jointly in causing turnover in vascular plants. For most subsets, the systematic change in composition is slightly more controlled by environmental conditions than by geographic distance. With increasing geographic distance, however, compositional similarity generally approaches very low values and environmental variables become increasingly expendable for predicting turnover. Topographical site properties such as elevation and area did not have important effects at the studied spatial grain and extent. Also with regard to the question whether niche-based or neutral processes drive global turnover, it is important to consider scaling effects. Chase (2014) provides compelling arguments why the importance of neutral processes should diminish with increasing spatial scale. At such scales, unique effects of geographic distance are more likely the product of evolutionary history than of neutral immigration-extinction dynamics. Moreover, it is important to note that dispersal, although partly stochastic, is not necessarily a neutral process: Species clearly differ in their dispersal ability which (1) violates the assumption of species equivalence in neutral models and (2) adds a deterministic component to the process of dispersal (Lowe & McPeek, 2014). Equating distance-related effects on species composition with the importance of neutral processes would therefore disregard the signal of non-random species filtering. According to our results, global turnover in vascular plants is structured by a combination of niche-based processes related to environmental factors and predominantly non-neutral dispersal- and speciation-related processes to geographical distance.

3.5.4 Methodological strengths and limitations

The nature of large-scale beta diversity data presents some statistical challenges due to the high number of low and zero-values. In contrast to the distance decay models, generalized dissimilarity modeling was relatively robust to this idiosyncrasy and proved to be a powerful tool for analyzing turnover, with our predictions of mainland compositional similarity (Figure 3.4) showcasing only one application. A mechanistic understanding of beta diversity is increasingly recognized as a key requirement for the effective protection of global biodiversity (Socolar *et al.*, 2016). Complex statistical models such as GDM are a promising way to further develop this understanding and to derive accurate predictions of ecosystem responses under future conditions (Mokany *et al.*, 2016). A next promising step towards this objective is to integrate phylogenetic information into the modeling framework (Graham & Fine, 2008; Rosauer *et al.*, 2014). This may reduce the problem of zero inflation, increase the informational resolution of the data, and allow for disentangling the role of evolutionary history and dispersal limitation in the spatial signal of compositional similarity.

3.6 Conclusion

Our results demonstrate a complex interplay of geographical and environmental factors as well as species characteristics in determining the magnitude and structure of turnover. We found strong evidence that (1) the geographical setting affects turnover via dispersal and environmental filtering and (2) the ability to pass these filters varies among taxonomic and functional groups and is largely congruent with expectations based on group specific attributes, particularly with dispersal ability. However, while the relationship between species attributes and turnover has received a fair amount of attention during the last decades, our findings suggest that the geographical context may play an equally, if not more important role in shaping patterns of compositional turnover. Addressing the specific ecological filters that are associated with different geographical settings may help to better understand large-scale patterns in beta diversity. For this purpose, the application of powerful statistical models such as GDM in combination with phylogenetic data offers great potential.

4 Source pools and disharmony of the world's island floras

König, Christian, Patrick Weigelt, Amanda Taylor, Anke Stein, Wayne Dawson, Franz Essl, Jan Pergl, Petr Pyšek, Mark van Kleunen, Marten Winter, Cyrille Chatelain, Jan J. Wieringa, Pavel Krestov and Holger Kreft

4.1 Abstract

Disharmony is a key concept in island biology that describes the biased representation of higher taxa on islands compared to their mainland source regions. Although differential colonization success of taxa is predicted by selective dispersal, environmental and biotic filtering, the empirical evidence for disharmony remains largely anecdotal. Here, we develop a novel method for delineating island source regions and present the first global quantitative assessment of island disharmony. We analyzed the overall compositional bias of 320 island floras and examined the over- or under-representation on islands for 450 plant families. We found that the compositional bias of island floras is strongly predicted by geographical and climatic island characteristics (isolation, area, geologic origin, temperature, precipitation), whereas the representation of individual families is only weakly related to family-specific characteristics (family size, family age, taxonomic group, functional traits), indicating that the taxonomic scope of the disharmony concept has historically limited its wider applicability. Our results provide a strong foundation for integrating disharmony with quantitative functional and phylogenetic approaches in order to gain a deeper understanding of assembly processes on islands.

4.2 Introduction

Islands are renowned for their unique biotas, often characterized by high levels of endemism (Kier *et al.*, 2009), species radiations (Böhle *et al.*, 1996; Givnish *et al.*, 2009), relictual taxa (Vargas, 2007) or peculiar shifts in species' life history and morphology (Carlquist, 1965). These features can be attributed to the isolated nature of islands (Weigelt & Kreft, 2013), which strongly affects the fundamental processes controlling species diversity: immigration, speciation, and extinction. Thus, research on island systems has stimulated many seminal contributions to evolutionary (Darwin, 1859; Wallace, 1881) and ecological theory (MacArthur & Wilson, 1963, 1967; Hubbell, 2001; Whittaker *et al.*, 2008). The island biogeographical studies by Carlquist(1965, 1967, 1974) are undoubtedly among these seminal contributions, providing substantial insights into processes such as long-distance dispersal or adaptive radiation, and inspiring island research to this date (Traveset *et al.*, 2015). In contrast to the strictly analytical approach of MacArthur & Wilson's equilibrium theory of island biogeography (MacArthur & Wilson, 1963, 1967), Carlquist's work offers mostly a natural history perspective. While this perspective does not allow for quantitative predictions of e.g. species richness, it is a powerful framework for understanding qualitative features of island biota such as taxonomic composition or morphological adaptations (Midway & Hodge, 2012). One such feature is the striking taxonomic "imbalance" of many island biotas– a phenomenon known as disharmony (Carlquist, 1965, 1974).

Island disharmony refers to the biased representation of higher taxa (e.g. families) in island biotas compared to their mainland source regions (Whittaker & Fernández-Palacios, 2007). It is the result of selective assembly mechanisms – dispersal filtering, environmental filtering and biotic filtering (Carlquist, 1974; Keddy, 1992; Weiher *et al.*, 2011; Kraft *et al.*, 2015) – acting with particular rigor in island systems, thus permitting only a subset of the mainland flora to successfully colonize islands (Weigelt *et al.*, 2015). The interplay between geographical setting and taxon-specific colonization success highlights two distinct aspects of island disharmony. On the one hand, the overall compositional bias of island floras relative to their mainland source regions (compositional disharmony) should reflect the impact of ecological filters during their assembly, and thus show predictable variation with island-specific characteristics such as isolation, area, climatic conditions, or geological origin. On the other hand, the selectivity of these filters should result in a predictable over- or under-representation of individual taxa on islands (representational disharmony) that is associated with taxon-specific attributes related to e.g. dispersal ability or environmental tolerances. Indeed, numerous studies demonstrate that community composition of island floras is strongly dependent on geographical setting (Cabral *et al.*, 2014; König *et al.*, 2017) and taxon-specific attributes (Burns, 2005; Olesen *et al.*, 2010).

While the theoretical underpinnings of island disharmony are well established, the concept itself has been applied rather inconsistently and lacks a quantitative basis. In particular, the specification of mainland source regions is not trivial and often very coarse (Bernardello *et* *al.*, 2006) and the taxonomic bias of island floras is usually illustrated by means of anecdotal evidence rather than objective quantitative measures (Francisco-Ortega *et al.*, 2010). In addition, there has been little work on whether the over- or under-representation of certain plant taxa on islands is globally consistent or not (but see e.g. Kreft *et al.*, 2010), and whether representational deviations are linked to taxon-specific attributes that supposedly affect colonization success (but see e.g. Grossenbacher *et al.*, 2017). Consequently, the empirical evidence for island disharmony remains fragmentary.

Here, we provide the first quantitative analysis of island disharmony for vascular plants using an unprecedented dataset of 320 island and 611 mainland floras including a total of 225,053 species. First, we present a novel method for estimating island-specific source regions and develop two mathematical indices that quantify compositional and representational disharmony. This enabled us to examine the phenomenon of disharmony from an island- and a taxon-centred perspective, and thus to disentangle its island-specific and taxon-specific drivers. Accordingly, we analyze compositional disharmony as a function of island isolation, area, geological origin, and climatic conditions, and representational disharmony as a function of families' species richness, age, range size, and predominant functional characteristics related to colonization success. In particular, we are interested in the importance of dispersal, environmental and biotic filtering in creating disharmonic island floras. If dispersal filtering is the dominant cause of disharmony (Carlquist, 1967, 1974), we would expect strong positive effects of isolation on compositional disharmony as well as a consistently positive effect of dispersal-related traits on representational disharmony. Alternatively, if environmental or biotic filtering processes play an important role (Carvajal-Endara *et al.*, 2017; Grossenbacher *et al.*, 2017), we expect to find a strong response of compositional and representational disharmony to island climatic variables and pollination or competition-related traits, respectively.

4.3 Methods

We examined the phenomenon of island disharmony from both an island- and a taxon-centred perspective (Figure 4.1). First, we assessed compositional disharmony, i.e. the phenomenon of island floras being taxonomically biased compared to their mainland source regions. Second, we investigated representational disharmony, i.e. the role of individual taxa in creating compositional disharmony by assessing their global representation on islands compared to the mainland. In both cases, we chose families as the focal taxonomic level. Families provide a reasonable balance between ecological uniformity and taxonomic aggregation. In contrast, higher taxonomic levels such as orders encompass too many species that are too heterogeneous in their attributes to be ecologically meaningful study units, whereas lower levels such as genera are too numerous to be harmonically represented in any island flora.

All analyzes were performed in the R statistical framework version 3.44 (R Core Team, 2017).

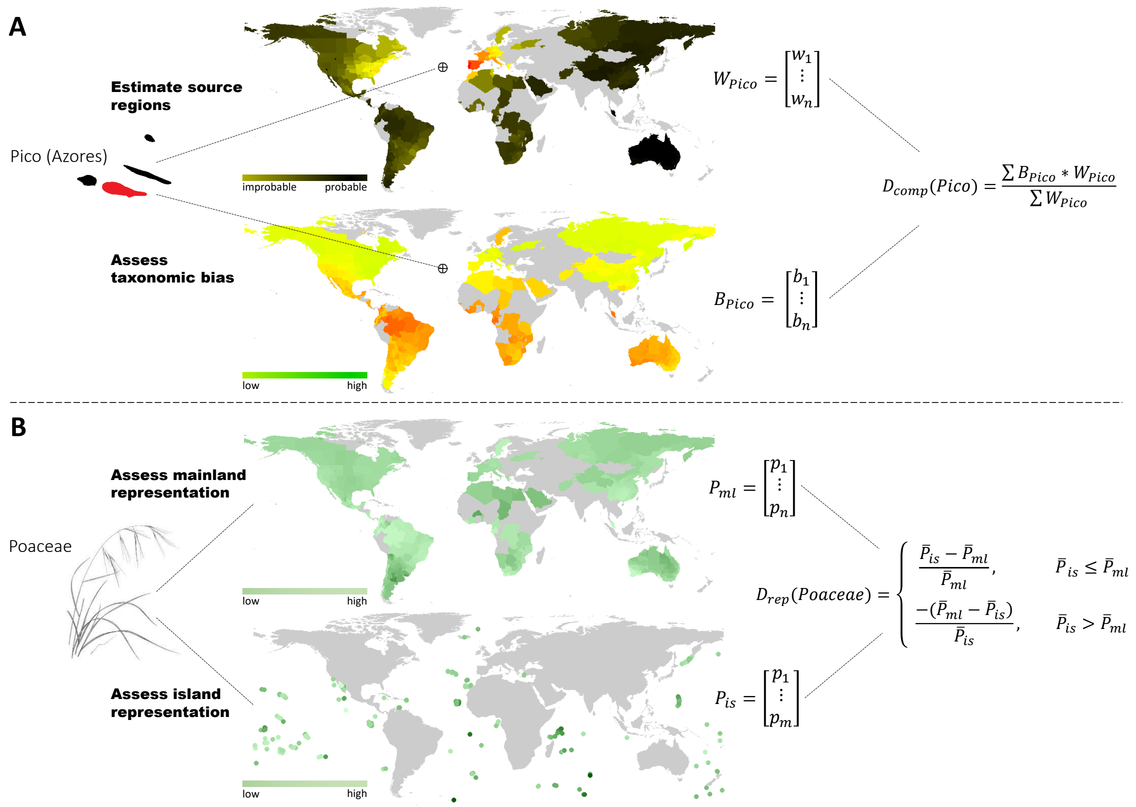


Figure 4.1: Schematic representation of the quantification of compositional and representational disharmony. (A) Calculation of compositional disharmony by the example of Pico Island (Azores). Source regions were estimated based on predictions of species turnover between the focal island and all mainland units (W, upper plot). The taxonomic bias between the focal island and all mainland units was quantified using Bray-Curtis dissimilarity on relative proportions of plant families (B, lower plot). The compositional disharmony of the focal island (D_{comp}) was then calculated as the mean taxonomic bias relative to all mainland regions, weighted by their respective source pool weight. (B) Estimation of representational disharmony by the example of Poaceae. Representational disharmony was quantified based on the mean proportion of the focal taxon in mainland floras ($\bar{P}_{ml}$) and island floras ($\bar{P}_{is}$). The corresponding index (D_{rep}) transforms the ratio between these two components to a range between -1 (family occurs on the mainland only) and 1 (family occurs on islands only).

4.3.1 Data collection

All ecological and environmental data were obtained from the Global Inventory of Floras and Traits database (GIFT, see Chapter 2), a novel resource for macroecological analyses of global plant diversity. The primary data type in GIFT are regional plant checklists that are integrated with additional information at the level of taxa (e.g. functional traits, taxonomic placement, phylogenetic relationships, or floristic status) and geographical units (e.g. climatic and environmental conditions, socioeconomic factors or physical geographic properties).

We extracted all checklists from GIFT that indicate which of the listed species are native to the corresponding geographical region. Based on this information, we excluded all non-native species from the analyses. Checklists referring to the same geographical unit were combined. We then removed all geographical units that were not covered by either a complete checklist of vascular plants or by several checklists that add up to a complete checklist of vascular plants (e.g. separate lists for pteridophytes and seed plants). Checklists were considered complete if the number of species fell into a reasonable range for the given area, biome and taxon (Kreft & Jetz, 2007). The resulting dataset contained checklists of native vascular plants for 611 mainland and 320 islands units (see Table A4.1).

4.3.2 Compositional disharmony

Quantification

The quantification of compositional disharmony requires estimates of island-specific source regions and an objective measure of taxonomic bias (Figure 4.1A). We based our method for estimating source regions on the fact that geographic distance and environmental gradients produce distinct and predictable patterns in species turnover (Fitzpatrick *et al.*, 2013; König *et al.*, 2017). Species turnover is a richness-insensitive measure of compositional similarity that quantifies the proportion of shared species between assemblages (Baselga, 2010). This makes turnover a crucial concept for constructing biogeographical species pools and delineating source regions (Carstensen *et al.*, 2013).

We used generalized dissimilarity modelling (Ferrier, 2002; Ferrier *et al.*, 2007) to model species turnover (β_{sim}, Koleff *et al.*, 2003) among geographical units as a function of pairwise geographic distance and differences in mean annual temperature, mean annual precipitation, temperature seasonality and precipitation seasonality. These covariates are strong predictors of large-scale species turnover (König *et al.*, 2017). The model was fitted using mainland data only (deviance explained = 80.5%), because island floras exhibit strong imprints of ecological filtering, which would mask the very effects we aim to quantify in this study. We derived model predictions of species turnover between each island and a global equal-area, equidistant hexagon grid (~ 23,300 km^2 per cell) covering all continents and important island source regions. The predictions were then aggregated to the level of the GIFT mainland units by calculating the area-weighted mean of grid values per unit (Figure 4.1A). The resulting $m \times n$ matrix, W, contained the expected proportion of shared species (1-β_{sim}) between all m islands and all n mainland units from a “mainland perspective”, i.e. assuming that island floras assemble under the same conditions as floras on the mainland. The matrix W can thus be interpreted as reflecting the importance of a given mainland unit as a source region for a given island, while excluding the effects of modified filtering during island colonization. We set $\mathsf{W} = 0$ for mainland units with very low values, i.e. highly improbable source region for a given island, to ensure a balanced estimation of compositional disharmony (see Supplementary text A4.1 for details). To validate our method, we compared the results with empirical source pool reconstructions based on floristic and phylogenetic relationships to the mainland.

To assess the actual taxonomic bias of an island flora, we converted the species checklists to relative family proportions in order to account for species richness differences. We then calculated the Bray-Curtis dissimilarity for all pairwise island-mainland combinations, yielding an $m \times n$ matrix, B. The Bray-Curtis index is the abundance-based version of the Sørensen index that relates the summed differences between two variables to their overall sum. Although usually applied to count data, the index also works with relative proportions (Greenacre & Primicerio, 2013), making it an appropriate measure of taxonomic bias.

Finally, we calculated the compositional disharmony of each island, $\mathsf{D}_{\mathsf{comp}}$, as the mean taxonomic bias of a given island i relative to all mainland units (B_{i}), weighted by their respective importance as a source region for the specific island (W_{i}). $\mathsf{D}_{\mathsf{comp}}$ ranges between 0 and 1, with higher values indicating more disharmonic floras (Figure A4.1).

Analysis

For each island, we calculated area (AREA; km²) and extracted averaged values for mean annual temperature (MAT; °C), mean annual precipitation (MAP; mm/a), temperature seasonality (TVAR; °C), and precipitation seasonality (PVAR; mm/a) from CHELSA climate layers (Karger *et al.*, 2017). We additionally calculated the minimum distance to the nearest mainland (DIST), assessed the geological origin based on pertinent literature (GEO, distinguishing between shelf, continental fragment, volcanic, tectonic uplift and atoll, see Chapter 3.3.2 for definitions), and respective island group (ARCH) for each island.

We $\log_{10}$-transformed AREA, DIST, MAP, TVAR and PVAR because of strongly skewed distributions. For MAT, no adequate transformation was found. We then scaled all continuous predictor variables to $\mu = 0$ and $\sigma = 1$ in order to standardize model coefficients. Although our response variable ($\mathsf{D}_{\mathsf{comp}}$) is, in principle, a proportional value ranging from 0 to 1, it was effectively bound between 0.19 and 0.87, and approximately normally distributed within that interval. We therefore performed the analysis of compositional disharmony using standard linear mixed effects models. Following Bunnefeld & Phillimore (2012), we specified archipelago (ARCH) as random effect. Moreover, we included interactions between DIST and GEO, and AREA and GEO in order to reflect potentially varying effects of island isolation and area on compositional disharmony, depending on whether an island has once been connected to the mainland or not. Based on the full model ($\mathsf{D}_{\mathsf{comp}} \sim \mathsf{DIST\colon GEO} + \mathsf{AREA\colon GEO} + \mathsf{MAT} + \mathsf{MAP} + \mathsf{TVAR} + \mathsf{PVAR} + \mathsf{GEO} + (1|\mathsf{ARCH})$), we (1) fitted models for all possible combinations of predictor variables, (2) selected those models with the most empirical support ($\Delta\text{AIC} < 2$), and (3) performed model averaging on this set of models (Grueber *et al.*, 2011; Barton, 2016). Unless stated otherwise, all reported effects are based on the averaged model results.

4.3.3 Representational disharmony

Quantification

For the quantification of representational disharmony, we developed a second index ($\mathsf{D}_{\mathsf{rep}}$) that is based on the mean proportional representation of the focal family across all mainland units ($\bar{P}_{ml}$) and all oceanic islands ($\bar{P}_{is}$, see Figure 4.1B). We focused on oceanic islands (volcanic islands, tectonic islands and atolls) because their floras reflect the effects of dispersal, environmental and biotic filtering most clearly (Whittaker & Fernández-Palacios, 2007) and thus allow for the most rigorous assessment of taxon-specific drivers of disharmony. The index is symmetric and obtains values in the interval $(-1, 1)$, with the sign indicating whether the focal family is proportionally more abundant on islands or on the mainland (Figure A4.1). For example, a family with $D_{rep} = 0.5$ has, on average, a two times higher proportion on islands as compared to the mainland, whereas a family with $D_{rep} = -0.9$ has a 10-times higher proportion on the mainland. The special cases of $D_{rep} = -1$, $D_{rep} = 0$ and $D_{rep} = 1$ indicate a family's restriction to the mainland, equal proportional representation on islands and the mainland, and restriction to islands, respectively.

Analysis

We obtained the total number of species in each family (SPEC) from The Plant List (2013) and supplemented it with values from Christenhusz & Byng (2016) where The Plant List did not resolve the respective family. The assignment of taxonomic supergroups (GROUP, distinguishing between angiosperms, gymnosperms and pteridophytes), family age estimates (AGE) and all further phylogenetic analyses were based on a recent megaphylogeny of vascular plants (Qian & Jin, 2016). For seed plants only, we selected six functional traits reflecting dispersal ability (dispersal syndrome, fruit type, seed mass), life history (growth form) and reproductive characteristics (pollination syndrome, self-compatibility). For each trait, we aggregated all species-level information available in GIFT (see Table A4.1) to the family-level: for categorical traits, we assigned a value to a family when the majority (> 66 %) of the species level values in GIFT were identical, i.e. a predominant trait syndrome was evident for the respective family; for numerical traits, we calculated the median across all species in the respective family with information on the respective trait.

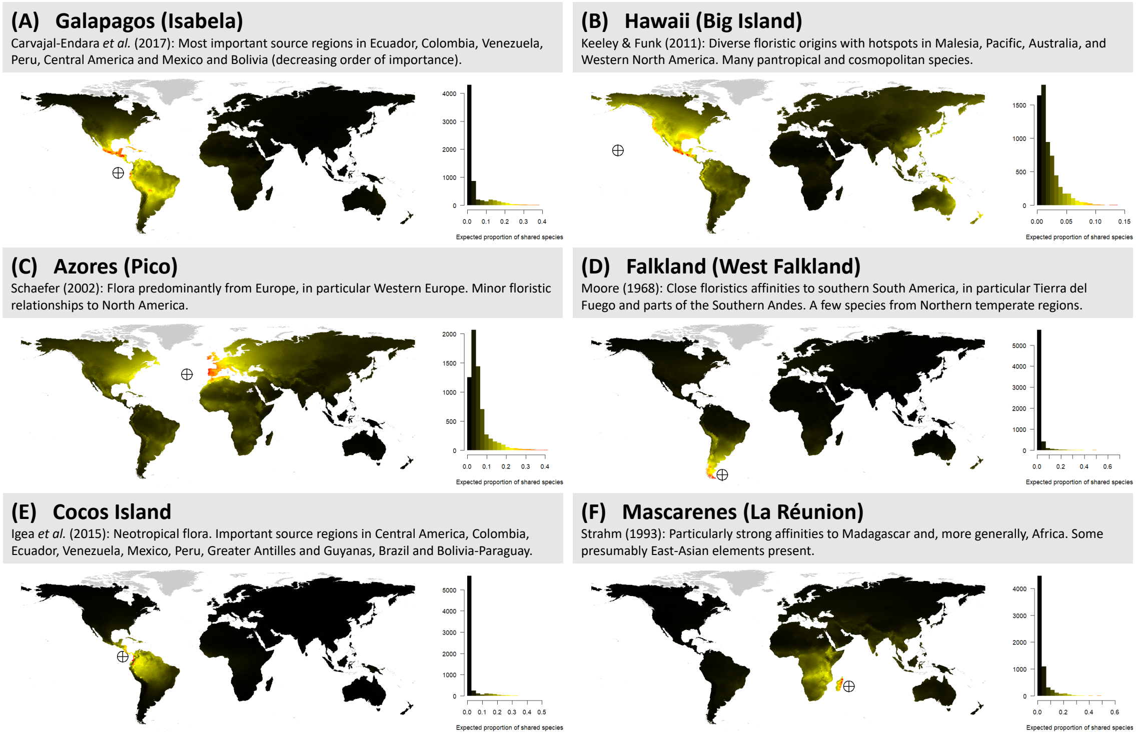
We used multiple linear regression to analyze representational disharmony (D_{rep}) as a function of GROUP, AGE, and SPEC. We had to omit all functional trait variables because missing data points would have drastically reduced the statistical power of the model. Instead, we analyzed the impact of traits on D_{rep} individually, using single-predictor linear models for continuous traits and one-way ANOVAs for categorical traits. For significant categorical variables, we performed a Tukey HSD post-hoc test to identify systematic variation among variable levels.

To test for phylogenetic signal in D_{rep}, i.e. whether closely related taxa tend to be similarly over- or under-represented on islands, we calculated Abouheif's C_{mean} using the *phylosignal* R-package (Keck, 2015). The C_{mean} index is a measure of phylogenetic autocorrelation that quantifies the squared differences between values (in this case D_{rep}) of neighboring tips in a phylogeny (Münkemüller *et al.*, 2012).

4.4 Results

4.4.1 Source region estimation

Our method for estimating island source regions showed strong agreement with empirical source pool reconstructions (Figure 4.2). Accordingly, most island floras sample from a limited set of nearby and climatically similar mainland regions. The estimated source regions for the Falkland Islands, for instance, are highly concentrated in the nearby non-tropical parts of South America, which corresponds closely to the account given by Moore (1968) (Figure 4.2D). However, with increasing isolation from the mainland, the distribution of island source regions became more diffuse in both the statistical and empirical reconstructions. For example, we estimated the source regions of the Hawaiian flora to be Circum-Pacific (Figure 4.2B). While this wide distribution of source regions is generally in agreement with the empirical reconstruction by Keeley & Funk (2011), our method puts more emphasis on North and Central America as source regions of the Hawaiian flora. Despite such minor discrepancies, the overall congruency between empirical reconstructions and our statistical estimates of island source regions demonstrates the robustness of our method.



(A) Galapagos (Isabela)
Carvajal-Endara *et al.* (2017): Most important source regions in Ecuador, Colombia, Venezuela, Peru, Central America and Mexico and Bolivia (decreasing order of importance).

(B) Hawaii (Big Island)
Keeley & Funk (2011): Diverse floristic origins with hotspots in Malesia, Pacific, Australia, and Western North America. Many pantropical and cosmopolitan species.

(C) Azores (Pico)
Schaefer (2002): Flora predominantly from Europe, in particular Western Europe. Minor floristic relationships to North America.

(D) Falkland (West Falkland)
Moore (1968): Close floristics affinities to southern South America, in particular Tierra del Fuego and parts of the Southern Andes. A few species from Northern temperate regions.

(E) Cocos Island
Igea *et al.* (2015): Neotropical flora. Important source regions in Central America, Colombia, Ecuador, Venezuela, Mexico, Peru, Greater Antilles and Guyanas, Brazil and Bolivia-Paraguay.

(F) Mascarenes (La Réunion)
Strahm (1993): Particularly strong affinities to Madagascar and, more generally, Africa. Some presumably East-Asian elements present.

Figure 4.2: Exemplary comparison of empirically reconstructed and statistically modelled source pools for six islands. Grey boxes summarize the results of previous source pool reconstructions based on floristic or phylogenetic affinities. Maps show corresponding statistical source pool estimates (proportion of shared species, 1-β_{sim}) between the focal island and 6505 equal-area grid cells (~23,300 km² each). Predictions were derived from a generalized dissimilarity model (Ferrier, 2002; Ferrier *et al.*, 2007) fitted with geographical and climatic characteristics of 611 mainland floras worldwide. Histograms show the distribution of predicted values for each focal island (note that the range of values differs among islands).

4.4.2 Compositional and representational disharmony

Compositional disharmony (D_{comp}) ranged between 0.19 (Corsica, Mediterranean Sea) and 0.87 (Clipperton Island, East Pacific). Overall, the most harmonic island floras were found in the Mediterranean Basin and off the shores of temperate continental regions (East Asia, Europe, North America). Particularly disharmonic floras were located in the Arctic and Antarctic seas and the Central Pacific (Figure 4.3A). In agreement with our expectations, compositional disharmony increased with island isolation and decreased with island area (Table 4.1, Figure 4.3B). Although compositional disharmony differed strongly among island types (Figure 4.3B), geological origin (GEO) was not an important variable in the multi-predictor setting – after controlling for the effect of island area, isolation and climatic characteristics, only atolls had a positive effect on D_{comp}. (Table 4.1). Correspondingly, none of the most supported models included an interaction term. Climatic variables had consistently negative and unexpectedly strong effects on compositional disharmony. Measured by the slope of standardized regression coefficients, precipitation seasonality (PVAR) was the second most important predictor of D_{comp} after area (AREA) – and before island isolation (DIST, Table 4.1). The explanatory power of geographical and climatic variables for compositional disharmony was relatively high ($R^2_{marginal} = 0.44$). However, the floras of e.g. Singapore (tropical, medium-sized shelf island, species-rich, not isolated, $D_{comp} = 0.45$) and South Georgia (Antarctic, large continental fragment, species-poor, highly isolated, $D_{comp} = 0.78$) indicate the existence of additional island-specific factors in creating compositional disharmony, which is also reflected by the substantial contribution of the random variable (ARCH) to the overall model fit ($R^2_{conditional} = 0.87$).

Table 4.1: Statistical model results for compositional disharmony (D_{comp}) and representational disharmony (D_{rep}). Variable abbreviations are given in the methods section. Variables are highlighted in bold if their confidence intervals ($CI_{2.5}$, $CI_{97.5}$) do not include zero.

Compositional disharmony (D_{comp})					Representational disharmony (D_{rep})			
	estimate	std. error	$CI_{2.5}$	$CI_{97.5}$		estimate	std. error	p-value
Intercept	0.547	0.012	0.499	0.557	Intercept	-0.307	0.279	0.272
AREA	**-0.060**	0.005	-0.070	-0.051	AGE	-0.244	0.149	0.102
DIST	**0.027**	0.009	0.009	0.046	SPEC	**0.160**	0.027	< 0.001
MAT	-0.020	0.018	-0.058	0.000	GROUP	-	-	-
TVAR	**-0.052**	0.015	-0.080	-0.024	Angiosperms	-	-	-
MAP	**-0.025**	0.010	-0.045	-0.005	Gymnosperms	0.073	0.200	0.714
PVAR	**-0.027**	0.010	-0.047	-0.007	Pteridophytes	**0.886**	0.113	< 0.001
GEO	-							
Shelf	0	-	-	-				
Fragment	-0.014	0.0230	-0.059	0.031				
Floor	0.023	0.019	-0.015	0.061				
Volcanic	0.015	0.014	-0.012	0.042				
Atoll	**0.050**	0.022	0.007	0.093				
model type	Gaussian Linear Mixed Effects Model				model type	Gaussian Linear Model		
variance explained	$R^2_{MARGINAL} = 0.44$ $R^2_{CONDITIONAL} = 0.87$				variance explained	$R^2 = 0.19$		
sample size	320				sample size	451		

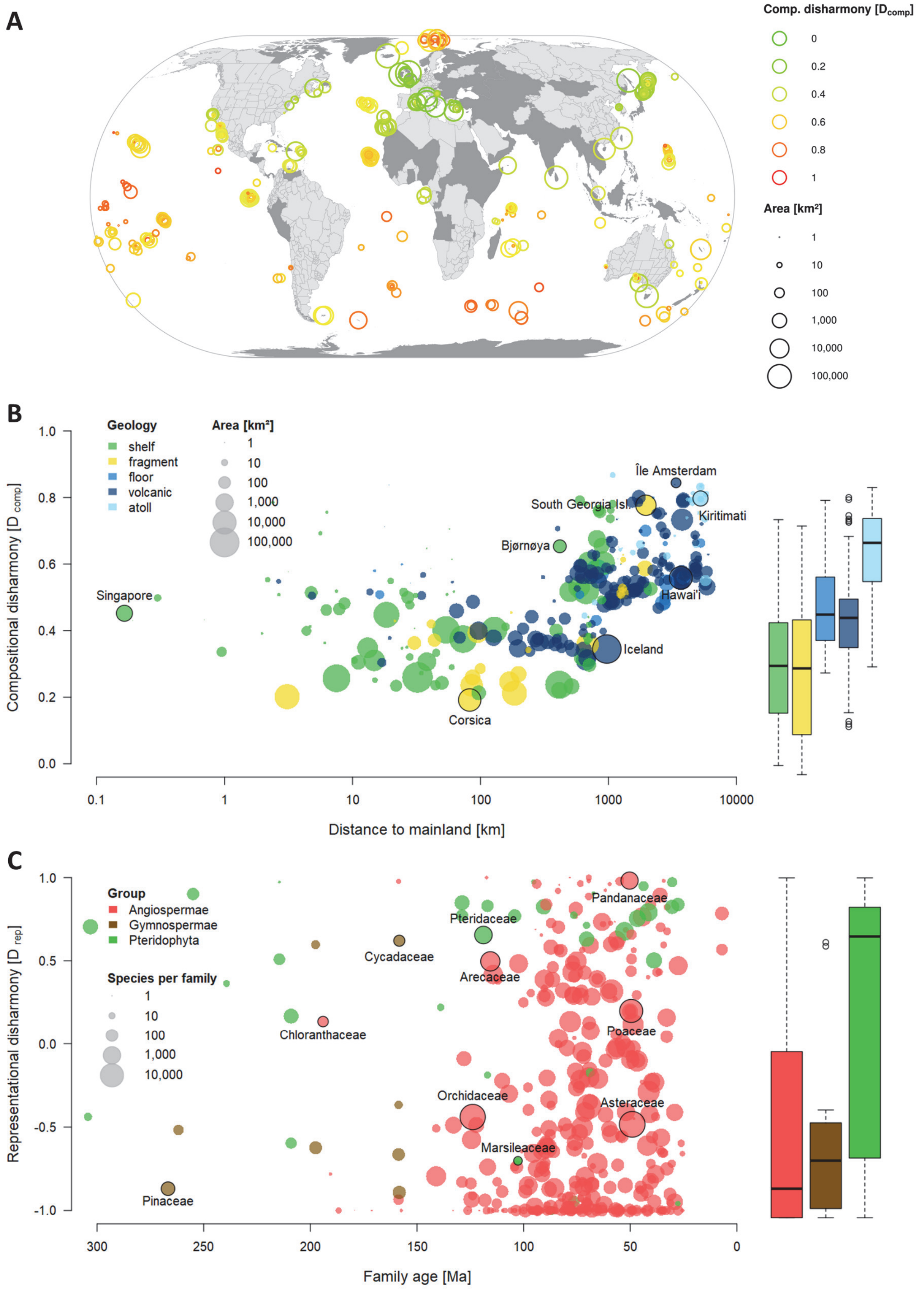


Figure 4.3: Global patterns in floristic disharmony from an island- and a taxon-centred perspective. **(A)** Compositional disharmony of island floras worldwide (n = 320). Mainland regions that were used for floristic comparisons (n = 611) are colored in light grey. **(B)** Compositional disharmony of island floras as a function of island isolation (x-axis), area (dot size), and geological origin (dot color). Higher values of D_{comp} indicate higher compositional dissimilarity relative to potential source regions on the mainland. **(C)** Representational disharmony of vascular plant families (n = 450) as a function of family age (x-axis), species per family (dot size), and taxonomic group (dot color). Values of D_{rep} above and below zero indicate a proportional over- and under-representation on islands, respectively.

Representational disharmony varied widely among vascular plant families (Figure 4.3C). Most notably, pteridophyte families were greatly over-represented on islands, whereas angiosperm and especially gymnosperm families tended to be under-represented. A few families such as Cycadaceae (gymnospermae, D_{rep} = 0.62) or Marsileaceae (pteridophytes, D_{rep} = -0.70) deviated from this general pattern. According to the multi-predictor model (Table 4.1), family species richness and taxonomic affiliation, but not family age had significant effects on representational disharmony. However, the explanatory power of the model was relatively low (R^2 = 0.19). Moreover, the relationship between family-level functional traits and representational disharmony was weak. Only in the case of pollination syndrome, we found significantly higher values of D_{rep} for predominantly wind-pollinated families compared to predominantly insect-pollinated families (Figure A4.2). We did not find significant effects of growth form, self-compatibility, seed mass, fruit type and dispersal syndrome (Figure A4.2). In agreement with these results, we did not find a significant phylogenetic signal in D_{rep} for seed plants (C_{mean} = 0.05, p = 0.071), but only for all vascular plants including the strongly over-represented pteridophyte clade (C_{mean} = 0.16, p = 0.001, Figure A4.3).

4.5 Discussion

Using a novel approach for estimating island source regions, we demonstrated that island disharmony can be assessed quantitatively. Our results show that compositional disharmony is a common feature of island floras worldwide, and that the magnitude of this effect clearly depends on the classical biogeographical variables of isolation and area, but is modulated by climatic conditions and geological history of an island. We found less clear relationships between family-specific characteristics and representational disharmony, i.e. the proportional over- or under-representation of individual taxa on islands relative to the mainland. Here, the most important predictor variable was a simple categorization of families into angiosperms, gymnosperms, and pteridophytes. In addition, species richness and pollination syndrome had significant effects on representational disharmony, whereas all other functional traits (growth form, dispersal syndrome, ability to self-pollinate, seed mass and fruit type) remained without effect.

4.5.1 A new approach for estimating floristic source regions

One key innovation of the present study is the outlined method for estimating floristic source regions based on a statistical model of species turnover. Source regions are typically defined as discrete, relatively large geographical units such as continents, biogeographical regions or countries (Fosberg, 1992; Keppel *et al.*, 2009; Keeley & Funk, 2011). This coarse-grained approach is often necessitated by the broad geographical scope of floristic data sources (e.g. regional checklists, inventories, Floras), but, nonetheless, has proven to be a valuable tool for understanding the origins of island floras. However, the recent explosion in ecological data availability (Hampton *et al.*, 2013) has enabled more highly resolved, quantitative approaches to delineate species pools (Carstensen *et al.*, 2013). Currently, the most versatile methods are based on ensembles of empirical (Graves & Rahbek, 2005) or modelled (Karger *et al.*, 2016) species distributions. Such species-level approaches are theoretically well suited for identifying source regions of a given island, but practically require quasi-complete knowledge on the spatial distribution of all species occurring there. At the moment, however, complete knowledge of global plant distributions is still beyond reach (Meyer *et al.*, 2016). Our approach, in contrast, models the proportion of shared species directly, instead of generating it from a compilation of complex species-level information. Consequently, source pool estimates based on species turnover are considerably less data-intensive than methods that require individual species distributions, while offering much finer spatial grain sizes and larger extents than empirical source pool reconstructions (see Figure 4.2). Considering that for the majority of islands no quantitative estimates of floristic source regions are available, the proposed method might add valuable aspects to island biogeographical research.

Island colonization is an accumulative process, acting over millions of years. The extended temporal dimension introduces uncertainties to the estimation of island source regions because the climate, habitat distribution, position, size and shape of both islands (Whittaker *et al.*, 2008; Weigelt *et al.*, 2016) and source regions(Galley & Linder, 2006; Pokorny *et al.*, 2015) may have changed considerably since colonization. Cronk (1987) illustrates this for the flora of Saint Helena, which is mostly derived from a now-extinct wet forest flora that occupied large parts of Southern Africa during the Miocene. Moreover, the effective isolation of an island is difficult to quantify and depends not only on the distance to the mainland, but also on the availability of stepping stones and the direction of predominant sea and wind currents (Cook & Crisp, 2005; Weigelt & Kreft, 2013), as well as the dispersal abilities of the focal taxon. This is illustrated by the overestimated importance of Central American regions in our prediction of Hawaiian source regions (see Figure 4.2). Although Hawaii is situated closer to North America than to Asia and Australia, it is effectively less isolated from the latter continents because of interspersed Pacific islands that facilitate stepping-stone colonization (Carlquist, 1967). Another potential source of uncertainty is that different habitats or elevational zones of an island may sample from distinct source pools on the mainland, and thus vary in their degree of isolation (Steinbauer *et al.*, 2012). A stronger consideration of the above-mentioned factors would certainly yield more accurate estimates of floristic source regions on the one hand, but adds much complexity on the other. Further methodological refinements and an extensive validation against empirical source pool reconstructions may help to find the appropriate balance between complexity and utility.

Turnover-based estimates of species source regions may be applied to research questions beyond island disharmony. In conservation planning, for example, taxon-specific source region estimates could be derived from highly resolved distributional data such as vegetation plots or small-scale checklists to inform the design of regional ecological corridors and protected areas (Socolar *et al.*, 2016). At larger scales, our approach could help to identify (potential) sources of species invasions and to implement measures preventing further introductions from such regions. Island conservation might particularly benefit from these tools, since island biotas are disproportionally affected by species introductions and naturalizations (van Kleunen *et al.*, 2015; Bellard *et al.*, 2016; Pyšek *et al.*, 2017).

4.5.2 Determinants of compositional and representational disharmony

Dispersal filtering has long been regarded as the predominant process in the assembly of island biotas, and therefore the main driver of disharmony (Carlquist, 1966a, 1967). Our results reveal that this is only partly true. On the one hand, the strong effect of isolation on island disharmony indeed suggests a major role of dispersal filtering in removing less dispersive taxa from the set of potential colonizers of an island. All gymnosperms except for Araucariaceae and Cycadaceae were under-represented on oceanic islands, which seems to fit the classical notion of gymnosperms as poor dispersers. Likewise, pteridophytes – possessing superior long-distance dispersal capabilities via ultra-light spores – were found to be strongly over-represented in island floras. These findings are in line with previous studies, which interpreted these broad taxonomic patterns as the outcome of selective dispersal filtering (Kreft *et al.*, 2010; Rumeu *et al.*, 2014; Weigelt *et al.*, 2015). On the other hand, the strong effects of area, temperature seasonality, and geological origin on compositional disharmony, as well as the relationship between representational disharmony and pollination- but not dispersal-related traits, indicate an important role of non-dispersal related processes. Pollination is increasingly recognized as a critical factor for the colonization of islands (Olesen *et al.*, 2010; Alsos *et al.*, 2015; Grossenbacher *et al.*, 2017). Given the general scarcity of animal pollinators on islands, abiotic pollination syndromes and the ability to self-pollinate are advantageous compared to biotic pollination or strict outcrossing (Baker, 1955). Indeed, we found wind-pollinated families to be much more evenly represented on islands than predominantly insect-pollinated families, whereas all other family-level traits had no detectable effects on representational disharmony. The lacking effect of classical dispersal traits such as seed mass, dispersal syndrome or fruit type seems to contradict longstanding assumptions regarding their relevance for island colonization (Carlquist, 1974; Howe & Smallwood, 1982; Portnoy & Willson, 1993). Indeed, while a relationship between such traits and dispersal distance is evident at small scales up to a few kilometers (Tackenberg *et al.*, 2003), long-distance dispersal seems to operate with such high levels of stochasticity and complexity that this relationship vanishes (Higgins *et al.*, 2003; Nathan, 2006; Nogales *et al.*, 2012; Le Roux *et al.*, 2014). Moreover, many large-seeded species that are adapted to dispersal by birds or seawater are successful long-distance dispersers, defying the common notion of small seeds as indicator of good dispersibility. Other studies even find no relationship at all between dispersal traits and colonization success (Heleno & Vargas, 2014).

Abiotic factors such as climate or soil substrate also act as filters for colonizing plant species, as recently demonstrated for the Galapagos Islands (Carvajal-Endara *et al.*, 2017) and New

Caledonia (Isnard *et al.*, 2016). The strong effect of temperature-related variables in our (Table 4.1) seems to support (abiotic) environmental filtering as an important driver of island disharmony. However, the climatic variables in our models did not reflect climatic similarity to the mainland (which would be a plausible cause of disharmony, but was accounted for during source pool estimation), but the average conditions of the islands themselves. Potential explanations for the positive relationship between compositional disharmony and island temperature and temperature seasonality include (1) stronger environmental filtering on islands with temperate or polar climates, (2) higher speciation rates on warm tropical islands, or (3) glacial dynamics limiting the available time for colonization on cold, high-latitude islands. In addition, separating abiotic and biotic drivers of community assembly is often difficult because competitors may preclude the establishment of colonizing species in generally suitable habitats, thus creating artificial environmental gradients in the composition of communities (Kraft *et al.*, 2015).

Even though the above considerations provide plausible explanations for the role of ecological factors in creating island disharmony, we want to stress that compositional disharmony, i.e. the overall compositional bias of island floras relative to their mainland source regions, is affected by neutral sampling effects related to species richness. Island floras can never be a perfect compositional representation of the much larger mainland source pool, because the number of families on an island is constrained by the number of species. Thus, extremely small proportions that arise, for example, in the case of rare families in species-rich mainland floras cannot be reproduced on islands, which inevitably increases compositional disharmony with decreasing species number (Figure A4.4). This interpretation moreover implies that compositional disharmony is also dependent on the size and spatial extent of the mainland source pool, as larger source pools usually include a higher number of taxa and are thus less likely to be represented harmonically in an island flora. This dependency might provide a further piece in the puzzle of understanding the disharmonic floras of highly isolated islands such as Hawaii or the Azores (see Figure 4.2B,C), whose source pools often encompass different biogeographical regions and continents.

4.5.3 Disharmony – a necessarily vague concept?

If the overall bias of island floras is rather accurately predicted by geographical and climatic island features, why does the representation of individual families on islands seem so unrelated to their functional traits? We consider two aspects relevant here. First, we approximated family-level traits based on species-level information of varying completeness (Figure A4.2). Missing data is a common problem in trait-based ecology (Taugourdeau *et al.*, 2014; Penone *et al.*, 2014) and a major source of uncertainty and bias in the characterization of ecological patterns (Hortal *et al.*, 2015). We therefore caution that our findings on representational disharmony depict trends rather than a definitive assessment. Second, a taxonomic perspective – especially when focusing on a fixed taxonomic level – is not an ideal framework for examining the outcomes of the complex ecological processes that produce disharmonic island floras. In the scientific literature, examples of disharmonic floristic elements range from small genera (e.g. *Metrosideros* in Carlquist, 1966a) to large taxonomic clades (e.g. pteridophytes in Braithwaite, 1975). Some studies assemble several such examples for a particular island or island group in order to arrive at a more general conclusion (Carlquist, 1967; Whittaker *et al.*, 1997). In fact, this is a reasonable approach because the degree to which taxa are consistently over- or under-represented on islands depends on their uniformity in terms of colonization success, and thus in terms of dispersal abilities, environmental tolerances and degree of biotic specialization. These parameters may vary greatly even within small families (Howe & Smallwood, 1982), but on the other hand show remarkable consistency within large taxonomic groups (Farjon, 2010), such that the required taxonomic level of observation is variable. Consequently, the classical concept of island disharmony *sensu* Carlquist (1965, 1974) – albeit immensely useful and influential – generalizes poorly across taxonomic groups.

4.5.4 Conclusion

We demonstrated that island disharmony is a common feature of insular floras worldwide, and that the traditional concept can be put on a quantitative footing. The generality and predictive power of the concept of island disharmony has historically been limited by its focus on taxonomic categorizations. However, given the rapid advances in ecological data availability and analytical tools, the approximation of ecological characteristics by means of taxonomic affiliation will eventually become obsolete. Instead, the assembly mechanisms that disharmony aims to reflect and explain can be investigated more directly using quantitative methods that are informed by functional and phylogenetic data. The outlined approach for the estimation of island source regions as well as the mathematical indices for quantifying island disharmony provide a big step towards this objective, and offer robust empirical insights into the factors shaping the composition of island floras.

General Discussion

Summary and contribution of this thesis

The present thesis contributes to the fields of macroecology and biogeography on two different levels.

Firstly, I provide important conceptual and practical advances on the subject of biodiversity data integration. I highlighted the significance of aggregated data types as unique sources of plant-ecological information and made the case for a targeted integration of diversity data across multiple domains (e.g. species distributions, functional traits, phylogenies) and levels of aggregation (e.g. point occurrences, vegetation plots, checklists). I showed that aggregated data types improve the representativeness and completeness of biodiversity data in general and constitute a crucial component in building a robust global baseline for plant diversity research (Chapter 1). I substantiated these points with three case studies that provide new insights into long-standing biogeographical questions regarding the global distribution of plant growth forms, the latitudinal gradient of seed mass, and the global prevalence of insular woodiness. All case studies were based on the Global Inventory of Floras and Traits database (GIFT, Chapter 2), which integrates hundreds of ecological data resources to support comprehensive analyses of taxonomic, functional and phylogenetic plant diversity at global scales. GIFT achieves substantially higher data coverage than established data providers in terms of coarse-grained species distributions (regional checklists) and functional traits (categorical traits on plant morphology and life history) and thus proves the effectiveness of the concepts outlined in Chapter 1. As such, GIFT fills a previously unoccupied niche in the macroecological data landscape and will continue to contribute to the knowledge of global plant diversity in the future.

Secondly, I make empirical contributions towards a better understanding of assembly processes on islands. Leveraging the high data coverage provided by GIFT, I investigated global patterns and drivers of island beta diversity (Chapter 3) and disharmony (Chapter 4). In Chapter 3, I showed that plant speciese assemblages on islands generally tend to be more similar – i.e. exhibit lower beta diversity – than on the mainland, and that this homogenization effect is most pronounced for geographic settings characterized by strong ecological filtering (e.g. remote, environmentally homogeneous atolls) and for plant groups that are relatively insensitive to these filters (e.g. highly dispersive pteridophytes). These findings add to the understanding of assembly processes on islands and demonstrate how beta diversity patterns can be used to evaluate the relative importance of ecological filters in various contexts. In Chapter 4, I addressed the phenomenon of island disharmony, which describes the biased representation of higher taxa on islands compared to their mainland source regions. I approached this subject from two perspectives, differentiating between island-specific factors potentially related to the overall disharmony of island floras (e.g. isolation, area, etc.), and taxon-specific characteristics potentially related to the overall representation of individual families on islands wordlwide (e.g. species number, dispersal traits, etc.). Representing the first quantitative, macroecological assessment of island disharmony to this date, my results show that the compositional bias of island floras is rather accurately predicted by island area, isolation, climate, whereas the representation of families on islands is only weakly related to family-specific characteristics. These findings suggest that a focus on taxa – a cornerstone of the classical concept of disharmony – is insufficient to capture the outcomes of complex ecological and evolutionary processes, rendering the increasing utilization of functional and phylogenetic approaches in biogeography a necessary next step.

Challenges and future perspectives

Data

GIFT is specifically designed to support large-scale analysis of plant diversity at high levels of geographical, taxonomic and functional coverage (see Chapters 1.4.1, 1.4.2, 1.4.3, 3 and 4). However, broad data coverage generally comes at the expense of fine-grained resolution (Mora *et al.*, 2008; Hortal *et al.*, 2015), which puts questions concerning local species abundances, community dynamics, infraspecific trait variation or biotic interactions currently beyond the scope of GIFT. Consequently, many of the pressing challenges imposed by global change cannot be conclusively addressed with GIFT.

Since the mid-twentieth century alone, human activity has caused an estimated 0.65 °C increase in global mean temperature (Ribes *et al.*, 2017), more than 17,000 introductions of alien species (Seebens *et al.*, 2017), and the loss of more than 2.3 million square kilometres of tropical rainforest (Rosa *et al.*, 2016). The accelerating pace of these and other anthropogenic pressures on our biosphere is projected to further amplify the detrimental impacts on populations, species and ecosystems worldwide (Tilman *et al.*, 1994; Vitousek *et al.*, 1997; Hoegh-Guldberg, 1999; Serreze & Barry, 2011; Newbold *et al.*, 2015; Hallmann *et al.*, 2017). Since the effects of human activity occur first at local scales before they emerge as consistent global patterns (Cardinale *et al.*, 2012; Steffen *et al.*, 2015), any strategy to counteract the ecological impacts of global change effectively has to operate at high spatiotemporal resolutions.

Recent initiatives such as PREDICTS (Hudson *et al.*, 2017) or BioTIME (Dornelas *et al.*, 2018) integrate highly resolved primary data that has been collected with a clearly defined purpose. Specifically, these two projects focus on disaggregated datasets (mostly at the plot-level) that contrast the biodiversity of multiple sites under different land-use regimes (PREDICTS) or document changes in the biodiversity of individual sites over time (BioTIME). Having amassed several million species-by-sites records each, PREDICTS and BioTIME represent substantial advancements towards establishing clear cause-effect relationships between human activities and local ecological responses. Complementary approaches towards this objective are collaborative research networks such as the International Long Term Ecological Research Network (ILTER, Vanderbilt & Gaiser, 2017) or the recently proposed Global Island Monitoring Scheme (GIMS, Borges *et al.*, 2018). Other than projects that integrate data that has already been collected, these initiatives can implement highly standardized sampling protocols and specifically target taxa, ecosystems or geographical regions of high ecological importance. This is particularly relevant with respect to island biotas, which are highly threatened due to small population sizes, low functional redundancy and highly specialized endemic species (Fordham & Brook, 2010), but generally under-represented in global biodiversity databases or established research networks (Borges *et al.*, 2016). Considering that islands are hotspots of historical species extinctions (Diamond *et al.*, 1989) and will continue to be disproportionately affected by climate change, habitat loss and species invasions (Sax & Gaines, 2008; Caujapé-Castells *et al.*, 2010; van Kleunen *et al.*, 2015; Harter *et al.*, 2015), island-focused initiatives such as GIMS are urgently needed.

Despite the high data coverage of GIFT compared to initiatives focussing on the disaggregated end of the data spectrum (see Chapter 1), there are remaining data gaps (see Chapter 2). The current priority of the GIFT initiative is therefore to further improve the coverage of distributional and functional trait data in order to support increasingly *accurate* inferences about global plant diversity. The difference in checklist coverage between Chapter 3 (258 islands and 346 mainland units as of August 2016) and Chapter 4 (320 islands and 611 mainland units as of May 2018) demonstrates the significant progress in this regard. Once GIFT has reached near-global coverage in terms of coarse-grained checklist data and simple functional traits, further efforts will be directed towards increasing data resolution, i.e. focussing data collection on species checklists for small geographical units or numerical functional traits. This will allow for increasingly *precise* inferences about global plant diversity and hopefully close the gap in resolution to complementary disaggregated ecological data resources.

Methodology

A recurring theme in this thesis is the call for a stronger utilization of functional and phylogenetic approaches to enhance classical taxonomy-based concepts in biogeography. The fundamental drawback of taxonomy is that it is not an accurate reflection of eco-evolutionary reality. The hierarchical structure of taxonomic classifications implies that organisms fall into well-defined groups of common descent (Hennig, 1966) or common appearance (Simpson, 1961; Mayr, 1969) – ideally both. In reality, however, this is rarely the case (de Queiroz, 1996; Ereshefsky, 2002). Higher taxonomic groups (e.g. genera or families) are constantly extended, merged or split based on rather flexible and subjective criteria (Dubois, 2007; Humphreys & Linder, 2009). Also relatively stable taxa of the same rank may vary widely in their age, diversification rate and phenotypic diversity (Magallón & Castillo, 2009; Cornwell *et al.*, 2014). Even the most fundamental taxonomic unit – the species – defies an objective, universal definition to this date (Wiley, 1978; Mallet, 1995; Aldhebiani, 2018, but see e.g. Barraclough & Humphreys, 2015).

The inconsistencies of taxonomic classifications may affect ecological and evolutionary inferences. For example, biogeographical regionalizations have been based on geographical patterns in (taxonomic) endemism for more than a century (Morrone, 2015). Regional boundaries were defined either qualitatively (e.g. Wallace, 1876; Takhtajan, 1986) or quantitatively (Kreft *et al.*, 2010), but generally based on the amount and rank (species, genus, family, etc.) of endemic taxa in a given area. However, as laid out above, taxonomy is not an ideal proxy for the ancestral relationships – and hence biogeographical affinities – of species assemblages. Phylogenetic approaches offer continuous and evolutionarily more realistic estimates of assemblage dissimilarity (Graham & Fine, 2008), thus allowing for a more objective and robust delineation of biogeographical regions (Holt *et al.*, 2013; Hattab *et al.*, 2015; Daru *et al.*, 2015). With my study on geographical patterns in species turnover, I examined a closely related topic in Chapter 3. Although the focus of this study was not a biogeographical regionalization, here too, the use of a phylogenetic measure would have increased the informational resolution and allowed for a differentiation between biogeographical (deep-time) and ecological (shallow-time) drivers of species turnover (see e.g. Rosauer *et al.*, 2014).

The lack of biological realism in taxonomic classifications has also been a major reason for the rise of functional approaches (Keddy, 1992; Woodward & Cramer, 1996; Lavorel *et al.*, 2007). Functional ecology aims to approximate fundamental organismal properties such as competitive ability or dispersal capacity by means of easily measurable traits such as vegetative height or seed mass (Lavorel & Garnier, 2002). Based on these traits, species can be grouped into functionally consistent groups that – in theory – respond similarly under a given set of conditions (Díaz & Cabido, 1997; Duckworth *et al.*, 2000). That such a grouping is not reliably achievable based on taxonomy was clearly shown in my research on island disharmony (Chapter 4). I found that the representation of plant families on islands was only weakly related to family-level functional traits that are widely considered to be associated with colonization success. This lack of a relationship was most plausibly explained by the high functional variation within families, suggesting that taxonomic groups are not always a suitable unit of observation for understanding complex biogeographical patterns such as island disharmony. The potential of focussing on functional rather than taxonomic biogeographical patterns is showcased in Chapters 1.4.1 and 1.4.2, where I derive global predictions of plant growth from composition and assess the latitudinal gradient in seed mass. Indeed, the recent increase in functional biogeographical studies indicates a growing recognition of such approaches (Violle *et al.*, 2014; Whittaker *et al.*, 2014; Engemann *et al.*, 2016; Zanne *et al.*, 2018).

Combining functional with phylogenetic approaches offers an immensely powerful toolkit to answer questions regarding the past, present and future of biodiversity. In a seminal study, Zanne *et al.* (2014) reconstructed the evolution of woodiness, leaf phenology and conduit diameter to understand the radiation of angiosperms into freezing environments. Their study demonstrated impressively the interplay of physiological, ecological, and evolutionary processes in shaping biogeographical patterns. Similar approaches can be applied in island biogeography, e.g. to assess the direction and rate of adaptive processes on islands. I touched on this subject in Chapter 1.4.3, where I provided strong evidence for a globally consistent evolutionary trend towards increased woodiness in island plant assemblages. A generalization of this case study could reconstruct key functional traits (Díaz *et al.*, 2016) on a representative set of islands based on phylogenetic relationships, to then identify general temporal patterns in island functional diversity and trait space dimensions. The results would be a valuable addition to existing studies at smaller taxonomic and geographical scales (e.g. Givnish *et al.*, 2009; Lens *et al.*, 2013) and help resolve competing hypothesis concerning the origin of island syndromes such as insular woodiness (Darwin, 1859; Wallace, 1881; Carlquist, 1965). Another interesting example for the synergies between phylogenetic and functional methods is given by Swenson *et al.* (2017), who predicted geographic variations in functional traits of the European tree flora based on phylogenetic and environmental relationships of the North American one. Although the authors acknowledged the imperfect predictability of some traits due to a weak phylogenetic signal, their rigorously predictive approach should serve as a model in other contexts. For example, it could be adapted to assess biogeographical theory by testing predictions about the properties of island floras based on the phylogenetic and functional characteristics of their source pools (see Chapter 4) and a hypothesized, island-specific set of ecological filters.

Theory

Although the increasing availability of data and methodological frameworks has sparked substantial progress in ecology, a predictive understanding of ecological systems ultimately requires a robust theoretical grounding (Maurer, 2000; Marquet *et al.*, 2014; Shou *et al.*, 2015). In general, robust theories should be derived from first principles, i.e. well-established axioms that clearly link a specific mechanism to a given class of phenomena, and generate as many predictions from as few assumptions as possible (Marquet *et al.*, 2014). Macroecology features several theories that generally match these criteria. The Metabolic Theory of Ecology (Brown *et al.*, 2004), for example, predicts ecological patterns at multiple scales, including population dynamics, statistical distributions of life-history traits, and latitudinal gradients in species diversity and ecosystem productivity from scaling relationships between metabolic rate, body size and temperature. The Neutral Theory of Biodiversity (Hubbell, 2001) produces realistic predictions of species-area relationships and abundance distributions from stochastic demographic processes (immigration, speciation, extinction) of ecologically identical "species". The Maximum Entropy Theory of Ecology (Harte, 2011) uses principles from information theory to predict accurate statistical distributions of e.g. abundance, body size, or the geographic density of species using only four variables to describe the state of an ecological system (number of individuals, number of species, total metabolic rate, area).

None of the above-mentioned theories explains all (or even the majority of) observable ecological phenomena. However, they provide logically consistent starting points for extensions and conceptual syntheses, as e.g. demonstrated by Enquist *et al.* (2015), who recast metabolic theory in terms of functional trait distributions, or Worm & Tittensor (2018), who synthesized aspects from neutral, metabolic and niche theory into a global theory of biodiversity. Most importantly, theoretical frameworks generate testable hypotheses about the core components and processes in ecological systems (see e.g. McGill, 2003, Hawkins *et al.*, 2007, Xiao *et al.*, 2015 for tests of the above-mentioned theories) and, in turn, depend critically on the empirical validation or falsification of these hypotheses (Maurer, 2000; Scheiner, 2013). This positive feedback between theory and empiricism can be greatly accelerated by integrated databases such as GIFT, which provide increasingly comprehensive and representative biodiversity data for rapid and conclusive hypothesis testing. Thus, the present thesis contributes to island biogeographical research not only through the detailed study of particular aspects of island plant diversity (Chapters 3 and 4), but also in a broader sense by outlining and implementating a valuable resource for future empirical studies (Chapters 1 and 2).

References

Abouheif, E. (1999). A method for testing the assumption of phylogenetic independence in comparative data. *Evolutionary Ecology Research*, **1**, 895–909.

Acevedo-Rodríguez, P. & Strong, M.T. (2007). Catalogue of the seed plants of the West Indies Website (last accessed: November 05, 2014).

Achen, C.H. (1982). *Interpreting and using regression.* Sage Publications.

Agarwal, S. & Varshavsky, A. (eds.) (2013). *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications.* ACM.

Aldhebiani, A.Y. (2018). Species concept and speciation. *Saudi Journal of Biological Sciences*, **25**, 437–440.

Alsos, I.G., Ehrich, D., Eidesen, P.B., Solstad, H., Westergaard, K.B., Schönswetter, P., Tribsch, A., Birkeland, S., Elven, R. & Brochmann, C. (2015). Long-distance plant dispersal to North Atlantic islands. Colonization routes and founder effect. *AoB PLANTS*, **7**.

Arrhenius, O. (1921). Species and area. *Journal of Ecology*, **9**, 95.

Azaele, S., Maritan, A., Cornell, S.J., Suweis, S., Banavar, J.R., Gabriel, D. & Kunin, W.E. (2015). Towards a unified descriptive theory for spatial ecology: predicting biodiversity patterns across spatial scales. *Methods in Ecology and Evolution*, **6**, 324–332.

Baker, H.G. (1955). Self-compatibility and establishment after "long-distance" dispersal. *Evolution*, **9**, 347–349.

Baldeck, C.A., Harms, K.E., Yavitt, J.B., et al. (2013). Soil resources and topography shape local tree community structure in tropical forests. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20122532.

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models. How, where and how many? *Methods in Ecology and Evolution*, **3**, 327–338.

Barnard, P. & Thuiller, W. (2008). Introduction. Global change and biodiversity. Future challenges. *Biology Letters*, **4**, 553–555.

Barraclough, T.G. & Humphreys, A.M. (2015). The evolutionary reality of species and higher taxa in plants: a survey of post ‐ modern opinion and evidence. *New Phytologist*, **207**, 291–296.

Barthlott, W., Burstedde, K., Geffert, J.L., Ibisch, Pierre L., Korotkova, N., Miebach, A., Rafiqpoor, M.D., Stein, A. & Mutke, J. (2015). *Biogeography and biodiversity of cacti.* Univ.-Verlag Isensee, Germany.

Barthlott, W., Mutke, J., Rafiqpoor, D., Kier, G. & Kreft, H. (2005). Global centers of vascular plant diversity. *Nova Acta Leopoldina NF*, **92**, 61–83.

Barton, K. (2016). *MuMIn: Multi-model inference (R-Package).*

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.

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*, 7–17.

Beck, J., Ballesteros-Mejia, L., Buchmann, C.M., Dengler, J., Fritz, S.A., Gruber, B., Hof, C., Jansen, F., Knapp, S., Kreft, H., Schneider, A.-K., Winter, M. & Dormann, C.F. (2012). What's on the horizon for macroecology? *Ecography*, **35**, 673–683.

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*, 1–36.

Beentje, H. (2016). *The Kew Plant Glossary. An illustrated dictionary of plant terms*, 2nd edn. Royal Botanic Gardens, Kew.

Bellard, C., Cassey, P. & Blackburn, T.M. (2016). Alien species as a driver of recent extinctions. *Biology Letters*, **12**, 20150623.

Bello, F. de, Lavorel, S., Díaz, S., et al. (2010). Towards an assessment of multiple ecosystem processes and services via functional traits. *Biodiversity and Conservation*, **19**, 2873–2893.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D.L. (2005). GenBank. *Nucleic acids research*, **33**, D34-8.

Bernardello, G., Anderson, G.J., Stuessy, T.F. & Crawford, D.J. (2006). The angiosperm flora of the Archipelago Juan Fernandez (Chile). Origin and dispersal. *Canadian Journal of Botany*, **84**, 1266–1281.

BGCI (2017). GlobalTreeSearch online database (last accessed: July 31, 2017).

Bin, Y., Wang, Z., Wang, Z., Ye, W., Cao, H. & Lian, J. (2010). The effects of dispersal limitation and topographic heterogeneity on beta diversity and phylobetadiversity in a subtropical forest. *Plant Ecology*, **209**, 237–256.

BirdLife International (2018). Taxonomic Checklist of the Birds of the World (last accessed: February 20, 2018).

Blonder, B., Lamanna, C., Violle, C. & Enquist, B.J. (2014). The n -dimensional hypervolume. *Global Ecology and Biogeography*, **23**, 595–609.

Blonder, B., Nogués-Bravo, D., Borregaard, M.K., Donoghue, J.C., Jørgensen, P.M., Kraft, N.J.B., Lessard, J.-P., Morueta-Holme, N., Sandel, B., Svenning, J.-C., Violle, C., Rahbek, C. & Enquist, B.J. (2015). Linking environmental filtering and disequilibrium to biogeography with a community climate framework. *Ecology*, **96**, 972–985.

Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K. & Mace, G.M. (2010). Distorted Views of Biodiversity. Spatial and Temporal Bias in Species Occurrence Data. *PLoS Biology*, **8**, e1000385.

Böhle, U.R., Hilger, H.H. & Martin, W.F. (1996). Island colonization and evolution of the insular woody habit in Echium L. (Boraginaceae). *Proceedings of the National Academy of Sciences of the United States*, **93**, 11740–11745.

Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M., Novak, M., Rudolf, V.H.W., Schreiber, S.J., Urban, M.C. & Vasseur, D.A. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology & Evolution*, **26**, 183–192.

Borcard, D., Legendre, P. & Drapeau, P. (1992). Partialling out the Spatial Component of Ecological Variation. *Ecology*, **73**, 1045–1055.

Borges, P.A.V., Cardoso, P., Gabriel, R., Ah-Peng, C. & Emerson, B.C. (2016). Challenges, advances and perspectives in island biogeography. *Frontiers of Biogeography*, **8.2**, e29136.

Borges, P.A.V., Cardoso, P., Kreft, H., et al. (2018). Global Island Monitoring Scheme (GIMS). A proposal for the long-term coordinated survey and monitoring of native island forest biota. *Biodiversity and Conservation*, **39**, 1645.

Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J.A., Mozzherin, D., Rees, T., Matasci, N., Narro, M.L., Piel, W.H., Mckay, S.J., Lowry, S., Freeland, C., Peet, R.K. & Enquist, B.J. (2013). The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics*, **14**, 16.

Brach, A.R. & Song, H. (2006). eFloras. New directions for online floras exemplified by the Flora of China Project. *Taxon*, **55**, 188–192.

Braithwaite, A.F. (1975). The phytogeographical relationships and origin of the New Hebrides fern flora. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **272**, 293.

Bramwell, D. & Caujapé-Castells, J. (eds.) (2011). *The biology of island floras.* Cambridge Univ. Press, Cambridge.

Brown, J.H. (1971). Mammals on Mountaintops. Nonequilibrium Insular Biogeography. *The American Naturalist*, **105**, 467–478.

Brown, J.H. (1995). *Macroecology.* University of Chicago Press, Chicago.

Brown, J.H., Gillooly, J.F., Allen, A.P., van Savage, M. & West, G.B. (2004). Toward a metabolic theory of ecology. *Ecology*, **85**, 1771–1789.

Brown, J.H. & Lomolino, M.V. (1998). *Biogeography*, 2nd edn. Sunderland MA: Sinauer Associates.

Brown, J.H. & Maurer, B.A. (1989). Macroecology: the division of food and space among species on continents. *Science*, **243**, 1145–1150.

Browne, R.A. (1981). Lakes as Islands. Biogeographic Distribution, Turnover Rates, and Species Composition in the Lakes of Central New York. *Journal of Biogeography*, **8**, 75.

Brunette, W., Sundt, M., Dell, N., Chaudhri, R., Breit, N. & Borriello, G. (2013). Open data kit 2.0. Expanding and refining information services for developing regions. *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications* (ed. by S. Agarwal and A. Varshavsky). ACM.

Bunnefeld, N. & Phillimore, A.B. (2012). Island, archipelago and taxon effects. Mixed models as a means of dealing with the imperfect design of nature's experiments. *Ecography*, **35**, 15–22.

Burns, K.C. (2005). A multi-scale test for dispersal filters in an island plant community. *Ecography*, **28**, 552–560.

Butler, E.E., Datta, A., Flores-Moreno, H., et al. (2017). Mapping local and global variability in plant trait distributions. *Proceedings of the National Academy of Sciences of the United States*, 201708984.

Cabral, J., Weigelt, P., Kissling, W.D. & Kreft, H. (2014). Biogeographic, climatic and spatial drivers differentially affect α-, β- and γ-diversities on oceanic archipelagos. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20133246.

Cadotte, M.W., Carscadden, K. & Mirotchnick, N. (2011). Beyond species: Functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology*, **48**, 1079–1087.

Cameron, R.A.D., Triantis, K.A., Parent, C.E., Guilhaumon, F., Alonso, M.R., Ibáñez, M., Frias Martins, A.M. de, Ladle, R.J., Whittaker, R.J. & McGeoch, M. (2013). Snails on oceanic islands. Testing the general dynamic model of oceanic island biogeography using linear mixed effect models. *Journal of Biogeography*, **40**, 117–130.

Canadell, J.G., Pataki, D.E. & Pitelka, L.F. (eds.) (2007). *Terrestrial Ecosystems in a Changing World.* Springer Berlin Heidelberg, Berlin, Heidelberg.

Cardillo, M., Gittleman, J.L. & Purvis, A. (2008). Global patterns in the phylogenetic structure of island mammal assemblages. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 1549–1556.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., et al. (2012). Biodiversity loss and its impact on humanity. *Nature*, **486**, 59.

Carlquist, S. (1965). *Island life. A natural history of the islands of the world*, 1st edn. Natural History Press, Garden City, N.Y.

Carlquist, S. (1966a). The biota of long-distance dispersal. I. Principles of dispersal and evolution. *The Quarterly Review of Biology*, **41**, 247–270.

Carlquist, S. (1966b). The biota of long - distance dispersal. II. Loss of dispersibility in Pacific Compositae. *Evolution*, **20**, 30–48.

Carlquist, S. (1966c). The biota of long-distance dispersal. III. Loss of dispersibility in the Hawaiian flora. *Brittonia*, **18**, 310–335.

Carlquist, S. (1966d). The biota of long - distance dispersal. IV. Genetic systems in the floras of oceanic islands. *Evolution*, **20**, 433–455.

Carlquist, S. (1967). The biota of long-distance dispersal. V. Plant dispersal to Pacific Islands. *Bulletin of the Torrey Botanical Club*, **94**, 129–162.

Carlquist, S. (1970). Wood anatomy of insular species of Plantago and the problem of raylessness. *Bulletin of the Torrey Botanical Club*, 353–361.

Carlquist, S. (1974). *Island Biology.* Columbia University Press, New York.

Carstensen, D.W., Lessard, J.-P., Holt, B.G., Krabbe Borregaard, M. & Rahbek, C. (2013). Introducing the biogeographic species pool. *Ecography*, **36**, 1310–1318.

Carvajal-Endara, S., Hendry, A.P., Emery, N.C. & Davies, T.J. (2017). Habitat filtering not dispersal limitation shapes oceanic island floras. Species assembly of the Galápagos archipelago. *Ecology Letters*, **20**, 495–504.

Caujapé-Castells, J., Tye, A., Crawford, D.J., Santos-Guerra, A., Sakai, A., Beaver, K., Lobin, W., Vincent Florens, F.B., Moura, M., Jardim, R., Gómes, I. & Kueffer, C. (2010). Conservation of oceanic island floras: Present and future global challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **12**, 107–129.

Cayuela, L., La Granzow-de Cerda, Í., Albuquerque, F.S. & Golicher, D.J. (2012). taxonstand. An r package for species names standardisation in vegetation databases. *Methods in Ecology and Evolution*, **3**, 1078–1083.

Chamberlain, S.A. & Szöcs, E. (2013). taxize. Taxonomic search and retrieval in R. *F1000Research.*

Chase, J.M. (2014). Spatial scale resolves the niche versus neutral theory debate. *Journal of Vegetation Science*, **25**, 319–322.

Christenhusz, M.J.M. & Byng, J.W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa*, **261**, 201–217.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, **12**, 2493–2537.

Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., Núñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H.C., Losos, E. & Hubbell, S.P. (2002). Beta-diversity in tropical forest trees. *Science*, **295**, 666–669.

Connor, E.F. & McCoy, E.D. (1979). The statistics and biology of the species-area relationship. *The American Naturalist*, 791–833.

Cook, L.G. & Crisp, M.D. (2005). Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography. *Journal of Biogeography*, **32**, 741–754.

Copernicus Global Land Service (2018). Global layers of biosphysical variables, Available at: https://land.copernicus.eu/global/products/.

Corney, D., Clark, J.Y., Tang, H.L. & Wilkin, P. (2012). Automatic extraction of leaf characters from herbarium specimens. *Taxon*, **61**, 231–244.

Cornwell, W.K., Westoby, M., Falster, D.S., et al. (2014). Functional distinctiveness of major plant lineages. *Journal of Ecology*, **102**, 345–356.

Cramer, W., Kicklighter, D.W., Bondeau, A., Iii, B.M., Churkina, G., Nemry, B., Ruimy, A., Schloss, A.L. & Intercomparison, T.P.O.T.P.N.M. (1999). Comparing global models of terrestrial net primary productivity (NPP). Overview and key results. *Global Change Biology*, **5**, 1–15.

Cronk, Q.C.B. (1987). The history of endemic flora of St. Helena: A relictual series. *New Phytologist*, **105**, 509–520.

Danielson, J.J. & Gesch, D.B. (2011). *Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010),* Reston, Virginia.

Daru, B.H., van der Bank, M., Maurin, O., Yessoufou, K., Schaefer, H., Slingsby, J.A. & Davies, T.J. (2015). A novel phylogenetic regionalization of phytogeographical zones of southern Africa reveals their hidden evolutionary affinities. *Journal of Biogeography*, **43**, 155–166.

Darwin, C.R. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, 1st edn. John Murray, London.

Darwin, C.R. & Wallace, A.R. (1858). *On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection,* London.

de Queiroz, K. (1996). The Linnaean Hierarchy and the Evolutionization of Taxonomy, with Emphasis on the Problem of Nomenclature. *Aliso*, **15**, 125–144.

Diamond, J.M., Ashmole, N.P. & Purves, P.E. (1989). The Present, Past and Future of Human-Caused Extinctions [and Discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **325**, 469–477.

Díaz, S. & Cabido, M. (1997). Plant functional types and ecosystem function in relation to global change. *Journal of Vegetation Science*, **8**, 463–474.

Díaz, S., Kattge, J., Cornelissen, J.H.C., et al. (2016). The global spectrum of plant form and function. *Nature*, 167–171.

Díaz, S., Lavorel, S., Bello, F. de, Quetier, F., Grigulis, K. & Robson, T.M. (2007). Incorporating plant functional diversity effects in ecosystem service assessments. *Proceedings of the National Academy of Sciences of the United States*, **104**, 20684–20689.

Dornelas, M., Antão, L.H., Moyes, F., et al. (2018). BioTIME: A database of biodiversity time series for the Anthropocene. *Global Ecology and Biogeography*, **27**, 760–786.

Doxsey-Whitfield, E., MacManus, K., Adamo, S.B., Pistolesi, L., Squires, J., Borkovska, O. & Baptista, S.R. (2015). Taking Advantage of the Improved Availability of Census Data. A First Look at the Gridded Population of the World, Version 4. *Papers in Applied Geography*, **1**, 226–234.

Dubois, A. (2007). Phylogeny, taxonomy and nomenclature: the problem of taxonomic categories and of nomenclatural ranks. *Zootaxa*, **1519**, 27–68.

Duckworth, J.C., Kent, M. & Ramsay, P.M. (2000). Plant functional types: an alternative to taxonomic plant community description in biogeography? *Progress in Physical Geography*, **24**, 515–542.

Elbroch, M., Mwampamba, T.H., Santos, M.J., Zylberberg, M., Liebenberg, L., Minye, J., Mosser, C. & Reddy, E. (2011). The value, limitations, and challenges of employing local experts in conservation research. *Conservation biology : the journal of the Society for Conservation Biology*, **25**, 1195–1202.

Elith, J. & Leathwick, J.R. (2009). Species Distribution Models. Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Eller, F., Skálová, H., Caplan, J.S., et al. (2017). Cosmopolitan Species As Models for Ecophysiological Responses to Global Change. The Common Reed Phragmites australis. *Frontiers in Plant Science*, **8**.

Engemann, K., Enquist, B.J., Sandel, B., Boyle, B., Jørgensen, P.M., Morueta-Holme, N., Peet, R.K., Violle, C. & Svenning, J.-C. (2015). Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution*, **5**, 807–820.

Engemann, K., Sandel, B., Enquist, B.J., Jørgensen, P.M., Kraft, N.J.B., Marcuse-Kubitza, A., McGill, B., Morueta-Holme, N., Peet, R.K., Violle, C., Wiser, S. & Svenning, J.-C. (2016). Patterns and drivers of plant functional group dominance across the Western Hemisphere. A macroecological re-assessment based on a massive botanical dataset. *Botanical Journal of the Linnean Society*, **180**, 141–160.

Enquist, B.J., Condit, R., Peet, R.K., Schildhauer, M. & Thiers, B. (2016). Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. *PeerJ Preprints*, **4**, e2615v2.

Enquist, B.J., Norberg, J., Bonser, S.P., Violle, C., Webb, C.T., Henderson, A., Sloat, L.L. & Savage, V.M. (2015). Scaling from Traits to Ecosystems. Developing a general Trait Driver Theory via integrating trait-based and metabolic scaling theories. *Trait-Based Ecology - From Structure to Function* (ed. by S. Pawar, G. Woodward and A.I. Dell), pp. 249–318. Elsevier Science.

Ereshefsky, M. (2002). Linnaean Ranks: Vestiges of a Bygone Era. *Philosophy of Science*, **69**, S305-S315.

Fargione, J., Brown, C.S. & Tilman, D. (2003). Community assembly and invasion: An experimental test of neutral versus niche processes. *Proceedings of the National Academy of Sciences of the United States*, **100**, 8916–8920.

Farjon, A. (2010). *A Handbook of the World's Conifers*, 1st edn. Brill Academic Publishing.

Farjon, A. & Filer, D. (2013). *An atlas of the world's conifers. an analysis of their distribution, biogeography, diversity and conservation status.* Brill, Leiden, Netherlands.

Fattorini, S. (2010). The influence of geographical and ecological factors on island beta diversity patterns. *Journal of Biogeography*, **37**, 1061–1070.

Fegraus, E.H., Andelman, S., Jones, M.B. & Schildhauer, M. (2005). Maximizing the value of ecological data with structured metadata. An introduction to Ecological Metadata Language (EML) and principles for metadata creation. *Bulletin of the Ecological Society of America.*

Ferrier, S. (2002). Mapping Spatial Pattern in Biodiversity for Regional Conservation Planning: Where to from Here? *Systematic Biology*, **51**, 331–363.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, **13**, 252–264.

FitzJohn, R.G., Pennell, M.W., Zanne, A.E., Stevens, P.F., Tank, D.C. & Cornwell, W.K. (2014). How much of the world is woody? *Journal of Ecology*, **102**.

Fitzpatrick, M.C., Sanders, N.J., Normand, S., Svenning, J.-C., Ferrier, S., Gove, A.D. & Dunn, R.R. (2013). Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20131201.

Flann, C. (2009). Global Compositae Checklist (last accessed: July 03, 2018).

Fordham, D.A. & Brook, B.W. (2010). Why tropical island endemics are acutely susceptible to global change. *Biodiversity and Conservation*, **19**, 329–342.

Fosberg, F.R. (1992). Vegetation of the Society Islands. *Pacific Science.*

Francisco-Ortega, J., Wang, Z.-S., Wang, F.-G., Xing, F.-W., Liu, H., Xu, H., Xu, W.-X., Luo, Y.-B., Song, X.-Q., Gale, S., Boufford, D.E., Maunder, M. & An, S.-Q. (2010). Seed Plant Endemism on Hainan Island. A Framework for Conservation Actions. *The Botanical Review*, **76**, 346–376.

Franklin, J., Serra-Diaz, 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.

Frodin, D.G. (2001). *Guide to Standard Floras of the World.* Cambridge University Press, Cambridge.

Galley, C. & Linder, H.P. (2006). Geographical affinities of the Cape flora, South Africa. *Journal of Biogeography*, **33**, 236–250.

Gallinat, A.S., Russo, L., Melaas, E.K., Willis, C.G. & Primack, R.B. (2018). Herbarium specimens show patterns of fruiting phenology in native and invasive plant species across New England. *American Journal of Botany*, **97**, 1296.

García-Verdugo, C., Baldwin, B.G., Fay, M.F. & Caujapé-Castells, J. (2014). Life history traits and patterns of diversification in oceanic archipelagos. A meta-analysis. *Botanical Journal of the Linnean Society*, **174**, 334–348.

Garnier, E., Stahl, U., Laporte, M.-A., et al. (2017). Towards a thesaurus of plant characteristics. An ecological contribution. *Journal of Ecology*, **105**, 298–309.

GBIF (2018). The Global Biodiversity Information Faculty, Available at: http://www.gbif.org (last accessed: July 27, 2017).

Gewin, V. (2016). Data sharing. An open mind on open data. *Nature*, **529**, 117–119.

Gillespie, R.G. (2007). Oceanic Islands: Models of diversity. *Encyclopedia of Biodiversity*, 590–599.

Givnish, T.J., Millam, K.C., Mast, A.R., Paterson, T.B., Theim, T.J., Hipp, A.L., Henss, J.M., Smith, J.F., Wood, K.R. & Sytsma, K.J. (2009). Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society B: Biological Sciences*, **276**, 407–416.

Gonzalez, A., Cardinale, B.J., Allington, G.R.H., Byrnes, J., Endsley, K.A., Brown, D.G., Hooper, D.U., Isbell, F., O'Connor, M.I. & Loreau, M. (2016). Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. *Ecology*, **97**, 1949–1960.

Graham, C.H., Carnaval, A.C., Cadena, C.D., Zamudio, K.R., Roberts, T.E., Parra, J.L., McCain, C.M., Bowie, R.C.K., Moritz, C., Baines, S.B., Schneider, C.J., VanDerWal, J., Rahbek, C., Kozak, K.H. & Sanders, N.J. (2014). The origin and maintenance of montane diversity. integrating evolutionary and ecological processes. *Ecography*, **37**, 711–719.

Graham, C.H. & Fine, P.V.A. (2008). Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, **11**, 1265–1277.

Graves, G.R. & Rahbek, C. (2005). Source pool geometry and the assembly of continental avifaunas. *Proceedings of the National Academy of Sciences of the United States*, **102**, 7871–7876.

Greenacre, M.J. & Primicerio, R. (2013). *Multivariate analysis of ecological data*, 1st edn. Fundación BBVA, Bilbao.

Grisebach, A.H.R. (1866). Die Vegetations-Gebiete der Erde, übersichtlich zusammengestellt. *Petermanns geographische Mitteilungen*, **12**, 45–56.

Grömping, U. (2006). Relative importance for linear regression in R: the package relaimpo. *Journal of Statistical Software*, **17**.

Grossenbacher, D.L., Brandvain, Y., Auld, J.R., et al. (2017). Self-compatibility is over-represented on islands. *New Phytologist*, **215**, 469–478.

Grueber, C.E., Nakagawa, S., Laws, R.J. & Jamieson, I.G. (2011). Multimodel inference in ecology and evolution. Challenges and solutions. *Journal of Evolutionary Biology*, **24**, 699–711.

Guralnick, R., Walls, R. & Jetz, W. (2017). Humboldt Core - toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. *Ecography*, **1297**, 8.

Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., Goulson, D. & Kroon, H. de (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE*, **12**, e0185809.

Hampton, S.E., Anderson, S.S., Bagby, S.C., Gries, C., Han, X., Hart, E.M., Jones, M.B., Lenhardt, W.C., MacDonald, A., Michener, W.K., Mudge, J., Pourmokhtarian, A., Schildhauer, M.P., Woo, K.H. & Zimmerman, N. (2015). The Tao of open science for ecology. *Ecosphere*, **6**, art120.

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S. & Porter, J.H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, **11**, 156–162.

Harris, L.D. (1984). *The fragmented forest. Island biogeography theory and the preservation of biotic diversity.* University of Chicago Press.

Harte, J. (2011). *Maximum entropy and ecology. A theory of abundance, distribution, and energetics.* Oxford University Press.

Harter, D.E.V., Irl, S.D.H., Seo, B., Steinbauer, M.J., Gillespie, R., Triantis, K.A., Fernández-Palacios, J.-M. & Beierkuhnlein, C. (2015). Impacts of global climate change on the floras of oceanic islands – Projections, implications and current knowledge. *Perspectives in Plant Ecology, Evolution and Systematics*, **17**, 160–183.

Hattab, T., Albouy, C., Lasram, F.B.R., Le Loc'h, F., Guilhaumon, F. & Leprieur, F. (2015). A biogeographical regionalization of coastal Mediterranean fishes. *Journal of Biogeography*, **42**, 1336–1348.

Hawkins, B.A., Albuquerque, F.S., Araújo, M.B., et al. (2007). A global evaluation of Metabolic Theory as an explanation for terrestrial species richness gradients. *Ecology*, **88**, 1877–1888.

Heleno, R. & Vargas, P. (2014). How do islands become green? *Global Ecology and Biogeography*.

Helme, N.A. & Trinder-Smith, T.H. (2006). The endemic flora of the Cape Peninsula, South Africa. *South African Journal of Botany*, **72**, 205–210.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., et al. (2017). SoilGrids250m. Global gridded soil information based on machine learning. *PLoS ONE*, **12**, e0169748.

Hennig, W. (1966). *Phylogenetic systematics.* University of Illinois Press, Urbana.

Higgins, S.I., Nathan, R. & Cain, M.L. (2003). Are long-distance dispersal events in plants usually caused by nonstandard means of dispersal? *Ecology*, **84**, 1945–1956.

Hijmans, R.J., Garcia, N., Wieczorek, J., Rala, A., Maunahan, A. & Kapoor, J. (2009). GADM database of Global Administrative Areas, Version 1 (last accessed: June 16, 2010).

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *The American Naturalist*, **163**, 192–211.

Hoegh-Guldberg, O. (1999). Climate change, coral bleaching and the future of the world's coral reefs. *Marine and freshwater research*, **50**, 839–866.

Hoehndorf, R., Alshahrani, M., Gkoutos, G.V., Gosline, G., Groom, Q., Hamann, T., Kattge, J., Oliveira, S.M. de, Schmidt, M., Sierra, S., Smets, E., Vos, R.A. & Weiland, C. (2016). The flora phenotype ontology (FLOPO): tool for integrating morphological traits and phenotypes of vascular plants. *Journal of Biomedical Semantics*, **7**, 65.

Hoekstra, H. & Fagan, W. (1998). Body size, dispersal ability and compositional disharmony: the carnivore-dominated fauna of the Kuril Islands. *Diversity and Distributions*, **4**, 135–149.

Holt, B.G., Lessard, J.-P., Borregaard, M.K., Fritz, S.A., Araújo, M.B., Dimitrov, D., Fabre, P.-H., Graham, C.H., Graves, G.R., Jønsson, K.A., Nogués-Bravo, D., Wang, Z., Whittaker, R.J., Fjeldså, J. & Rahbek, C. (2013). An Update of Wallace's Zoogeographic Regions of the World. *Science*, **339**, 74–78.

Hortal, J. (2008). Uncertainty and the measurement of terrestrial biodiversity gradients. *Journal of Biogeography*, **35**, 1335–1336.

Hortal, J., Bello, F. de, 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.

Howe, H.F. & Smallwood, J. (1982). Ecology of seed dispersal. *Annual Review of Ecology, Evolution, and Systematics*, **13**, 201–228.

Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*, 1st edn. Princeton Univ. Press, Princeton, NJ.

Hudson, L.N., Newbold, T., Contu, S., et al. (2017). The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecology and Evolution*, **7**, 145–188.

Hughes, C. & Eastwood, R. (2006). Island radiation on a continental scale. Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences of the United States*, **103**, 10334–10339.

Humboldt, A.v. (1805-1834). *Voyage aux régions équinoxiales du Nouveau Continent. Fait en 1799, 1800, 1801, 1802, 1803 et 1804 par Al. de Humboldt et A. Bonpland.* F. Schoell, Paris.

Humphreys, A.M. & Linder, H.P. (2009). Concept versus data in delimitation of plant genera. *Taxon*, **58**, 1054–1074.

Huston, M.A. (1999). Local processes and regional patterns: appropriate scales for understanding variation in the diversity of plants and animals. *Oikos*, **86**, 393–401.

Hutchinson, G.E. (1957). Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

IPNI (2012). The International Plant Names Index (last accessed: February 20, 2018).

Isnard, S., L'huillier, L., Rigault, F. & Jaffré, T. (2016). How did the ultramafic soils shape the flora of the New Caledonian hotspot? *Plant and Soil*, **403**, 53–76.

IUCN (2018). IUCN Red List of Threatened Species (last accessed: February 20, 2018).

Jansen, F. & Dengler, J. (2010). Plant names in vegetation databases – a neglected source of bias. *Journal of Vegetation Science*, **21**, 1179–1186.

Jardim Botânico do Rio de Janeiro (2016). Flora do Brasil 2020 em construção (last accessed: May 09, 2016).

Jetz, W., Cavender-Bares, J., Pavlick, R., et al. (2016). Monitoring plant functional diversity from space. *Nature plants*, **2**, 16024.

Jetz, W., McPherson, J.M. & Guralnick, R.P. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution*, **27**, 151–159.

Johnson, P.C.D., Barry, S.J.E., Ferguson, H.M. & Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution*, **6**, 133–142.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P. & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, **4**, 170122 EP.

Karger, D.N., Cord, A.F., Kessler, M., Kreft, H., Kühn, I., Pompe, S., Sandel, B., Cabral, J., Smith, A.B., Svenning, J.-C., Tuomisto, H., Weigelt, P. & Wesche, K. (2016). Delineating probabilistic species pools in ecology and biogeography. *Global Ecology and Biogeography*, **25**, 489–501.

Kattge, J., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P.B. & Wright, I.J. (2011a). TRY – a global database of plant traits. *Global Change Biology*, **17**, 2905–2935.

Kattge, J., Díaz, S. & Wirth, C. (2014). Of carrots and sticks. *Nature Geoscience*, **7**, 778–779.

Kattge, J., Ogle, K., Bönisch, G., Díaz, S., Lavorel, S., Madin, J., Nadrowski, K., Nöllert, S., Sartor, K. & Wirth, C. (2011b). A generic structure for plant trait databases. *Methods in Ecology and Evolution*, **2**, 202–213.

Kearney, M., Simpson, S.J., Raubenheimer, D. & Helmuth, B. (2010). Modelling the ecological niche from functional traits. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3469–3483.

Keck, F. (2015). *phylosignal. Exploring the Phylogenetic Signal in Continuous Traits (R-package).*

Keddy, P.A. (1992). A Pragmatic Approach to Functional Ecology. *Functional Ecology*, **6**, 621–626.

Keeley, S.G. & Funk, V.A. (2011). Origin and evolution of Hawaiian endemics: new patterns revealed by molecular phylogenetic studes. *The biology of island floras* (ed. by D. Bramwell and J. Caujapé-Castells), pp. 57–88. Cambridge Univ. Press, Cambridge.

Kelling, S., Hochachka, W.M., Fink, D., Riedewald, M., Caruana, R., Ballard, G. & Hooker, G. (2009). Data-intensive Science. A New Paradigm for Biodiversity Studies. *BioScience*, **59**, 613–620.

Kent, M. (2005). Biogeography and macroecology. *Progress in Physical Geography*, **29**, 256–264.

Keppel, G., Lowe, A.J. & Possingham, H.P. (2009). Changing perspectives on the biogeography of the tropical South Pacific. Influences of dispersal, vicariance and extinction. *Journal of Biogeography*, **36**, 1035–1054.

Kerr, J.T., Kharouba, H.M. & Currie, D.J. (2007). The macroecological contribution to global change solutions. *Science*, **316**, 1581–1584.

Kier, G., Kreft, H., Lee, T.M., Jetz, W., Ibisch, P.L., Nowicki, C., Mutke, J. & Barthlott, W. (2009). A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences of the United States*, **106**, 9322–9327.

Knapp, A.K., Briggs, J.M., Collins, S.L., Archer, S.R., Bret-Harte, M.S., Ewers, B.E., Peters, D.P., Young, D.R., Shaver, G.R. & Pendall, E. (2008). Shrub encroachment in North American grasslands. Shifts in growth form dominance rapidly alters control of ecosystem carbon inputs. *Global Change Biology*, **14**, 615–623.

Koleff, P., Gaston, K.J. & Lennon, J.J. (2003). Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, **72**, 367–382.

König, C., Weigelt, P. & Kreft, H. (2017). Dissecting global turnover in vascular plants. *Global Ecology and Biogeography*, **26**, 228–242.

Kraft, N.J.B., Adler, P.B., Godoy, O., James, E.C., Fuller, S., Levine, J.M. & Fox, J. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, **29**, 592–599.

Kraft, N.J.B., Comita, L.S., Chase, J.M., et al. (2011). Disentangling the drivers of β diversity along latitudinal and elevational gradients. *Science*, **333**, 1755–1758.

Kreft, H. & Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences of the United States*, **104**, 5925–5930.

Kreft, H., Jetz, W., Mutke, J. & Barthlott, W. (2010). Contrasting environmental and regional effects on global pteridophyte and seed plant diversity. *Ecography*, **33**, 408–419.

Kreft, H., Jetz, W., Mutke, J., Kier, G. & Barthlott, W. (2008). Global diversity of island floras from a macroecological perspective. *Ecology Letters*, **11**, 116–127.

Kueffer, C., Drake, D.R. & Fernández-Palacios, J.M. (2014). Island biology. Looking towards the future. *Biology Letters*, **10**.

Kuenzer, C., Ottinger, M., Wegmann, M., Guo, H., Wang, C., Zhang, J., Dech, S. & Wikelski, M. (2014). Earth observation satellite sensors for biodiversity monitoring. Potentials and bottlenecks. *International Journal of Remote Sensing*, **35**, 6599–6647.

La Salle, J., Williams, K.J. & Moritz, C. (2016). Biodiversity analysis in the digital era. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **371**.

Lamanna, C., Blonder, B., Violle, C., et al. (2014). Functional trait space and the latitudinal diversity gradient. *Proceedings of the National Academy of Sciences of the United States*, **111**, 13745–13750.

Lavorel, S., Díaz, S., Cornelissen, J.H.C., Garnier, E., Harrison, S.P., McIntyre, S., Pausas, J.G., Pérez-Harguindeguy, N., Roumet, C. & Urcelay, C. (2007). Plant Functional Types: Are We Getting Any Closer to the Holy Grail? *Terrestrial Ecosystems in a Changing World* (ed. by J.G. Canadell, D.E. Pataki and L.F. Pitelka), pp. 149–164. Springer Berlin Heidelberg, Berlin, Heidelberg.

Lavorel, S. & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, **16**, 545–556.

Le Roux, J.J., Strasberg, D., Rouget, M., Morden, C.W., Koordom, M. & Richardson, D.M. (2014). Relatedness defies biogeography. The tale of two island endemics (Acacia heterophylla and A. koa). *New Phytologist*, **204**, 230–242.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, **521**, 436–444.

Legendre, P. (2008). Studying beta diversity. Ecological variation partitioning by multiple regression and canonical analysis. *Journal of Plant Ecology*, **1**, 3–8.

Legendre, P. (2014). Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, 1324–1334.

Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005). Analyzing beta diversity. Partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.

Leibold, M.A. & McPeek, M.A. (2006). Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, **87**, 1399–1410.

Leishman, M.R. & Westoby, M. (1992). Classifying plants into groups on the basis of associations of individual traits--evidence from Australian semi-arid woodlands. *Journal of Ecology*, **80**, 417–424.

Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2001). The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology*, **70**, 966–979.

Lens, F., Davin, N., Smets, E. & del Arco, M. (2013). Insular Woodiness on the Canary Islands. A Remarkable Case of Convergent Evolution. *International Journal of Plant Sciences*, **174**, 992–1013.

Lenzner, B., Weigelt, P., Kreft, H., Beierkuhnlein, C. & Steinbauer, M.J. (2017). The general dynamic model of island biogeography revisited at the level of major flowering plant families. *Journal of Biogeography*, **44**, 1029–1040.

Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, **10**, 707–710.

Levin, D.A. (1974a). The oil content of seeds: An ecological perspective. *The American Naturalist*, **108**, 193–206.

Levin, S.A. (1974b). Dispersion and Population Interactions. *The American Naturalist*, **108**, 207–228.

Lichstein, J.W. (2007). Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology*, **188**, 117–131.

Lobo, J.M., Jiménez-Valverde, A. & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, **33**, 103–114.

Lomolino, M.V., Riddle, B.R. & Whittaker, R.J. (2016). *Biogeography. Biological diversity across space and time*, 5th edn. Sinauer Associates, Sunderland, Massechusetts, U.S.A.

López-Martínez, J.O., Sanaphre-Villanueva, L., Dupuy, J.M., Hernández-Stefanoni, J.L., Meave, J.A. & Gallardo-Cruz, J.A. (2013). β-Diversity of functional groups of woody plants in a tropical dry forest in Yucatan. *PLoS ONE*, **8**, e73660.

Lowe, W.H. & McPeek, M.A. (2014). Is dispersal neutral? *Trends in Ecology & Evolution*, **29**, 444–450.

Mabberley, D.J. (2008). *Mabberley's plant-book. A portable dictionary of plants, their classifications and uses*, 3rd edn. Cambridge Univ. Press, Cambridge u.a.

MacArthur, R.H. & Wilson, E.O. (1963). An equilibrium theory of insular zoogeography. *Evolution*, **17**, 373–387.

MacArthur, R.H. & Wilson, E.O. (1967). *The theory of island biogeography*, 1st edn. Princeton Univ. Press, Princeton, NJ.

Madrigal-González, J., Rios, R.S., Aragón, C.F. & Gianoli, E. (2017). Indirect facilitation by a liana might explain the dominance of a small tree in a temperate forest. *Journal of Plant Ecology*, rtx012.

Magallón, S. & Castillo, A. (2009). Angiosperm diversification through time. *American Journal of Botany*, **96**, 349–365.

Maitner, B.S., Boyle, B., Casler, N., et al. (2018). The bien r package. A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods in Ecology and Evolution*, **9**, 373–379.

Mallet, J. (1995). A species definition for the modern synthesis. *Trends in Ecology & Evolution*, **10**, 294–299.

Manly, B.F.J. (1986). Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. *Researches on Population Ecology*, **28**, 201–218.

Marquet, P.A., Allen, A.P., Brown, J.H., et al. (2014). On Theory in Ecology. *BioScience*, **64**, 701–710.

Maurer, B.A. (2000). Ecology needs theory as well as practice. *Nature*, **408**, 768 EP -.

Mayr, E. (1969). *Principles of systematic zoology*, 1st edn. McGraw-Hill, New York, London.

McGill, B.J. (2003). A test of the unified neutral theory of biodiversity. *Nature*, **422**, 881–885.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, **21**, 178–185.

McGlone, M.S., Richardson, S.J., Jordan, G.J. & Perry, G.L.W. (2015). Is there a 'suboptimal' woody species height? A response to Scheffer et al. *Trends in Ecology & Evolution*, **30**, 4–5.

Merow, C., Allen, J.M., Aiello-Lammens, M. & Silander, J.A. (2016). Improving niche and range estimates with Maxent and point process models by integrating spatially explicit information. *Global Ecology and Biogeography*, **25**, 1022–1036.

Meyer, C., Weigelt, P. & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, **19**, 992–1006.

Meyer, S.T., Ptacnik, R., Hillebrand, H., et al. (2018). Biodiversity–multifunctionality relationships depend on identity and number of measured functions. *Nature Ecology & Evolution*, **2**, 44–49.

Michener, W., Vieglais, D., Vision, T., Kunze, J., Cruse, P. & Janée, G. (2011). DataONE. Data Observation Network for Earth — Preserving Data and Enabling Innovation in the Biological and Environmental Sciences. *D-Lib Magazine*, **17**.

Michener, W.K. (2015a). Ecological data sharing. *Ecological Informatics*, **29**, 33–44.

Michener, W.K. (2015b). Ten Simple Rules for Creating a Good Data Management Plan. *PLoS Computational Biology*, **11**, e1004525.

Michener, W.K. & Jones, M.B. (2012). Ecoinformatics. Supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, **27**, 85–93.

Midway, S.R. & Hodge, A.-M.C. (2012). Carlquist revisited: history, success, and applicability of a natural history model. *Biology & Philosophy*, **27**, 497–520.

Mills, J.A., Teplitsky, C., Arroyo, B., et al. (2015). Archiving Primary Data: Solutions for Long-Term Studies. *Trends in Ecology & Evolution*, **30**, 581–589.

Mokany, K., Ferrier, S., Connolly, S.R., Dunstan, P.K., Fulton, E.A., Harfoot, M.B., Harwood, T.D., Richardson, A.J., Roxburgh, S.H., Scharlemann, J.P.W., Tittensor, D.P., Westcott, D.A. & Wintle, B.A. (2016). Integrating modelling of biodiversity composition and ecosystem function. *Oikos*, **125**, 10–19.

Moles, A.T. (2005). A brief history of seed size. *Science*, **307**, 576–580.

Moles, A.T., Ackerly, D.D., Tweddle, J.C., Dickie, J.B., Smith, R., Leishman, M.R., Mayfield, M.M., Pitman, A., Wood, J.T. & Westoby, M. (2007). Global patterns in seed size. *Global Ecology and Biogeography*, **16**, 109–116.

Moles, A.T., Perkins, S.E., Laffan, S.W., et al. (2014). Which is a better predictor of plant traits: temperature or precipitation? *Journal of Vegetation Science*, **25**, 1167–1180.

Moles, A.T., Warton, D.I., Warman, L., Swenson, N.G., Laffan, S.W., Zanne, A.E., Pitman, A., Hemmings, F.A. & Leishman, M.R. (2009). Global patterns in plant height. *Journal of Ecology*, **97**, 923–932.

Moore, A.L. & McCarthy, M.A. (2016). Optimizing ecological survey effort over space and time. *Methods in Ecology and Evolution*, **7**, 891–899.

Moore, D.M. (1968). The vascular flora of the Falkland Islands. *British Antarctic Survey Scientific Reports*, **60**.

Mora, C., Tittensor, D.P. & Myers, R.A. (2008). The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 149–155.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H. & Donoghue, P.C.J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences of the United States*, **115**, E2274-E2283.

Morrone, J.J. (2015). Biogeographical regionalisation of the world: a reappraisal. *Australian Systematic Botany*, **28**, 81.

Morueta-Holme, N., Enquist, B.J., McGill, B.J., et al. (2013). Habitat area and climate stability determine geographical variation in plant species range sizes. *Ecology Letters*, **16**, 1446–1454.

Moser, D., Dullinger, S., Englisch, T., Niklfeld, H., Plutzar, C., Sauberer, N., Zechmeister, H.G. & Grabherr, G. (2005). Environmental determinants of vascular plant species richness in the Austrian Alps. *Journal of Biogeography*, **32**, 1117–1127.

Mouquet, N., Lagadeuc, Y., Devictor, V., et al. (2015). Predictive ecology in a changing world. *Journal of Applied Ecology*, **52**, 1293–1310.

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K. & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, **3**, 743–756.

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

Nathan, R. (2006). Long-Distance Dispersal of Plants. *Science*, **313**, 786–788.

Nekola, J.C. & White, P.S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Ecology*, **26**, 867–878.

Newbold, T., Hudson, L.N., Hill, S.L.L., et al. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, **520**, 45.

Nogales, M., Heleno, R., Traveset, A. & Vargas, P. (2012). Evidence for overlooked mechanisms of long ‐ distance seed dispersal to and between oceanic islands. *New Phytologist*, **194**, 313–317.

OECD (2013). Glossary of statistical terms. Definition of "Data imputation", Available at: https://stats.oecd.org/glossary/detail.asp?ID=3406 (last accessed: October 10, 2017).

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Sólymos, P., M. Henry H. Stevens & Wagner, H. (2013). *vegan: Community Ecology Package, Available at:* http://CRAN.R-project.org/package=vegan.

Olesen, J.M., Carstensen, D.W. & Dupont, Y.L. (2010). Island pollination. Ways to solve the species poverty problem. *Recerca*, **19**, 149–171.

Ostertag, R., Warman, L., Cordell, S., Vitousek, P.M. & Lewis, O. (2015). Using plant functional traits to restore Hawaiian rainforest. *Journal of Applied Ecology*, **52**, 805–809.

Pachauri, R.K., Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q. & Dasgupta, P. (2014). *Climate change 2014. Synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change.* IPCC.

Page, R.D.M. (2008). Biodiversity informatics. The challenge of linking data and the role of shared identifiers. *Briefings in Bioinformatics*, **9**, 345–354.

Pagel, M., Meade, A., Barker, D. & Thorne, J. (2004). Bayesian Estimation of Ancestral Character States on Phylogenies. *Systematic Biology*, **53**, 673–684.

Paine, C.E.T., Baraloto, C., Chave, J. & Hérault, B. (2011). Functional traits of individual trees reveal ecological constraints on community assembly in tropical rain forests. *Oikos*, **120**, 720–727.

Palmer, M.W. & Richardson, J.C. (2012). Biodiversity Data in the Information Age. Do 21st Century Floras Make the Grade? *Castanea*, **77**, 46–59.

Patiño, J., Whittaker, R.J., Borges, P.A.V., et al. (2017). A roadmap for island biology. 50 fundamental questions after 50 years of The Theory of Island Biogeography. *Journal of Biogeography*, **41**, 1227.

Pawar, S., Woodward, G. & Dell, A.I. (eds.) (2015). *Trait-Based Ecology - From Structure to Function.* Elsevier Science.

Pearse, W.D., Cadotte, M.W., Cavender-Bares, J., Ives, A.R., Tucker, C.M., Walker, S.C. & Helmus, M.R. (2015). pez. phylogenetics for the environmental sciences. *Bioinformatics*, **31**, 2888–2890.

Pearson, R.G. & Dawson, T.P. (2003). Predicting the impacts of climate change on the distribution of species. Are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.

Pennisi, E. (2005). What determines species diversity? *Science*, **309**, 90.

Penone, C., Davidson, A.D., Shoemaker, K.T., Di Marco, M., Rondinini, C., Brooks, T.M., Young, B.E., Graham, C.H., Costa, G.C. & Freckleton, R. (2014). Imputation of missing data in life-history trait datasets: which approach performs the best? *Methods in Ecology and Evolution*, **5**, 961–970.

Pensoft (2017). *Biodiversity Data Journal. a peer review open-access journal.*

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., et al. (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany*, **61**, 167–234.

Petchey, O.L. & Gaston, K.J. (2002). Functional diversity (FD), species richness and community composition. *Ecology Letters*, **5**, 402–411.

Peters, D.P.C., Havstad, K.M., Cushing, J., Tweedie, C., Fuentes, O. & Villanueva-Rosales, N. (2014a). Harnessing the power of big data. Infusing the scientific method with machine learning to transform ecology. *Ecosphere*, **5**, art67.

Peters, D.P.C., Loescher, H.W., SanClements, M.D. & Havstad, K.M. (2014b). Taking the pulse of a continent. Expanding site-based research infrastructure for regional- to continental-scale ecology. *Ecosphere*, **5**, art29.

Peterson, A.T. (2011). *Ecological niches and geographic distributions.* Princeton University Press.

Piel, W.H., Chan, L., Dominus, M.J., Ruan, J., Vos, R.A. & Tannen, V. (2009). TreeBASE v. 2: A Database of Phylogenetic Knowledge. *e-BioSphere 2009.*

Pitman, N.C.A., Terborgh, J., Silman, M.R. & Nuñez V., P. (1999). Tree species distributions in an upper Amazonian forest. *Ecology*, **80**, 2651–2661.

Pokorny, L., Riina, R., Mairal, M., Meseguer, A.S., Culshaw, V., Cendoya, J., Serrano, M., Carbajal, R., Ortiz, S., Heuertz, M. & Sanmartín, I. (2015). Living on the edge. Timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. *Frontiers in Genetics*, **6**, 154.

Poncet, V., Munoz, F., Munzinger, J., Pillon, Y., Gomez, C., Couderc, M., Tranchant-Dubreuil, C., Hamon, S. & Kochko, A. de (2013). Phylogeography and niche modelling of the relict plant Amborella trichopoda (Amborellaceae) reveal multiple Pleistocene refugia in New Caledonia. *Molecular Ecology*, **22**, 6163–6178.

Portnoy, S. & Willson, M.F. (1993). Seed dispersal curves. Behavior of the tail of the distribution. *Evolutionary Ecology*, **7**, 25–44.

Powney, G.D. & Isaac, N.J.B. (2015). Beyond maps. a review of the applications of biological records. *Biological Journal of the Linnean Society*, **115**, 532–542.

Prentice, I.C., Bondeau, A., Cramer, W., Harrison, S.P., Hickler, T., Lucht, W., Sitch, S., Smith, B. & Sykes, M.T. (2007). Dynamic Global Vegetation Modeling. Quantifying Terrestrial Ecosystem Responses to Large-Scale Environmental Change. *Terrestrial Ecosystems in a Changing World* (ed. by J.G. Canadell, D.E. Pataki and L.F. Pitelka), pp. 175–192. Springer Berlin Heidelberg, Berlin, Heidelberg.

Pyšek, P., Pergl, J., Essl, F., et al. (2017). Naturalized alien flora of the world. *Preslia*, **89**, 203–274.

Qian, H. (2009). Beta diversity in relation to dispersal ability for vascular plants in North America. *Global Ecology and Biogeography*, **18**, 327–332.

Qian, H. & Jin, Y. (2016). An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *Journal of Plant Ecology*, **9**, 233–239.

Qian, H. & Ricklefs, R.E. (2007). A latitudinal gradient in large-scale beta diversity for vascular plants in North America. *Ecology Letters*, **10**, 737–744.

Qian, H., Swenson, N.G. & Zhang, J. (2013). Phylogenetic beta diversity of angiosperms in North America. *Global Ecology and Biogeography*, **22**, 1152–1161.

Queenborough, S.A. & Porras, C. (2014). Expanding the coverage of plant trait databases–A comparison of specific leaf area derived from fresh and dried leaves. *Plant Ecology & Diversity*, **7**, 383–388.

R Core Team (2017). *R: A language and environment for statistical computing,* Vienna, Austria, Available at: http://www.R-project.org.

Ragni, L. & Greb, T. (2018). Secondary growth as a determinant of plant shape and form. *Seminars in cell & developmental biology*, **79**, 58–67.

Raunkiær, C. (1934). *The life forms of plants and statistical plant geography.* Clarendon press, Oxford.

Reichman, O.J., Jones, M.B. & Schildhauer, M.P. (2011). Challenges and opportunities of open data in ecology. *Science*, **331**, 703–705.

Reichstein, M., Bahn, M., Mahecha, M.D., Kattge, J. & Baldocchi, D.D. (2014). Linking plant and ecosystem functional biogeography. *Proceedings of the National Academy of Sciences of the United States*, **111**, 13697–13702.

Ribes, A., Zwiers, F.W., Azaïs, J.-M. & Naveau, P. (2017). A new statistical approach to climate change detection and attribution. *Climate Dynamics*, **48**, 367–386.

Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H. & Possingham, H.P. (2006). Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters*, **9**, 1136–1145.

Rosa, I.M.D., Smith, M.J., Wearn, O.R., Purves, D. & Ewers, R.M. (2016). The Environmental Legacy of Modern Tropical Deforestation. *Current biology : CB*, **26**, 2161–2166.

Rosauer, D.F., Ferrier, S., Williams, K.J., Manion, G., Keogh, J.S. & Laffan, S.W. (2014). Phylogenetic generalised dissimilarity modelling. A new approach to analysing and predicting spatial turnover in the phylogenetic composition of communities. *Ecography*, **37**, 21–32.

Royal Botanic Gardens Kew (2016). Seed information database (SID). v7.1, Available at: http://data.kew.org/sid/ (last accessed: March 22, 2013).

Rumeu, B., Afonso, V., Fernández-Palacios, J.M. & Nogales, M. (2014). Diversity, distribution and conservation status of island conifers: a global review. *Diversity and Distributions*, **20**, 272–283.

Ruokolainen, K., Tuomisto, H., Macía, M.J., Higgins, M.A. & Yli-Halla, M. (2007). Are floristic and edaphic patterns in Amazonian rain forests congruent for trees, pteridophytes and Melastomataceae? *Journal of Tropical Ecology*, **23**, 13.

Sandel, B., Gutiérrez, A.G., Reich, P.B., Schrodt, F., Dickie, J., Kattge, J. & Schmidtlein, S. (2015). Estimating the missing species bias in plant trait measurements. *Journal of Vegetation Science*, **26**, 828–838.

Santos, A.M.C., Cianciaruso, M.V. & Marco, P. de (2015). Global patterns of functional diversity and assemblage structure of island parasitoid faunas. *Global Ecology and Biogeography*.

Santos, A.M.C., Field, R. & Ricklefs, R.E. (2016). New directions in island biogeography. *Global Ecology and Biogeography*, **25**, 751–768.

Santos, A.M.C., Jones, O.R., Quicke, D.L.J. & Hortal, J. (2010). Assessing the reliability of biodiversity databases: identifying evenly inventoried island parasitoid faunas (Hymenoptera: Ichneumonoidea) worldwide. *Insect Conservation and Diversity*, **3**, 72–82.

Sax, D.F. & Gaines, S.D. (2008). Species invasions and extinction. The future of native biodiversity on islands. *Proceedings of the National Academy of Sciences of the United States*, **105**, 11490–11497.

Scheffer, M., Vergnon, R., Cornelissen, J.H.C., Hantson, S., Holmgren, M., van Nes, E.H. & Xu, C. (2014). Why trees and shrubs but rarely trubs? *Trends in Ecology & Evolution*, **29**, 433–434.

Scheiner, S.M. (2013). The ecological literature, an idea-free distribution. *Ecology Letters*, **16**, 1421–1423.

Scheiter, S., Langan, L. & Higgins, S.I. (2013). Next-generation dynamic global vegetation models. Learning from community ecology. *New Phytologist*, **198**, 957–969.

Schrodt, F., Kattge, J., Shan, H., 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.

Seebens, H., Blackburn, T.M., Dyer, E.E., et al. (2017). No saturation in the accumulation of alien species worldwide. *Nature Communications*, **8**, 14435.

Serreze, M.C. & Barry, R.G. (2011). Processes and impacts of Arctic amplification: A research synthesis. *Global and Planetary Change*, **77**, 85–96.

Shou, W., Bergstrom, C.T., Chakraborty, A.K. & Skinner, F.K. (2015). Theory, models and biology. *eLife*, **4**.

Simberloff, D.S. (1969). Experimental Zoogeography of Islands. A Model for Insular Colonization. *Ecology*, **50**, 296–314.

Simberloff, D.S. (1974). Equilibrium Theory of Island Biogeography and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, **5**, 161–182.

Simpson, G.G. (1961). *Principles of Animal Taxonomy.* Columbia University Press.

Smith, S.A. & Brown, J.W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, **105**, 1–13.

Socolar, J.B., Gilroy, J.J., Kunin, W.E. & Edwards, D.P. (2016). How should beta-diversity inform biodiversity conservation? *Trends in Ecology & Evolution*, **31**, 67–80.

Soininen, J., McDonald, R. & Hillebrand, H. (2007). The distance decay of similarity in ecological communities. *Ecography*, **30**, 3–12.

Steffen, W., Richardson, K., Rockström, J., et al. (2015). Sustainability. Planetary boundaries: guiding human development on a changing planet. *Science*, **347**, 1259855.

Stegen, J.C. & Swenson, N.G. (2009). Functional trait assembly through ecological and evolutionary time. *Theoretical Ecology*, **2**, 239–250.

Steinbauer, M.J., Otto, R., Naranjo-Cigala, A., Beierkuhnlein, C. & Fernández-Palacios, J.-M. (2012). Increase of island endemism with altitude - speciation processes on oceanic islands. *Ecography*, **35**, 23–32.

Stevens, G.C. (1989). The latitudinal gradient in geographical range. How so many species coexist in the tropics. *The American Naturalist*, **133**, 240–256.

Stevens, P.F. (2001 onwards). Angiosperm Phylogeny Website v.12, Available at: http://www.mobot.org/MOBOT/research/APweb/ (last accessed: August 26, 2014).

Stoddard, D.R. (1992). Biogeography of the tropical Pacific. *Pacific Science*, **46**, 276–293.

Stuart, Y.E., Losos, J.B. & Algar, A.C. (2012). The island-mainland species turnover relationship. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 4071–4077.

Sutherland, W.J., Freckleton, R.P., Godfray, H.C.J., et al. (2013). Identification of 100 fundamental ecological questions. *Journal of Ecology*, **101**, 58–67.

Swenson, N.G., Enquist, B.J., Pither, J., et al. (2012). The biogeography and filtering of woody plant functional diversity in North and South America. *Global Ecology and Biogeography*, **21**, 798–808.

Swenson, N.G., Weiser, M.D., Mao, L., Araújo, M.B., Diniz ‐ Filho, J.A.F., Kollmann, J., Nogués ‐ Bravo, D., Normand, S., Rodríguez, M.A., García ‐ Valdés, R., Valladares, F., Zavala, M.A. & Svenning, J. ‐ C. (2017). Phylogeny and the prediction of tree functional diversity across novel continental settings. *Global Ecology and Biogeography*, **26**, 553–562.

Syfert, M.M., Joppa, L., Smith, M.J., Coomes, D.A., Bachman, S.P. & Brummitt, N.A. (2014). Using species distribution models to inform IUCN Red List assessments. *Biological Conservation*, **177**, 174–184.

Tackenberg, O., Poschlod, P. & Bonn, S. (2003). Assessment of wind dispersal potential in plant species. *Ecological Monographs*, **73**, 191–205.

Takhtajan, A. (1986). *Floristic regions of the world.* University of California Press, Berkeley, CA.

Taugourdeau, S., Villerd, J., Plantureux, S., Huguenin-Elie, O. & Amiaud, B. (2014). Filling the gap in functional trait databases: use of ecological hypotheses to replace missing data. *Ecology and Evolution*, **4**, 944–958.

TDWG (2007). World geographical scheme for recording plant distributions (last accessed: December 22, 2017).

Tello, J.S., Myers, J.A., Macia, M.J., Fuentes, A.F., Cayola, L., Arellano, G., Loza, M.I., Torrez, V., Cornejo, M., Miranda, T.B. & Jørgensen, P.M. (2015). Elevational gradients in beta-diversity reflect variation in the strength of local community assembly mechanisms across spatial scales. *PLoS ONE*, **10**, e0121458.

The Angiosperm Phylogeny Group (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, **181**, 1–20.

The Plant List (2013). The Plant List v1.1, Available at: http://www.theplantlist.org/.

Tilman, D., May, R.M., Lehman, C.L. & Nowak, M.A. (1994). Habitat destruction and the extinction debt. *Nature*, **371**, 65.

Traveset, A., Fernández-Palacios, J.M., Kueffer, C., Bellingham, P.J., Morden, C. & Drake, D.R. (2015). Introduction to the Special Issue. Advances in island plant biology since Sherwin Carlquist's Island Biology. *AoB PLANTS*, **8**.

Tuomisto, H., Ruokolainen, K. & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, **299**, 241–244.

Tutin, T.G., Heywood, V.H., Burges, N.A., Moore, D.M., Valentine, D.H., Walters, S.M. & Webb, D.A. (eds.) (1964–1980). *Flora Europaea. Volume 1-5.* Cambridge University Press, Cambridge, UK.

Ulloa Ulloa, C., Acevedo-Rodríguez, P., Beck, S., et al. (2017). An integrated assessment of the vascular plant species of the Americas. *Science*, **358**, 1614–1617.

UNEP-WCMC (2013). *Global distribution of islands. Global Island Database (version 2). Based on Open Street Map data (© OpenStreetMap contributors),* Cambridge, UK.

UNEP-WCMC (2014). *Data Standards for the World Database on Protected Areas (WDPA),* Cambridge, UK.

van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice. Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, **45**, 1–67.

van Kleunen, M., Dawson, W., Essl, F., et al. (2015). Global exchange and accumulation of non-native plants. *Nature*, **525**, 100–103.

Vanderbilt, K. & Gaiser, E. (2017). The International Long Term Ecological Research Network. A platform for collaboration. *Ecosphere*, **8**.

VanDerWal, J., Shoo, L.P., Graham, C. & Williams, S.E. (2009). Selecting pseudo-absence data for presence-only distribution modeling. How far should you stray from what you know? *Ecological Modelling*, **220**, 589–594.

Vargas, P. (2007). Are Macaronesian islands refugia of relict plant lineages? A molecular survey. *Phylogeography of Southern European Refugia. Evolutionary perspectives on the origins and conservation of European biodiversity* (ed. by S. Weiss and N. Ferrand), pp. 297–314. Springer Netherlands.

Vellend, M., Baeten, L., Myers-Smith, I.H., Elmendorf, S.C., Beauséjour, R., Brown, C.D., Frenne, P.D., Verheyen, K. & Wipf, S. (2013). Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences of the United States*, **110**, 19456–19459.

Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S*, Fourth. Springer, New York.

Violle, C., Reich, P.B., Pacala, S.W., Enquist, B.J. & Kattge, J. (2014). The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences of the United States*, **111**, 13690–13696.

Vitousek, P.M. (2002). Oceanic islands as model systems for ecological studies. *Journal of Ecology*, **29**, 573–582.

Vitousek, P.M., D'antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R. (1997). Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology*, 1–16.

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.* Macmillan and Co, London.

Wallace, A.R. (1878). *Tropical nature and other essays.* Macmillan, London.

Wallace, A.R. (1881). *Island life, or, the phenomena and causes of insular faunas and floras. Including a revision and attempted solution of the problem of geological climates.* Harper & Bros., New York.

Walther, B.A. & Moore, J.L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, **28**, 815–829.

Watson, J.E.M., Jones, K.R., Fuller, R.A., Di Marco, M., Segan, D.B., Butchart, S.H.M., Allan, J.R., McDonald-Madden, E. & Venter, O. (2016). Persistent Disparities between Recent Rates of Habitat Conversion and Protection and Implications for Future Global Conservation Targets. *Conservation Letters*, **9**, 413–421.

Weigelt, P., Jetz, W. & Kreft, H. (2013). Bioclimatic and physical characterization of the world's islands. *Proceedings of the National Academy of Sciences of the United States*, **110**, 15307–15312.

Weigelt, P., Kissling, W.D., Kisel, Y., Fritz, S.A., Karger, D.N., Kessler, M., Lehtonen, S., Svenning, J.-C. & Kreft, H. (2015). Global patterns and drivers of phylogenetic structure in island floras. *Scientific Reports*, **5**, 12213.

Weigelt, P., König, C. & Kreft, H. (2018). The Global Inventory of Floras and Traits (GIFT) database, Available at: http://gift.uni-goettingen.de.

Weigelt, P. & Kreft, H. (2013). Quantifying island isolation - insights from global patterns of insular plant species richness. *Ecography*, **36**, 417–429.

Weigelt, P., Steinbauer, M.J., Cabral, J. & Kreft, H. (2016). Late Quaternary climate change shapes island biodiversity. *Nature*, **532**, 99–102.

Weiher, E., Freund, D., Bunton, T., Stefanski, A., Lee, T. & Bentivenga, S. (2011). Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2403–2413.

Weiss, S. & Ferrand, N. (eds.) (2007). *Phylogeography of Southern European Refugia. Evolutionary perspectives on the origins and conservation of European biodiversity.* Springer Netherlands.

Westoby, M., Leishman, M.R., Lord, J., Poorter, H. & Schoen, D.J. (1996). Comparative ecology of seed size and dispersal. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **351**, 1309–1318.

Whitlock, M.C. (2011). Data archiving in ecology and evolution. Best practices. *Trends in Ecology & Evolution*, **26**, 61–65.

Whittaker, R.H. (1956). Vegetation of the Great Smoky Mountains. *Ecological Monographs*, **26**, 1–80.

Whittaker, R.H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, **30**, 279–338.

Whittaker, R.J. & Fernández-Palacios, J.M. (2007). *Island biogeography. Ecology, evolution, and conservation*, 2nd edn. Oxford Univ. Press, Oxford.

Whittaker, R.J., Fernández-Palacios, J.M., Matthews, T.J., Borregaard, M.K. & Triantis, K.A. (2017). Island biogeography. Taking the long view of nature's laboratories. *Science*, **357**, eaam8326.

Whittaker, R.J., Jones, S.H. & Partomihardjo, T. (1997). The rebuilding of an isolated rain forest assemblage. How disharmonic is the flora of Krakatau? *Biodiversity and Conservation*, **6**, 1671–1696.

Whittaker, R.J., Rigal, F., Borges, P.A.V., Cardoso, P., Terzopoulou, S., Casanoves, F., Pla, L., Guilhaumon, F., Ladle, R.J. & Triantis, K.A. (2014). Functional biogeography of oceanic islands and the scaling of functional diversity in the Azores. *Proceedings of the National Academy of Sciences of the United States*, **111**, 13709–13714.

Whittaker, R.J., Triantis, K.A. & Ladle, R.J. (2008). A general dynamic theory of oceanic island biogeography. *Journal of Ecology*, **35**, 977–994.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T. & Vieglais, D. (2012). Darwin Core. An evolving community-developed biodiversity data standard. *PLoS ONE*, **7**, e29715.

Wiley, E.O. (1978). The Evolutionary Species Concept Reconsidered. *Systematic Zoology*, **27**, 17.

Willis, K.J. (ed.) (2017). *State of the World's Plants 2017. Report.* Royal Botanic Gardens, Kew, London.

Winter, M., Schweiger, O., Klotz, S., et al. (2009). Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. *Proceedings of the National Academy of Sciences of the United States*, **106**, 21721–21725.

Woodward, F.I. & Cramer, W. (1996). Plant functional types and climatic change: introduction. *Journal of Vegetation Science*, **7**, 306–308.

Worm, B. & Tittensor, D.P. (2018). *A theory of global biodiversity.* Princeton University Press, Princeton.

Wullschleger, S.D., Epstein, H.E., Box, E.O., Euskirchen, E.S., Goswami, S., Iversen, C.M., Kattge, J., Norby, R.J., van Bodegom, P.M. & Xu, X. (2014). Plant functional types in Earth system models. Past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. *Annals of Botany*, **114**, 1–16.

Xiao, X., McGlinn, D.J. & White, E.P. (2015). A strong test of the maximum entropy theory of ecology. *The American Naturalist*, **185**, E70-80.

Zanne, A.E., Pearse, W.D., Cornwell, W.K., McGlinn, D.J., Wright, I.J. & Uyeda, J.C. (2018). Functional biogeography of angiosperms. Life at the extremes. *New Phytologist*, **218**, 1697–1709.

Zanne, A.E., Tank, D.C., Cornwell, W.K., et al. (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature*, **506**, 89–92.

Zotz, G. (2013). The systematic distribution of vascular epiphytes - a critical update. *Botanical Journal of the Linnean Society*, **171**, 453–481.

Zuloaga, F.O., Morrone, O. & Belgrano, M.J. (2004). Catálogo de las Plantas Vasculares del Cono Sur (last accessed: March 16, 2015).

Acknowledgements

First, I want to thank my supervisor Prof. Dr. Holger Kreft. It was your lectures that sparked my interest in macroecology, your administrative and financial support that enabled my PhD studies, and your profound knowledge and dedicated supervision that were critical for completing this thesis. Moreover, I want to thank Dr. Patrick Weigelt, who provided invaluable advice in countless discussions of new ideas, preliminary results and statistical problems. Being your successor as a PhD student meant I had some big shoes to fill, but you always supported me in doing so. I am also grateful to my second supervisor, Prof. Dr. Erwin Bergmeier, for his much-appreciated scientific guidance.

Many thanks to my collaborators, especially the GloNAF team and Jens Kattge from TRY, who provided crucial input to Chapters 1 and 4. I am also thankful to the many people who directly or indirectly contributed data to the GIFT database and thus supported both my PhD thesis and a great project for biogeographical research.

At this point, I was going to thank my lab mates and friends, but there is no need to make that distinction. You all were great company during every single day in office and many, many coffee breaks, nights at the *Willi*, crazy motto parties and exciting conference trips. I am really thankful to have been part of the gang. Many thanks of course to all my friends outside the university.

My family supported me in so many ways not only throughout my PhD studies, but throughout the last ten years that I have been studying at the university. I am therefore most grateful to my father and his partner, Manuela, for always having my back, and always encouraging me to follow my interests.

Finally, I am grateful for the most precious finding of my PhD – that of my soulmate and love, Yue. We first met in the corridor of our lab, and while our scientific paths have diverged in the meantime, I am looking forward to explore every other path in life together with you. Thank you for everything.

Selbstständigkeitserklärung

Ich versichere hiermit, dass ich die vorliegende Dissertation ohne fremde Hilfe selbstständig verfasst und nur die von mir angegebenen Quellen und Hilfsmittel verwendet habe. Wörtlich oder sinngemäß aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht. Die Richtlinien zur Sicherung der guten wissenschaftlichen Praxis an der Universität Göttingen wurden von mir beachtet.

Christian König

Göttingen, den 20. August 2018

Appendix

A1 Supplementary information to Chapter 1

Box A1.1: Summary of the GIFT database.[1]

GIFT is a new database that integrates distributional, functional, and phylogenetic plant diversity data for macroecological analyses (Weigelt *et al.*, 2018). The project focusses on the aggregated end of the biodiversity data spectrum (Figure 1.1) and builds primarily on data that has been mobilized from the floristic literature and online databases. GIFT collaborates closely with other initiatives that aggregate and integrate global plant diversity data, e.g. GLoNAF (Global Naturalized Alien Flora, (van Kleunen *et al.*, 2015)) and the TRY database (Kattge *et al.*, 2011a).

The three fundamental units of information in GIFT are data sources, species and geographical entities. These units are tightly interconnected and integrated with additional descriptive data such as, e.g., bibliographic information (data sources), functional traits, phylogenetic relationships and distributions (species), or climatic characteristics (geographical units). All data in GIFT are subjected to semi-automated import and processing workflows that ensure full data integrity and traceability. Species names undergo an orthographical and taxonomic standardisation based on The Plant List version 1.1 (The Plant List, 2013) and additional taxonomic resources used by the Taxonomic Name Resolution Service (Boyle *et al.*, 2013). Functional trait data is standardized according to relevant trait literature (Pérez-Harguindeguy *et al.*, 2013; Garnier *et al.*, 2017) and further augmented by means of logical imputation (Figure 1.2). Geographical entities are linked to spatial polygons from established providers of geographical information, e.g. GADM (https://gadm.org) for administrative areas, TDWG (https://www.kew.org/tdwg-world-geographical-scheme-recording-plant-distributions) for ecoregions, or WDPA (https://www.protectedplanet.net) for protected areas. Unique identifiers allow exploring complex relationships among data, e.g. global range size distributions or the variation of trait syndromes across taxa.

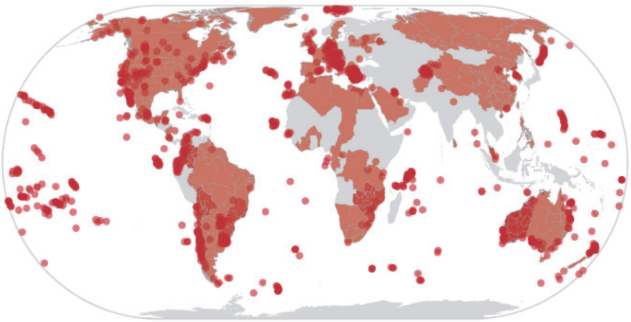


Figure A1.1: Geographical coverage of GIFT for native vascular plant checklists

GIFT currently comprises 322,002 species of vascular plants (Tracheophyta) and 2,893 geographical units, including administrative and natural entities such as countries, islands, protected areas and natural landscapes (). A full inventory of (native) vascular plants is available for 2,062 geographical units (Figure A1.1), spanning around eleven orders of magnitude in area (10 m2 – 3,069,765 km2). Inventory coverage is almost complete for the Americas and Australia, but still has gaps in Central Africa and large parts of Asia. With 1,273 checklists, island floras are particularly well covered in GIFT. With respect to functional traits, 155 data sources provide information on 83 standardized traits for 548,886 unstandardized species names, amounting to 3,475,337 original trait records. After logical imputation and subsequent aggregation, GIFT holds 2,307,100 species-trait-combinations for 267,978 standardized plant species. GIFT focusses on whole-plant traits and consequently has the highest coverage for morphological traits (e.g. woodiness: 234,214 spp., growth form: 213,372 spp.) and life history traits (e.g. Raunkiær life form: 100,607 spp., life cycle: 84,206 spp.). More complex categorical traits such as photosynthetic pathway (31,534 spp.), dispersal syndrome (8,204 spp.), or pollination syndrome (4,511 spp.), as well as numerical traits such as maximum/mean plant height (61,551 spp.), mean seed mass (23,874 spp.), or mean specific leaf area (2,304 spp.) are generally rarer.

---

[1] This box is part of the main text in the original manuscript. For the purpose of this dissertation, it has been moved to the appendix because Chapter 2 provides the same information in more detail.

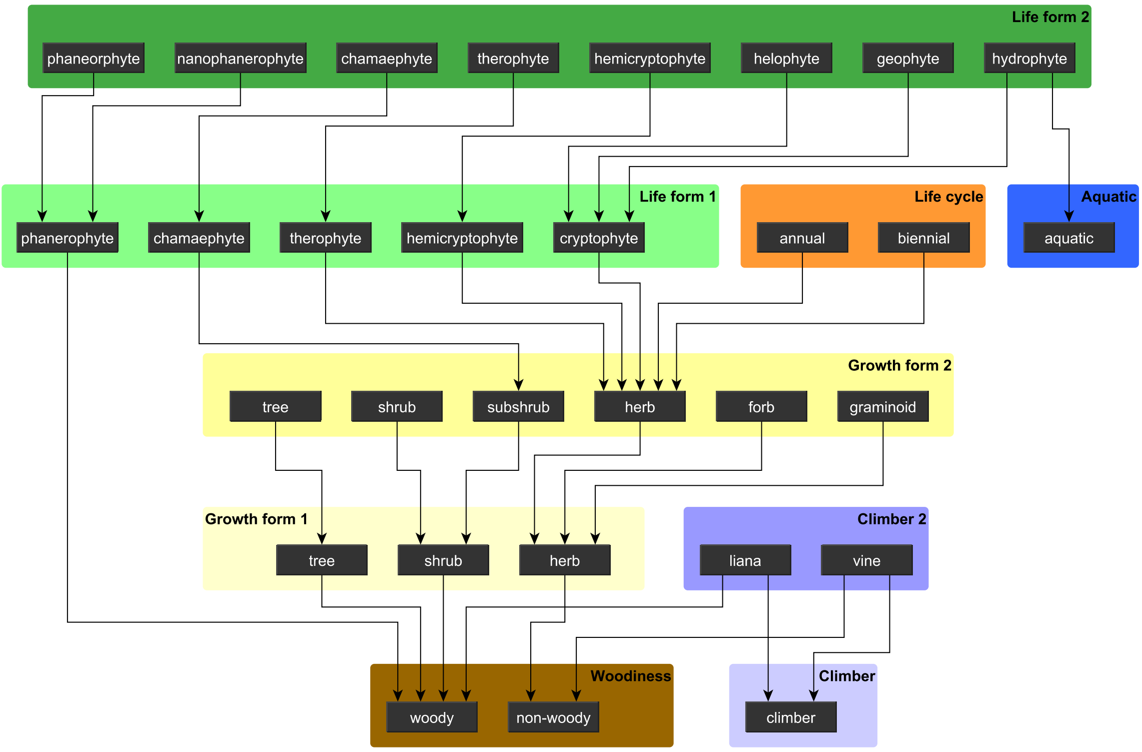


Figure A1.2: Main module of the directed graph used for hierarchical trait derivation in GIFT, defining unambiguous relationships among 30 categorical levels from five functional traits (life form, life cycle, aquatic, growth form, climber and woodiness). Some traits are represented in multiple versions (e.g. growth form) to account for varying levels of detail of original information. The full list of parent-child relationships used for trait derivation in GIFT (71 connections among 89 categorical levels) is given in Table A2.2.

A2 Supplementary information to Chapter 2

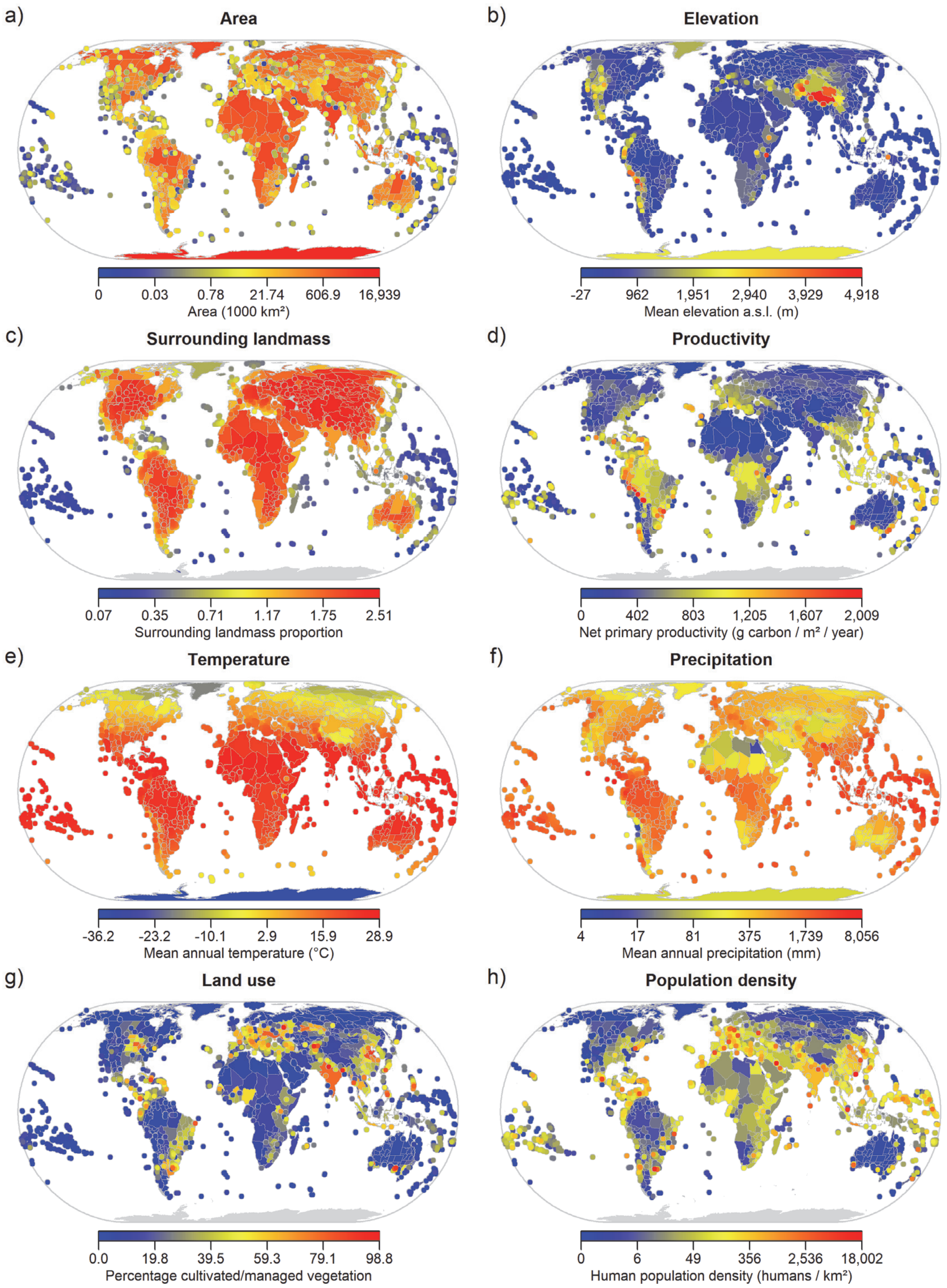


Figure A2.1: Geographical summary of selected environmental variables in GIFT. For a full list of geographic, environmental and socio-economic variables and source references see Table A2.3. Regions <25,000 km² are plotted as points.

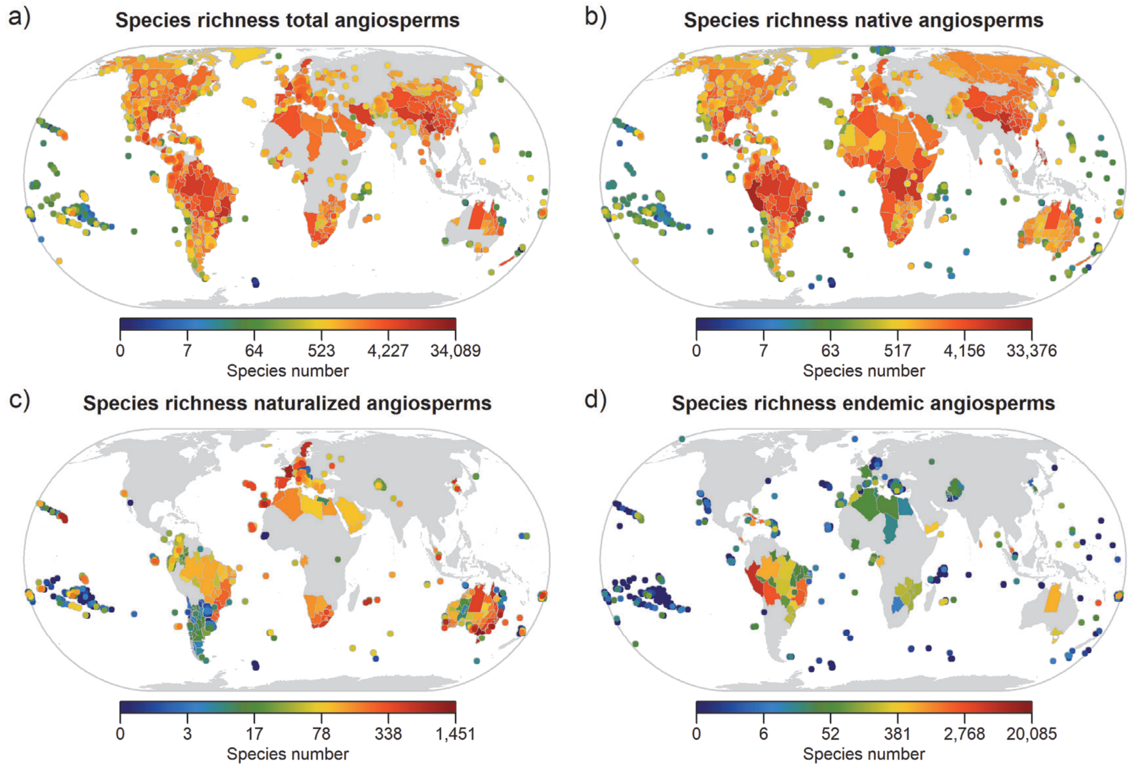


Figure A2.2: Spatial coverage of floristic subsets in GIFT. Species richness of total angiosperms (a) refers to regions with information on both native and introduced angiosperm species or species with unresolved floristic status. Most commonly resources in GIFT include information on native species (b), while information on introduced naturalized (c) and endemic species (d) is considerably rarer. Endemism information is most common for island regions. Regions <25,000 km² are plotted as points.

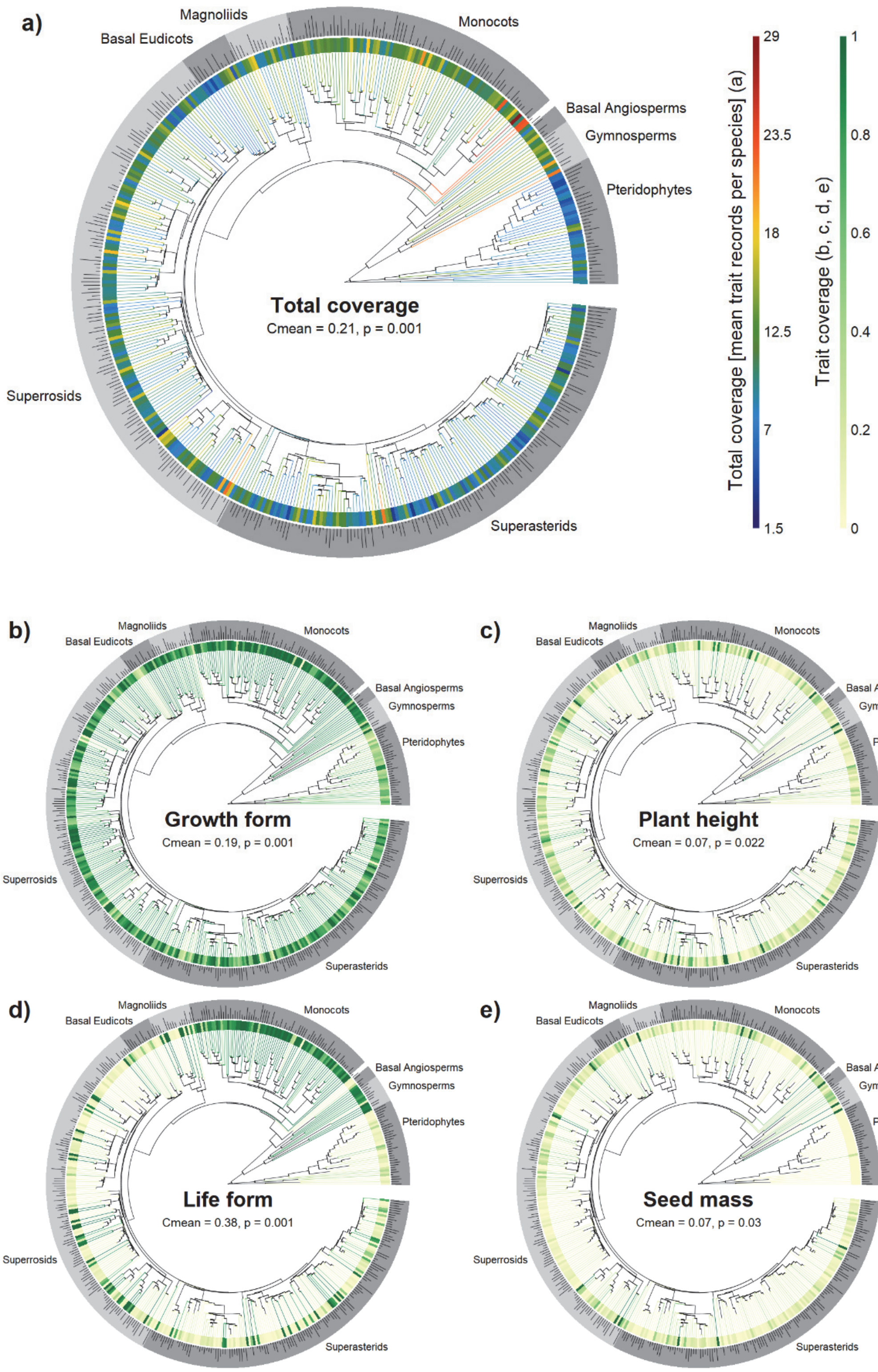


Figure A2.3: Taxonomic trait coverage of GIFT across all functional traits at the family level (a; mean trait records per species) and for four selected functional traits individually (b-e; number of species with trait information/total number of species). Tip color and inner ring color denote trait coverage per family, outer ring delimits major clades of vascular plants. The height of bars in the outer ring is proportional to $\log_{10}$ family size. Phylogenetic signal in taxonomic coverage was assessed as Abouheif's Cmean, a measure of phylogenetic autocorrelation based on the sum of the successive squared differences between values of neighbouring tips in the phylogeny (Abouheif, 1999).

Table A2.1: Functional traits in GIFT, their broader trait categories, type, units and factor levels respectively and the number of species covered in GIFT.

Trait-ID	Category	Trait	Trait variant	Type	Unit/Level	Species count
1.1.1	Morphology	Woodiness	Woodiness 1	categorial	woody, non-woody, variable	234214
1.2.1	Morphology	Growth form	Growth form 1	categorial	herb, shrub, tree, other	213372
1.2.2	Morphology	Growth form	Growth form 2	categorial	herb, graminoid, forb, subshrub, shrub, tree, palm, other	165540
1.3.1	Morphology	Epiphyte	Epiphyte 1	categorial	obligatory, facultative, terrestrial	206699
1.3.2	Morphology	Epiphyte	Epiphyte 2	categorial	holoepiphyte, hemiepiphyte, prim_hemiepiphyte, sec_hemiepiphyte, facultative, terrestrial	70816
1.4.1	Morphology	Climber	Climber 1	categorial	obligatory, facultative, self-supporting	223280
1.4.2	Morphology	Climber	Climber 2	categorial	liana, vine, self-supporting	93098
1.5.1	Morphology	Parasite	Parasite 1	categorial	obligatory, facultative, independent	209609
1.6.1	Morphology	Plant height	Plant height min	numeric	m	28859
1.6.2	Morphology	Plant height	Plant height max	numeric	m	53449
1.6.3	Morphology	Plant height	Plant height mean	numeric	m	16127
1.7.1	Morphology	Aquatic	Aquatic 1	categorial	aquatic, semiaquatic, terrestrial	148432
1.8.2	Morphology	DBH	DBH max	numeric	cm	25
1.8.3	Morphology	DBH	DBH mean	numeric	cm	25
2.1.1	Life history	Lifecycle	Lifecycle 1	categorial	annual, biennial, perennial, variable	84206
2.3.1	Life history	Life form	Life form 1	categorial	phanerophyte, chamaephyte, hemicryptophyte, cryptophyte, therophyte	81198
2.3.2	Life history	Life form	Life form 2	categorial	nanophanerophyte, phanerophyte, chamaephyte, hemicryptophyte, geophyte, hydrophyte, helophyte, therophyte, lithophyte	100607
2.4.1	Life history	Deciduousness	Deciduousness 1	categorial	deciduous, evergreen, variable	8069
3.1.1	Reproduction	Self fertilization	Self fertilization 1	categorial	present, absent	3538
3.10.1	Reproduction	Seed length	Seed length min	numeric	mm	1196
3.10.2	Reproduction	Seed length	Seed length max	numeric	mm	1344
3.10.3	Reproduction	Seed length	Seed length mean	numeric	mm	535
3.11.1	Reproduction	Seed width	Seed width min	numeric	mm	100
3.11.2	Reproduction	Seed width	Seed width max	numeric	mm	277
3.12.1	Reproduction	Seed height	Seed height min	numeric	mm	7
3.12.2	Reproduction	Seed height	Seed height max	numeric	mm	148
3.13.1	Reproduction	Fruit length	Fruit length min	numeric	cm	962
3.13.2	Reproduction	Fruit length	Fruit length max	numeric	cm	1096
3.13.3	Reproduction	Fruit length	Fruit length mean	numeric	cm	520
3.14.1	Reproduction	Fruit width	Fruit width min	numeric	cm	380
3.14.2	Reproduction	Fruit width	Fruit width max	numeric	cm	432
3.14.3	Reproduction	Fruit width	Fruit width mean	numeric	cm	561
3.15.1	Reproduction	Fruit height	Fruit height min	numeric	cm	54
3.15.2	Reproduction	Fruit height	Fruit height max	numeric	cm	54
3.15.3	Reproduction	Fruit height	Fruit height mean	numeric	cm	54
3.16.1	Reproduction	Fruit type	Fruit type 1	categorial	achene, baccate, berry, capsule, drupe, follicle, lomentum, nut, pod, pome, schizocarp, siliqua, utricle, other	5883
3.17.1	Reproduction	Dehiscence	Dehiscence 1	categorial	dehiscent, indehiscent	5513
3.18.1	Reproduction	Fruit dryness	Fruit dryness 1	categorial	dry, fleshy	2559
3.19.1	Reproduction	Seed volume	Seed volume min	numeric	mm^3	2805
3.19.2	Reproduction	Seed volume	Seed volume max	numeric	mm^3	2805
3.19.3	Reproduction	Seed volume	Seed volume mean	numeric	mm^3	2805
3.2.1	Reproduction	Seed mass	Seed mass min	numeric	g	22655
3.2.2	Reproduction	Seed mass	Seed mass max	numeric	g	22655
3.2.3	Reproduction	Seed mass	Seed mass mean	numeric	g	23874
3.20.1	Reproduction	Monocarpy	Monocarpy	categorial	monocarp, polycarp	88
3.3.1	Reproduction	Dispersal syndrome	Dispersal syndrome 1	categorial	anemochorous, zoochorous, autochorous, hydrochorous, unspecialized	8204
3.3.2	Reproduction	Dispersal syndrome	Dispersal syndrome 2	categorial	anemochorous, anthropochorous, autochorous, endozoochorous, epizoochorous, hydrochorous,	3003

					myrmecochorous, unspecialized, zoochorous	
3.4.1	Reproduction	Reproduction_sexual	Reproduction sexual 1	categorial	dioecious, monoecious, bisexual	2699
3.5.1	Reproduction	Reproduction_asexual	Reproduction asexual 1	categorial	present, absent	1703
3.5.2	Reproduction	Reproduction_asexual	Reproduction asexual 2	categorial	above-ground, below-ground	12
3.6.1	Reproduction	Pollination syndrome	Pollination syndrome 1	categorial	wind, water, insect, bird, bat, other	4511
3.6.2	Reproduction	Pollination syndrome	Pollination syndrome 2	categorial	wind, water, bee, beetle, ant, butterfly, moth, fly, insect, bird, bat, other	4014
3.7.1	Reproduction	Flowering time	Flowering start	categorial	Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, variable	7013
3.7.2	Reproduction	Flowering time	Flowering end	categorial	Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, variable	6102
3.8.1	Reproduction	Fruiting time	Fruiting start	categorial	Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, variable	514
3.8.2	Reproduction	Fruiting time	Fruiting end	categorial	Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, variable	512
3.9.1	Reproduction	Seeds_per_fruit	Seeds per fruit	categorial	0, 1-10, 11-100, 101-1000, >1000	783
4.1.1	Physiology	Specific Leaf Area (SLA)	SLA min	numeric	cm^2/g	252
4.1.2	Physiology	Specific Leaf Area (SLA)	SLA max	numeric	cm^2/g	252
4.1.3	Physiology	Specific Leaf Area (SLA)	SLA mean	numeric	cm^2/g	2304
4.10.1	Physiology	Carnivory	Carnivory	categorial	carnivorous, non-carnivorous	1769
4.2.1	Physiology	Photosynthetic pathway	Photosynthetic pathway	categorial	C3, C4, CAM	31534
4.3.1	Physiology	Stem specific density (SSD)	SSD min	numeric	mg/cm^3	371
4.3.2	Physiology	Stem specific density (SSD)	SSD max	numeric	mg/cm^3	371
4.3.3	Physiology	Stem specific density (SSD)	SSD mean	numeric	mg/cm^3	1094
4.4.3	Physiology	Leaf size	Leaf size mean	numeric	cm^2	1708
4.5.1	Physiology	Nitrogen fixer	Nitrogen fix 1	categorial	yes, no	10143
4.6.1	Physiology	Leaf length	Leaf length min	numeric	cm	5787
4.6.2	Physiology	Leaf length	Leaf length max	numeric	cm	6581
4.6.3	Physiology	Leaf length	Leaf length mean	numeric	cm	132
4.7.1	Physiology	Leaf width	Leaf width min	numeric	cm	5203
4.7.2	Physiology	Leaf width	Leaf width max	numeric	cm	5791
4.7.3	Physiology	Leaf width	Leaf width mean	numeric	cm	132
4.8.3	Physiology	Leaf thickness	Leaf thickness mean	numeric	mm	134
4.9.1	Physiology	Leaf dry matter content (LDMC)	LDMC min	numeric	mg/g	483
4.9.2	Physiology	Leaf dry matter content (LDMC)	LDMC max	numeric	mg/g	483
4.9.3	Physiology	Leaf dry matter content (LDMC)	LDMC mean	numeric	mg/g	483
5.1.1	Genetics	Chromosome number	Chromosome number	text		730
6.1.1	Ecology	Elevational range	Elevational range min	numeric	m AMSL	53024
6.1.2	Ecology	Elevational range	Elevational range max	numeric	m AMSL	54767
6.1.3	Ecology	Elevational range	Elevational range mean	numeric	m AMSL	8170
6.2.1	Ecology	Bedrock	Bedrock 1	text		290
6.3.1	Ecology	Habitat	Habitat 1	text		35542

Table A2.2: Links between parent traits and derived traits used in the hierarchical trait derivation in GIFT.

Trait-ID parent	Trait parent	Trait value parent	Trait-ID child	Trait child	Trait value child
3.6.2	Pollination syndrome 2	ant	3.6.1	Pollination syndrome 1	insect
3.6.2	Pollination syndrome 2	wind	3.6.1	Pollination syndrome 1	wind
3.6.2	Pollination syndrome 2	water	3.6.1	Pollination syndrome 1	water
3.6.2	Pollination syndrome 2	bee	3.6.1	Pollination syndrome 1	insect
3.6.2	Pollination syndrome 2	beetle	3.6.1	Pollination syndrome 1	insect
3.6.2	Pollination syndrome 2	butterfly	3.6.1	Pollination syndrome 1	insect
3.6.2	Pollination syndrome 2	moth	3.6.1	Pollination syndrome 1	insect
3.6.2	Pollination syndrome 2	bird	3.6.1	Pollination syndrome 1	bird
3.6.2	Pollination syndrome 2	bat	3.6.1	Pollination syndrome 1	bat
3.3.2	Dispersal syndrome 2	anemochorous	3.3.1	Dispersal syndrome 1	anemochorous
3.3.2	Dispersal syndrome 2	zoochorous	3.3.1	Dispersal syndrome 1	zoochorous
3.3.2	Dispersal syndrome 2	endozoochorous	3.3.1	Dispersal syndrome 1	zoochorous
3.3.2	Dispersal syndrome 2	epizoochorous	3.3.1	Dispersal syndrome 1	zoochorous
3.3.2	Dispersal syndrome 2	myrmecochorous	3.3.1	Dispersal syndrome 1	zoochorous
3.3.2	Dispersal syndrome 2	autochorous	3.3.1	Dispersal syndrome 1	autochorous
3.3.2	Dispersal syndrome 2	hydrochorous	3.3.1	Dispersal syndrome 1	hydrochorous
3.3.2	Dispersal syndrome 2	anthropochorous	3.3.1	Dispersal syndrome 1	unspecialized
3.3.2	Dispersal syndrome 2	unspecialized	3.3.1	Dispersal syndrome 1	unspecialized
2.3.2	Life form 2	nanophanerophyte	2.3.1	Life form 1	phanerophyte
2.3.2	Life form 2	phanerophyte	2.3.1	Life form 1	phanerophyte
2.3.2	Life form 2	chamaephyte	2.3.1	Life form 1	chamaephyte
2.3.2	Life form 2	hemicryptophyte	2.3.1	Life form 1	hemicryptophyte
2.3.2	Life form 2	geophyte	2.3.1	Life form 1	cryptophyte
2.3.2	Life form 2	hydrophyte	2.3.1	Life form 1	cryptophyte
2.3.2	Life form 2	hydrophyte	1.7.1	Aquatic 1	aquatic
2.3.2	Life form 2	helophyte	2.3.1	Life form 1	cryptophyte
2.3.2	Life form 2	therophyte	2.3.1	Life form 1	therophyte
2.3.1	Life form 1	chamaephyte	1.2.2	Growth form 2	subshrub
2.3.1	Life form 1	hemicryptophyte	1.2.2	Growth form 2	herb
2.3.1	Life form 1	cryptophyte	1.2.2	Growth form 2	herb
2.3.1	Life form 1	therophyte	1.2.2	Growth form 2	herb
1.4.2	Climber 2	vine	1.4.1	Climber 1	obligatory
1.4.2	Climber 2	vine	1.1.1	Woodiness 1	non-woody
1.4.2	Climber 2	liana	1.4.1	Climber 1	obligatory
1.4.2	Climber 2	liana	1.1.1	Woodiness 1	woody
1.4.2	Climber 2	self-supporting	1.4.1	Climber 1	self-supporting
1.3.2	Epiphyte 2	holoepiphyte	1.3.1	Epiphyte 1	obligatory
1.3.2	Epiphyte 2	hemiepiphyte	1.3.1	Epiphyte 1	obligatory

1.3.2	Epiphyte 2	prim hemiepiphyte	1.3.1	Epiphyte 1	obligatory
1.3.2	Epiphyte 2	sec hemiepiphyte	1.3.1	Epiphyte 1	obligatory
1.3.2	Epiphyte 2	facultative	1.3.1	Epiphyte 1	facultative
1.3.2	Epiphyte 2	terrestrial	1.3.1	Epiphyte 1	terrestrial
1.2.2	Growth form 2	herb	1.2.1	Growth form 1	herb
1.2.2	Growth form 2	graminoid	1.2.1	Growth form 1	herb
1.2.2	Growth form 2	forb	1.2.1	Growth form 1	herb
1.2.2	Growth form 2	subshrub	1.2.1	Growth form 1	shrub
1.2.2	Growth form 2	shrub	1.2.1	Growth form 1	shrub
1.2.2	Growth form 2	tree	1.2.1	Growth form 1	tree
1.2.2	Growth form 2	palm	1.2.1	Growth form 1	other
1.2.2	Growth form 2	other	1.2.1	Growth form 1	other
1.2.1	Growth form 1	herb	1.1.1	Woodiness 1	non-woody
1.2.1	Growth form 1	shrub	1.1.1	Woodiness 1	woody
1.2.1	Growth form 1	tree	1.1.1	Woodiness 1	woody
2.1.1	Lifecycle 1	biennial	1.2.2	Growth form 2	herb
2.1.1	Lifecycle 1	annual	1.2.2	Growth form 2	herb
2.3.1	Life form 1	phanerophyte	1.1.1	Woodiness 1	woody
3.6.2	Pollination syndrome 2	insect	3.6.1	Pollination syndrome 1	insect
3.6.2	Pollination syndrome 2	other	3.6.1	Pollination syndrome 1	other
3.6.2	Pollination syndrome 2	fly	3.6.1	Pollination syndrome 1	insect
1.5.1	Parasite 1	obligatory	1.3.2	Epiphyte 2	terrestrial
3.16.1	Fruit type 1	achene	3.17.1	Dehiscence 1	indehiscent
3.16.1	Fruit type 1	berry	3.17.1	Dehiscence 1	indehiscent
3.16.1	Fruit type 1	drupe	3.17.1	Dehiscence 1	indehiscent
3.16.1	Fruit type 1	capsule	3.17.1	Dehiscence 1	dehiscent
3.16.1	Fruit type 1	baccate	3.17.1	Dehiscence 1	indehiscent
3.16.1	Fruit type 1	drupe	3.17.1	Dehiscence 1	indehiscent
3.16.1	Fruit type 1	lomentum	3.17.1	Dehiscence 1	indehiscent
3.16.1	Fruit type 1	nut	3.17.1	Dehiscence 1	indehiscent
3.16.1	Fruit type 1	pod	3.17.1	Dehiscence 1	dehiscent
3.16.1	Fruit type 1	siliqua	3.17.1	Dehiscence 1	dehiscent
3.16.1	Fruit type 1	utricle	3.17.1	Dehiscence 1	indehiscent

Table A2.3: Groups of physical geographical, environmental and socio-economic variables in GIFT. Metrics of datasets of type miscellaneous are calculated based on the regions polygons and eventually additional resources cited under "References". For resources of type raster, summary statistics (15 quantiles including minimum, median and maximum, mean, standard deviation, mode, number of unique values, Shannon diversity and number of cells) are calculated for all raster cells that fall into a region or are crossed by its border.

Dataset	Type	Reference
Accessibility to cities	Raster	Weiss *et al.* 2018
Archipelago	Miscellaneous	
Area of ice cover today and during last glacial maximum	Miscellaneous	Ehlers *et al.* 2011
Area of region and largest included landmass	Miscellaneous	
Biome	Miscellaneous	Dinerstein *et al.* 2017
Canopy height	Raster	Simard *et al.* 2011
Centroid latitude and longitude coordinates	Miscellaneous	
Chelsa 1.2 bioclimatic variables	Raster	Karger *et al.* 2017
Consensus land cover	Raster	Tuanmu & Jetz 2014
Distance to nearest mainland	Miscellaneous	Weigelt *et al.* 2013
GDP from satellite	Raster	Ghosh *et al.* 2010
Geographic extent of region and largest included landmass	Miscellaneous	
Global aridity and potential evapotranspiration	Raster	Zomer *et al.* 2008
Global cloud cover	Raster	Wilson & Jetz 2016
Global gridded soil information	Raster	Hengl *et al.* 2017
Global habitat heterogeneity	Raster	Tuanmu & Jetz 2015
Global multi-resolution terrain elevation data	Raster	USGS 2011
Global soil water balance	Raster	Trabucco & Zomer 2010
Gridded population of the world 4	Raster	Doxsey-Whitfield *et al.* 2015
Human footprint	Raster	Venter *et al.* 2016
Last glacial maximum mainland connection, area and number of landmass entities	Miscellaneous	Weigelt *et al.* 2016, Weatherall *et al.* 2015
Last of the wild human influence	Raster	WCS & CIESIN 2005a, 2005b
MODIS gross and net primary productivity	Raster	Zhao & Running 2010
Number of landmass entities comprised by region	Miscellaneous	
Past climate change velocity	Raster	Weigelt *et al.* 2013, Hijmans *et al.* 2005
Species richness at nearest mainland	Miscellaneous	Weigelt *et al.* 2013, Kreft & Jetz 2007
Surrounding landmass proportion	Miscellaneous	Weigelt *et al.* 2013
Takhtajan floristic region	Miscellaneous	Takhtajan 1986
TDWG region	Miscellaneous	Brummitt 2001
Tree density	Raster	Crowther *et al.* 2015
WorldClim 1.4 last glacial maximum climate	Raster	Hijmans *et al.* 2005
WorldClim 2.0 bioclimatic variables	Raster	Fick & Hijmans 2017

Brummitt, R. K. (2001) *World geographical scheme for recording plant distributions. Plant taxonomic database standards No. 2. (with assistance from F. Pando, S. Hollis, N. A. Brummitt and others)*. Edition 2. Hunt Institute for Botanical Documentation, Carnegie Mellon University, Pittsburgh, US.

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.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E., et al. (2017) An ecoregion-based approach to protecting half the terrestrial realm. *BioScience* **67**, 534-545.

Doxsey-Whitfield, E., MacManus, K., Adamo, S.B., Pistolesi, L., Squires, J., Borkovska, O. & Baptista, S.R. (2015) Taking advantage of the improved availability of census data: A first look at the gridded population of the world, version 4. *Papers in Applied Geography* **1**, 226-234.

Ehlers, J., Gibbard, P.L. & Hughes, P.D. (2011) *Quaternary glaciations - extent and chronology. A closer look.* Elsevier, Amsterdam, The Netherlands.

Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302-4315.

Ghosh, T., Powell, R.L., Elvidge, C.D., Baugh, K.E., Sutton, P.C. & Anderson, S. (2010) Shedding light on the global distribution of economic activity. *The Open Geography Journal* **3**, 147-160.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017) SoilGrids250m. Global gridded soil information based on machine learning. *PloS one* **12**, e0169748. DOI: 10.1371/journal.pone.0169748.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.

Karger, D. Nikolaus, Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., et al. (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data* **4**, 170122. DOI: 10.1038/sdata.2017.122.

Kreft, H. & Jetz, W. (2007) Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences* **104**, 5925-5930.

Simard, M., Pinto, N., Fisher, J.B. & Baccini, A. (2011) Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research* **116**. G04021. DOI: 10.1029/2011JG001708.

Takhtajan, A. (1986) *Floristic regions of the world.* University of California Press, Berkeley, US.

Trabucco, A. & Zomer, R. J. (2010) *Global Soil Water Balance Geospatial. CGIAR Consortium for Spatial Information.* Available at: http://www.cgiar-csi.org/data/global-high-resolution-soil-water-balance (accessed 18 April 2016).

Tuanmu, M.-N. & Jetz, W. (2014) A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography* **23**, 1031-1045.

Tuanmu, M.-N. & Jetz, W. (2015) A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Global Ecology and Biogeography* **24**, 1329-1339.

USGS (2011) *Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010).* U.S. Geological Survey. Virginia. Available at https://lta.cr.usgs.gov/GMTED2010 (accessed 2 May 2017).

Venter, O., Sanderson, E.W., Magrach, A., Allan, J.R., Beher, J., Jones, K., et al. (2016) Global terrestrial Human Footprint maps for 1993 and 2009. *Scientific Data* **3**, 160067. DOI: 10.1038/sdata.2016.67.

WCS & CIESIN (2005a) *Last of the Wild Project, Version 2, 2005 (LWP-2). Global Human Footprint Dataset (Geographic).* NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, US. DOI: 10.7927/H4M61H5F.

WCS & CIESIN (2005b) *Last of the Wild Project, Version 2, 2005 (LWP-2). Global Human Influence Index (HII) Dataset (Geographic).* NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, US. DOI: 10.7927/H4BP00QC.

Weatherall, P., Marks, K.M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J.E., et al. (2015) A new digital bathymetric model of the world's oceans. *Earth and Space Science* **2**, 331-345. DOI: 10.1002/2015EA000107.

Weigelt, P., Jetz, W. & Kreft, H. (2013) Bioclimatic and physical characterization of the world's islands. *Proceedings of the National Academy of Sciences* **110**, 15307-15312. DOI: 10.1073/pnas.1306309110.

Weigelt, P., Steinbauer, M.J., Cabral, J.S. & Kreft, H. (2016) Late Quaternary climate change shapes island biodiversity. *Nature* **532**, 99-102.

Weiss, D.J., Nelson, A., Gibson, H.S., Temperley, W., Peedell, S., Lieber, A., et al. (2018) A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333. DOI: 10.1038/nature25181.

Wilson, A.M. & Jetz, W. (2016) Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. *PLOS Biology* **14**, e1002415. DOI: 10.1371/journal.pbio.1002415.

Zhao, M. & Running, S.W. (2010) Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science* **329**, 940-943.

Zomer, R.J., Trabucco, A., Bossio, D.A. & Verchot, L.V. (2008) Climate change mitigation. A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems & Environment* **126**, 67-80.

A3 Supplementary information to Chapter 3

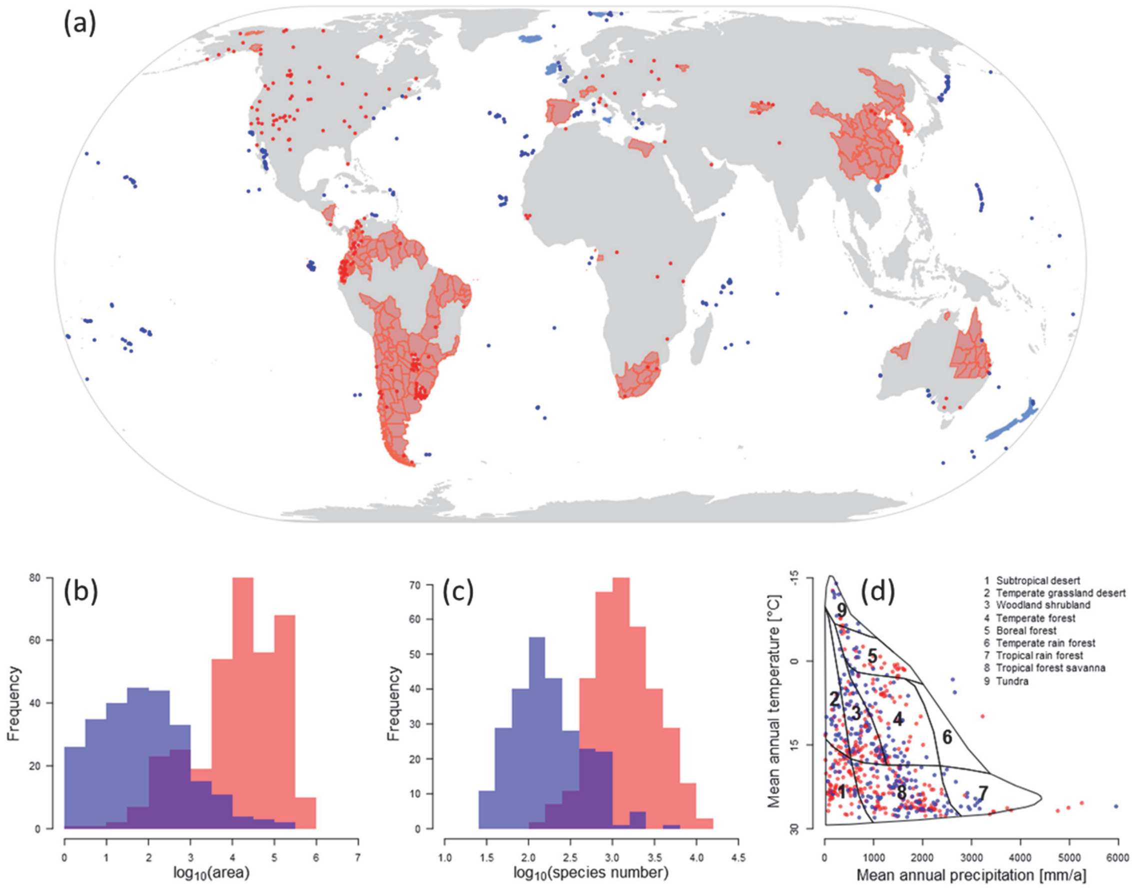


Figure A3.1: Summary of the operational geographical units (OGUs) analysed in Chapter 4. In all panels, island OGUs are depicted blue and mainland OGUs are depicted red. (a) Geographical distribution, Projection: Eckert IV, (b) Histogram of area per OGU, (c) Histogram of species number per OGU and (d) Major climatic characteristics of OGUs plotted in Whittaker biome space (Whittaker, R.H. (1970). *Communities and ecosystems*. Macmillan, New York).

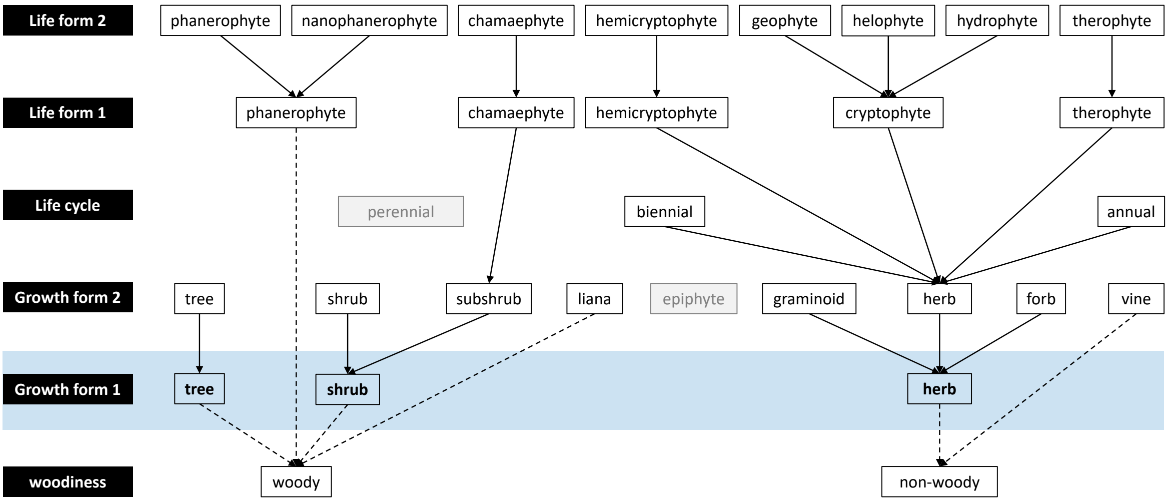


Figure A3.2: Graph structure used to derive plant growth form from other available functional traits. Greyed out fields indicate that there is no meaningful derivation for the respective trait. Dotted lines indicate derivations that were not used in this study.

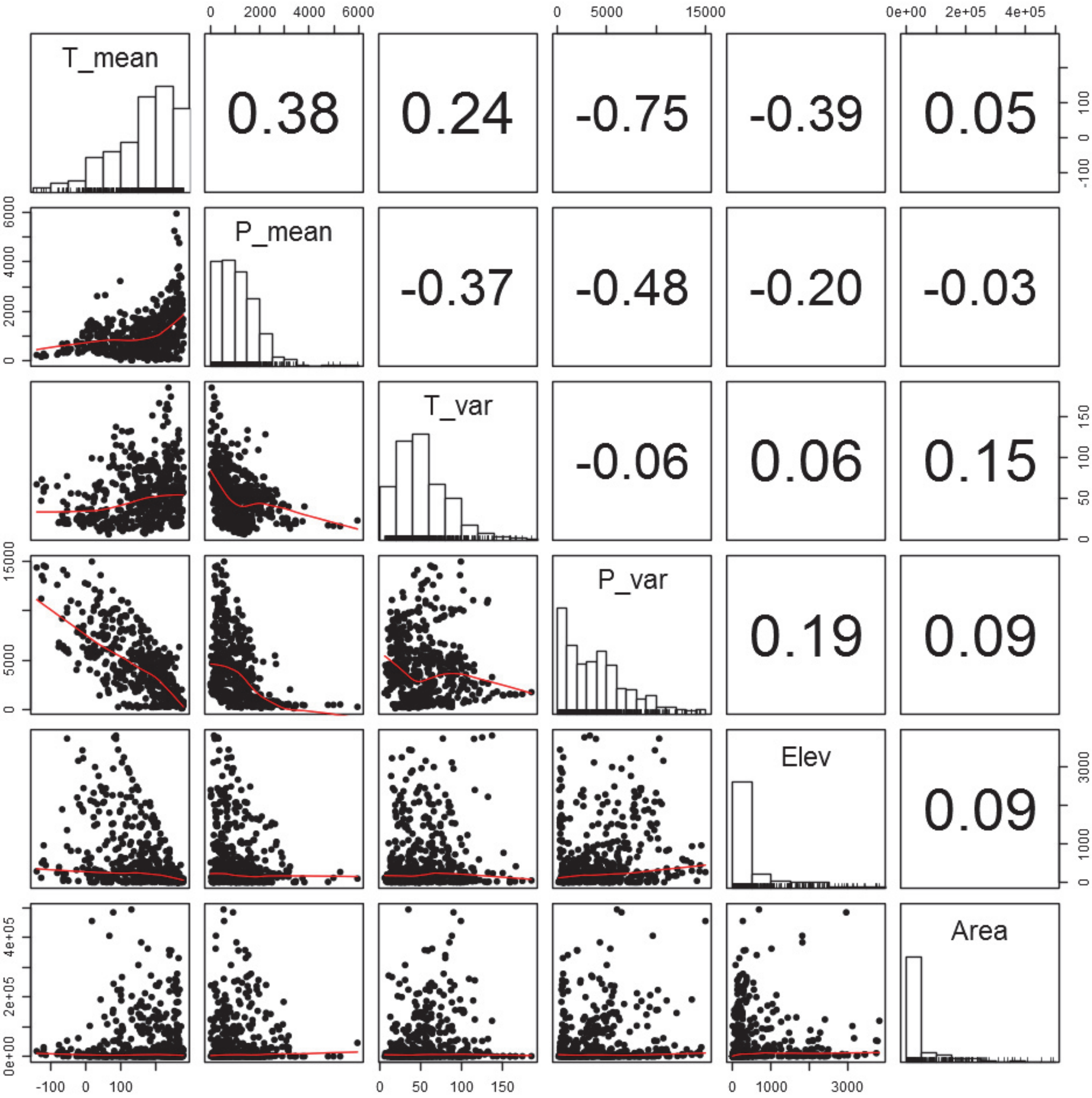


Figure A3.3: Pairwise correlation of predictor variables used in GDM. Lower left triangle: Scatterplot for pairs of variables and LOESS-fit (red line), Diagonal: Histogram and variable abbreviations, Upper right triangle: Pearson correlation coefficients between pairs of variables. The highest observed variance inflation factor (not shown) was 3.25 for mean annual temperature (T_mean), suggesting no problematic amounts of multicollinearity among predictors.

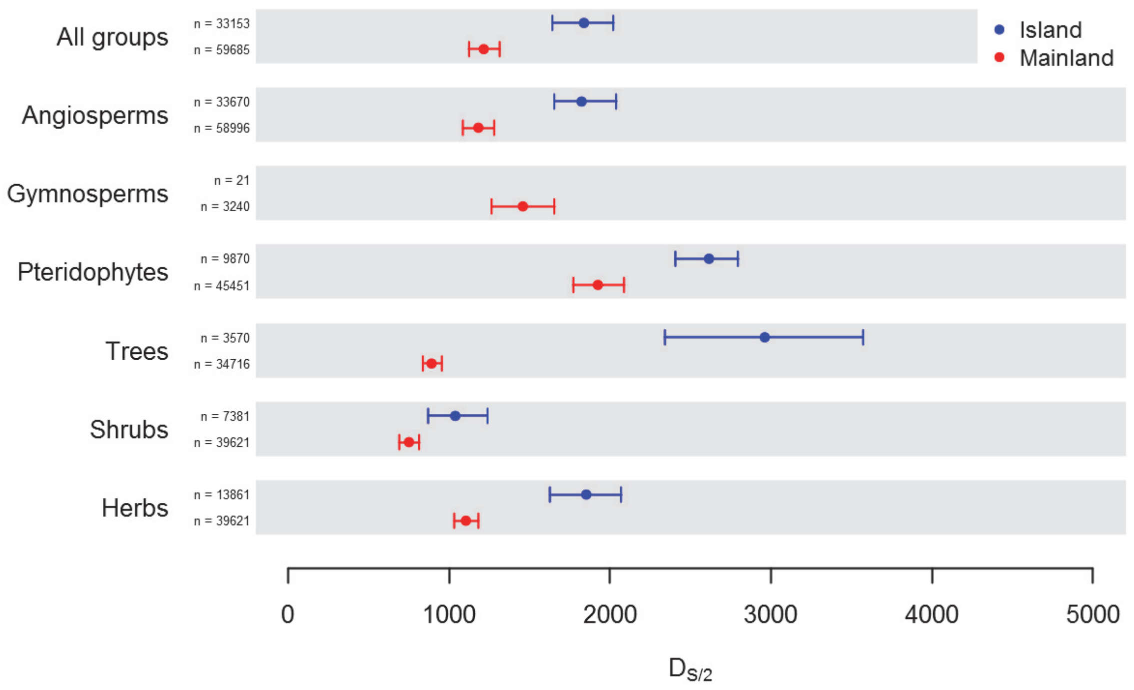


Figure A3.4: Distance decay for taxonomic and functional groups (island vs. mainland comparison). $D_{S/2}$ denotes the distance after which similarity decreases by 50 % (halving distance) and n denotes the number of unique pairwise combinations within each subset. Model coefficients (dots) were obtained using a log-binomial generalized linear model (GLM) with an intercept of 1. Confidence intervals (whiskers) were computed by subsampling the data 250 times, refitting the model and taking the 2.5 and 97.5 percentiles of the sampling distribution of coefficient estimates. No significant model was obtained for subset 'gymnosperms × island'.

Table A3.1: Distance decay model summaries. Halving distance ($D_{S/2}$) denotes the distance after which similarity decreases by 50 % and n denotes the number of unique pairwise combinations within each subset. Model coefficients ('beta_hat') were obtained using a log-binomial generalized linear model (GLM) with an intercept of 1. Confidence intervals ('beta_CI_min' and 'beta_CI_max') were computed by subsampling the data 1000 times, refitting the model and taking the 2.5 and 97.5 percentiles of the sampling distribution of coefficient estimates. Significance levels: *** – $p < 0.001$; n.s. – not significant

subset	n	halving distance ($D_{S/2}$)	beta_hat	beta_CI_min	beta_CI_max	significance
all entities	182106	1576	-0.00044	-0.000469	-0.000417	***
angiosperms	182106	1544	-0.000449	-0.000479	-0.000425	***
angiosperms × island	33670	1826	-0.00038	-0.000419	-0.00034	***
angiosperms × mainland	58996	1184	-0.000586	-0.000638	-0.000541	***
atolls	406	5967	-0.000116	-0.000147	-0.000092	***
continental islands	4005	1384	-0.000501	-0.000569	-0.000454	***
gymnosperms	3828	1476	-0.000469	-0.00055	-0.000413	***
gymnosperms × island	21	1099	-0.000631	-0.013594	-0.000263	n.s.
gymnosperms × mainland	3240	1459	-0.000475	-0.000548	-0.000419	***
herbs	100576	1523	-0.000455	-0.000487	-0.000421	***
herbs × island	13861	1853	-0.000374	-0.000426	-0.000335	***
herbs × mainland	39621	1106	-0.000627	-0.000671	-0.000586	***
islands	33153	1840	-0.000377	-0.000422	-0.000343	***
mainland	59685	1216	-0.00057	-0.000616	-0.000527	***
oceanic islands	9870	1628	-0.000426	-0.000482	-0.000383	***
pteridophytes	97903	2156	-0.000322	-0.000345	-0.000306	***
pteridophytes × island	9870	2612	-0.000265	-0.000288	-0.000248	***
pteridophytes × mainland	45451	1925	-0.00036	-0.000391	-0.000332	***
shrubs	81406	826	-0.000839	-0.000907	-0.000777	***
shrubs × island	7381	1039	-0.000667	-0.000796	-0.000559	***
shrubs × mainland	39621	752	-0.000922	-0.001002	-0.000851	***
trees	60726	1163	-0.000596	-0.00066	-0.000543	***
trees × island	3570	2959	-0.000234	-0.000296	-0.000194	***
trees × mainland	34716	892	-0.000777	-0.000827	-0.000725	***

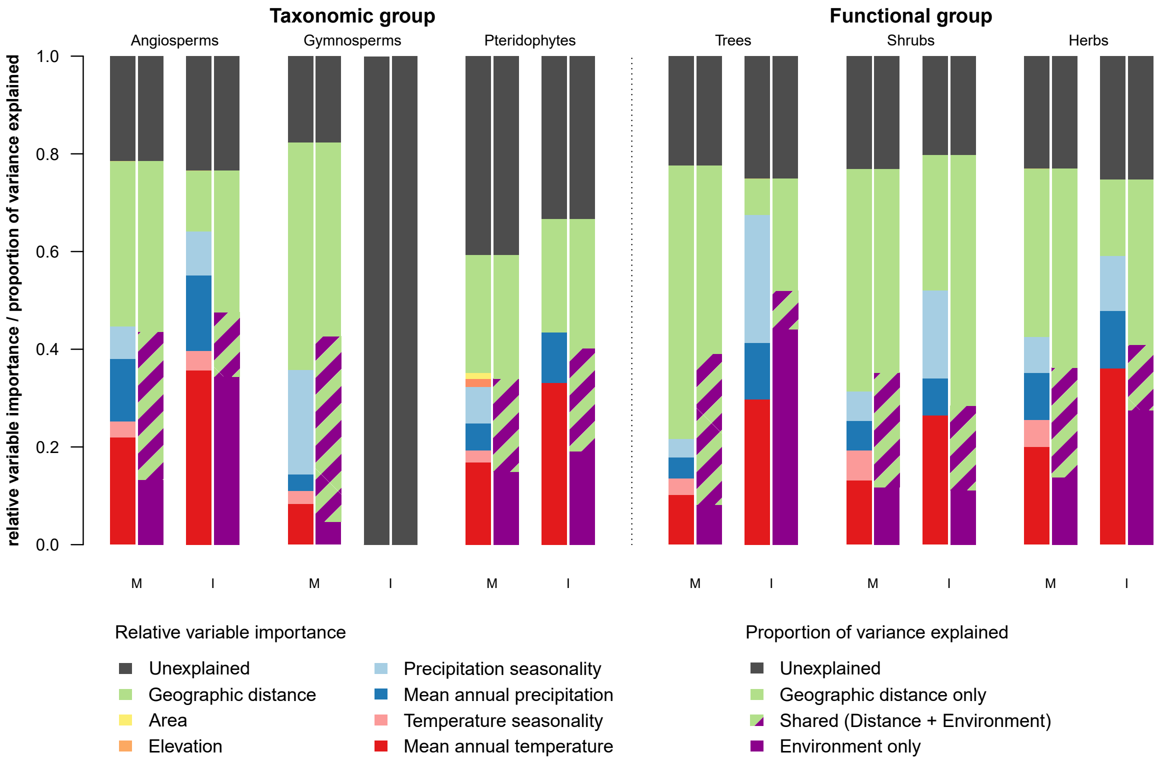
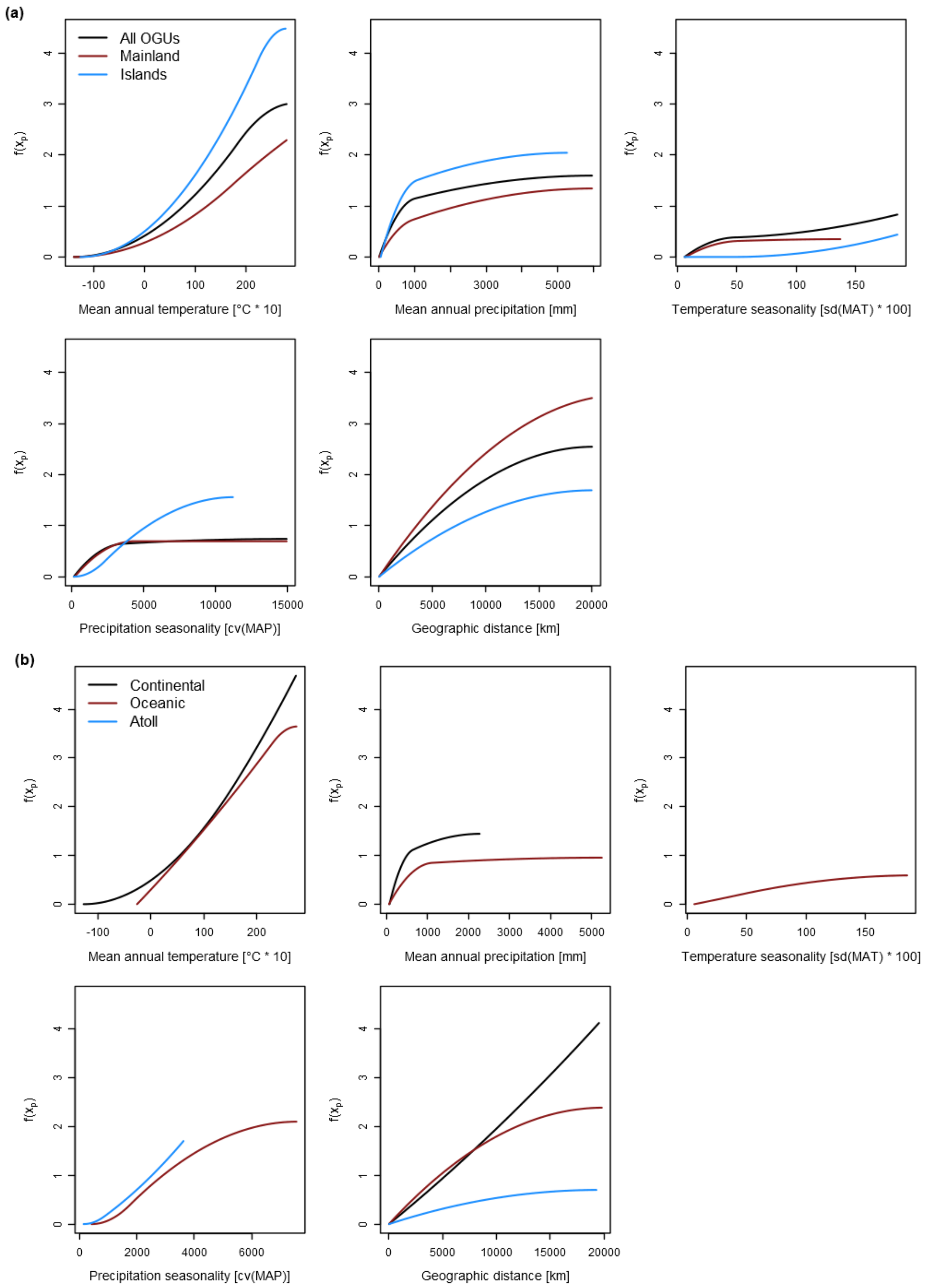


Figure A3.5: Turnover partitioning for taxonomic and functional groups (mainland vs. island comparison) using generalized dissimilarity modelling. Results are shown for different subsets of the entire dataset based on geographical setting, island type, taxonomic, and functional group. For each subset, two alternative measures of variable importance are presented where the left-hand side is based on the height of GDM transformation functions and the right-hand side is based on variation partitioning. No GDM could be fitted for gymnosperms on islands.


(a)
All OGUs
Mainland
Islands
f(x_p)
0
1
2
3
4
Mean annual temperature [°C * 10]
-100
0
100
200
Mean annual precipitation [mm]
0
1000
3000
5000
Temperature seasonality [sd(MAT) * 100]
0
50
100
150
Precipitation seasonality [cv(MAP)]
0
5000
10000
15000
Geographic distance [km]
0
5000
10000
15000
20000
(b)
Continental
Oceanic
Atoll
Mean annual temperature [°C * 10]
-100
0
100
200
Mean annual precipitation [mm]
0
1000
2000
3000
4000
5000
Temperature seasonality [sd(MAT) * 100]
0
50
100
150
Precipitation seasonality [cv(MAP)]
0
2000
4000
6000
Geographic distance [km]
0
5000
10000
15000
20000

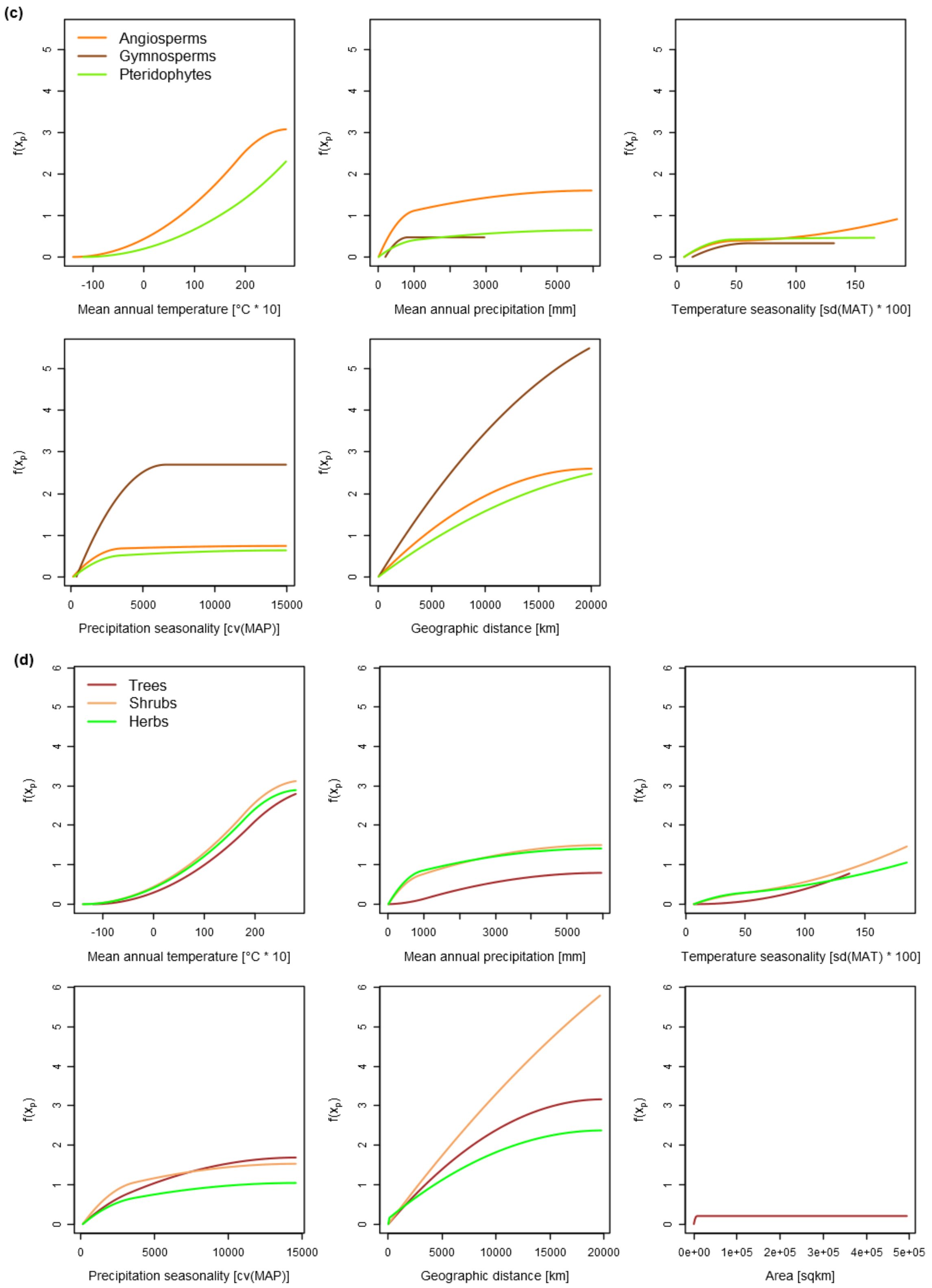


Figure A3.6: GDM transformation functions for subsets based on (a) geographical setting, (b) island type, (c) taxonomic and (d) functional group. Units of measurement were adopted unchanged from WorldClim (Hijmans *et al.*, 2005).

Table A3.2: Data references for Chapter 3.

	distribution	traits
Academica Sinica (ed.) (1998). *Proc Int Symp on Rare, Threatened, and Endangered Floras of Asia and the Pacific.* Institute of Botany, Taiwan.	1	0
Alves, R. (1998). *Ilha da Trindade e Arquipélago Martin Vaz. um ensaio geobotânico.* Serviço de Documentação da Marinha, Rio de Janeiro, Brasil.	1	0
Arechavaleta, M., Rodríguez, S., Zurita, N. & García, A. (2009). *Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres.* Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias, Santa Cruz de Tenerife, Spain.	1	0
Arechavaleta, M., Zurita, N., Marrero, M. & Martín, J. (2005). *Lista preliminar de especies silvestres de Cabo Verde (hongos, plantas y animales terrestres).* Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias, Santa Cruz de Tenerife, Spain.	1	0
Athens, J.S., Blinn, D.W. & Ward, J.V. (2007). Vegetation history of Laysan Island, Northwestern Hawaiian Islands. *Pacific Science*, **61**, 17–37.	1	0
Baker, M. & Duretto, M. (2011). *A census of the vascular plants of Tasmania.* Tasmanian Herbarium, Tasmanian Museum and Art Gallery, Hobart, Australia.	1	0
Barker, W.R., Barker, R.M., Jessop, J.P. & Vonow, H.P. (2005). Census of South Australian vascular plants. *Journal of the Adelaide Botanic Gardens Supplement*, **1**, 1–396.	1	0
Belhacene, L. (2010). Catalogue 2010 des plantes vasculaires du département de la Haute-Garonne. *Supplément à Isaatis*, **10**.	1	0
Bernal, R., Gradstein, S.R. & Celis, M. (2015). Catálogo de plantas y líquenes de Colombia, Available at: http://catalogoplantasdecolombia.unal.edu.co/ (last accessed: January 15, 2016).	1	1
BioScripts (2014). Flora Vascular, Available at: http://www.floravascular.com/ (last accessed: May 25, 2014).	1	1
Borges, P., Abreu, C., Aguiar, A., Carvalho, P., Jardim, R., Melo, I., Oliveira, P., Sérgio, C., Serrano, A. & Vieira, P. (2008). *Listagem dos fungos, flora e fauna terrestres dos arquipélagos da Madeira e Selvagens.* Direcção Regional do Ambiente da Madeira and Universidade dos Açores, Funchal and Angra do Heroísmo, Portugal.	1	0
Borges, P., Costa, A., Cunha, R., Gabriel, R., Gonçalves, V., Martins, A.F., Melo, I., Parente, M., Raposeiro, P., Rodrigues, P., Santos, R.S., Silva, L., Vieira, P. & Vieira, V. (2010). *A list of the terrestrial and marine biota from the Azores.* Princípia, Cascais.	1	0
Botanical Garden Tel Aviv (2015). *Israel Flora.* (unpublished). Tel Aviv University.	1	1
Bowdoin Scientific Station (2011). Vascular plants of Kent Island, Available at: https://www.bowdoin.edu/kent-island/species/plants.shtml (last accessed: September 14, 2011).	1	0
Brennan, K. (1996). *An annotated checklist of the vascular plants of the Alligator Rivers Region, Northern Territory, Australia,* Barton, Australia.	1	1
Brofas, G., Karetsos, G., Panitsa, M. & Theocharopoulos, M. (2001). The flora and vegetation of Gyali Island, SE Aegean, Greece. *Willdenowia*, **31**, 51–70.	1	0
Broughton, D.A. & McAdam, J.H. (2005). A checklist of the native vascular flora of the Falkland Islands (Islas Malvinas). new information on the species present, their ecology, status and distribution. *The Journal of the Torrey Botanical Society*, **132**, 115–148.	1	0
Burton, R.M. (1991). A check-list and evaluation of the flora of Nisyros (Dodecanese, Greece). *Willdenowia*, **20**, 15–38.	1	0
Butler, B.J., Barclay, J.S. & Fisher, J.P. (1999). Plant communities and flora of Robins Island (Long Island), New York. *Journal of the Torrey Botanical Society*, **126**, 63–76.	1	0
Byrd, G.V. (1984). Vascular vegetation of Buldir Island, Aleutian Islands, Alaska, compared to another Aleutian Island. *Arctic*, **37**, 37–48.	1	0
CARMABI (2009). Dutch Caribbean Biodiversity Explorer, Available at: http://www.dcbiodata.net/explorer/home (last accessed: June 24, 2011).	1	0

	distribution	traits
Cascante-Marín, A. & Estrada-Chavarría, A. (2012). Las plantas vasculares de El Rodeo, Costa Rica. *Brenesia*, **77**, 71–128.	1	1
Case, T., Cody, M. & Ezcurra, E. (2002). *A new island biogeography of the Sea of Cortés.* Oxford University Press, New York, NY.	1	0
Catarino, L., Martins, E.S., Basto, M.F. & Diniz, M.A. (2008). An annotated checklist of the vascular flora of Guinea-Bissau (West Africa). *Blumea-Biodiversity, Evolution and Biogeography of Plants*, **53**, 1–222.	1	0
Chang, C.-S., Kim, H. & Chang, K. (2014). *Provisional Checklist of the Vascular Plants for the Korea Peninsular Flora (KPF). version 1.0,* Korea.	1	0
Charters, M. (2007). Flora of Bermuda, Available at: http://www.calflora.net/floraofbermuda/.	1	0
Chernyaeva, A. (1973). Flora of Onekotan Island. *Bulletin of Main Botanical Garden*, **87**, 21–29.	1	0
Chiapella, J. & Ezcurra, C. (1999). La flora del parque provincial Tromen, provincia de Neuquén, Argentina. *Multequina*, **8**, 51–60.	1	0
Chinese Virtual Herbarium (2016). The Flora of China v. 5.0, Available at: http://www.cvh.org.cn/ (last accessed: January 15, 2016).	1	1
Christmas Island National Park (2002). *Third Christmas Island national park management plan.* Parks Australia North, Christmas Island, Australia.	1	0
Christodoulakis, D. (1996). The flora of Ikaria (Greece, E. Aegean Islands). *Phyton*, **36**, 63–91.	1	0
Clark, J.L., Neill, D.A. & Asanza, M. (2006). Floristic checklist of the Mache-Chindul mountains of Northwestern Ecuador. *Contributions from the United States National Herbarium*, 1–180.	1	1
Conti, F., Abbate, G., Alessandrini, A. & Blasi, C. (2005). *Annotated Checklist of the Italian Vascular Flora.* Palombi Editori, Roma, Italy.	1	0
Conti, F. & Bartolucci, F. (2015). *The Vascular Flora of the National Park of Abruzzo, Lazio and Molise (Central Italy).* Springer International Publishing, Cham.	1	0
Cronk, Q.C.B. (1989). The past and present vegetation of St Helena. *Journal of Biogeography*, **16**, 47–64.	1	0
D'Arcy, W. (1971). The island of Anegada and its flora. *Atoll Research Bulletin*, **139**, 1–21.	1	0
Desmet, P. & Brouillet, L. (2013). Database of Vascular Plants of Canada (VASCAN): a community contributed taxonomic checklist of all vascular plants of Canada, Saint Pierre and Miquelon, and Greenland. *Phytokeys*, **25**, 55–67	0	1
Directorate of Wrangel Island Reserve (2003). *Natural System of Wrangel Island Reserve.* The World Heritage Committee, Paris, France.	1	0
Domínguez, E., Marticorena, C., Elvebakk, A. & Pauchard, A. (2004). Catálogo de la flora vascular del Parque Nacional Pali Aike. XII Región, Chile. *Gayana Botánica*, **61**, 67–72.	1	1
Doroftei, M., Oprea, A., Ştefan, N. & Sârbu, I. (2011). Vascular wild flora of Danube Delta Biosphere Reserve. *Sci. Annals of Danube Delta Institute*, 15–52.	1	0
Du Puy, D. (1993). Christmas Island. species lists, Available at: http://www.anbg.gov.au/abrs/online-resources/flora/ (last accessed: April 06, 2011).	1	0
Egea, J. de, Peña-Chocarro, M., Espada, C. & Knapp, S. (2012). Checklist of vascular plants of the Department of Ñeembucú, Paraguay. *Phytokeys*, 15–179.	1	1
Egorova, E. (1964). Flora of Shiashkotan Island. *Bulletin of the Main Botanical Garden*, **54**, 114–120.	1	0
Esler, A. (1978). Botanical features of the Mokohinau Islands. *TANE*, **24**, 187–197.	1	0
Evarts-Bunders, P., Evarte-Bundere, G., Bāra, J. & Nitcis, M. (2013). The flora of vascular plants in nature reserve „Eglone". *Acta Biologica Universitatis Daugavpiliensis.*	1	0
Exell, A. (1944). *Catalogue of the vascular plants of S. Tome (with Principe and Annobon).* Trustees of the British Museum, London, UK.	1	0

	distribution	traits
Fern, K. & Fern, A. (2015). Useful Tropical Plants Database (last accessed: July 25, 2015)	0	1
Figueroa-C., Y. & Galeano, G. (2007). Lista comentada de las plantas vasculares del enclave seco interandino de La Tatacoa (Huila, Colombia). *Caldasia*, **29**, 263–281.	1	1
Fischer, E., Rembold, K., Althof, A., Obholzer, J., Malombe, I., Mwachala, G., Onyango, J.C., Dumbo, B. & Theisen, I. (2010). Annotated checklist of the vascular plants of Kakamega Forest, Western Province, Kenya. *Journal of East African Natural History*, **99**, 129–226.	1	1
Florence, J., Chevillotte, H., Ollier, C. & Meyer, J.-Y. (2007). Base de données botaniques Nadeaud de l'Herbier de la Polynésie française (PAP), Available at: http://www.herbier-tahiti.pf/ (last accessed: July 01, 2011).	1	0
Fosberg, F., Renvoize, S. & Townsend, C. (1980). *The flora of Aldabra and neighbouring islands.* HMSO, London, UK.	1	0
Fosberg, F. & Sachet, M. (1987). Flora of Maupiti, Society Islands. *Atoll Research Bulletin*, **294**, 1–70.	1	0
Fosberg, F., Stoddart, D., Sachet, M.-H. & Spellman, D. (1982). Plants of the Belize cays. *Atoll Research Bulletin*, **258**, 1–77.	1	0
Franklin, J., Keppel, G. & Whistler, W. (2008). The vegetation and flora of Lakeba, Nayau and Aiwa islands, central Lau Group, Fiji. *Micronesica*, **40**, 169–225.	1	0
Funk, V.A., Hollowell, T., Berry, P., Kelloff, C. & Alexander, S.N. (2007). *Checklist of the plants of the Guiana Shield (Venezuela: Amazonas, Bolivar, Delta Amacuro; Guyana, Surinam, French Guiana).* Department of Botany, National Museum of Natural History, Washington, DC.	1	0
Gabrielsen, G., Brekke, B., Alsos, I. & Hansen, J. (1997). *Natur-og kulturmiljøet på Jan Mayen.* Norsk Polarinstitutt, Oslo, Norway.	1	0
Gage, S., Joneson, S., Barkalov, V., Eremenko, N. & Takahashi, H. (2006). A newly compiled checklist of the vascular plants of the Habomais, the Little Kurils. *Bulletin of the Hokkaido University Museum*, **3**, 67–91.	1	0
Gerlach, J. (2003). The biodiversity of the granitic islands of Seychelles. *Phelsuma*, **11** (Supplement A), 1–47.	1	0
Green, P. (1994). Lord Howe Island. species lists, Available at: http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/49/index.html (last accessed: April 06, 2011).	1	0
Greene, S.W. & Walton, D.W. (1975). An annotated check list of the sub-Antarctic and Antarctic vascular flora. *Polar Record*, **17**, 473–484.	1	0
Harris, D.J. (2002). *The vascular plants of the Dzanga-Sangha Reserve, Central African Republic.* Royal Botanic Garden Edinburgh, Edinburgh, Scotland, UK.	1	1
Hauenstein, E., Muñoz-Pedreros, A., Yánez, J., Sánchez, P., Möller, P., Guiñez, B. & Gil, C. (2009). Flora y vegetación de la Reserva Nacional Lago Peñuelas, Reserva de la Biósfera, Región de Valparaíso, Chile. *Bosque (Valdivia)*, **30**.	1	0
Hill, M. (ed.) (2002). *Biodiversity surveys and conservation potential of inner Seychelles islands.* Smithsonian Institution, Washington, DC.	1	0
Hill, S.R. (1986). An annotated checklist of the vascular flora of Assateague Island (Maryland and Virginia). *Castanea*, **51**, 265–305.	1	0
Hnatiuk, R. (1993). Subantarctic Islands. species lists, Available at: http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/50/index.html (last accessed: April 07, 2011).	1	0
Imada, C.T. (ed.) (2012). *Hawaiian Native and Naturalized Vascular Plants Checklist*, (December 2012 update). Hawaii Biological Survey, Bishop Museum, Honolulu, Hawai'i.	1	0
Instituto de Botánica Darwinion (2014). Catálogo de las Plantas Vasculares del Cono Sur, Available at: http://www.darwin.edu.ar/Proyectos/FloraArgentina/fa.htm (last accessed: March 16, 2015).	1	1
Jackes, B. (2010). *Plants of Magnetic Island*, 3rd edn. James Cook University, Townsville, Australia.	1	0
Jahn, R. & Schönfelder, P. (1995). *Exkursionsflora für Kreta.* Ulmer (Eugen), Stuttgart, Germany.	1	0

	distribution	traits
Jaramillo Díaz, P. & Guézou, A. (2011). CDF checklist of Galapagos vascular plants. Bungartz, F., Herrera, H., Jaramillo, P., Tirado, N., Jímenez-Uzcategui, G., Ruiz, D., Guézou, A. & Ziemmeck, F. (eds.), Available at: http://www.darwinfoundation.org/datazone/checklists/vascular-plants/ (last accessed: February 02, 2011).	1	0
Jardim Botânico do Rio de Janeiro. Flora do Brasil 2020 em construção, Available at: http://floradobrasil.jbrj.gov.br/ (last accessed: May 09, 2016).	1	1
Johnson, P.N. & Campbell, D.J. (1975). Vascular plants of the Auckland Islands. *New Zealand journal of botany*, **13**, 665–720.	1	0
Johnston, I. (1931). The flora of the Revillagigedo Islands. *Proceedings of the California Academy of Sciences*, **20**, 9–104.	1	0
Jordano, P. (2008). FRUBASE, Available at: http://ebd10.ebd.csic.es/mywork/frubase/frubase.html (last accessed: 26.06.15)	0	1
Junak, S., Philbrick, R., Chaney, S. & Clark, R. (1997). *A checklist of vascular plants of Channel Islands National Park*, 2nd edn. Southwest Parks and Monuments Association, Tucson, Arizona.	1	0
Kamari, G., Phitos, D., Snogerup, B. & Snogerup, S. (1988). Flora and vegetation of Yioura, N Sporades, Greece. *Willdenowia*, **17**, 59–85.	1	0
Kattge, J., Diaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P.B. & Wright, I.J. (2011). TRY – a global database of plant traits. *Global Change Biology*, **17**, 2905–2935	0	1
Kelloff, C.L. & Funk, V.A. (1998). Preliminary checklist of the plants of Kaieteur National Park, Guyana. *Biological Diversity of the Guianas*.	1	0
Kerguelen, M. (2005). Base de Données Nomenclaturales de la Flore de France, Available at: http://www.tela-botanica.org/page:284# (last accessed: January 13, 2012).	1	0
Kirchner, F., Picot, F., Merceron, E. & Gigot, G. (2010). *Flore vasculaire de La Réunion*. Conservatoire Botanique National de Mascarin, Réunion, France.	1	0
Kleyer, M., Bekker, R.M., Knevel, I.C., et al. (2008). The LEDA Traitbase: a database of life-history traits of the Northwest European flora. *Journal of Ecology*, **96**, 1266–1274	0	1
Koltzenburg, M. (2011). Checkliste der Gefässpflanzen Irlands, Available at: http://www.saxifraga.de/eire/irl_artenlisten_gp.html (last accessed: April 04, 2011).	1	0
Kraaij, T. (2011). The flora of the Bontebok National Park in regional perspective. *South African Journal of Botany*, **77**, 455–473.	1	1
Kristinsson, H. (2008). *Checklist of the vascular plants of Iceland*. Náttúrufræðistofnun Íslands, Reykjavík, Iceland.	1	0
Lange, P.J. de & Cameron, E.K. (1999). The vascular flora of Aorangi Island, Poor Knights Islands, northern New Zealand. *New Zealand journal of botany*, **37**, 433–468.	1	0
Lazkov, G.A. & Sultanova, B.A. (2011). *Checklist of vascular plants of Kyrgyzstan*. Botanical Museum, Finnish Museum of Natural History, Helsinki.	1	0
Le Houerou, H.N. (2004). Plant diversity in Marmarica (Libya & Egypt): a catalogue of the vascular plants reported with their biology, distribution, frequency, usage, economic potential, habitat and main ecological features, with an extensive bibliography. *Candollea*, **59**, 259–308.	1	1
Lee, R.-Y., Jang, H.-D., Kim, Y.-Y., Yang, S.-G., Choi, H.-J., Ji, S.-J. & Oh, B.-U. (2014). Flora of vascular1 plants in the Chilgapsan Provincial Park, Korea. *Journal of Asia-Pacific Biodiversity*, **7**, 237–247.	1	0
Lester-Garland, L. (1903). *A flora of the islands of Jersey. with a list of the plants of the Channel Islands in general, and remarks upon their distribution and geographical affinities*. West, Newman & Co, London, UK.	1	0
Levin, G. & Moran, R. (1989). The vascular flora of Socorro, Mexico. *Memoirs of the San Diego Society of Natural History*, **16**, 1–71.	1	0
Lipkin, R. (2005). *Aniakchak National Monument and Preserve, vascular plant inventory: final technical report*. National Park Service, Southwest Alaska Network Inventory & Monitroing Program, Anchorage, USA.	1	0

	distribution	traits
Luke, Q. (2005). Annotated Checklist of the Plants of the Shimba Hills, Kwale District, Kenya. *Journal of East African Natural History*, **94**, 5–120.	1	1
Marquand, E. (1901). *Flora of Guernsey and the lesser Channel Islands. namely Alderney, Sark, Herm, Jethou, and the adjacent islets.* Dulau & Co, London, UK.	1	0
Marticorena, C., Squeo, F.A., Arancio, G. & Muñoz, M. (2008). Catálogo de la flora vascular de la IV Región de Coquimbo. *Libro rojo de la flora nativa y de los sitios prioritarios para su conservación: Región de Atacama* (ed. by F.A. Squeo, G. Arancio and J.R. Gutiérrez), pp. 105–142. Ediciones Universidad de La Serena La Serena.	1	1
Marticorena, C., Stuessy, T.F. & Baeza, C.M. (1998). Catalogue of the vascular flora of the Robinson Crusoe or Juan Fernández islands, Chile. *Gayana Botánica*, **55**, 187–211.	1	1
Masharabu, T. *Flore et végétation du Parc National de la Ruvubu au Burundi: diversité, structure et implications pour la conservation.* Editions universitaires europeennes.	1	1
McClatchey, W., Thaman, R. & Vodonaivalu, S. (2000). A preliminary checklist of the flora of Rotuma with Rotuman names. *Pacific Science*, **54**, 345–363.	1	0
McCrea, J. (2003). *Inventory of the land conservation values of the Houtman Abrolhos Islands.* Department of Fisheries, Government of Western Australia, Perth, Australia.	1	0
Medina, R., Reina-E, M., Herrera, E., Ávila, F.A., Chaparro, O. & Cortés-B., R. (2010). Catálogo preliminar da flora vascular dos bosques subandinos da cuchilla El Fara (Santander-Colômbia). *Colombia Forestal*, **13**, 55–85.	1	1
Medjahdi, B., Ibn Tattou, M., Barkat, D. & Benabedli, K. (2009). La flore vasculaire des monts des Trara (nord ouest algérien). *Acta Botanica Malacitana*, **34**, 57–75.	1	0
Miller, A. & Morris, M. (2004). *Ethnoflora of the Soqotra Archipelago.* Royal Botanic Garden, Edinburgh, UK.	1	0
Moran, R. (1996). *The flora of Guadalupe Island, Mexico.* California Academy of Sciences, San Francisco, CA.	1	0
Nationalpark Eifel (2015). Artenliste Farne und Blütenpflanzen, Available at: http://www.nationalpark-eifel.de/go/artenliste.html (last accessed: March 11, 2015).	1	0
New Zealand Plant Conservation Network (2011). New Zealand's Flora, Available at: http://nzpcn.org.nz/page.asp?flora (last accessed: March 18, 2011).	1	0
Norton, J., Majid, S.A., Allan, D., Al Safran, M., Böer, B. & Richer, R.A. (2009). *An illustrated checklist of the flora of Qatar.* Browndown Publications Gosport, Gosport, UK.	1	1
Notov, A.A. (2010). *National park "Zavidovo": vascular plants, bryophyte, lichens,* Moscow.	1	0
NPS (2015). NPSpecies. Information on Species in National Parks, Available at: https://irma.nps.gov/NPSpecies/ (last accessed: April 09, 2015).	1	0
Oggero, A.J. & Arana, M.D. (2012). Inventario de las plantas vasculares del sur de la zona serrana de Córdoba, Argentina. *Hoehnea*, **39**, 171–199.	1	0
Pal, D., Kumar, A. & Dutt, B. (2014). Floristic diversity of Theog Forest Division, Himachal Pradesh, Western Himalaya. *Check list*, **10**, 1083–1103.	1	1
Pandža, M. (2010). Flora parka prirode Papuk (Slavonija, Hrvatska). *Šumarski list*, **134**, 25–43.	1	0
Parks Canada (2015). Biotics Web Explorer, Available at: http://www.pc.gc.ca/apps/bos/BOSIntro_e.asp (last accessed: April 13, 2015).	1	0
Peña-Chocarro, M.d.C. (2010). *Updated checklist of vascular plants of the Mbaracayú Forest Nature Reserve (Reserva Natural del Bosque Mbaracayú), Paraguay.* Magnolia Press, Auckland, N.Z.	1	1
Pôle Flore Habitats (2015). Catalogue de la flore vasculaire de Rhône-Alpes, Available at: http://www.pifh.fr/pifhcms/index.php (last accessed: January 12, 2015).	1	0
Proctor, G. (1980). Checklist of the plants of Little Cayman. Geography and ecology of Little Cayman. *Atoll Research Bulletin*, **241**, 71–80.	1	0
Queensland Government (2014). Census of the Queensland flora 2014, Available at: https://data.qld.gov.au/dataset/census-of-the-queensland-flora-2014 (last accessed: February 05, 2015).	1	0

	distribution	traits
Rakov, N.S., Saksonov S.V., Senator S.A. & Vasjukov V.M. (2014). *Vascular plants of Ulyanovsk Region*. Russian Academy of Sciences, Togliatti.	1	0
Raulerson, L. (2006). Checklist of plants of the Mariana Islands. *University of Guam Herbarium Contribution*, **37**, 1–69.	1	0
Real Jardín Botánico de Madrid (2016). Flora de Guinea Equatorial, Available at: http://www.floradeguinea.com/ (last accessed: February 14, 2016).	1	0
Renvoize, S.A. (1975). A floristic analysis of the western Indian Ocean coral islands. *Kew Bulletin*, **30**, 133–152.	1	0
Robinson, A.C., Canty, P.D. & Fotheringham, D. (2008). Investigator group expedition 2006. flora and vegetation. *Transactions of the Royal Society of South Australia*, **132**, 173–220.	1	0
Robinson, A.C., Canty, P.D., Wace, N.M. & Barker, R.M. (2003). The encounter 2002 expedition to the isles of St Francis, South Australia. flora and vegetation. *Transactions of the Royal Society of South Australia*, **127**, 107–128.	1	0
Rossetto, Elson Felipe Sandoli & Vieira, Ana Odete Santos (2013). Vascular Flora of the Mata dos Godoy State Park, Londrina, Paraná, Brazil. *Check list*, **9**, 1020–1034.	1	1
Rundel, P.W., Dillon, M.O. & Palma, B. (1996). Flora and Vegetation of Pan de Azúcur National Park in the Atacama desert of Northern Chile. *Gayana Bot*, **53**, 295–315.	1	1
SANBI (2014). Plants of Southern Africa. An online checklist, Available at: http://posa.sanbi.org (last accessed: March 09, 2015).	1	1
Sandbakk, B.E., Alsos, I.G., Arnesen, G. & Elven, R. (1996). The flora of Svalbard, Available at: http://svalbardflora.no/ (last accessed: March 16, 2011).	1	0
Searle, J. & Madden, S. (2006). *Flora assessment of South Stradbroke Island*. Gold Coast City Council, Gold Coast City, Australia.	1	0
Selvi, F. (2010). A critical checklist of the vascular flora of Tuscan Maremma (Grosseto province, Italy). *Fl. Medit*, **20**, 47–139.	1	1
Shaw, J.D., Spear, D., Greve, M. & Chown, S.L. (2010). Taxonomic homogenization and differentiation across Southern Ocean Islands differ among insects and vascular plants. *Journal of Biogeography*, **37**, 217–228.	1	0
Slik, F.J.W., Arroyo-Rodríguez, V., Aiba, S.-I., 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	0	1
SLUFG (2015). *Rote Liste und Artenliste Sachsen. Farn- und Samenpflanzen*. Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie, Dresden.	1	0
Stace, C., Ellis, R., Kent, D. & McCosh, D. (2003). *Vice-county Census Catalogue of the vascular plants of Great Britain, the Isle of Man and the Channel Islands*. Botanical Society of the British Isles, London, UK.	1	0
Stalmans, M. (2006). *Tinley's plant species list for the Greater Gorongosa ecosystem, Moçambique*. Unpublished report by International Conservation Services to the Carr Foundation and the Ministry of Tourism.	1	1
Stalter, R. & Lamont, E.E. (2006). The historical and extant flora of Sable Island, Nova Scotia, Canada. *Journal of the Torrey Botanical Society*, **133**, 362–374.	1	0
Sykes, W. (1970). Contributions to the flora of Niue. *Bulletin. Department of Scientific and Industrial Research, New Zealand*, **200**, 321.	1	0
Sykes, W., West, C.J., Beever, J.E. & Fife, A.J. (2000). *Kermadec Islands flora-special edition. a compilation of modern material about the flora of the Kermadec Islands*. Manaaki Whenua Press, Landcare Research.	1	0
Takahashi, H., Barkalov, V., Gage, S., Joneson, S., Ilushko, M. & Zhuravlev, Y. (2002). A floristic study of the vascular plants of Raikoke, Kuril Islands. *Acta Phytotaxonomica et Geobotanica*, **53**, 17–33.	1	0
Takahashi, H., Barkalov, V., Gage, S., Semsrott, B., Ilushko, M. & Zhuravlev, Y. (1999). A preliminary checklist of the vascular plants of Chirinkotan, Kuril Islands. *Journal of Phytogeography and Taxonomy*, **47**, 131–137.	1	0
Takahashi, H., Barkalov, V., Gage, S., Semsrott, B., Ilushko, M. & Zhuravlev, Y. (2006). A floristic study of the vascular plants of Kharimkotan, Kuril Islands. *Bulletin of the Hokkaido University Museum*, **3**, 41–66.	1	0

	distribution	traits
Takahashi, H., Barkalov, V., Gage, S. & Zhuravlev, Y. (1997). A preliminary study of the flora of Chirpoi, Kuril Islands. *Acta Phytotaxonomica et Geobotanica*, **48**, 31–42.	1	0
Telford, I. (1993). Cocos (Keeling) Islands. Species lists, Available at: http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/50/ (last accessed: April 07, 2011).	1	0
Thaman, R., Fosberg, F., Manner, H. & Hassall, D. (1994). The flora of Nauru. *Atoll Research Bulletin*, **392**, 1–233.	1	0
The State of Victoria, Department of Environment and Primary Industries (2015). Victorian Biodiversity Atlas (last accessed: 17.02.15).	1	1
Tropicos (2015a). Catálogo de las Plantas Vasculares de Bolivia. *Tropicos*. Missouri Botanical Garden, St. Louis, Available at: http://www.tropicos.org/Project/BC.	1	1
Tropicos (2015b). Catalogue of the Vascular Plants of Ecuador. *Tropicos*. Missouri Botanical Garden, St. Louis, Available at: http://www.tropicos.org/Project/CE.	1	1
Tropicos (2015c). Flora de Nicaragua. *Tropicos* . Missouri Botanical Garden, St. Louis, Available at: http://www.tropicos.org/Project/FN.	1	0
UIB (2007). Herbario virtual del Mediterráneo Occidental, Available at: http://herbarivirtual.uib.es/cas-med/ (last accessed: August 07, 2012).	1	0
University of Kent (2012). Cook Islands Biodiversity and Ethnobiology Database, Available at: http://cookislands.pacificbiodiversity.net/cibed/dbs/search.html (last accessed: April 12, 2012).	1	0
van Vreeswyk, A., Payne, A., Leighton, K.A. & Hennig, P. (2004). *An inventory and condition survey of the Pilbara region, Western Australia.* Department of Agriculture.	1	1
USDA & NRCS (2015). The PLANTS Database. National Plant Data Team (ed.), Available at: http://plants.usda.gov (last accessed: April 24, 2015)	0	1
Vander Velde, N. (2003). The vascular plants of Majuro atoll, Republic of the Marshall Islands. *Atoll Research Bulletin*, **503**.	1	0
Vanderplank, S. (2010). *The Vascular Flora of Greater San Quintín, Baja California, Mexico.*	1	1
Viciani, D., Gonnelli, V., Sirotti, M. & Agostini, N. (2010). An annotated check-list of the vascular flora of the "Parco Nazionale delle Foreste Casentinesi, Monte Falterona e Campigna"(Northern Apennines Central Italy). *Webbia*, **65**, 3–131.	1	1
Vogt, C. (2011). Composición de la Flora Vascular del Chaco Boreal, Paraguay. I. Pteridophyta y Monocotiledoneae. *Steviana*, **3**, 13–47.	1	1
Wace, N.M. (1961). The vegetation of Gough Island. *Ecological Monographs*, **31**, 337–367.	1	0
Wace, N.M. & Dickson, J.H. (1965). The terrestrial botany of the Tristan da Cunha Islands. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, **249**, 273–360.	1	0
Wagner, W., Herbst, D. & Lorence, D. (2005). Flora of the Hawaiian Islands website, Available at: http://botany.si.edu/pacificislandbiodiversity/hawaiianflora/ (last accessed: October 16, 2010).	1	1
WCSP (2014). World Checklist of Selected Plant Families, Available at: http://apps.kew.org/wcsp/home.do (last accessed: December 01, 2014)	0	1
Wellington Botanical Society (2008). *Native vascular plants of Great Barrier Island.* Wellington Botanical Society, Wellington, New Zealand.	1	0
Whistler, W. (1983). Vegetation and flora of the Aleipata Islands, Western Samoa. *Pacific Science*, **37**, 227–249.	1	0
Whistler, W. (1996). Botanical survey of Diego Garcia, Chagos Archipelago, British Indian Ocean Territory. Appendix E1. *Isle Botanica*, 1–72.	1	0
Whistler, W. (1998). *A study of the rare plants of American Samoa.* US Fish and Wildlife Service, Honolulu, Hawaii.	1	0
ZDSF & SKEW (2014). Info Flora. Artenliste Schweiz 5x5 km, Available at: https://www.infoflora.ch/de/daten-beziehen/artenliste-5x5-km.html (last accessed: February 13, 2015).	1	0

	distribution	traits
Zvyagintseva, K.O. (2015). *An annotated checklist of the urban flora of Kharkiv,* Kharkiv National University.	1	0
Евстигнеев, О.И. & Федотов, Ю.П. (2007). *Флора сосудистых растений заповедника "Брянский лес"*. [Гос. природ. биосфер. заповедник" Брян. лес"] Брянск.	1	0
Хапугин, А.А. (2013). *Сосудистые растения Ромодановского района Республики Мордовия (конспект флоры),* Saransk.	1	0

A4 Supplementary information to Chapter 4

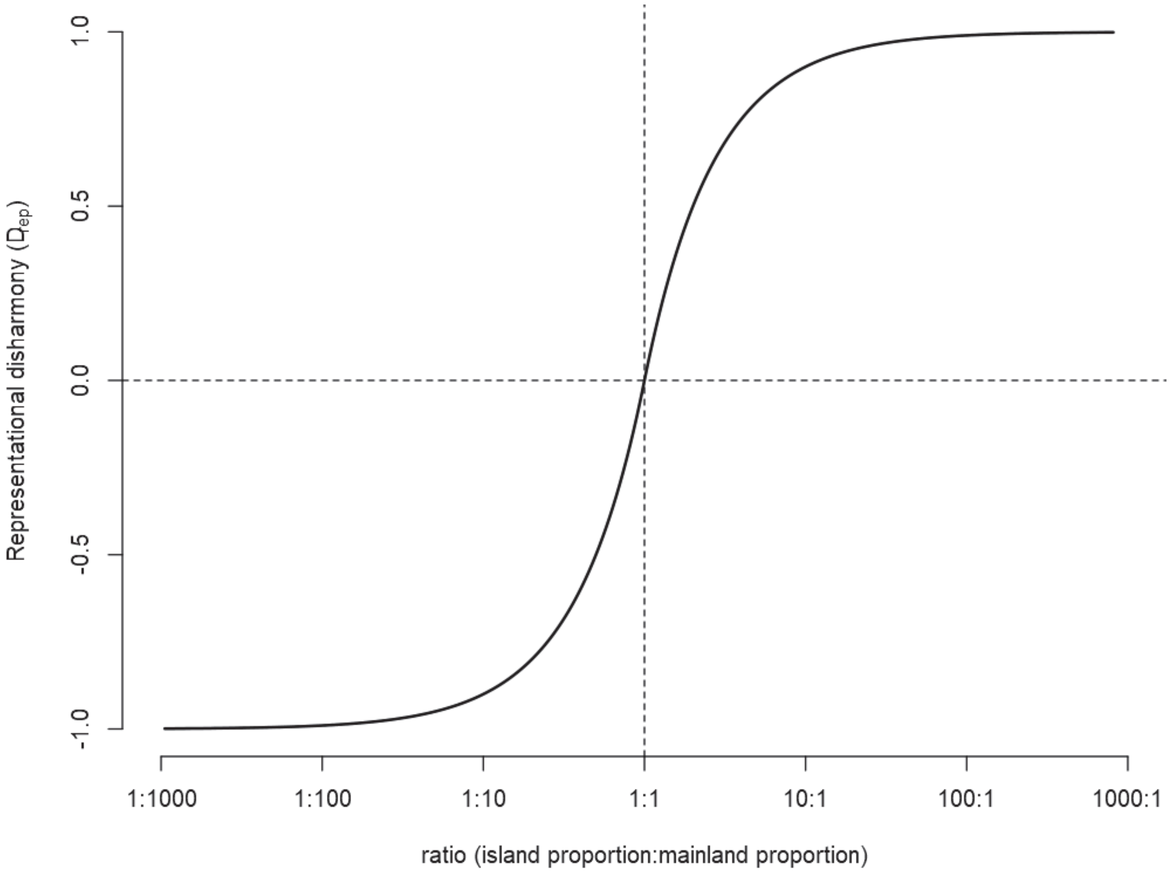


Figure A4.1: Representational disharmony (D_{rep}) as a function of the ratio between mean proportional representation in island- and mainland floras.

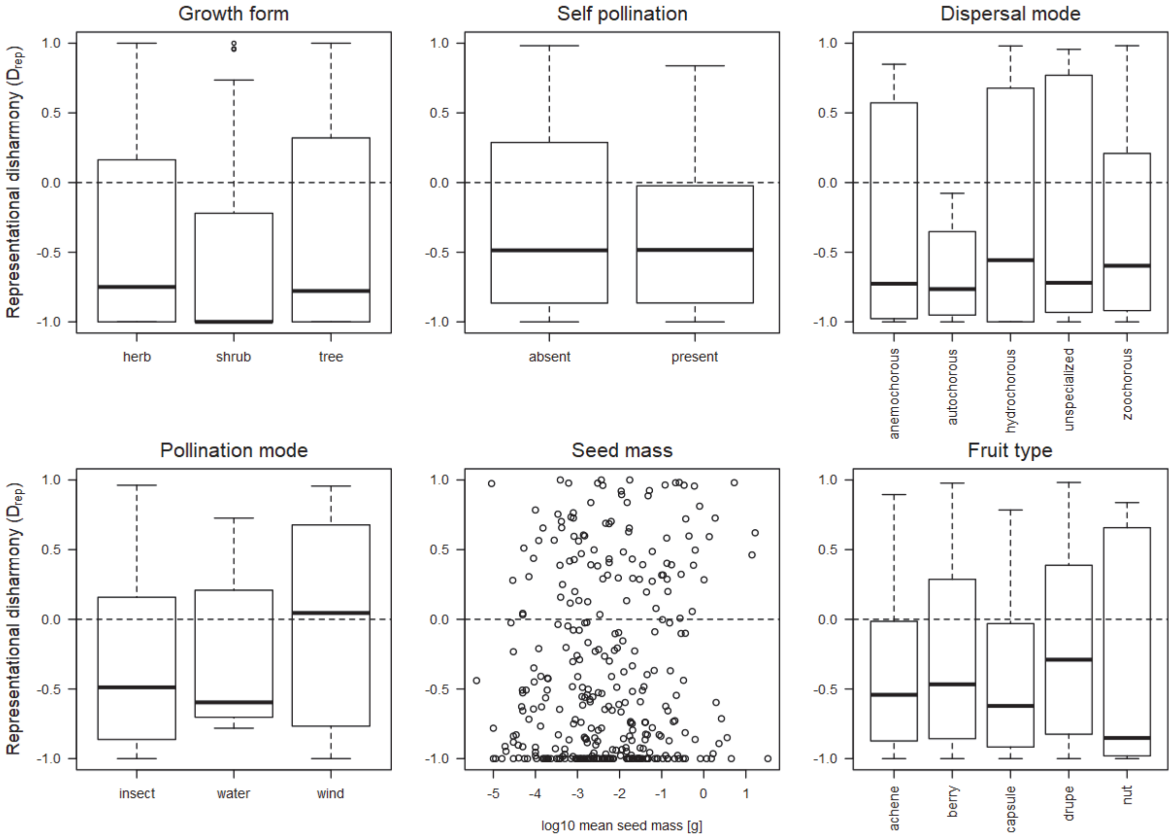


Figure A4.2: Relationship of representational disharmony (D_{rep}) and family-level functional traits. Family-level traits were aggregated from species-level information of varying availability (growth form: 213,317 spp., self-pollination: 3,538 spp., dispersal mode: 8,208 spp., pollination mode: 4,511 spp., seed mass: 23,874 spp., fruit type: 5,967 spp.) using the median (seed mass) or majority of values (all other traits) per family.

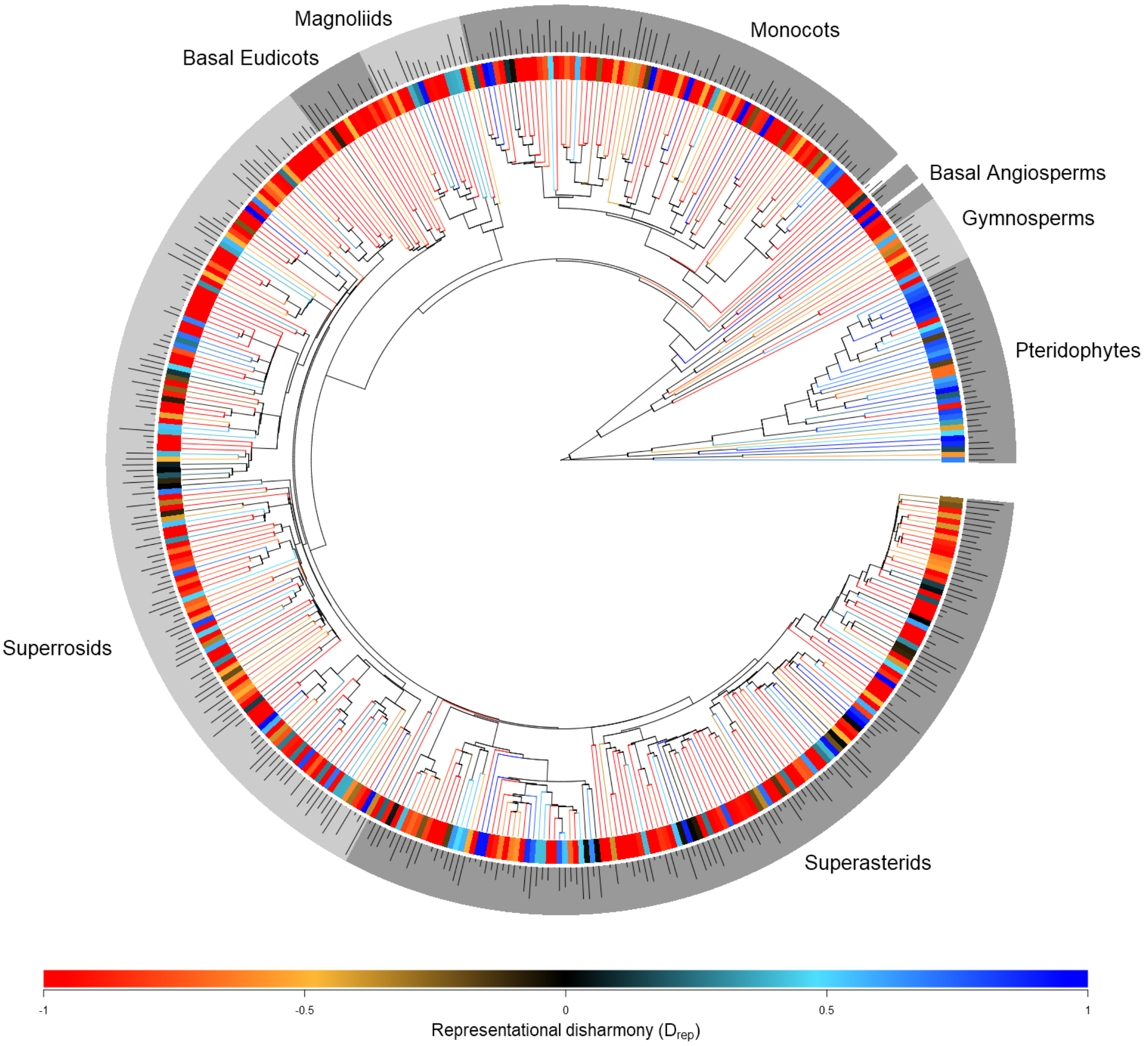


Figure A4.3: Representational disharmony (D_{rep}) of 450 vascular plant families. Negative values indicate a proportional under-representation of the respective family in island floras compared to the mainland, whereas positive values indicate an over-representation. $D_{rep} = -1$, $D_{rep} = 0$ and $D_{rep} = 1$ respectively denote a restriction to the mainland, even representation in island and mainland floras, and a restriction to islands.

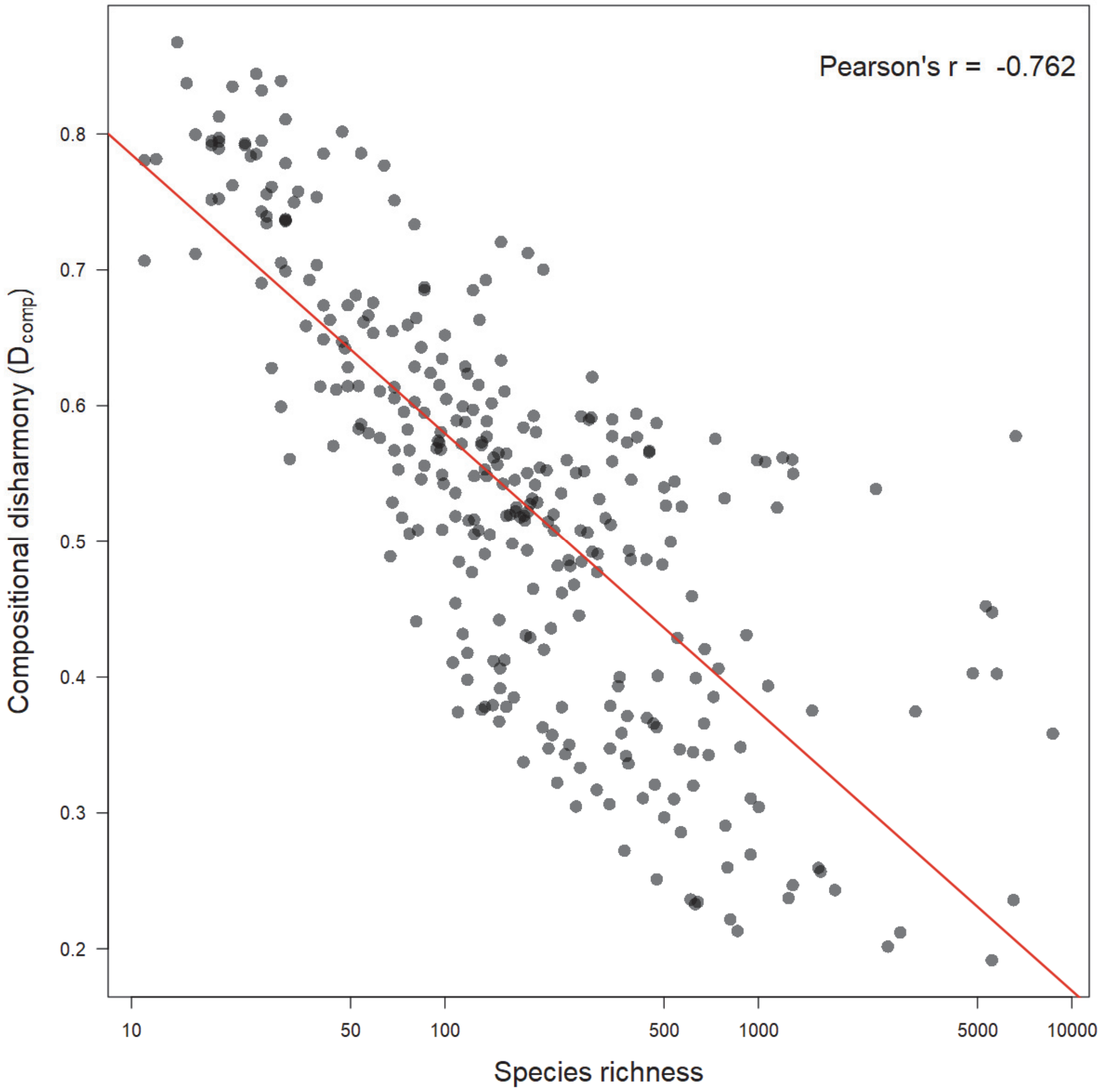


Figure A4.4: Correlation between compositional disharmony (D_{comp}) and $\log_{10}$(species richness).

Supplementary text A4.1: Truncation of mainland weights vector W

The presented measure of compositional disharmony is calculated as the mean Bray-Curtis dissimilarity of family proportions of a given island i relative to all mainland floras (B_i), weighted by the mainland floras' estimated importance as source regions for the respective island (W_i):

$$D_{comp}(i) = \frac{\sum B_i * W_i}{\sum W_i}$$

We obtained W_i from predictions of species turnover between island i and all mainland units. The model we used to predict species turnover (Generalized dissimilarity modelling, Ferrier, 2002; Ferrier *et al.*, 2007) converges against, but never actually reaches 1 (complete species turnover) at large geographical and/or environmental distances. This is generally a realistic and useful property, because even remote or environmentally dissimilar regions might share a small fraction of their species. However, in our particular use case, the combined weight of many mainland regions with relatively low values of W_i could still introduce a substantial bias into D_{comp}. This problem was especially pronounced for islands having only very few important mainland source regions, e.g. due to the lack of floristic data for nearby continental regions in our dataset.

To avoid a dependency of D_{comp} on the amount and geographical distribution of available mainland floras, we set $W_i = 0$ for mainland units falling below a certain island-specific value. To determine this threshold, we performed the following steps:

1. Scale the source pool weights vector W_i for a given island i to a range of 0 to 1.
2. Sort W_i in decreasing order, i.e. rank all mainland units according to their importance as source region for island i (Figure A4.5, black line).
3. Fit a curve to the ordered values in W_i to obtain a smooth and continuous approximation of the weights distribution (Figure A4.5, blue line). We used the `smooth.spline`-function in R with a smoothing parameter of `spar = 0.75`.
4. Take the derivative W_i' to quantify the rate of change in W_i.
5. Define the threshold for island i as that value in W_i, after which W_i' permanently falls below 0.001, i.e. no more substantial changes in the importance of mainland source regions are predicted (Figure A4.5, red lines). The value of 0.001 was chosen after carefully inspecting curves of W_i and W_i' for a representative set of island floras.

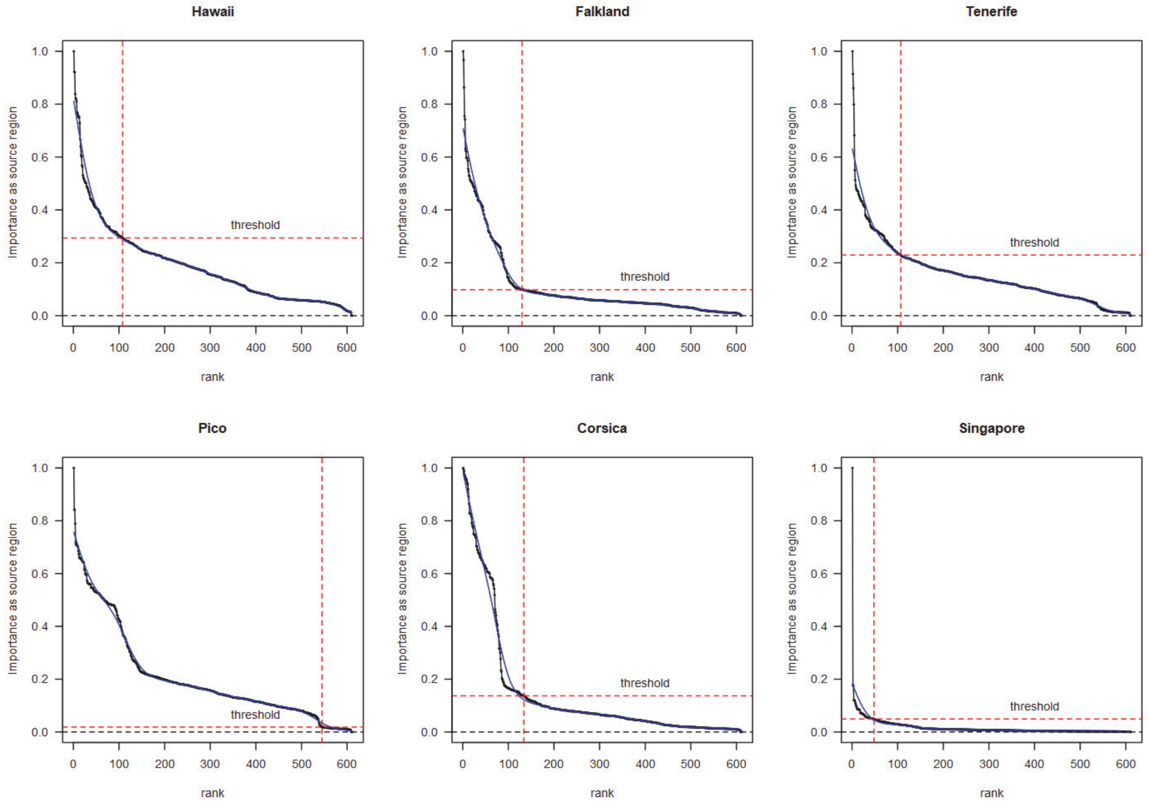


Figure A4.5: Determination of threshold value for weights vector W, exemplarily for six islands.

Table A4.1: Data references for Chapter 4.

	distribution	traits
3D Environmental (2013). *Profile for management of the habitats and related ecological and cultural resource values of Boigu Island*	0	1
Acevedo-Rodríguez, P. & Strong, M.T. (2007). Catalogue of the seed plants of the West Indies Website, Available at: http://botany.si.edu/antilles/WestIndies/catalog.htm (last accessed: March 01, 2011)	1	0
Alves, R.J.V. (1998). *Ilha da Trindade & Arquipélago Martin Vaz. Um ensaio geobotânico.* Serviço de Documentação da Marinha, Rio de Janeiro, Brasil	1	0
Arakaki, M. & Cano, A. (2003). Composición florística de la cuenca del río Ilo-Moquegua y Lomas de Ilo, Moquegua, Peru. *Revista Peruana de Biología*, **10**, 5–19	0	1
Arechavaleta, M., Rodríguez, S., Zurita, N. & García, A. (2009). *Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres.* Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias, Santa Cruz de Tenerife, Spain	1	0
Arechavaleta, M., Zurita, N., Marrero, M.C. & Martín, J.L. (2005). *Lista preliminar de especies silvestres de Cabo Verde (hongos, plantas y animales terrestres).* Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias, Santa Cruz de Tenerife, Spain	1	0
Ashmole, P. & Ashmole, M. (2000). *St Helena and Ascension Island. A natural history.* Anthony Nelson Ltd, Oswestry, Shropshire, UK	1	0
Athens, J.S., Blinn, D.W. & Ward, J.V. (2007). Vegetation history of Laysan Island, Northwestern Hawaiian Islands. *Pacific Science*, **61**, 17–37	1	0
Baker, M.L. & Duretto, M.F. (2011). *A census of the vascular plants of Tasmania.* Tasmanian Herbarium, Tasmanian Museum and Art Gallery, Hobart, Australia	1	0
Barker, W.R., Barker, R.M., Jessop, J.P. & Vonow, H.P. (2005). Census of South Australian vascular plants. *Journal of the Adelaide Botanic Gardens Supplement*, **1**, 1–396	1	0
Belhacene, L. (2010). Catalogue 2010 des plantes vasculaires du département de la Haute-Garonne. *Supplément à Isaatis*, **10**, 1–145	1	0
Benito, B.M., Lorite, J., Pérez-Pérez, R., Gómez-Aparicio, L., Peñas, J. & Robertson, M. (2014). Forecasting plant range collapse in a mediterranean hotspot. When dispersal uncertainties matter. *Diversity and Distributions*, **20**, 72–83	0	1
Bernal, R., Gradstein, S.R. & Celis, M. (2015). Catálogo de plantas y líquenes de Colombia, Available at: http://catalogoplantasdecolombia.unal.edu.co/ (last accessed: January 15, 2016)	1	1
BGCI (2017). GlobalTreeSearch online database, Available at: www.bgci.org/globaltree_search.php (last accessed: August 14, 2017)	1	1
Bingham, M.G., Willemen, A., Wursten, B.T., Ballings, P. & Hyde, M.A. (2016). Flora of Zambia, Available at: http://www.zambiaflora.com/ (last accessed: November 14, 2016)	1	0
BioScripts (2014). Flora Vascular, Available at: http://www.floravascular.com/ (last accessed: May 25, 2014)	1	1
Borges, P.A.V., Abreu, C., Aguiar, A.M.F., Carvalho, P., Jardim, R., Melo, I., Oliveira, P., Sérgio, C., Serrano, A.R.M. & Vieira, P. (2008). *Listagem dos fungos, flora e fauna terrestres dos arquipélagos da Madeira e Selvagens.* Direcção Regional do Ambiente da Madeira and Universidade dos Açores, Funchal and Angra do Heroísmo, Portugal	1	0
Borges, P.A.V., Costa, A., Cunha, R., Gabriel, R., Gonçalves, V., Martins, A.F., Melo, I., Parente, M., Raposeiro, P., Rodrigues, P., Santos, R.S., Silva, L., Vieira, P. & Vieira, V. (2010). *A list of the terrestrial and marine biota from the Azores.* Princípia, Cascais	1	0
Botanical Garden Tel Aviv. *Israel Flora.* Tel Aviv University	1	1
Bowdoin Scientific Station (2011). Vascular plants of Kent Island, Available at: https://www.bowdoin.edu/kent-island/species/plants.shtml (last accessed: September 14, 2011)	1	0

	distribution	traits
Breckle, S.-W., Hedge, I.C. & Rafiqpoor, M.D. (2013). *Vascular plants of Afghanistan. An augmented checklist.* Scientia Bonnensis, Bonn	1	0
Brennan, K. (1996). *An annotated checklist of the vascular plants of the Alligator Rivers Region, Northern Territory, Australia.* Supervising Scientist, Barton, Australia	1	1
Bridgewater, S.G.M., Harris, D.J., Whitefoord, C., Monro, A.K., Penn, M.G., Sutton, D.A., Sayer, B., Adams, B., Balick, M.J., Atha, D.H., Solomon, J. & Holst, B.K. (2006). A Preliminary Checklist of the vascular plants of the Chiquibul Forest, Belize. *Edinburgh Journal of Botany*, **63**, 269–321	0	1
Brofas, G., Karetsos, G., Panitsa, M. & Theocharopoulos, M. (2001). The flora and vegetation of Gyali Island, SE Aegean, Greece. *Willdenowia*, **31**, 51–70	1	0
Broughton, D.A. & McAdam, J.H. (2005). A checklist of the native vascular flora of the Falkland Islands (Islas Malvinas). new information on the species present, their ecology, status and distribution. *The Journal of the Torrey Botanical Society*, **132**, 115–148	1	0
Brundu, G. & Camarda, I. (2013). The Flora of Chad: a checklist and brief analysis. *Phytokeys*, **23**, 1–17	1	0
Bundesamt für Naturschutz (2016). Floraweb, Available at: www.floraweb.de (last accessed: September 15, 2016)	1	0
Burton, R.M. (1991). A check-list and evaluation of the flora of Nisyros (Dodecanese, Greece). *Willdenowia*, **20**, 15–38	1	0
Butler, B.J., Barclay, J.S. & Fisher, J.P. (1999). Plant communities and flora of Robins Island (Long Island), New York. *Journal of the Torrey Botanical Society*, **126**, 63–76	1	0
Byrd, G.V. (1984). Vascular vegetation of Buldir Island, Aleutian Islands, Alaska, compared to another Aleutian Island. *Arctic*, **37**, 37–48	1	0
CARMABI (2009). Dutch Caribbean Biodiversity Explorer, Available at: http://www.dcbiodata.net/explorer/home (last accessed: June 24, 2011)	1	0
Cascante-Marín, A. & Estrada-Chavarría, A. (2012). Las plantas vasculares de El Rodeo, Costa Rica. *Brenesia*, **77**, 71–128	1	1
Case, T.J., Cody, M.L. & Ezcurra, E. (2002). *A new island biogeography of the Sea of Cortés.* Oxford University Press, New York, NY	1	0
Catarino, L., Martins, E.S., Basto, M.F. & Diniz, M.A. (2008). An annotated checklist of the vascular flora of Guinea-Bissau (West Africa). *Blumea-Biodiversity, Evolution and Biogeography of Plants*, **53**, 1–222	1	0
Chang, C.-S., Kim, H. & Chang, K. *Provisional Checklist of the Vascular Plants for the Korea Peninsular Flora (KPF). Version 1.0,* Korea	1	0
Chang, C.-S., Kim, H. & Chang, K. *Provisional Checklist of the Vascular Plants for the Korea Peninsular Flora (KPF). Version 1.0,* Korea	1	0
Chawla, A., Parkash, O., Sharma, V., Rajkumar, S., Lal, B., Gopichand, Singh, R.D. & Thukral, A.K. (2012). Vascular plants, Kinnaur, Himachal Pradesh, India. *Check list*, **8**, 321–348	0	1
Cheffings, C.M. & Farrell, L. (eds.) (2005). *The vascular plant red data list for Great Britain.* Joint Nature Conservation Committee, Peterborough	1	0
Chernyaeva, A.M. (1973). Flora of Onekotan Island. *Bulletin of Main Botanical Garden*, **87**, 21–29	1	0
Chiapella, J. & Ezcurra, C. (1999). La flora del parque provincial Tromen, provincia de Neuquén, Argentina. *Multequina*, **8**, 51–60	1	0
Chinese Virtual Herbarium (2016). The Flora of China v. 5.0, Available at: http://www.cvh.org.cn/ (last accessed: January 15, 2016)	1	1
Chong, K.Y., Tan, T.W.H. & Corlett, R.T. (2009). *A checklist of the total vascular plant flora of Singapore. Native, Naturalised and Cultivated Species.* Raffles Museum of Biodiversity Research, Singapore	1	1
Christmas Island National Park (2002). *Third Christmas Island national park management plan.* Parks Australia North, Christmas Island, Australia	1	0

	distribution	traits
Christodoulakis, D. (1996). The flora of Ikaria (Greece, E. Aegean Islands). *Phyton*, **36**, 63–91	1	0
Clark, J.L., Neill, D.A. & Asanza, M. (2006). Floristic checklist of the Mache-Chindul mountains of Northwestern Ecuador. *Contributions from the United States National Herbarium*, **54**, 1–180	1	1
Cochard, R. & Bloesch, U. (2007). Electronic plant species database of the Saadani National Park, coastal Tanzania, Available at: http://www.wildlife-baldus.com/saadani.html (last accessed: November 14, 2016)	1	1
CONABIO (2016). Sistema Nacional de Información sobre Biodiversidad, Available at: https://www.gob.mx/conabio (last accessed: April 11, 2016)	1	0
Conti, F. & Bartolucci, F. (2015). *The Vascular Flora of the National Park of Abruzzo, Lazio and Molise (Central Italy).* Springer International Publishing, Cham	1	0
Conti, F., Abbate, G., Alessandrini, A. & Blasi, C. (2005). *Annotated Checklist of the Italian Vascular Flora.* Palombi Editori, Roma, Italy	1	0
Convey, P., Lewis Smith, R.I., Hodgson, D.A. & Peat, H.J. (2000). The flora of the South Sandwich Islands, with particular reference to the influence of geothermal heating. *Journal of Biogeography*, **27**, 1279–1295	1	0
Costion, C. & Lorence, D. (2012). The endemic plants of Micronesia. A geographical checklist and commentary. *Micronesica*, **43**, 51–100	1	1
Cronk, Q.C.B. (1989). The past and present vegetation of St Helena. *Journal of Biogeography*, **16**, 47–64	1	0
Da Vela, M., Frignani, F., Bonari, G. & Angiolini, C. (2013). La flora vascolare della diserva naturale "La Pietra" (Toscana meridionale). *Micologia e vegetazione mediterranea*, **28**, 135–160	0	1
Danihelka, J., Chrtek, J. & Kaplan, Z. (2012). Checklist of vascular plants of the Czech Republic. *Preslia*, **84**, 647–811	1	0
D'Arcy, W.G. (1971). The island of Anegada and its flora. *Atoll Research Bulletin*, **139**, 1–21	1	0
Dauby, G., Leal, M. & Stevart, T. (2008). Vascular plant checklist of the coastal National Park of Pongara, Gabon. *Systematics and geography of plants*, **78**, 155–216	0	1
Dauby, G., Zaiss, R., Blach-Overgaard, A., et al. (2016). RAINBIO. A mega-database of tropical African vascular plants distributions. *PhytoKeys*, **74**, 1–18	1	1
de la Luz, León, Rebman, J., Domínguez-León, M. & Domínguez-Cadena, R. (2008). The vascular flora and floristic relationships of the sierra de la Giganta in Baja California Sur, México. *Revista mexicana de biodiversidad*, **79**, 29–65	0	1
Desmet, P. & Brouillet, L. (2013). Database of Vascular Plants of Canada (VASCAN): a community contributed taxonomic checklist of all vascular plants of Canada, Saint Pierre and Miquelon, and Greenland. *Phytokeys*, **25**, 55–67	0	1
Dimopoulos, P., Raus, T., Bergmeier, E., Constantinidis, T., Iatrou, G., Kokkini, S., Strid, A. & Tzanoudakis, D. (2013). *Vascular plants of Greece. An annotated checklist.* Botanischer Garten und Botanisches Museum Berlin-Dahlem, Freie Universität Berlin; Hellenic Botanical Society, Berlin, Athens	1	0
Domínguez, E., Marticorena, C., Elvebakk, A. & Pauchard, A. (2004). Catálogo de la flora vascular del Parque Nacional Pali Aike. XII Región, Chile. *Gayana Botánica*, **61**, 67–72	1	1
Doroftei, M., Oprea, A., Ştefan, N. & Sârbu, I. (2011). Vascular wild flora of Danube Delta Biosphere Reserve. *Sci. Annals of Danube Delta Institute*, **17**, 15–52	1	0
Dowhan, J.J. & Rozsa, R. (1989). Flora of Fire Island, Suffolk County, New York. *Bulletin of the Torrey Botanical Club*, **116**, 265–282	1	0
Du Puy, D.J. (1993). Christmas Island. species lists, Available at: http://www.anbg.gov.au/abrs/online-resources/flora/ (last accessed: April 06, 2011)	1	0
Egea, J. de, Peña-Chocarro, M., Espada, C. & Knapp, S. (2012). Checklist of vascular plants of the Department of Ñeembucú, Paraguay. *Phytokeys*, **9**, 15–179	1	1

	distribution	traits
Egorova, E.M. (1964). Flora of Shiashkotan Island. *Bulletin of the Main Botanical Garden*, **54**, 114–120	1	0
Ejaz-ul-Islam Dar, M., Cochard, R., Shrestha, R.P. & Ahmad, S. (2012). Floristic composition of Machiara National Park, District Muzaffarabad Azad Kashmir, Pakistan. *International Journal of Biosciences*, **2**, 28–45	0	1
Engemann, K., Sandel, B., Boyle, B.L., Enquist, B.J., Jørgensen, P.M., Kattge, J., McGill, B.J., Morueta-Holme, N., Peet, R.K., Spencer, N.J., Violle, C., Wiser, S.K. & Svenning, J.-C. (2016). A plant growth form dataset for the New World. *Ecology*, **97**, 3243	0	1
Esler, A.E. (1978). Botanical features of the Mokohinau Islands. *TANE*, **24**, 187–197	1	0
Espinosa-Jiménez, J.A., Pérez-Farrera, M.Á. & Martínez-Camilo, R. (2011). Inventario florístico del Parque Nacional Cañón del Sumidero, Chiapas, México. *Boletín de la Sociedad Botánica de México*, **89**, 37–82	0	1
Evarts-Bunders, P., Evarte-Bundere, G., Bāra, J. & Nitcis, M. (2013). The flora of vascular plants in nature reserve „Eglone”. *Acta Biologica Universitatis Daugavpiliensis*, **13**, 21–38	1	0
Exell, A.W. (1944). *Catalogue of the vascular plants of S. Tome (with Principe and Annobon).* Trustees of the British Museum, London, UK	1	0
Fern, K. & Fern, A. (2015). Useful Tropical Plants Database (last accessed: July 25, 2015)	0	1
Ferreira, A.L., Coutinho, B.R., Pinheiro, H.T. & Thomaz, L.D. (2007). Composição florística e formações vegetais da Ilha dos Franceses, Espírito Santo. *Bol. Mus. Biol. Mello Leitao*, **22**, 25–44	0	1
Figueiredo, E., Paiva, J., Stevart, T., Oliveira, F. & Smith, G.F. (2011). Annotated catalogue of the flowering plants of São Tomé and Príncipe. *Bothalia*, **41**, 41–82	1	1
Figueroa-C., Y. & Galeano, G. (2007). Lista comentada de las plantas vasculares del enclave seco interandino de La Tatacoa (Huila, Colombia). *Caldasia*, **29**, 263–281	1	1
Fischer, E., Rembold, K., Althof, A., Obholzer, J., Malombe, I., Mwachala, G., Onyango, J.C., Dumbo, B. & Theisen, I. (2010). Annotated checklist of the vascular plants of Kakamega Forest, Western Province, Kenya. *Journal of East African Natural History*, **99**, 129–226	1	1
Fischer, M.A., Adler, W. & Oswald, K. (2008). *Exkursionsflora für Österreich, Liechtenstein und Südtirol. Bestimmungsbuch für alle in der Republik Österreich, im Fürstentum Liechtenstein und in der Autonomen Provinz Bozen*, 3., verb. u. erw. Aufl. der "Exkursionsflora von Österreich". OÖ Landesmuseum, Linz	1	0
Florence, J., Chevillotte, H., Ollier, C. & Meyer, J.-Y. (2007). Base de données botaniques Nadeaud de l'Herbier de la Polynésie française (PAP), Available at: http://www.herbier-tahiti.pf/ (last accessed: July 01, 2011)	1	0
Fosberg, F.R., Renvoize, S.A. & Townsend, C.C. (1980). *The flora of Aldabra and neighbouring islands.* HMSO, London, UK	1	0
Franklin, J., Keppel, G. & Whistler, W.A. (2008). The vegetation and flora of Lakeba, Nayau and Aiwa islands, central Lau Group, Fiji. *Micronesica*, **40**, 169–225	1	0
Funk, V.A., Hollowell, T., Berry, P., Kelloff, C. & Alexander, S.N. (2007). *Checklist of the plants of the Guiana Shield (Venezuela: Amazonas, Bolivar, Delta Amacuro; Guyana, Surinam, French Guiana).* Department of Botany, National Museum of Natural History, Washington, DC	1	0
Gabrielsen, G.W., Brekke, B., Alsos, I.G. & Hansen, J.R. (1997). *Natur-og kulturmiljoet på Jan Mayen.* Norsk Polarinstitutt, Oslo, Norway	1	0
Gage, S., Joneson, S.L., Barkalov, V.Y., Eremenko, N.A. & Takahashi, H. (2006). A newly compiled checklist of the vascular plants of the Habomais, the Little Kurils. *Bulletin of the Hokkaido University Museum*, **3**, 67–91	1	0
Gamit, S.B., Maurya, R.R., Qureshimatva, U.M. & Solanki, H.A. (2015). Check list of flowering plants in Tapi District, Gujarat, India. *International Journal of advanced Research*, **3**, 1104–1123	0	1
Gerlach, J. (2003). The biodiversity of the granitic islands of Seychelles. *Phelsuma*, **11** (Supplement A), 1–47	1	0
Gray, A., Pelembe, T. & Stroud, S. (2005). The conservation of the endemic vascular flora of Ascension Island and threats from alien species. *Oryx*, **39**, 449–453	1	0

	distribution	traits
Green, P.S. (1994). Lord Howe Island. species lists, Available at: http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/49/index.html (last accessed: April 06, 2011)	1	0
Green, P.S. (1994). Norfolk Island. Species lists, Available at: http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/49/index.html (last accessed: April 06, 2011)	1	0
Greene, S.W. & Walton, D.W.H. (1975). An annotated check list of the sub-Antarctic and Antarctic vascular flora. *Polar Record*, **17**, 473–484	1	0
Harris, D.J. (2002). *The vascular plants of the Dzanga-Sangha Reserve, Central African Republic.* Royal Botanic Garden Edinburgh, Edinburgh, Scotland, UK	1	1
Hauenstein, E., Muñoz-Pedreros, A., Yánez, J., Sánchez, P., Möller, P., Guiñez, B. & Gil, C. (2009). Flora y vegetación de la Reserva Nacional Lago Peñuelas, Reserva de la Biósfera, Región de Valparaíso, Chile. *Bosque (Valdivia)*, **30**, 159–179	1	0
Hawkins, B.A., Rueda, M., Rangel, T.F., Field, R., Diniz-Filho, J.A.F. & Linder, P. (2013). Community phylogenetics at the biogeographical scale: cold tolerance, niche conservatism and the structure of North American forests. *Journal of Biogeography*, **41**, 23–38	0	1
Heberling, J.M., Jo, I., Kozhevnikov, A., Lee, H. & Fridley, J.D. (2017). Biotic interchange in the Anthropocene. Strong asymmetry in East Asian and eastern North American plant invasions. *Global Ecology and Biogeography*, **26**, 447–458	0	1
Hill, M.J. (ed.) (2002). *Biodiversity surveys and conservation potential of inner Seychelles islands.* Smithsonian Institution, Washington, DC	1	0
Hill, S.R. (1986). An annotated checklist of the vascular flora of Assateague Island (Maryland and Virginia). *Castanea*, **51**, 265–305	1	0
Hnatiuk, R.J. (1993). Subantarctic Islands. species lists, Available at: http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/50/index.html (last accessed: April 07, 2011)	1	0
Hyde, M.A., Wursten, B.T., Ballings, P. & Coates Palgrave, M. (2016). Flora of Botswana, Available at: http://www.botswanaflora.com/ (last accessed: November 14, 2016)	1	0
Hyde, M.A., Wursten, B.T., Ballings, P. & Coates Palgrave, M. (2016). Flora of Zimbabwe (last accessed: November 14, 2016)	1	0
Hyde, M.A., Wursten, B.T., Ballings, P. & Coates Palgrave, M. (2016). Flora of Mozambique, Available at: http://www.mozambiqueflora.com/ (last accessed: November 14, 2016)	1	0
Hyde, M.A., Wursten, B.T., Ballings, P. & Coates Palgrave, M. (2016). Flora of Malawi, Available at: http://www.malawiflora.com/index.php (last accessed: November 14, 2016)	1	0
Imada, C.T. (ed.) (2012). *Hawaiian Native and Naturalized Vascular Plants Checklist*, (December 2012 update). Hawaii Biological Survey, Bishop Museum, Honolulu, Hawai'i	1	0
INBIO (2000). Lista de planta de Costa Rica. With updates by Eduardo Chacón, Available at: http://www.inbio.ac.cr/papers/manual_plantas/index.html (last accessed: February 05, 2016)	1	0
Jackes, B.R. (2010). *Plants of Magnetic Island*, 3rd edn. James Cook University, Townsville, Australia	1	0
Jahn, R. & Schönfelder, P. (1995). *Exkursionsflora für Kreta.* Ulmer (Eugen), Stuttgart, Germany	1	0
Jaramillo Díaz, P. & Guézou, A. (2011). CDF checklist of Galapagos vascular plants. Bungartz, F., Herrera, H.W., Jaramillo, P., Tirado, N., Jímenez-Uzcategui, G., Ruiz, D., Guézou, A. & Ziemmeck, F. (eds.), Available at: http://www.darwinfoundation.org/datazone/checklists/vascular-plants/ (last accessed: February 02, 2011)	1	0
Jardim Botânico do Rio de Janeiro (2016). Flora do Brasil 2020 em construção, Available at: http://floradobrasil.jbrj.gov.br/ (last accessed: May 09, 2016)	1	1
Jiménez, J.E., Juárez, P. & Díaz, A. (2016). Checklist of the vascular flora of Reserva Biológica San Luis, Costa Rica. *Check list*, **12**, 1859	1	1

	distribution	traits
Johnson, P.N. & Campbell, D.J. (1975). Vascular plants of the Auckland Islands. *New Zealand journal of botany*, **13**, 665–720	1	0
Johnston, I.M. (1931). The flora of the Revillagigedo Islands. *Proceedings of the California Academy of Sciences*, **20**, 9–104	1	0
Jordano, P. (2008). FRUBASE, Available at: http://ebd10.ebd.csic.es/mywork/frubase/frubase.html (last accessed: 26.06.15)	0	1
Junak, S., Philbrick, R., Chaney, S. & Clark, R. (1997). *A checklist of vascular plants of Channel Islands National Park*, 2nd edn. Southwest Parks and Monuments Association, Tucson, Arizona	1	0
Kabuye, C.H.S., Mungai, G.M. & Mutangah, J.G. (1986). Flora of Kora National Reserve. *Kora: An Ecological Inventory of the Kora National Reserve, Kenya. Kora Research Project 1982-85 : a Joint Venture Between the National Museums of Kenya and the Royal Geographical Society* (ed. by M. Coe and N.M. Collins), pp. 57–104. Royal Geographical Society	0	1
Kamari, G., Phitos, D., Snogerup, B. & Snogerup, S. (1988). Flora and vegetation of Yioura, N Sporades, Greece. *Willdenowia*, **17**, 59–85	1	0
Karlsson, T. & Agestam, M. (2014). Checklist of Nordic vascular plants. Sweden Checklist, Available at: http://www.euphrasia.nu/checklista/index.eng.html (last accessed: February 11, 2016)	1	0
Kattge, J., Diaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P.B. & Wright, I.J. (2011). TRY – a global database of plant traits. *Global Change Biology*, **17**, 2905–2935	0	1
Kelloff, C.L. & Funk, V.A. (1998). *Preliminary checklist of the plants of Kaieteur National Park, Guyana.* National Museum of Natural History, Smithsonian Institution, Washington	1	0
Kelly, L. (2006). The vascular flora of Huggins Island, Onslow County, North Carolina. *Castanea*, **71**, 295–311	1	0
Kemenes, A. (2003). Distribuição espacial da flora terrestre fanerogâmica do Parque Nacional Marinho de Abrolhos, BA. *Revista Brasil. Bot*, **26**, 141–150	0	1
Keppel, G. (2005). *Summary report on forests of the Mataqali Nadicake Kilaka, Kubulau district, Bua, Vanua Levu*	0	1
Keppel, G., Gillespie, T.W., Ormerod, P. & Fricker, G.A. (2016). Habitat diversity predicts orchid diversity in the tropical south-west Pacific. *Journal of Biogeography*, **43**, 2332–2342	1	0
Kerguelen, M. (2005). Base de Données Nomenclaturales de la Flore de France, Available at: http://www.tela-botanica.org/page:284# (last accessed: January 13, 2012)	1	0
Khanina, L., Zaugolnova, L., Smirnova, O., Shovkun, M. & Glukhova, E. (2016). Flora of vascular plants in the Central European Russia, Available at: http://www.impb.ru/eco/index.php?l=en (last accessed: September 06, 2017)	1	0
Kim, H.-J., Ji, S.-J., Jung, S.-Y., Park, S.H., Lee, S.-G., Lee, C.-W. & Chang, K.S. (2015). Flora of Vascular Plants in Deokjeokdo (Ongjin-gun) and Its Adjacent Regions, Korea. *Korean Journal of Plant Resources*, **28**, 487–510	1	0
Kingston, N., Waldren, S. & Bradley, U. (2003). The phytogeographical affinities of the Pitcairn Islands – a model for south-eastern Polynesia? *Journal of Biogeography*, **30**, 1311–1328	1	1
Kirchner, F., Picot, F., Merceron, E. & Gigot, G. (2010). *Flore vasculaire de La Réunion.* Conservatoire Botanique National de Mascarin, Réunion, France	1	0
Kleyer, M., Bekker, R.M., Knevel, I.C., et al. (2008). The LEDA Traitbase: a database of life-history traits of the Northwest European flora. *Journal of Ecology*, **96**, 1266–1274	0	1
Koltzenburg, M. (2011). Checkliste der Gefässpflanzen Irlands, Available at: http://www.saxifraga.de/eire/irl_artenlisten_gp.html (last accessed: April 04, 2011)	1	0

	distribution	traits
Kraaij, T. (2011). The flora of the Bontebok National Park in regional perspective. *South African Journal of Botany*, **77**, 455–473	1	1
Kraft, T.S., Wright, S.J., Turner, I., Lucas, P.W., Oufiero, C.E., Supardi Noor, M.N., Sun, I.-F., Dominy, N.J. & Whitney, K. (2015). Seed size and the evolution of leaf defences. *Journal of Ecology*, **103**, 1057–1068	0	1
Kristinsson, H. (2008). *Checklist of the vascular plants of Iceland.* Náttúrufræðistofnun Íslands, Reykjavík, Iceland	1	0
Kumar, A., Bajpai, O., Mishra, A.K., Sahu, N., Behera, S.K., Bargali, S.S. & Chaudhary, L.B. (2015). A checklist of the flowering plants of Katerniaghat Wildlife Sanctuary, Uttar Pradesh, India. *Journal of Threatened Taxa*, **7**, 7309–7408	0	1
Kuzmenkova, S.M. (2015). Plants of Belarus, Available at: http://hbc.bas-net.by/plantae/eng/default.php (last accessed: February 15, 2016)	1	0
Lange, P.J. de & Cameron, E.K. (1999). The vascular flora of Aorangi Island, Poor Knights Islands, northern New Zealand. *New Zealand journal of botany*, **37**, 433–468	1	0
Lange, P.J. de, Heenan, P.B. & Rolfe, J.R. (2011). *Checklist of vascular plants recorded from Chatham Islands*	1	0
Lazkov, G.A. & Sultanova, B.A. (2011). *Checklist of vascular plants of Kyrgyzstan.* Botanical Museum, Finnish Museum of Natural History, Helsinki	1	0
Le Houerou, H.N. (2004). Plant diversity in Marmarica (Libya & Egypt): a catalogue of the vascular plants reported with their biology, distribution, frequency, usage, economic potential, habitat and main ecological features, with an extensive bibliography. *Candollea*, **59**, 259–308	1	1
Lee, R.-Y., Jang, H.-D., Kim, Y.-Y., Yang, S.-G., Choi, H.-J., Ji, S.-J. & Oh, B.-U. (2014). Flora of vascular plants in the Chilgapsan Provincial Park, Korea. *Journal of Asia-Pacific Biodiversity*, **7**, 237–247	1	0
Lester-Garland, L.V. (1903). *A flora of the islands of Jersey. with a list of the plants of the Channel Islands in general, and remarks upon their distribution and geographical affinities.* West, Newman & Co, London, UK	1	0
Letcher, S.G., Lasky, J.R., Chazdon, R.L., et al. (2015). Environmental gradients and the evolution of successional habitat specialization. A test case with 14 Neotropical forest sites. *Journal of Ecology*, **103**, 1276–1290	0	1
Levin, G.A. & Moran, R. (1989). The vascular flora of Socorro, Mexico. *Memoirs of the San Diego Society of Natural History*, **16**, 1–71	1	0
Limbu, D., Koirala, M. & Shang, Z. (2013). A Checklist of Angiospermic Flora of Tinjure-Milke-Jaljale, Eastern Nepal. *Nepal Journal of Science and Technology*, **13**, 87–96	0	1
Linhart, Y.B. (1980). Local biogeography of plants on a Caribbean atoll. *Journal of Biogeography*, **7**, 159–171	0	1
Lipkin, R. (2005). *Aniakchak National Monument and Preserve, vascular plant inventory: final technical report.* National Park Service, Southwest Alaska Network Inventory & Monitroing Program, Anchorage, USA	1	0
Lopez-Martinez, J.O., Sanaphre-Villanueva, L., Dupuy, J.M., Hernandez-Stefanoni, J.L., Meave, J.A. & Gallardo-Cruz, J.A. (2013). Beta-Diversity of functional groups of woody plants in a tropical dry forest in Yucatan. *PloS one*, **8**, e73660	0	1
Lord, J.M. (2015). Patterns in floral traits and plant breeding systems on Southern Ocean Islands. *AoB Plants*, **7**	1	1
Lorite, J. (2016). An updated checklist of the vascular flora of Sierra Nevade (SE Spain). *Phytotaxa*, **261**, 1–57	1	0
Luke, Q. (2005). Annotated Checklist of the Plants of the Shimba Hills, Kwale District, Kenya. *Journal of East African Natural History*, **94**, 5–120	1	1
Luna-Jorquera, G., Fernández, C.E. & Rivadeneira, M.M. (2012). Determinants of the diversity of plants, birds and mammals of coastal islands of the Humboldt current systems. Implications for conservation. *Biodiversity and Conservation*, **21**, 13–32	1	0
Marquand, E.D. (1901). *Flora of Guernsey and the lesser Channel Islands. namely Alderney, Sark, Herm, Jethou, and the adjacent islets.* Dulau & Co, London, UK	1	0

	distribution	traits
Marticorena, C., Squeo, F.A., Arancio, G. & Muñoz, M. (2008). Catálogo de la flora vascular de la IV Región de Coquimbo. *Libro rojo de la flora nativa y de los sitios prioritarios para su conservación: Región de Atacama* (ed. by F.A. Squeo, G. Arancio and J.R. Gutiérrez), pp. 105–142. Ediciones Universidad de La Serena La Serena	1	1
Marticorena, C., Stuessy, T.F. & Baeza, C.M. (1998). Catalogue of the vascular flora of the Robinson Crusoe or Juan Fernández islands, Chile. *Gayana Botánica*, **55**, 187–211	1	1
Masharabu, T. *Flore et végétation du Parc National de la Ruvubu au Burundi: diversité, structure et implications pour la conservation.* Editions universitaires europeennes	1	1
McClatchey, W., Thaman, R. & Vodonaivalu, S. (2000). A preliminary checklist of the flora of Rotuma with Rotuman names. *Pacific Science*, **54**, 345–363	1	0
McCrea, J. (2003). *Inventory of the land conservation values of the Houtman Abrolhos Islands.* Department of Fisheries, Government of Western Australia, Perth, Australia	1	0
Medina, R., Reina-E, M., Herrera, E., Ávila, F.A., Chaparro, O. & Cortés-B., R. (2010). Catálogo preliminar da flora vascular dos bosques subandinos da cuchilla El Fara (Santander-Colômbia). *Colombia Forestal*, **13**, 55–85	0	1
Medjahdi, B., Ibn Tattou, M., Barkat, D. & Benabedli, K. (2009). La flore vasculaire des Monts des Trara (Nord Ouest Algérien). *Acta Botanica Malacitana*, **34**, 57–75	1	0
Miller, A.G. & Morris, M. (2004). *Ethnoflora of the Soqotra Archipelago.* Royal Botanic Garden, Edinburgh, UK	1	0
Ministry of Environment (2012). *The national red list 2012 of Sri Lanka. Conservation status of the fauna and flora,* Colombo, Sri Lanka	1	0
Moran, R. (1996). *The flora of Guadalupe Island, Mexico.* California Academy of Sciences, San Francisco, CA	1	0
Morat, P., Jaffré, T., Tronchet, F., Munziger, J., Pillon, Y., Veillon, J.-M., Chalopin, M., Birnbaum, P., Rigault, F., Dagostini, G., Tinerl, J. & Lowry II, P.P. (2012). The taxonomic reference base Florical and characteristics of the native vascular flora of New Caledonia. *Adansonia*, **34**, 179–221	1	0
Nationalpark Eifel (2015). Artenliste Farne und Blütenpflanzen, Available at: http://www.nationalpark-eifel.de/go/artenliste.html (last accessed: March 11, 2015)	1	0
Nikolić, T. (2016). Flora Croatica Database, Available at: http://hirc.botanic.hr/fcd (last accessed: February 12, 2016)	1	0
Norton, J., Majid, S.A., Allan, D., Al Safran, M., Böer, B. & Richer, R.A. (2009). *An illustrated checklist of the flora of Qatar.* Browndown Publications Gosport, Gosport, UK	1	1
Notov, A.A. (2010). *National park "Zavidovo": vascular plants, bryophyte, lichens,* Moscow	1	0
NPS (2015). NPSpecies. Information on Species in National Parks, Available at: https://irma.nps.gov/NPSpecies/ (last accessed: April 09, 2015)	1	0
Oggero, A.J. & Arana, M.D. (2012). Inventario de las plantas vasculares del sur de la zona serrana de Córdoba, Argentina. *Hoehnea*, **39**, 171–199	1	0
Oliveira-Filho, A.T. (2014). NeoTropTree. Flora arbórea da Região Neotropical: Um banco de dados envolvendo biogeografia, diversidade e conservação, Available at: http://www.icb.ufmg.br/treeatlan/ (last accessed: February 12, 2016)	0	1
Pal, D., Kumar, A. & Dutt, B. (2014). Floristic diversity of Theog Forest Division, Himachal Pradesh, Western Himalaya. *Check list*, **10**, 1083–1103	1	1
Pandža, M. & Skvorc, Z. (2002). The flora of some uninhabited Sibenik Archipelago islands (Dalmatia, Croatia). *Natura Croatica*, **11**, 367–385	1	0
Pandža, M. (2002). Flora of the small islands of Murter. *Natura Croatica*, **11**, 77–101	1	0
Pandža, M. (2003). Flora of the island of Zirje and the small islands around it (eastern Adriatic coast, Croatia). *Acta Botica Croatia*, **62**, 115–139	1	0
Pandža, M. (2010). Flora parka prirode Papuk (Slavonija, Hrvatska). *Šumarski list*, **134**, 25–43	1	0

	distribution	traits
Parks Canada (2015). Biotics Web Explorer, Available at: http://www.pc.gc.ca/apps/bos/BOSIntro_e.asp (last accessed: April 13, 2015)	1	0
Paula, S., Arianoutsou, M., Kazanis, D., Tavsanoglu, Ç., Lloret, F., Buhk, C., Ojeda, F., Luna, B., Moreno, J.M. & Rodrigo, A. (2009). Fire-related traits for plant species of the Mediterranean Basin. *Ecology*, **90**, 1420	0	1
Peña-Chocarro, M.d.C. (2010). *Updated checklist of vascular plants of the Mbaracayú Forest Nature Reserve (Reserva Natural del Bosque Mbaracayú), Paraguay.* Magnolia Press, Auckland, N.Z.	1	1
Pôle Flore Habitats (2015). Catalogue de la flore vasculaire de Rhône-Alpes, Available at: http://www.pifh.fr/pifhcms/index.php (last accessed: January 12, 2015)	1	0
Price, J.P. & Wagner, W.L. (2011). A phylogenetic basis for species–area relationships among three Pacific Island floras. *American Journal of Botany*, **98**, 449–459	1	0
Proctor, G.R. (1980). Checklist of the plants of Little Cayman. Geography and ecology of Little Cayman. *Atoll Research Bulletin*, **241**, 71–80	1	0
Proctor, G.R. (1989). *Ferns of Puerto Rico and the Virgin Islands.* Memoirs of the New York Botanical Garden, New York, NY	1	0
Queensland Government (2014). Census of the Queensland flora 2014, Available at: https://data.qld.gov.au/dataset/census-of-the-queensland-flora-2014 (last accessed: February 05, 2015)	1	0
Rahman, A.H.M.M. (2013). Angiospermic Flora of Rajshahi District, Bangladesh. *American Journal of Life Sciences*, **1**, 105	0	1
Rahman, M.S., Hossain, G.M., Khan, S.A. & Uddin, S.N. (2015). An annotated Checklist of the Vascular Plants of Sundarban Mangrove Forest of Bangladesh. *Bangladesh Journal of Plant Taxonomy*, **22**, 17–41	0	1
Rakov, N.S., Saksonov S.V., Senator S.A. & Vasjukov V.M. (2014). *Vascular plants of Ulyanovsk Region.* Russian Academy of Sciences, Togliatti	1	0
Raulerson, L. (2006). Checklist of plants of the Mariana Islands. *University of Guam Herbarium Contribution*, **37**, 1–69	1	0
Renvoize, S.A. (1975). A floristic analysis of the western Indian Ocean coral islands. *Kew Bulletin*, **30**, 133–152	1	0
Robinson, A.C., Canty, P.D. & Fotheringham, D. (2008). Investigator group expedition 2006. flora and vegetation. *Transactions of the Royal Society of South Australia*, **132**, 173–220	1	0
Robinson, A.C., Canty, P.D., Wace, N.M. & Barker, R.M. (2003). The encounter 2002 expedition to the isles of St Francis, South Australia. flora and vegetation. *Transactions of the Royal Society of South Australia*, **127**, 107–128	1	0
Rossetto, E.F.S. & Vieira, A.O.S. (2013). Vascular Flora of the Mata dos Godoy State Park, Londrina, Paraná, Brazil. *Check list*, **9**, 1020–1034	1	1
Roux, J.P. (2009). *Synopsis of the Lycopodiophyta and Pteridophyta of Africa, Madagascar and neighbouring islands.* South African National Biodiversity Institute, Cape Town, South Africa	1	0
Royal Botanic Gardens and Domain Trust (2017). PlantNET - The NSW Plant Information Network System, Available at: http://plantnet.rbgsyd.nsw.gov.au (last accessed: September 16, 2016)	1	1
Royal Botanic Gardens Kew (2016). Seed information database (SID). v7.1, Available at: http://data.kew.org/sid/ (last accessed: March 22, 2013)	0	1
Rundel, P.W., Dillon, M.O. & Palma, B. (1996). Flora and Vegetation of Pan de Azúcur National Park in the Atacama desert of Northern Chile. *Gayana Bot*, **53**, 295–315	1	1
Sachet, M.-H. (1962). Flora and vegetation of Clipperton Island. *Proceedings of the California Academy of Sciences*, **31**, 249–307	1	0
Sáez, L. & Rosselló, J.A. (2001). *Llibre vermell de la flora vascular de les Illes Balears.* Direcció General de Biodiversitat, Conselleria de Medi Ambient, Govern de les Illes Balears, Palma de Mallorca, Spain	1	0

	distribution	traits
SANBI (2014). Plants of Southern Africa. An online checklist, Available at: http://posa.sanbi.org (last accessed: March 09, 2015)	1	1
Sandbakk, B.E., Alsos, I.G., Arnesen, G. & Elven, R. (1996). The flora of Svalbard, Available at: http://svalbardflora.no/ (last accessed: March 16, 2011)	1	0
Schaefer, H., Hardy, O.J., Silva, L., Barraclough, T.G. & Savolainen, V. (2011). Testing Darwin's naturalization hypothesis in the Azores. *Ecology Letters*, **14**, 389–396	0	1
Schönfelder, P. & Schönfelder, I. (1997). *Die Kosmos-Kanarenflora. Über 850 Arten der Kanarenflora und 48 tropische Ziergehölze*. Franckh-Kosmos, Stuttgart	0	1
Searle, J. & Madden, S. (2006). *Flora assessment of South Stradbroke Island.* Gold Coast City Council, Gold Coast City, Australia	1	0
Selvi, F. (2010). A critical checklist of the vascular flora of Tuscan Maremma (Grosseto province, Italy). *Fl. Medit*, **20**, 47–139	1	1
Senterre, B., Chew, M.Y. & Chung, R.C.K. (2015). Flora and vegetation of Pulau Babi Tengah, Johor, Peninsular Malaysia. *Check list*, **11**, 1714	0	1
Shaw, J.D., Spear, D., Greve, M. & Chown, S.L. (2010). Taxonomic homogenization and differentiation across Southern Ocean Islands differ among insects and vascular plants. *Journal of Biogeography*, **37**, 217–228	1	0
Short, P.S., Albrecht, D.E., Cowie, I.D., Lewis, D.L. & Stuckey, B.M. (2011). *Checklist of the vascular plants of the Northern Territory.* Department of Natural Resources, Environment, The Arts and Sport, Darwin	1	0
Silaeva, T.B., Chugunov, G.G., Kiryukhin, I.V., Ageeva, A.M., Vargot, E.V., Grishutkina, G.A. & Khapugin, A.A. *Flora of the national park "Smolny". Mosses and vascular plants: annotated list of species [In Russian].* Commission of RAS for the Conservation of Biological Diversity [Комиссия РАН по сохранению биологического разнообразия]	1	0
Singh, A. (2015). Observations on the vascular flora of Banaras Hindu University Main Campus, India. *International Journal of Modern Biology and Medicine*, **6**, 48–87	1	1
Slik, F.J.W., Arroyo-Rodríguez, V., Aiba, S.-I., 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	0	1
SLUFG (2015). *Rote Liste und Artenliste Sachsen. Farn- und Samenpflanzen.* Sächsisches Landesamt für Umwelt, Landwirtschaft und Geologie, Dresden	1	0
Smith, A.C. (1979-1996). *Flora Vitiensis nova. a new Flora of Fiji (spermatophytes only).* Pacific Tropical Botanical Garden (Lawaii, Hawaii)	1	1
Stace, C.A., Ellis, R.G., Kent, D.H. & McCosh, D.J. (2003). *Vice-county Census Catalogue of the vascular plants of Great Britain, the Isle of Man and the Channel Islands.* Botanical Society of the British Isles, London, UK	1	0
Stalmans, M. (2006). *Tinley's plant species list for the Greater Gorongosa ecosystem, Moçambique.* Unpublished report by International Conservation Services to the Carr Foundation and the Ministry of Tourism	1	1
Stalter, R. & Lamont, E.E. (2006). The historical and extant flora of Sable Island, Nova Scotia, Canada. *Journal of the Torrey Botanical Society*, **133**, 362–374	1	0
Stoddart, D.R. & Fosberg, F.R. (1994). Flora of the Phoenix Islands, central Pacific. *Atoll Research Bulletin*, **393**, 1–60	1	0
Strahm, W.A. *The conservation and restoration of the flora of Mauritius and Rodrigues*	1	0
Sykes, W.R. (1970). Contributions to the flora of Niue. *Bulletin. Department of Scientific and Industrial Research, New Zealand*, **200**, 321	1	0
Sykes, W.R., West, C.J., Beever, J.E. & Fife, A.J. (2000). *Kermadec Islands flora-special edition. a compilation of modern material about the flora of the Kermadec Islands.* Manaaki Whenua Press, Landcare Research	1	0
Takahashi, H., Barkalov, V.Y., Gage, S. & Zhuravlev, Y.N. (1997). A preliminary study of the flora of Chirpoi, Kuril Islands. *Acta Phytotaxonomica et Geobotanica*, **48**, 31–42	1	0

	distribution	traits
Takahashi, H., Barkalov, V.Y., Gage, S., Joneson, S., Ilushko, M. & Zhuravlev, Y.N. (2002). A floristic study of the vascular plants of Raikoke, Kuril Islands. *Acta Phytotaxonomica et Geobotanica*, **53**, 17–33	1	0
Takahashi, H., Barkalov, V.Y., Gage, S., Semsrott, B., Ilushko, M. & Zhuravlev, Y.N. (1999). A preliminary checklist of the vascular plants of Chirinkotan, Kuril Islands. *Journal of Phytogeography and Taxonomy*, **47**, 131–137	1	0
Takahashi, H., Barkalov, V.Y., Gage, S., Semsrott, B., Ilushko, M. & Zhuravlev, Y.N. (2006). A floristic study of the vascular plants of Kharimkotan, Kuril Islands. *Bulletin of the Hokkaido University Museum*, **3**, 41–66	1	0
Tamis, W.L.M., van der Meijden, R., Runhaar, J., Bekker, R.M., Ozinga, W.A., Odé, B. & Hoste, I. (2004). Standard List of the Flora of the Netherlands 2003. *Gorteria*, **30**, 101–195	1	1
Tatewaki, M. (1957). Geobotanical studies on the Kurile Islands. *Acta Horti Gotoburgensis*, **21**, 43–123	1	0
Taylor, R. (2006). *Straight through from London. the Antipodes and Bounty Islands, New Zealand.* Heritage Expeditions New Zealand, Christchurch, New Zealand	1	0
Tela Botanica (2015). Base de données des Trachéophytes de France métropolitaine (bdtfx), Available at: http://www.tela-botanica.org/bdtfx	1	0
Ter Steege, H., Vaessen, R.W., Cardenas-Lopez, D., Sabatier, D., Antonelli, A., Oliveira, S.M. de, Pitman, N.C.A., Jorgensen, P.M. & Salomao, R.P. (2016). The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Scientific reports*, **6**, 29549	0	1
Thaman, R.R., Fosberg, F.R., Manner, H.I. & Hassall, D.C. (1994). The flora of Nauru. *Atoll Research Bulletin*, **392**, 1–233	1	0
Thomas, J. (2011). Plant diversity of Saudi Arabia. Flora checklist, Available at: http://plantdiversityofsaudiarabia.info/biodiversity-saudi-arabia/flora/Checklist/Cheklist.htm (last accessed: January 14, 2016)	1	0
Tonkov, S., Pavlova, D., Atanassova, J., Nedelcheva, A. & Marinova, E. (2004). *Floristis catalogue of the nature reserve Rilomanastirska Gora (Central Rila Mountains). I. The locality Kirilova.* University of Sofia	0	1
Tropicos (2015). Catálogo de las Plantas Vasculares de Bolivia. *Tropicos* . Missouri Botanical Garden, St. Louis, Available at: http://www.tropicos.org/Project/BC	1	1
Tropicos (2015). Catalogue of the Vascular Plants of Ecuador, Available at: http://www.tropicos.org/Project/CE (last accessed: October 22, 2015)	1	1
Tropicos (2015). Catalogue of the Vascular Plants of the Department of Antioquia (Colombia). *Tropicos* . Missouri Botanical Garden, St. Louis, Available at: http://www.tropicos.org/Project/CV	0	1
Tropicos (2015). Flora de Nicaragua. *Tropicos* . Missouri Botanical Garden, St. Louis, Available at: http://www.tropicos.org/Project/FN	1	0
Tropicos (2015). Panama Checklist. *Tropicos* . Missouri Botanical Garden, St. Louis, Available at: http://www.tropicos.org/Project/PAC	0	1
Tropicos (2015). Paraguay Checklist. *Tropicos* . Missouri Botanical Garden, St. Louis, Available at: http://www.tropicos.org/Project/Paraguay	0	1
Tropicos (2015). Peru Checklist. *Tropicos* . Missouri Botanical Garden, St. Louis, Available at: http://www.tropicos.org/Project/PEC	0	1
Trusty, J.L., Kesler, H.C. & Delgado, G.H. (2006). Vascular flora of Isla del Coco, Costa Rica. *Proceedings of the California Academy of Sciences*, **57**, 247–355	1	0
Turner, I.M. (1995). *A catalogue of the vascular plants of Malaya.* Gardens' Bulletin, Singapore	1	1
Tutul, E., Uddin, M.Z., Rahman, M.O. & Hassan, M.A. (2009). Angiospermic flora of Runctia sal forest, Bangladesh. I. Liliopsida (Monocots). *Bangladesh Journal of Plant Taxonomy*, **16**, 83–90	0	1
UIB (2007). Herbario virtual del Mediterráneo Occidental, Available at: http://herbarivirtual.uib.es/cas-med/ (last accessed: August 07, 2012)	1	0

	distribution	traits
University of Greifswald (2010). FloraGREIF - Virtual Flora of Mongolia, Available at: http://greif.uni-greifswald.de/floragreif/ (last accessed: February 02, 2016)	0	1
University of Kent (2012). Cook Islands Biodiversity and Ethnobiology Database, Available at: http://cookislands.pacificbiodiversity.net/cibed/dbs/search.html (last accessed: April 12, 2012)	1	0
USDA & NRCS (2015). The PLANTS Database. National Plant Data Team (ed.), Available at: http://plants.usda.gov (last accessed: April 24, 2015)	1	1
van Vreeswyk, A.M.E., Payne, A.L., Leighton, K.A. & Hennig, P. (2004). *An inventory and condition survey of the Pilbara region, Western Australia.* Department of Agriculture	0	1
Vanderplank, S.E. (2010). *The Vascular Flora of Greater San Quintín, Baja California, Mexico.* CGU Theses & Dissertations	1	1
Velarde, E., Wilder, B.T., Felcer, R.S. & Ezcurra, E. (2014). Floristic diversity and dynamics of Isla Rasa, Gulf of California - A globally important seabird island. *Botanical Sciences*, **92**, 89–101	0	1
Velarde, E.P., Guzmán, R.C. & Koch, S.D. (2008). Plantas vasculares y vegetación de la parte alta del Arroyo Agua Fría, municipio de Minatitlán, Colima, México. *Acta Botanica Mexicana*, **84**, 25–72	0	1
Velayos Rodríguez, M. (2016). Flora de Guinea Equatorial, Available at: http://www.floradeguinea.com/ (last accessed: February 14, 2016)	1	0
Velayos, M., Barberá, P., Cabezas, F.J., La Estrella, M.d., Fero, M. & Aedo, C. (2014). Checklist of the Vascular Plants of Annobón (Equatorial Guinea). *Phytotaxa*, **171**, 1–78	1	0
VicFlora (2016). Flora of Victoria, Available at: http://vicflora.rbg.vic.gov.au (last accessed: November 30, 2016)	1	0
VicFlora (2016). Flora of Victoria, Available at: http://vicflora.rbg.vic.gov.au (last accessed: November 30, 2016)	1	1
Viciani, D., Gonnelli, V., Sirotti, M. & Agostini, N. (2010). An annotated check-list of the vascular flora of the "Parco Nazionale delle Foreste Casentinesi, Monte Falterona e Campigna"(Northern Apennines Central Italy). *Webbia*, **65**, 3–131	1	1
Vogt, C. (2011). Composición de la Flora Vascular del Chaco Boreal, Paraguay. I. Pteridophyta y Monocotiledoneae. *Steviana*, **3**, 13–47	1	1
Wace, N.M. & Dickson, J.H. (1965). The terrestrial botany of the Tristan da Cunha Islands. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, **249**, 273–360	1	0
Wace, N.M. (1961). The vegetation of Gough Island. *Ecological Monographs*, **31**, 337–367	1	0
Wagner, W.L. & Lorence, D.H. (2002). Flora of the Marquesas Islands website, Available at: http://botany.si.edu/pacificislandbiodiversity/marquesasflora/index.htm (last accessed: November 28, 2014)	1	0
Wagner, W.L., Herbst, D.R. & Lorence, D.H. (2005). Flora of the Hawaiian Islands website, Available at: http://botany.si.edu/pacificislandbiodiversity/hawaiianflora/ (last accessed: October 16, 2010)	1	1
WCSP (2014). World Checklist of Selected Plant Families, Available at: http://apps.kew.org/wcsp/home.do (last accessed: December 01, 2014)	1	0
Webster, G.L. & Rhode, R.M. (2001). Plant diversity of an Andean cloud forest: inventory of the vascular plants of Maquipucuna, Ecuador. *Publications in Botany*, **82**, 1–228	0	1
Wellington Botanical Society (2008). *Native vascular plants of Great Barrier Island.* Wellington Botanical Society, Wellington, New Zealand	1	0
Wester, L. (1985). Checklist of the vascular plants of the northern Line Islands. *Atoll Research Bulletin*, **187**, 1–38	1	0
Western Australian Herbarium (2017). FloraBase - the Western Australian Flora. Western Australian Herbarium (ed.), Available at: https://florabase.dpaw.wa.gov.au/ (last accessed: September 16, 2016)	1	1

	distribution	traits
Whistler, W.A. (1983). Vegetation and flora of the Aleipata Islands, Western Samoa. *Pacific Science*, **37**, 227–249	1	0
Whistler, W.A. (1998). *A study of the rare plants of American Samoa.* US Fish and Wildlife Service, Honolulu, Hawaii	1	0
Whistler, W.A. (2012). Botanical survey of the Ringgold Islands, Fiji. *Allertonia*, **11**, 1–28	1	0
Whistler, W.A. *Botanical survey of Diego Garcia, Chagos Archipelago, British Indian Ocean Territory.* Appendix E1	1	0
Woodroffe, C.D. (1986). Vascular plant speciesarea relationships on Nui Atoll, Tuvalu, Central Pacific: a reassessment of the small island effect. *Australien Journal of Ecology*, **11**, 21–31	1	1
Yineger, H., Kelbessa, E., Bekele, T. & Lulekal, E. (2008). Floristic composition and structure of the dry afromontane forest at Bale Mountains National Park, Ethiopia. *SINET: Ethiopian Journal of Science*, **31**, 103–120	0	1
ZDSF & SKEW (2014). Info Flora. Artenliste Schweiz 5x5 km, Available at: https://www.infoflora.ch/de/daten-beziehen/artenliste-5x5-km.html (last accessed: February 13, 2015)	1	0
Zhang, S.T., Zhen Du, G. & Chen, J.K. (2004). Seed size in relation to phylogeny, growth form and longevity in a subalpine meadow on the east of the Tibetan plateau. *Folia Geobotanica*, **39**, 129–142	0	1
Zuloaga, F.O., Morrone, O. & Belgrano, M. (2014). Catálogo de las Plantas Vasculares del Cono Sur, Available at: http://www.darwin.edu.ar/Proyectos/FloraArgentina/fa.htm (last accessed: March 16, 2015)	1	1
Zvyagintseva, K.O. (2015). *An annotated checklist of the urban flora of Kharkiv.* Kharkiv National University, Kharkiv, Ukraine	1	0
Артемов, И.А. (2012). *Определитель растений Катунского биосферного заповедника.* Russian Academy of Sciences, БАРНАУЛ	0	1
Евстигнеев, О.И. & Федотов, Ю.П. (2007). *Флора сосудистых растений заповедника "Брянский лес".* Гос. природ. биосфер. заповедник Брян. лес, Брянск	1	0
Хапугин, А.А. (2013). *Сосудистые растения Ромодановского района Республики Мордовия (конспект флоры),* Saransk	1	0